# CONSTRUCTION IN PROGRESS:

# THE COGNITIVE ARCHITECTURE OF EPISODIC MEMORY AND SENSORY IMAGINATION

ANDREA BLOMKVIST

A thesis submitted in partial fulfilment of the requirements for the degree of
Doctor of Philosophy

The University of Sheffield
Faculty of Arts and Humanities
Department of Philosophy

2021

# CONSTRUCTION IN PROGRESS:

## THE COGNITIVE ARCHITECTURE OF EPISODIC MEMORY AND SENSORY IMAGINATION

ANDREA BLOMKVIST

A thesis submitted in partial fulfilment of the requirements for the degree of
Doctor of Philosophy

The University of Sheffield
Faculty of Arts and Humanities
Department of Philosophy

Submitted on 16/7/2021

# Abstract

This thesis develops a novel architecture of episodic memory and sensory imagination, which details the processes involved in producing such episodes. The proposed cognitive architecture – the Constructive Episodic Simulation Hypothesis+ (CESH+) – builds on its predecessor CESH, and adds new features of content retrieval. On top of the dissociable retrieval and recombination processes posited by CESH, the thesis makes an empirical case for adding (i) memory indices storing the addresses of content, (ii) a Bayesian ranking/selection mechanism, which functions to select the appropriate content to be retrieved, as well as (iii) more fine-grained retrieval processes. To garner support for the model, the thesis extensively discusses two recent problems in cognitive science. The first concerns how imagination could be a skill. As skills are both improvable by practice and controlled, a cognitive architecture needs to account for how this can be the case. The thesis argues that we can make sense of how this can be so by appealing to the computations carried about by the Bayesian ranking/selection mechanism. The second problem concerns a new condition called 'aphantasia', which is characterised by subjects' impaired mental imagery in both memory and imagination, and which currently lacks an adequate explanation. The thesis offers the first comprehensive overview of the impairments of aphantasia, and argues that CESH+ has the resources to explain these impairments. In the light of the explanatory power of CESH+, the thesis concludes that CESH+ paves the way to an improved understanding of the workings of memory and imagination.

# Table of Contents

**ABSTRACT** I

**TABLE OF CONTENTS** II

**LIST OF FIGURES** VI

**LIST OF TABLES** VII

**DECLARATION** VIII

**ACKNOWLEDGEMENTS** X

**INTRODUCTION** 1

PRECIS OF CHAPTERS 6

*Chapter 1* 6

*Chapter 2* 7

*Chapter 3* 8

*Chapter 4* 8

*Chapter 5* 9

**CHAPTER 1**

**RE-EXAMINING THE PERCEPTUAL VIEW OF VISUAL IMAGERY** 11

0. INTRODUCTION 11

1. REPRESENTATIONS: DISCURSIVE OR PICTORIAL? 12

2. THE PERCEPTUAL VIEW OF VISUAL IMAGERY 14

*2.1 Kosslyn's Architecture* 14

*2.2 The Perceptual View in Philosophy* 17

3. OBJECTIONS TO THE PERCEPTUAL VIEW 20

*3.1 Correlational Evidence* 20

*3.2 Dissociation Evidence* 22

3.3 Consequences for the Perceptual View 26

4. TOWARDS A NEW VIEW 27

5. CONCLUSION 29

**CHAPTER 2**

**TOWARDS A COGNITIVE ARCHITECTURE OF MEMORY AND IMAGINATION** 30

0. INTRODUCTION 30

1. PRELIMINARY DISTINCTIONS 32

2. THE PRESERVATIVE VIEW AND THE GENERATIVE VIEW OF EPISODIC MEMORY 35

*2.1 The Preservative View* *36*

*2.2 The Generative View* *40*

3. THE CONSTRUCTIVE EPISODIC SIMULATION HYPOTHESIS 42

*3.1 (i) The Construction Claim* *45*

3.1.1 (a) No Dissociation Between Episodic Memory and Imagination 46

3.1.2 (b) Memory and Aging 47

3.1.3 (c) Experimental Data from an Episodic Specificity Induction (ESI) Paradigm 48

*3.2 (ii) The Implementation Claim* *49*

*3.3 Taking Stock* *51*

5. CONCLUSION 52

**CHAPTER 3**

**THE CONSTRUCTIVE EPISODIC SIMULATION HYPOTHESIS +** **54**

0. INTRODUCTION 54

1. CESH+ 55

*1.1 Memory Indices* *56*

*1.2 Retrieval Processes and Memory Traces* *58*

*1.3 Memory Traces* *61*

*1.4 Differences between CESH+ and the Perceptual View* *63*

2. BAYESIAN INFERENCE AND BAYESIAN GENERATION 64

*2.1 Bayesian Generation for Content Selection* *67*

3. IS A BAYESIAN SELECTION MECHANISM COMPUTATIONALLY TRACTABLE? 72

4. CONCLUSION 73

**CHAPTER 4**

**IMAGINATION AS A SKILL** **75**

0. INTRODUCTION 75

1. THE HALLMARKS OF SKILL 76

2. KIND'S CASE FOR IMAGINATION AS A SKILL 79

3. IMPROVING IMAGINATION BY PRACTISING 81

4. CONTROL IN IMAGINATION 84

*4.1 A Theoretical Framework for Control* *85*

5. EMPIRICAL RESEARCH ON CONTROL IN IMAGINATION 87

*5.1 Offline Processing* *88*

*5.2 The Reality Constraint* *90*

*5.3 The Change Constraint* 93

*5.4 Competence or Performance?* 93

*5.5 Is Imagining a Skill?* 94

6. Improving and Controlling Imagination via Epistemic Feelings 95

*6.1 Epistemic Feelings* 96

*6.2 Epistemic Feelings and Imagination* 98

*6.3 The Possibility of Inaccurate Epistemic Feelings* 101

7. Conclusion 103

**CHAPTER 5**

**APHANTASIA: IN SEARCH OF A THEORY** **105**

0. Introduction 105

1. Definitions of Congenital Aphantasia 107

2. Empirical Data on Congenital Aphantasia 109

*2.1 Voluntary Visual Imagery* 110

*2.2 Non-Visual Imagery* 111

*2.3 Involuntary Imagery* 112

*2.4 Memory* 114

*2.5 Atemporal and Future Imagination* 115

*2.6 Spatial Imagery* 116

*2.7 Prosopagnosia* 117

*2.8 Taking Stock* 118

3. Objections to Current Theories 118

*3.1 Nanay's No Conscious Imagery Account* 119

3.1.1 The Account 119

3.1.2 Problems for the Account 121

*3.2 Ventral and Dorsal Streams of Visual Imagery* 123

3.2.1 The account 123

3.2.2 Problems for the Account 125

4. A New Theory 126

5. Conclusion 130

**CONCLUSION** **132**

*1.1 Experiment 1 – Testing the Perceptual View and CESH+ using an fMRI paradigm* 133

*1.2 Experiment 2 – Are erroneous imaginings accompanied by a feeling of error, and can this aid improvement?* 134

*1.3 Experiment 3 – Can aphantasics improve their ability to recall/imagine more details?* 134

**BIBLIOGRAPHY** **136**

# List of Figures


Figure 1. Kosslyn's Cognitive Architecture of Visual Imagery. 15

Figure 2. Schematic overview of de Gelder et al.'s experiment. 25

Figure 3. Memory structures. 33

Figure 4. Baddeley and Hitch's Working Memory Model. 33

Figure 5. A boxological representation of the Constructive Episodic Simulation Hypothesis. 45

Figure 6. A boxological representation of the Constructive Episodic Simulation Hypothesis+. 59

Figure 7. Figures used in a mental rotation task. 82

Figure 8. Trials from Onishi and Baillargeon (2005) false belief experiment with 15-month old infants. 89

Figure 9. Experimental design and results from Fernandez Cruz (2016). 100

Figure 10. Binocular rivalry paradigm and experimental timeline. 121

Figure 11. Graphical depiction of the cognitive processes related to mental imagery in non-aphantasic individuals. 125

Figure 12. A boxological depiction of the Constructive Episodic Simulation Hypothesis+. 126

Figure 13. Venn diagrams depicting possible relations of the voluntary imagery impairment and the involuntary imagery impairment in aphantasia. 128

# List of Tables

TABLE 1. LIKELIHOODS AND PRIORS FOR THE SALLY EXAMPLE. 66

TABLE 2. LIKELIHOODS AND PRIORS FOR THE CROATIA EXAMPLE. 70

TABLE 3. THE JOINT PROBABILITY DISTRIBUTION FOR THE CROATIA EXAMPLE. 70

TABLE 4. DEFINITIONS OF APHANTASIA USED IN THE LITERATURE. 108

# Declaration

I, the author, confirm that the Thesis is my own work. I am aware of the University's Guidance on the Use of Unfair Means (www.sheffield.ac.uk/ssid/unfair-means). This work has not been previously been presented for an award at this, or any other, university.

This work was supported by the Arts & Humanities Research Council (grant number 169547374) through the White Rose College of the Arts & Humanities.

# Acknowledgements

I came to Sheffield to learn how to do cognitive science, but as the cliché goes, I learned so much more than that. Just as cognitive science is a discipline that draws on the expertise in other disciplines, my time at Sheffield has been characterised by an engagement with a huge number of disciplines and research areas way beyond my initial interests. The experiences that I have gained here with the people at and around the university have hugely enriched my time here in Sheffield. I'd like to give thanks to some of those people here.

I am grateful to my primary supervisor, Luca Barlassina, who has guided me through the whole PhD, from the first tentative steps on imagination and affective forecasting, to the final draft of the thesis. It's been highly stimulating to work together, and though I have sometimes felt devastated when my pieces of writing have come back with red comments all over them, the attention to detail and clever objections have always been admirable, and they have definitely made me a better philosopher and scientist. I hope that I have not said (too many) things that are trivially true or obviously false in this thesis.

I am also grateful to my secondary supervisors, Dorothea Debus and Dominic Gregory. I am particularly indebted to them for the stimulating discussions on imagination as a skill, and mental imagery, respectively. I also want to thank Steve Laurence for offering me advice and guidance on teaching, as well as for my desk space in the Hang Seng Centre.

For highly stimulating discussions of cognitive science and philosophy, I thank the Cognitive Science Club at the University of Sheffield, and for contributing to the ideas in my thesis, I would like to thank the audiences of the various conferences at which I have presented my research. I have also had the pleasure of organising many conferences of my own, primarily with Alice Murphy, Jacopo Frascaroli, Deb Marber, and Simon Fokt.

I would also like to thank a number of people in the Philosophy Department at the University of Sheffield. I have made many friends when working both in the attic room and computer room over the years, and I have found the community incredibly supportive, open, and fun. I have loved all the shenanigans we have been up to, from the construction of the meme wall to the chess craze which took hold for months. In no particular order, I would like to thank Emma Bolton, Josh Black, Damiano La Manna, Miklos Kûrthy, Anton Eriksson, David Strohmaier, Will Morgan, Siri Romare, Rosa Vince, Matt Cull, Viktoriia Kononova, Yuhan (Felicity) Fu, Yifan Mei, James Lloyd, Giulia Casini, Maria Pietrini Sanchez, Anna Klieber, Nadia Medhi, James Lewis, Brendan Kelters, Kayleigh Doherty, and Barend Cornelis de Rooij. A particular thank you goes to Robbie Morgan for being such a great palleo, and for the never-ending and generous supply of bourbons. Isela González Vázquez also deserves a special thank you for all the fun times we had in the HRI office (especially the office-furniture-workouts!). And for providing me with studies on rats whenever I needed them, I would like to thank Richard Fitzpatrick and Alex Duval.

Back home, I would like to thank my parents for their unwaning support for me, especially in my academic life – from the time when I started Rosendalsskolan until now, when I'm finishing (what is likely to be) my final degree. Thank you also to Matilda, who has been there for me since we were little and I was too shy to be your friend.

Finally, I thank my partner, Will Hornett, for encouraging me to do 'that cognitive science' and for pointing out that I knew more than I thought about it. I have learned an incredible amount from our discussions about my thesis, and through numerous arguments about functionalism, cognitive architecture, reasons, actions, and whatnot. Thank you for helping me through all the difficult times, especially with adjusting to working from home during the pandemic. I could never have been this productive in our bubble at home without your support.

When I started this journey, I thought I would have two bound copies to hand in in person, that I would get that silly sticker, and I would no doubt go to the pub with all my friends. I did not think that I would submit my thesis from our spare-bedroom-turned-office and have the second COVID jab later that same day. But the most important thing remains; I actually have a thesis of which I am proud. Thank you all for making that possible.

Andrea Blomkvist
16/7/2021
Sheffield, England

# Introduction

In this thesis, I will try to convince you that memory and imagination are underpinned by the same cognitive architecture, and I will detail how this architecture functions. More precisely, my argument applies to *episodic memory* and *sensory imagination*. The former is the type of memory which deals with autobiographical content (Tulving, 1972, 1983) — the kind of memory you use when you remember, say, your last time at the seaside. The latter is the kind of imagination you use to imagine sensory details, like the image of an apple or what an instrument sounds like (Kosslyn, 1994; Goldman, 2006; Schacter and Addis, 2007; Nanay, 2018). The focus is thus not on *semantic memory*, which deals with conceptual knowledge, such as remembering *that* Paris is the capital of France, nor do I focus on *suppositional imagination*, such as imagining *that* Paris is *not* the capital of France (Balcerak Jackson, 2016; Williamson, 2016). My aim is rather to forward the claim that both episodic memory and sensory imagination are underwritten by the same architecture, an architecture which I develop in this thesis and call the 'Constructive Episodic Simulation Hypothesis+' (henceforth: CESH+).[1]

Let me start by giving you an intuitive sense of the proposal. Close your eyes and imagine being at the seaside. Got it? Now, try to remember your last experience of going to the seaside. Got that too? You might have noticed that your imagining and remembering had one thing in common: they both involved similar *mental imagery*. In both cases, you probably had a visual image of the sea, and maybe you could 'hear' the waves crashing in, or even 'taste' the salt in the air. That is, your mental imagery probably consisted of a mix of visual imagery (Kosslyn and Shwartz, 1977; Kosslyn, Thompson and Ganis, 2006; Bartolomeo *et al.*, 2020), auditory imagery (Halpern *et al.*, 2004; Zatorre, Halpern and Bouffard, 2010; Herholz, Halpern and Zatorre, 2012), olfactory imagery (Bensafi and Rouby, 2007; Arshamian *et al.*, 2008; Royet *et al.*, 2013; Young, 2019), gustatory imagery (Olivetti Belardinelli *et al.*, 2009), and maybe also motor imagery (Jeannerod, 1994; Munzert and Lorey, 2013), affective imagery (Blackwell, 2020), and temporal imagery (Viera and Nanay, 2020).

---

[1] Henceforth, I drop the prefixes 'episodic' and 'sensory' where it is clear from the context which kinds of memory and imagination are being discussed.

But so what? We might use mental imagery both in episodic memory and sensory imagination, but that does not show us that the two are products of the same cognitive architecture. Intuitively, the differences seem to outweigh these superficial similarities. Just consider what kinds of capacities memory and imagination are and the uses we put them to. On the face of it, memory is a capacity which encodes and stores conceptual knowledge and information about our *actual* experiences, whereas imagination is a capacity for creating *fantasy* and make-believe. That is, memory appears to be a *preservative* capacity, optimised for encoding, storing, and retrieving content (Martin and Deutscher, 1966; Bernecker, 2010), whereas imagination is a *generative* capacity, optimised not for *re*producing content, but for generating new content (Schacter and Addis 2007; Debus 2014; De Brigard et al. 2013).

Here, I want to convince you of an alternative story: memory and imagination are products of the same architecture, and this architecture is optimised for content *generation*. My proposal – the CESH+ architecture - can be seen as a development and expansion of the so-called Constructive Episodic Simulation Hypothesis (CESH), which posits that both rememberings and imaginings are produced by a semantic retrieval process, an episodic retrieval process, and a recombination process (Schacter and Addis, 2007, 2020; Addis and Schacter, 2008; Addis *et al.*, 2009; Suddendorf, Addis and Corballis, 2011; Addis, 2018). CESH has been developed as part of a recent trend in cognitive science, where psychologists have suggested that imagination and memory are underwritten by the same *episodic system* and enabled by the same processes (Schacter and Addis, 2007, 2020); neuroscientists are reconceiving of the function of neural areas that were thought to mainly be involved in memory recall – such as the hippocampus – as playing a pivotal role in imagination (Rosenbaum *et al.*, 2000; Addis, Wong and Schacter, 2007; Cooper *et al.*, 2011; Addis, 2018); and philosophers have argued that the memory and imagination are mental states which lie on a continuum (Perrin and Michaelian, 2017), as well as argued that episodic memory is generative capacity (Michaelian, 2011; Perrin and Michaelian, 2017). I call this family of theories the 'Generative View'. It is my purpose to defend and expand on the Generative View in cognitive science, by employing these recent advances in the psychology, philosophy, and neuroscience of memory and imagination to develop CESH+.

My argument for CESH+ proceeds as follows. What I do first in this thesis is to cast doubt on the Perceptual View – the received view of how *visual* imagery is produced – in order to pave the way for my new theory. The Perceptual View claims that the production of visual imagery relies on early visual processing (Kosslyn, 1980, 1994; Currie and Ravenscroft, 2002; Goldman, 2006; Kosslyn, Thompson and Ganis, 2006; Nanay, 2010, 2018). I argue against it by showing that one of its crucial predictions is false, and this opens up the possibility for exploring an alternative model, the Generative View. I then argue in favour of a specific version of the Generative View, namely, CESH, and I support it using data from research on amnesia (Klein, Loftus and Kihlstrom, 2002; Rosenbaum *et al.*, 2005), semantic dementia (Irish *et al.*, 2012; Irish and Piguet, 2013; Renoult *et al.*, 2019), and aging (Addis, Roberts and Schacter, 2011; Madore, Gaesser and Schacter, 2014; Madore, Jing and Schacter, 2016). In particular, I argue for (i) the Construction Claim – that episodic memory and sensory imagination are both dependent on the same constructive processes – and (ii) the Implementation Claim – that these processes are largely implemented in the same neural network.

At that point, I show that CESH is sketchy in a number of places: it does not explain how stored content is selected, retrieved, and recombined to form a remembering or imagining. For this reason, I propose CESH+. Firstly, in addition to the processes posited by CESH, CESH+ adds *memory indices*, which store the addresses of locations of content, enabling retrieval processes to 'know' where to retrieve content from. I argue for this addition by appealing to how memory is implemented in computers, and, drawing on neuroscientific research in rodents and humans, I show that this is a plausible proposal for human cognition as well (Teyler and DiScenna 1986; Rudy and O'Reilly 2001; Rudy, Huff, and Matus-Amat 2004; Teyler and Rudy 2007; Langille and Gallistel 2020; Goode et al. 2020). Secondly, on the basis of evidence from cognitive neuroscience, I also add *specialised retrieval processes* which have more fine-grained functions than the ones posited by CESH (Wheeler, Petersen and Buckner, 2000; Wheeler and Buckner, 2003, 2004; Wheeler *et al.*, 2006; Danker and Anderson, 2010). CESH+ posits eight dissociable retrieval processes, such as the visual retrieval process, auditory retrieval process, olfactory retrieval process, and so on. Thirdly, CESH does not tell us *how* certain content is selected for retrieval. That is, it does not specify the procedure for selecting content for retrieval. To rectify this, CESH+ posits a *Bayesian ranking and selection mechanism* (henceforth: Bayesian selection mechanism), which works by Bayesian generation to rank and select content for recombination, given what the agent is intending to remember/imagine. Here, my argument draws on other Bayesian models in cognitive science, which have successfully explained capacities for causal reasoning (Walker and Gopnik, 2013) and word recognition (Norris, 2006).

But despite a fancy model, you still might not believe me. After all, plenty of elaborate models have turned out to be false. So, in order to make my proposal even more convincing, I will demonstrate the explanatory power of CESH+ by showing how it can shed light on two recent problems in cognitive science. The first concerns whether imagination is a skill, and if so, how it *could* be. The domain of skills normally include various bodily actions, such as playing football or playing an instrument, and mental actions, such as performing mental arithmetic (Fridland, 2014; Buskell, 2015). Imagination might seem like an odd candidate to be considered as a skill, especially since it is hard to measure how good someone is at imagining. I will show that in order to determine which capacities are skills, we need to show that the capacities meet two hallmarks of skill: skills are improvable by practice, and skills can be controlled (Stanley and Williamson, 2001, 2017; Dreyfus, 2002; Stanley and Krakauer, 2013; Fridland, 2014; Buskell, 2015; Wu, 2016). I show that imagination can also be characterised thusly. But this still leaves a problem: how could imagination possibly be a skill which can be improved and controlled? Here, I show that CESH+ can give us a computational story of both improvability and control, since we can appeal to changes in the calculations performed by the Bayesian selection mechanism to explain how different content is selected for recombination at different times.

The second problem concerns a condition called 'aphantasia', which is characterised by subjects' impaired mental imagery in both episodic memory and sensory imagination (Zeman, Dewar and Della Sala, 2015; Bainbridge *et al.*, 2020; Dawes *et al.*, 2020; Zeman *et al.*, 2020; Milton *et al.*, 2021). Though research has recently taken off in this area and many studies have collected data on the condition, there has been little attempt to give a cognitive explanation of the condition. That is, theorists have not given much thought to the ways in which the mind malfunctions such as to produce these impairments seen in aphantasia. Two explanations have been put forward, both based on the Perceptual View (Pearson, 2019; Nanay, 2021). I argue that these cannot adequately explain the data from aphantasia, particularly since they do not have the resources to explain aphantasics' impairment in episodic memory, and they also lack the resources to explain the fact that aphantasics are impaired with respect to all kinds of mental imagery, not just visual imagery. I argue instead that CESH+ has the explanatory power to explain all of the data. This is a major contribution to this new research field, and showing that CESH+ can shed light on both of these issues demonstrates the explanatory power of the model.

In sum, my thesis does the following. In the first two chapters, a defence of CESH and the Generative View is offered. I start by arguing against the most prevalent architecture of visual imagery, and present CESH as an alternative. But CESH needs to be improved in a number of places, hence, CESH+ is motivated and introduced. The rest of the thesis tests the explanatory power of CESH+ by looking at imagination as a skill and aphantasia. I conclude that CESH+ represents a promising new architecture of episodic memory and sensory imagination. The precis of chapters (see below) details the aim of each chapter in more detail, the line of argument, as well as how the chapters contribute to the overall aim of the thesis. But before outlining this, I need to highlight two key assumptions of this thesis.

Firstly, following a great number of researchers in cognitive science and philosophy, I will assume a *representational-computational theory of mind* (Fodor, 1975; Dretske, 1981; Pylyshyn, 1984). According to this, the mind contains mental representations, that is, mental particulars that represent possible or actual objects, properties, and states of affairs. These features are captured by the *content* of a mental representation. For example, if one agent believes that it was rainy yesterday, and another agent believes that it was sunny yesterday, then they have mental representations with different contents. Not only can there be variation in mental contents, but mental representations can also take different *formats*, by being either discursive (Pylyshyn, 1973), imagistic (Kosslyn, 1973), or cartographic (Camp, 2007). Moreover, I also assume that the mind works by performing computations over these representations. That is, the mind manipulates representations according to certain specified rules, and this, ultimately, results in behaviour. For example, much like a digital computer or a calculator, the mind can represent the number 2 as '2', perform the operation '+1', and end up representing the answer as '3'. Settling on a particular way of cashing out the computational theory of mind will not be necessary for this thesis, and I believe that what I propose here is compatible with most versions of computationalism.

Secondly, I also assume that a representational-computational theory of mind is best developed when coupled with a multi-level approach to explaining cognition (Marr, 1982; Pylyshyn, 1984). The kind of multi-level approach that I adopt is inspired by the work of Marr and Pylyshyn. According to this kind of view, the highest level of explanation is the *computational level*, which concerns specifying the function of a system. For example, the function of a system could be to encode information. But there are many different ways in which this function can be performed.

The second level, the *algorithmic level*, specifies the structure of the cognitive architecture[2] needed to execute this function, and the processes by which it does so. It is possible that the same function is realised by multiple architectures. For example, the function of solving a mathematical equation might be realised differently in a human compared to in a digital computer. That is, the human and the digital computer might use different algorithms – step-by-step processes – to solve it, such as going from left to right, or starting with multiplication and division. The algorithms used will depend in part on differences in the architectures and the algorithms which the architecture enables. Finally, the third level is the *implementational level*, which specifies the medium that the cognitive architecture is implemented in. In humans, the cognitive architecture realising the function of encoding information is implemented in biological matter, whereas in a digital computer, it is implemented in electrical circuits. These levels can be thought about in isolation from each other (e.g., we can specify the function of a system without necessarily specifying how it is realised), but the levels also importantly constrain each other in various ways. For example, the implementational level can constrain what kinds of computations can be carried out by the system, and the architecture can constrain the possible algorithms. In this thesis, my main focus lies in detailing the cognitive architecture of episodic memory and sensory imagination and the processes carried out in the architecture I posit. At times, I will also touch on both the computational and implementational levels. This kind of approach to the mind has been taken by many cognitive scientists and philosophers who have developed architectures of the mind, such as Anderson's ACT-R theory (1993) Nichols and Stich's Cognitive Theory of Pretense (2000), and Carruthers' Massive Modularity theory (2006). Some of these attempts to give an architecture of the whole mind, whereas others are more limited; my architecture is limited to the episodic system.

## Precis of Chapters

### *Chapter 1*

The aim of this chapter is to argue against the most commonly accepted cognitive architecture of visual imagery, in order to suggest that exploring a new theory is a viable endeavour. The Perceptual Theory, most famously defended by Kosslyn, holds that visual imagery production

[2] Cognitive architectures are also sometimes called 'functional architectures' (Pylyshyn, 1984).

relies on early visual processing, and is implemented in the same neural areas as vision (particularly the primary visual cortex). Versions of this theory are common in neuroscience, but they have also been influential in philosophy, where philosophers such as Currie and Ravenscroft (2002), Goldman (2006), and Nanay (2010, 2021) are either inspired by them (Currie, Ravenscroft, and Goldman), or wholeheartedly endorse them (Nanay). I argue against the Perceptual Theory by expanding and improving upon a recent line of argument suggested by Cavedon-Taylor (*forthcoming a*; *forthcoming b*), where I show that there is a *double* dissociation between visual imagery and vision. This is contrary to the main prediction of the theory, and shows that it is unlikely that visual imagery production relies on early visual processing implemented in V1. This opens up a route to exploring a new theory of visual imagery, the Episodic View, which suggests that visual imagery is more reliant on higher cognitive brain areas, such as the fusiform gyrus, rather than early perceptual areas like V1 (Bartolomeo *et al.*, 2020; Spagna *et al.*, 2021). Interestingly, this account also shifts the focus from *visual* imagery production to *mental* imagery production in general, as the areas it suggests are not dedicated visual areas. Though this is promising, I note that this proposal as it currently stands is sketchy, as it only provides an account on the implementational level. Therefore, chapters 2 and 3 are dedicated to elaborating on the architectural claims.

*Chapter 2*

The aim of this chapter is to argue for the Generative View, which holds that both memory and imagination involve generative processes, and to suggest that CESH is a promising model of the cognitive architecture of memory and imagination. I give some preliminary distinctions between short-term/long-term memory, and episodic/semantic memory, and I home in on episodic memory as the most relevant for my thesis. To introduce the discussion of what cognitive architecture underwrites the production of episodic memory, I first focus on the computational level question of what the function of episodic memory is. I here contrast the Preservative View (Martin and Deutscher, 1966; Bernecker, 2010), which holds that the function is to preserve content, with the Generative View (De Brigard, 2014; Perrin and Michaelian, 2017), which holds that the function is to generate content. On the basis of empirical considerations, I suggest that the Generative View is plausible. Since the function of imagination is widely accepted as generating content, this indicates that the same architecture could underlie both memory and imagination. Moving to the architectural level, I defend CESH as a model of the cognitive architecture. In particular, CESH posits a semantic retrieval process, an episodic retrieval process, and a recombination process, and I show how empirical data supports the architecture. CESH also makes an implementational claim about which neural areas implement these processes, and I show how this too is corroborated by data. Still, despite CESH's initial plausibility as a cognitive architecture of memory and imagination, there are important issues on which it is silent, such as how exactly the retrieval processes operate and how memory traces are stored. Specifying this would make CESH a better model, and this is the project I undertake in chapter 3.

*Chapter 3*

The aim of this chapter is to make modifications to CESH, turning it into CESH+, which will give the theory more explanatory power. I make three modifications backed up by empirical evidence. Firstly, CESH does not detail any mechanisms which direct retrieval processes to where content is stored, thus leaving the question open of how the retrieval processes are able to retrieve content successfully. I add memory indices, an idea borrowed from computer science, where indices store the addresses of information. Similarly, I suggest that our cognitive architecture contains such indices. Secondly, CESH also posits very course-grained retrieval processes, but empirical evidence suggests that retrieval processes should be individuated at a finer grain. I add eight episodic retrieval processes, individuated according to the kind of information they function to retrieve, as well as two spatial retrieval processes. In line with this, I suggest that we should also endorse a distributed view of memory traces. Thirdly, CESH is also sketchy when it comes to how these retrieval processes operate. That is, CESH does not specify the step-by-step procedure which is followed when a retrieval processes retrieves any particular content. I put forward a Bayesian ranking/selection mechanism, and detail the procedure which is followed when content is selected to be retrieved to be recombined into a remembering/imagining. The result, I claim, is that CESH+ is a more explanatorily powerful model than CESH, something we can now put to test.

*Chapter 4*

The aim of this chapter is to argue that imagination is a skill, and to show that CESH+ can elucidate how this can be the case by providing a sub-personal account. To show that imagination is a skill, I focus on two hallmarks which have been universally taken to be indicate skill by both philosophers and psychologists (Anderson, 1982; Dreyfus, 2002; Stanley and Krakauer, 2013; Fridland, 2014; Wu, 2016). These are: (1) improvability by practice; and (2) control. (1) states that skills are improvable by practice, and (2) states that skills can be controlled, both on a personal and sub-personal level. Imagination has recently been argued to be a skill (Kind, 2020a), but I show that the argument is not strong enough to establish this conclusion, as it does not show that imagination meets (1). I provide new empirical data to support this case. To show that imagination also meets (2), I provide a theoretical account of what it would mean for imagination to be controlled on a personal level, where I argue that agents have to employ certain constraints, and I then provide empirical data from developmental psychology to demonstrate that agents indeed do employ these constraints even at an early age. I conclude that imagination is a skill. But there is still a pressing question of *how possibly* imagination could be a skill. That is, it is not clear how a cognitive architecture supports this. Here, I argue that CESH+ provides this architecture. Drawing upon CESH+, I suggest that imaginings are often accompanied by epistemic feelings (Michaelian and Arango-Muñoz, 2014) which play a causal role in changing the likelihood that certain memory content is chosen to be recombined into an imagining. Specifically, I suggest that a feeling of error about an imagining reliably indicates that the imagining is indeed accurate, and that this serves to change the likelihoods such that a more accurate imagining can be produced. Hence, not only is imagination a skill, but CESH+ can also give us a sub-personal account of how it works. This lends supports to CESH+ and the more general Generative View.

*Chapter 5*

The aim of this chapter is to use CESH+ to give a cognitive explanation of aphantasia. Aphantasia is a condition where people, amongst other symptoms, are impaired with respect to generating voluntary visual imagery (Greenberg and Knowlton, 2014; Zeman, Dewar and Della Sala, 2015; Fulford *et al.*, 2018; Jacobs, Schwarzkopf and Silvanto, 2018; Keogh and Pearson, 2018, 2021; Pearson, 2019; Bainbridge *et al.*, 2020; Dawes *et al.*, 2020; Zeman *et al.*, 2020; Milton *et al.*, 2021; Nanay, 2021). I provide an overview of the research to date on aphantasia, and I show that it robustly indicates six data points that need to be accounted for by a cognitive explanation. These are: (1) the impairment in generating voluntary visual imagery; (2) the impairment in generating mental imagery with respect to different sensory systems; (3) the differential impairment in producing voluntary imagery and involuntary imagery; (4) the impairment in recalling episodic memory details; (5) the impairment with respect to content generation for both atemporal events and future events; and (6) the retained ability to solve spatial imagery tasks and score averagely on spatial imagery questionnaires. I then consider two explanations that have been given (Pearson, 2019; Nanay, 2021), both of which are versions of the Perceptual View discussed in chapter 1. For the purposes of assessing the explanations here, I will disregard the arguments against the view I presented in chapter 1, and assess the explanations on their own terms. I find that both views fail to account for (1) – (6), and I argue that CESH+ can provide a complete cognitive explanation of aphantasia. This marks CESH+ out as the only theory which can currently explain all the impairments in aphantasia, and hence, as giving the best explanation of the condition. This fact indicates further that CESH+ is a plausible theory of the cognitive architecture of memory and imagination, and it favours the Generative View of memory.

# Chapter 1

# Re-examining the Perceptual View of Visual Imagery

## 0. Introduction

Experiencing visual imagery is phenomenally similar to seeing; what it is like to experience, say, visual imagery of an apple is similar to what it is like to actually see an apple. But many argue that this is not just a surface level connection between vision and visual imagery. According to the most the most popular theory of the cognitive architecture of visual imagery, the one put forward by Kosslyn (1980; 1994; Kosslyn et al. 2001), visual imagery and vision are produced by the same cognitive architecture, as well as are implemented in the same neural substrates – particularly, the theory claims that visual imagery and vision both rely on visual processing in area V1. This theory has been developed and defended for over 40 years in psychology and neuroscience, and philosophy has since also seen a host of theories derived from, and inspired by it (Currie, 1995; Goldman, 2006; Nanay, 2010). Following Cavedon-Taylor (*forthcoming a*; *forthcoming b*), I call this family of theories in psychology, philosophy, and neuroscience, the 'Perceptual View'.

My aim in this chapter is to cast doubt on this view, and to support a different model where visual imagery instead shares architecture and neural substrates with memory rather than with vision. Variations of this model has been recently put forward in neuroscience (Bartolomeo *et al.*, 2020; Schacter and Addis, 2020; Thorudottir *et al.*, 2020; Spagna *et al.*, 2021), and offers an alternative picture of visual imagery, where it is no longer 'built up' by low-level perceptual processes, but rather heavily reliant on by top-down cognitive processes. In this chapter, I will argue that there is robust empirical evidence to prefer this novel model to the traditional one. This argument is a central steppingstone for my thesis as a whole. Once the Perceptual View is out of the way, I will then argue for an alternative promising account of the cognitive architecture of mental imagery *as a whole*: mental imagery, I propose, is produced by the so-called *episodic system*, which underwrites both memory and imagination, and it is implemented in higher level areas than proposed by the Perceptual View.

The present chapter is structured as follows. Since Kosslyn's theory started out as a view about the format of imagery – Kosslyn famously argues that visual imagery is pictorial – I will briefly introduce the question of format in §1 to provide some relevant background, before moving on to how the theory has developed into also being a theory of cognitive architecture and neural implementation. In §2, I show that Kosslyn's theory is committed to this claim:

i. Visual imagery relies on the same neural substrates as early visual processing

Here, I also introduce the wider family of the Perceptual View within philosophy and I show that many of them, such as Currie and Ravenscroft (2002), Goldman (2006), and Nanay (2010, 2021) also hold (i). In §3, I provide arguments against (i). In particular, by expanding and elaborating upon a recent line of argument developed by Cavedon-Taylor (*forthcoming a*; *forthcoming b*), I show that there is a *double* dissociation between visual imagery and vision. It is thus looks unlikely that visual imagery relies on visual processing in V1, and Kosslyn's proposed architecture has been dealt a serious blow. This opens up a route to exploring a new theory of visual imagery, which I consider in §4. I suggest that the same neuroimaging studies which provide evidence against the Perceptual View, also provide evidence in favour an alternative view, whereby visual imagery is more reliant on higher cognitive brain areas, such as the fusiform gyrus, rather than early perceptual areas, such as V1.

## 1. Representations: Discursive or Pictorial?

Kosslyn's cognitive architecture of visual imagery production started out as a theory about the format of visual imagery, so briefly outlining the format question will be useful for the discussion further on. If one accepts the Representational Theory of Mind (Fodor, 1975; Dretske, 1981; Tye, 1995), a question which quickly arises is what format mental representations take. In the so-called Imagery Debate (for overviews, see: Block (1981) and Nigel (2021)), where this question was disputed, two formats have taken centre stage: the discursive format and the pictorial format.[1]

[1] These formats unhelpfully go under many different names. The discursive format is also known as 'propositional' or 'descriptive'. The pictorial format is also known as 'imagistic' or 'quasi-pictorial' (particularly in Kosslyn's later work). Moreover, sometimes the discursive format is referred to as 'digital' and the pictorial as 'analog'. However, whether something is digital or analog is potentially more a question about how information is *encoded*, and it is not clear that the discursive/pictorial distinction maps onto a digital/analog distinction, as a pictorial format can be encoded digitally (e.g., a photo as a digital JPEG file). More confusing still, there are two senses of analog which are often exploited, though rarely made explicit.

These formats represent in different ways. Roughly, a discursive format represents by using constitutive symbols, i.e., by using the '1' to represent true and '0' to represent false. Language is arguably the best example of a discursive format, where a sentence like 'there are three giraffes in the veld' represents the state of affairs *There are three giraffes in the veld* by different symbols referring to different elements, e.g., 'giraffe' referring to a giraffe. Note that the sentence looks nothing like three giraffes in the veld. This stands in contrast to the pictorial format, where a picture of three giraffes in the veld *mirrors* the represented state of affairs. Pictures obey the Parts Principle, meaning that they depict things by picturing their *parts* (Fodor, 2007). For example, a photo of three giraffes in the veld depicts this scene by having spatially arranged parts depict parts of that scene.[2]

The Imagery Debate is concerned with whether visual representations have a discursive format (Pylyshyn, 1973, 1981, 2002, 2003b, and see also Bartolomeo, 2002b) or a pictorial format (Kosslyn, 1980, 1983; Block, 1983; Farah, 1984; Tye, 1991; Kosslyn, Thompson and Ganis, 2006; Pearson, 2019). Kosslyn has argued that visual imagery is pictorial – or later, *quasi*-pictorial (see §2) – based on behavioural evidence from mental rotation experiments. In these experiments, participants are asked to match a picture of a 3D shape to the same 3D shape which has been rotated, and it is widely thought that people rotate shapes mentally to solve this task.[3] These experiments have found that participants take longer to rotate mental objects a greater angle, and Kosslyn argues that this is best explained by positing that the representational format is pictorial: just like spatial objects take longer to rotate a greater angle, so do representations, since they too have spatial properties (Shepard and Metzler 1971; Park 2019; Kosslyn and Shwartz 1977; Kosslyn, Ball, and Reiser 1978). In later years, Kosslyn has also developed a cognitive architecture, whereby

A representation could be analog if it represents in a continuous way. Another way in which a representation is taken to be analog is by standing in a mirroring relation to what it is representing. A mercury thermometer is analog in both senses, as it represents the temperature continuously, *and* it mirrors the temperature rising by the mercury bar rising (Beck, 2018). An example of an object which represents in an analog way only in the second sense is an analog clock, which mirrors the time by the angle of its hands, but nevertheless moves in discrete steps (Lewis, 1971). For more on the format of representations, see: Beck (2018) and Nigel (2021).

[2] As Fodor (2007) noted, what pictures represent is underdetermined. An image of three giraffes in the veld could be a representation of 'a family of giraffes, 'an odd number of Granny's favourite creatures', 'a number of Granny's favourite odd creatures', and so on, but the sentence 'there are three giraffes in the veld' clearly is not underdetermined in this way.

[3] There are alternative strategies to solving these tasks, for example by counting the number of squares on the shape in order to match it. For alternative strategies and a critique of inferring that mental imagery is used to solve these tasks, see Pylyshyn (2002).

visual imagery relies on early visual processing, and he takes this to further support the claim that the format is pictorial (Kosslyn *et al.*, 2001; Kosslyn, Thompson and Ganis, 2006). I will now turn to discuss this cognitive architecture, its implementation, and the philosophical theories inspired by it.

## 2. The Perceptual View of Visual Imagery

In this section, I will outline Kosslyn's cognitive architecture for visual imagery, which is committed to visual imagery and vision sharing the same architecture, and in particular holds the implementational claim (i) that visual imagery relies on the same neural substrates as early visual processing (Kosslyn, 1973, 1980, 1983, 1994; Kosslyn, Ball and Reiser, 1978; Kosslyn, Thompson and Ganis, 2006). Further, following Cavedon-Taylor ( *forthcoming a*; *forthcoming b*), I show that many theories in philosophy are committed to (i) as well. Simulation Theories, as developed by Currie and Ravenscroft (2002), and Goldman (2006), partially rely on neuroscientific evidence indicating that visual areas are active in visual imagery, though the theories are not necessarily committed to the full architecture that Kosslyn posits. Moreover, Nanay's more recent view of visual imagery is not only reliant on this neural evidence, but also committed to Kosslyn's claims about architecture and implementation (Nanay, 2010, 2018, 2021). This means that any argument against (i) is not just an argument against Kosslyn's view, but against the family of the Perceptual View, which commonly holds that early visual processing is necessary for visual imagery.

### *2.1 Kosslyn's Architecture*

As discussed in the Introduction, the level of the cognitive architecture is generally distinct from the implementation level (Marr, 1982), but it has become increasingly common to develop theories which are committed to an architecture *and* its neural implementation. Kosslyn's theory is such a theory, where he not only posits an architecture where visual imagery and vision are reliant on a common functional mechanism, but he also holds (i) that visual imagery relies on the same neural substrates as early visual processing (Kosslyn, 1994; Kosslyn *et al.*, 2001; Kosslyn, Thompson and Ganis, 2006). In fact, he relies on this claim as evidence for the cognitive architecture, so targeting this implementational claim targets a main reason to support the architecture as well. Let me start by outlining the architecture, and then see how Kosslyn supports it by appealing neuroimaging evidence.

Kosslyn's theory aims to model how vision and visual imagery are generated by appealing to a common architecture. His cognitive architecture includes a *visual buffer*, which is a functional space that is sometimes described as a display surface, and sometimes, more metaphorically, as a blackboard which supports read/write functions. The visual buffer is integral to both vision and visual imagery, as it takes input both from long term memory (when creating imagery) and the eyes (when there is a visual stimulus). Both vision and visual imagery can 'write' on this blackboard. Following input, a quasi-picture is created on the visual buffer, and it can then be inspected by a functional mind's eye, which reads and interprets the visual buffer's surface display, giving rise to a conscious experience. Though a 'functional mind's eye' sounds like a fancy posit, it is really a scanner whose function is to scan the content on the visual buffer to obtain information. The basic architecture is thus relatively simple, with a display that can take input from two sources and a function which can read off this display. It is important to note that this theory makes visual imagery generation a bottom-up process, much like vision. That is, visual imagery an vision both rely on low-level perceptual areas, rather than higher cognitive areas.

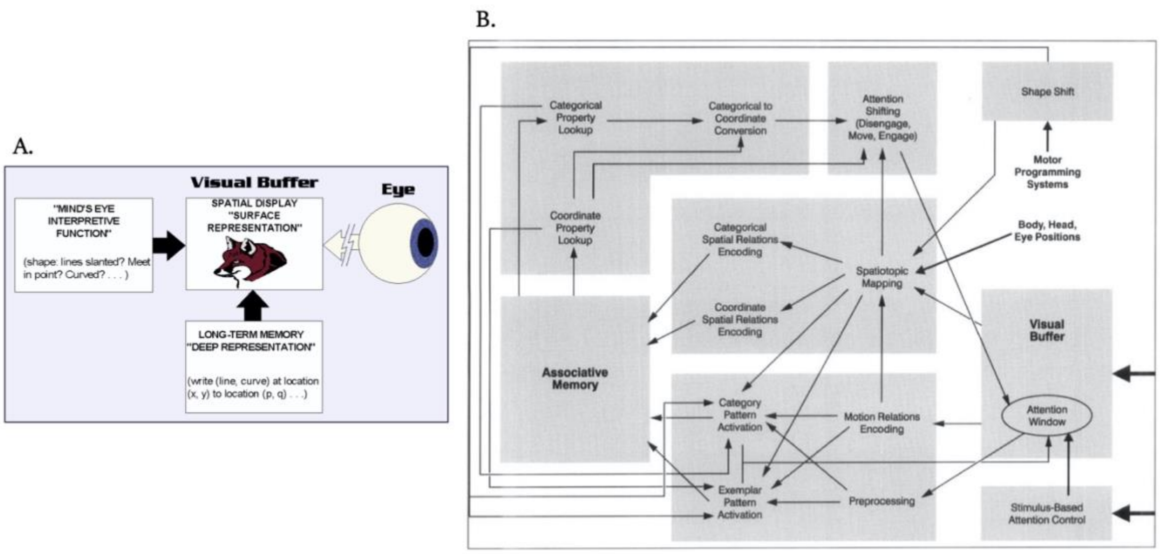


*Figure 1. Kosslyn's Cognitive Architecture of Visual Imagery. A. shows the cognitive architecture proposed by Kosslyn (1980), and B. shows the more elaborate architecture proposed by Kosslyn (1994).*

Kosslyn's notion of the visual buffer as a *functional space* is connected to his original idea of representations being pictorial (Kosslyn, 1980, 1994). Consider for example a picture of a horse. Here, each part of the picture represents a part of the horse (the picture of the horse's ears represents the horse's ears). The parts of the picture thus mirror the properties that the horse actually has, and this is clearly different from a non-pictorial representation (the letters making up the word ‘horse’ do not each represent a part of the horse).[4] But according to Kosslyn, visual images are not *literally* spatial, but they are *functionally* spatial (a visual representation of a banana is, for example, not curved).[5] That is, they operate *as though* they were laid out in space. This is why Kosslyn now calls these representations ‘quasi-pictorial’. More precisely, Kosslyn claims that when a visual image is stored in a firing pattern of neural populations, the firing pattern functions as if they constitute a psychical space.[6] The exact details of how this works need not detain us here, as the focus is on the general cognitive architecture.

Importantly, the posit of the visual buffer is integral to the claim that visual imagery relies on early visual processes. This brings us to Kosslyn’s implementational claim: the buffer is realised in V1 (Kosslyn *et al.*, 2001). Kosslyn singles out V1 as a particular area of importance because of its function in vision and its anatomical structure. Firstly, V1 is the first area which receives visual information from the retina via the optic nerve, and it is responsible for detecting movement, as well as the shape, size, and colour of an object. Secondly, the V1 area is retinotopically organised, meaning that it is spatially structured in a 2D grid. Corresponding to this retinotopic layout, the visual buffer is taken to be composed of multiple arrays, such that visual imagery on the buffer is laid out in a functional space, which can be read off by the functional mind’s eye.

Support for the buffer’s being realised in V1 comes from studies that have shown that early visual areas like V1 are activated both during visual imagery and during vision (Kosslyn, 1994; Kosslyn *et al.*, 2001; Kosslyn, Thompson and Ganis, 2006). Other supporting studies have shown that the content of visual imagery can be decoded from information in V1 because of its retinotopic organisation (Senden *et al.*, 2019), and that TMS interference to V1 impacts visual imagery

---

[4] But, as observed by Beck (2019), this proposal only makes sense if both the represented and the representation have parts, as pictures and horses indeed do. It is not clear in the case of mental representation that the representation has parts; in fact, this is exactly what is disputed, and hence it would be begging the question to assume that the representation has parts that mirrors the parts of the represented.

[5] This claim is related to Block’s (1983) No Seeum Objection. The objection states that visual imagery cannot be pictorial, because if we were to look in the brain, we would not see any pictures. However, given that Kosslyn’s claim now is that visual imagery is only functionally spatial, we should not expect to see anything that looks like a picture in the brain. It is more likely that, say, firing patterns which indicate a spatial representation have to be decoded before we can say that the format is indeed spatial (and therefore that the representation is pictorial). A recent study claims to indeed have decoded activity in V1 (Senden *et al.*, 2019).

[6] Firing patters *storing* information has recently been disputed, see Gallistel (2020).

(Kosslyn *et al.*, 1999). Together, these studies support the claim that (i) visual imagery relies on the same neural substrates as early visual processing. Specifically, as visual imagery is displayed on the visual buffer, it means that visual imagery is realised in V1. Moreover, based on the visual buffer being implemented in V1, Kosslyn also argues that the format for imagery is pictorial. The argument is that V1 is spatially structured, and V1 activation is necessary for visual imagery, therefore, there is strong evidence for thinking that visual imagery is spatially structured, that is, pictorial.

Kosslyn's theory predicts that we should not find any instances of visual imagery without V1 activation, since V1 activation is said to be *necessary* for visual imagery. It should be noted that it is possible to hold a view about an architecture similar to Kosslyn's, but without being committed to where this is implemented, and thus avoid this necessity claim. Possibly, the visual buffer could be realised elsewhere and both vision and visual imagery might rely on it. But the arguments I will forward in the next section casts doubt on this weaker claim too, as I show a double dissociation between vision and visual imagery as to make the architecture including a visual buffer implausible. But before I object to the cognitive architecture presented here, I will show that views similar to Kosslyn's are now widespread within philosophy too.

### *2.2 The Perceptual View in Philosophy*

Interestingly, there are many theories within philosophy inspired by, or very similar to, Kosslyn's theory. Amongst these are the Simulationist Theories, which are advocated both by Currie and Ravenscroft (1995; 2002), and Goldman (Goldman, 2006). Their main claim is that we *simulate* experiences when imagining — for example, what we do when we visually imagine an object is simulating *seeing* the object. More specifically, Goldman's Simulation Claim states that a process P is a simulation of another process P' only if P duplicates, replicates, or resembles P′ in some significant respects (relative to the purposes or function of the task). So, a process that produces visual imagery is a simulation of a process that produces vision only if it duplicates/replicates/resembles the second process in a significant respect, relative to the function of the second process.[7] Proponents of this kind of view often rely on neuroscientific evidence for their claim that the same processes are involved in vision and visual imagery. In fact, both

[7] Variations of the Simulation Claim and how to state it have been extensively discussed elsewhere, see: Barlassina and Gordon (2017).

Goldman and Currie rely on data collected by Kosslyn (1980; 1993) and Farah (1988) indicating shared neural substrates.

It should be noted that Goldman is cautious about just how strong the claim that vision and visual imagery are implemented in the same neural substrates should be, and he rightly points out that Kosslyn has had to weaken this claim in light of evidence showing that, for example, congenitally blind people have visual dream content (Bértolo *et al.*, 2003; Goldman, 2006). Somewhat more puzzling is that Goldman claims that his own Simulation Theory is *neutral* when it comes to cognitive architecture and neural implementation.[8] Now, it is true that Goldman's Simulation Claim, as stated above, is neutral with respect to this. However, the theory that Goldman goes on to develop is *not* neutral, as it relies on empirical evidence about shared neutral substrates for visual imagery and vision. Take away this evidence, and you have also taken away the evidence for the Simulation Claim. That is, we would no longer have a reason to believe Goldman's general Simulation Claim with respect to visual imagery if it were not for this empirical evidence. So, the particular theory that Goldman develops cannot be neutral, even though it is the case that the Simulation Claim itself is formulated in neutral terms. But in contrast to Kosslyn's theory, this means that Goldman's Simulation Theory cannot be falsified by showing that visual imagery and vision are *not* reliant on the same neural substrates. This is because Goldman does not stake his claims on V1 in particular; all his Simulation Theory needs is for a simulation process to be implemented *some*where. This stands in contrast to Kosslyn's theory. Still, showing that the Simulation Claim is false when it comes to vision and visual imagery is problematic for Goldman's view since this is the particular evidence used to support the view.

Nanay's theory is a philosophical view which relies even more heavily on Kosslyn's work, since it endorses Kosslyn's proposals about both architecture and areas of implementation. Nanay claims that mental imagery is perceptual processing that is *not* triggered by the corresponding sensory stimuli in the relevant sense modality.[9] By 'perceptual processing', Nanay means processing in the

---

[8] He also rightly points out that his theory is neutral about the format of visual imagery, and his sympathies lie with Pylyshyn's critique of Kosslyn's pictorialism (Goldman, 2006, pp. 155–156).

[9] The way Nanay's account is phrased raises a worry about multiple realisability (Putnam, 1967). Since Nanay *identifies* mental imagery with perceptual processing, the account entails that mental imagery is only realisable in organisms with a perceptual system. This does not tally well with classical concerns about mental states being realisable in different kind of organisms who might lack perceptual systems, such as Martians (Lewis, 1983). The worry is easily avoided though, by rephrasing the claim in terms of realisability

early sensory cortices, V1, V2, V4/V8, and MT (Nanay, 2010, 2018, 2021).[10] It interestingly departs from philosophical accounts by claiming that imagery is not necessarily consciously experienced, as philosophical views commonly claim (Currie, 1995; Kind, 2001; Kung, 2010). Instead, he follows Kosslyn (1995) and Pearson (2015) in claiming that it can also be *un*conscious, and he proposes that this can help us solve problems in cognitive science, such as providing an explanation for aphantasia. These arguments are not relevant at present, but will be discussed in depth in chapter 5.

The central claim of Nanay's account is that mental imagery relies on perceptual processing in early sensory cortices (Slotnick, Thompson and Kosslyn, 2005; Page, Duhamel and Crognale, 2011, but see also: Bridge *et al.*, 2012). Drawing extensively on the work of Kosslyn, Nanay claims that there is clear evidence for early visual processing's involvement in visual imagery, and indeed that there is great overlap in the neural substrates that realise vision and visual imagery (Nanay, 2016, 2018). He motivates the involvement of V1 in particular by appealing to how the content of visual imagery can be decoded from patterns of activation in this area (Nanay, 2018), taking advantage of BOLD activations in the area as measured by fMRI imaging (Naselaris *et al.*, 2015). Closely mimicking Kosslyn's architecture of the visual buffer, he describes this area as a functional 'blackboard' which is retinotopically organised, and onto which both visual imagery and vision can 'write' (Nanay, 2021).

We can thus see that the Perceptual View are common in philosophy. In fact, Cavedon-Taylor (*forthcoming a*; *forthcoming b*) argues that there are even more philosophical views that hold that there are dependencies between vision and visual imagery, such as Martin's (2002) and Soteriou's (2013) views, which claim that imagery states attempt to mimic perceptual states (Cavedon-Taylor dubs this 'weak perceptualism' as opposed to Kosslyn's 'strong perceptualism'). But as these claims are not concerned with cognitive architecture, I will not discuss them here. Now, I will turn my attention to arguing against the Perceptual View, with a main focus on falsifying the claim that visual imagery is reliant on V1 activation.

---

instead of identity, such that mental processing is realised by perceptual processing, rather than identical to it.

[10] Though Nanay's view concerns *mental* imagery as opposed to only *visual* imagery, Nanay's arguments almost exclusively focus on visual imagery (Nanay, 2010, 2015, 2016, 2021). This discussion is likewise narrowed to his account of visual imagery.

## 3. Objections to the Perceptual View

In this section, I argue against the cognitive architecture proposed by Kosslyn. This also serves as an argument against the philosophical theories discussed above, though to varying degrees depending on how committal they are to which neural areas are necessary for visual imagery. My argument here takes a two-part structure. Firstly, relying on correlational evidence first discussed in philosophy by Cavedon-Taylor (*forthcoming a*; *forthcoming b*), I show that there are non-clinical cases of visual imagery without V1 activation. Secondly, I show that there is causal evidence for a double dissociation between V1 activation and visual imagery. The first side of the double dissociation – cases of no visual imagery, despite V1 being functional – has already been demonstrated by Cavedon-Taylor, and I supplement his argument with further evidence. The argument for the other side of the double dissociation – cases of retained visual imagery despite a dysfunctional V1 – is novel and rests on a recent clinical case. Taken together, this evidence falsifies the claim (i) that visual imagery relies on early visual processing. I also briefly sketch how this could be used as an argument against the claim that the format of visual imagery is pictorial.

### *3.1 Correlational Evidence*

Cavedon-Taylor (*forthcoming a*; *forthcoming b*) briefly discusses a meta-analysis by Spagna et al. (2021), including 46 experiments across 41 fMRI studies, which examined whether early visual areas were activated in visual imagery conditions. Due to its importance, I will recount it more in depth here.

Spagna et al. investigated different conditions across different studies, such as a visual imagery condition vs. a perception condition, a visual imagery condition vs. control condition, and a visual imagery condition vs. a rest condition. The imagery tasks varied across experiments, and included imagining faces of familiar relatives (Boly *et al.*, 2007), mental rotation of 3D objects (Logie *et al.*, 2011), and imagining objects from auditory instructions (Handy *et al.*, 2004; Olivetti Belardinelli *et al.*, 2009). The Perceptual View predicts a differential activation of the early visual areas during visual imagery compared to rest/control, such that the areas show an increase in activation in the former. The theory further predicts that in the visual imagery condition vs. perception condition, the activation in the early visual areas should be similar, as both these activities are hypothesised to rely on early visual areas. Note that there does not need to be a complete overlap of activated areas, but we should expect V1 to be activated in both cases, as V1 activation is hypothesised to be necessary for both vision and visual imagery due to the involvement of the visual buffer.

But the meta-analysis did not carry out these predictions. Instead of early visual areas being more activated in imagery conditions compared to rest/control, results showed greater activation in the left fusiform gyrus, anterior regions of the left temporal lobe, and the medial temporal lobe – these are not visual areas. The visual imagery vs. perception condition showed an increase in activation for visual imagery in the insular cortex, and bilaterally in the supplementary motor area/anterior cingulate cortex. This puts pressure on V1's involvement in visual imagery, as it looks as though visual imagery can be experienced without the involvement of V1.[11]

This finding generates a puzzle, since we have a vast amount of previous evidence that does indeed indicate that V1 plays a (perhaps necessary) role in visual imagery. For example, one oft-cited study shows that TMS interference on V1 impacts visual imagery (Kosslyn *et al.*, 1999). This evidence is in need of an alternative explanation if one is to argue that V1 does not play a necessary role, or no role at all, in visual imagery. How can these two lines of evidence be reconciled? Spagna et al. offer the following explanation: the effect on visual imagery observed in Kosslyn et al. (1999) 'might have resulted from modulation of downstream visual areas' as 'a TMS pulse applied over V1/V2 can stimulate not only local neuronal assemblies, but also excitatory projecting neurons reaching through the entire visual system, and beyond [...] (Parkin, Ekhtiari and Walsh, 2015; Bergmann and Hartwigsen, 2020)'. The effect on visual imagery could thus be due to another area being disrupted, as the TMS pulse is not contained.

Another piece of evidence which needs an alternative explanation is a recent study, which shows that the content of visual imagery can be decoded from V1 activity (Senden *et al.*, 2019). Again, if V1 does not play a role in visual imagery, this is surprising. Here, Spagna et al. argue that since this study is correlational rather than causal, this finding might reflect a non-functional by-product, rather than the true basis of visual imagery. This alternative interpretation is supported by another decoding study, where visual imagery of faces was decoded, and it was found that the occipital regions did not support decoding above chance, whereas decoding from the temporal lobe was supported above chance (VanRullen and Reddy, 2019). TMS and decoding studies are

[11] The study also investigated whether motor imagery depends on the primary motor cortex, and results indicated that it does not.

perhaps the kind of correlational studies which cry out the most for an alternative explanation, but there are many more studies which would need to be accounted for too. However, this is too big a project for this chapter, and since alternative interpretations and further arguments against Kosslyn's theory have already been given by Bartolomeo (Bartolomeo, 2002a, 2002b, 2008), this will not be my focus here.[12]

Further, correlational studies do not tell us about the *causal* significance of different neural areas; as discussed above, it is possible that certain areas are activated as a by-product, rather than because of their causal role in visual imagery. To tell us about the causal role which we need to specify in a cognitive architecture, we need to look at dissociation evidence.

### *3.2 Dissociation Evidence*

Various recent studies together point towards a *double dissociation* between vision and visual imagery. Recall that Kosslyn's theory posits that the visual buffer, realised in V1, is necessary for both visual imagery and visual perception. It thus predicts that if the early visual areas are damaged, both vision and visual imagery should be affected. But this does not seem to be the case, as multiple case studies have shown a dissociation.

One side of this dissociation has already been discussed by Cavedon-Taylor (*forthcoming a*; *forthcoming b*), who argues that there are cases of people who experience no visual imagery, despite having intact vision. In particular, he discusses a case of a patient who retained good visual object recognition abilities, despite being unable to imagine objects (Farah, Levine and Calvanio, 1988), and one case of a patient who, following brain damage, lost the ability to visualise, but showed no impairments in vision (Moro *et al.*, 2008). However, both patients experienced other impairments as well; the patient described by Farah et al. also was impaired with respect to representing relational facts between objects, and the patient described by Moro et al. was

[12] Interestingly, Bartolomeo (2002a) also describes many cases that might now be classified as acquired aphantasia. For example, one patient M.G. had a visual imagery deficit following a left temporal lobe lesion and could not visualise shapes, buildings, or faces (Basso, Bisiach and Luzzatti, 1980), another patient DW had an imagery generation deficit following a left temporal-parietal lesion (Riddoch, 1990), and yet another patient JB could not re-visualise the shape or colour of objects and faces, or the shape of letters (Sirigu and Duhamel, 2001). Common to these cases was that the left temporal lobe was extensively damaged.

impaired with respect to practical reasoning and decision-making. This leaves open a possibility that these other impairments affected their ability to generate visual imagery.

But there are other cases which Cavedon-Taylor do not mention that further strengthen the case for this side of the dissociation (Bartolomeo *et al.*, 1998; Bartolomeo, Bachoud-Lévi and Thiebaut de Schotten, 2014; Thorudottir *et al.*, 2020). One particularly striking piece of evidence comes from a stroke patient, who frequently used visual imagery in his work as an architect, but found himself unable to do so after the stroke (Thorudottir *et al.*, 2020). The stroke mainly affected the left fusiform gyrus and part of the right lingual gyrus, but early visual areas, such as V1, remained intact, and the patient reported no problems with vision (for a discussion of further cases, see: Thorudottir et al. (2020)).[13] Since this patient were not impaired with respect to other functions, there is less doubt that something else could underlie the visual imagery impairment.

But one might think that this kind of evidence is still compatible with the Perceptual View. For example, though Kosslyn claims that V1 activation is necessary for visual imagery, he is not committed to the claim that *only* V1 activation is necessary for visual imagery. Indeed, his cognitive architecture tells us as much when he posits that information needs to be retrieved from memory, making visual imagery dependent on memory processes as well. So, though this dissociation evidence is a blow to the theory, it is not devastating, as the lack of visual imagery can be explained by appealing to other processes being impaired (presumably, which processes these are might differ from case to case). The more damning evidence for the claim that V1 activation is necessary for visual imagery would be evidence that shows visual imagery despite a *dys*functional V1.

Interestingly, there is support for this claim, as studies report stroke patients with bilateral occipital lesions, who have severe visual deficits, but who can nevertheless experience vivid visual imagery (Bartolomeo *et al.*, 1998; Bartolomeo, Bachoud-Lévi and Thiebaut de Schotten, 2014; de

[13] Moreover, a large population has been identified as having aphantasia, a condition where visual imagery is impaired, but vision is spared (Zeman et al. 2010; 2020; Dawes et al. 2020). Though only two fMRI studies have been conducted on this population (Fulford *et al.*, 2018; Milton *et al.*, 2021), it seems implausible that they could be impaired with respect to early visual areas, as vision is intact. This condition is discussed in depth in Chapter 5, so I will leave it aside for now.

Gelder *et al.*, 2015).[14] For example, de Gelder et al. report a case of a man – T.N. – who, following a bilateral lesion in V1, suffered total clinical blindness but could still produce vivid visual imagery.

Since this experiment is of pivotal importance for my argument, I will discuss it in depth. The study involved patient T.N., a 59-year-old man who had suffered two consecutive occipital strokes at the age of 52. The strokes destroyed nearly the complete primary visual cortices, though there was some residual tissue, but this appeared dysfunctional on MRI (Van den Stock, Tamietto, Zhan, et al., 2014). Multimodal magneto-as well as electro-diagnostic investigations were conducted, and did not reveal any evidence for residual functioning of V1, so T.N.'s primary visual cortices thus appear completely dysfunctional. Previous studies have confirmed blindness over the whole visual field, though interestingly, residual vision has been documented for navigation (de Gelder *et al.*, 2008), categorisation of body stimuli (Van den Stock *et al.*, 2014), and facial affect recognition (Pegna, Landis and Khateb, 2008).

de Gelder et al.'s study used an experimental design to test whether V1 was differentially activated during either visual perception or visual imagery tasks compared to a rest condition. In an fMRI scanner, T.N. was presented simultaneously with a visual stimulus and an auditory stimulus. There were two categories for the visual stimulus: scene or face. That is, T.N. was either presented with a picture of a scene (collapsed building) or a face (angry face). There were two visual modalities: the pictures could be presented either as scrambled or unscrambled. At the same time, T.N. was presented with one of two auditory stimuli: angry person imagery (a verbal instruction to imagine an angry person) or tree imagery (a verbal instruction to imagine a tree). For angry person imagery, T.N. was instructed to imagine walking down a street and being approached by an angry man, and for tree imagery, he was instructed to imagine that he was walking towards a bus station when he came across an open plain with a tree in the centre. This produced a 2 × 2 × 2 factorial-blocked design (see Figure 2). Each stimulation block lasted for 18 seconds, during which 4 stimuli were randomly presented, and the verbal instructions were given at the beginning

[14] In a response to Pearson (2019), these cases have been highlighted by Bartolomeo et al. (2020) to argue that the V1 area is not necessary for the production of visual imagery. In his response, Pearson (2020a) questions the methodology of studies, particularly focusing on that there is no quantitative evidence showing patients' vividness ratings of visual imagery before and after a lesion. Despite a lack of quantitative evidence, I believe that the qualitative evidence from patients is strong. These are not cases of a reduction of vividness, for which more exact qualitative evidence would be preferred, rather, they are much more extreme cases where patients claim to now have no vividness *at all*.

of each block. T.N. was instructed to keep his eyes open and look straight ahead, and after each stimulation block, he was instructed to imagine total darkness (rest condition). Two healthy male controls performed two runs of the same tasks, with their eyes closed.

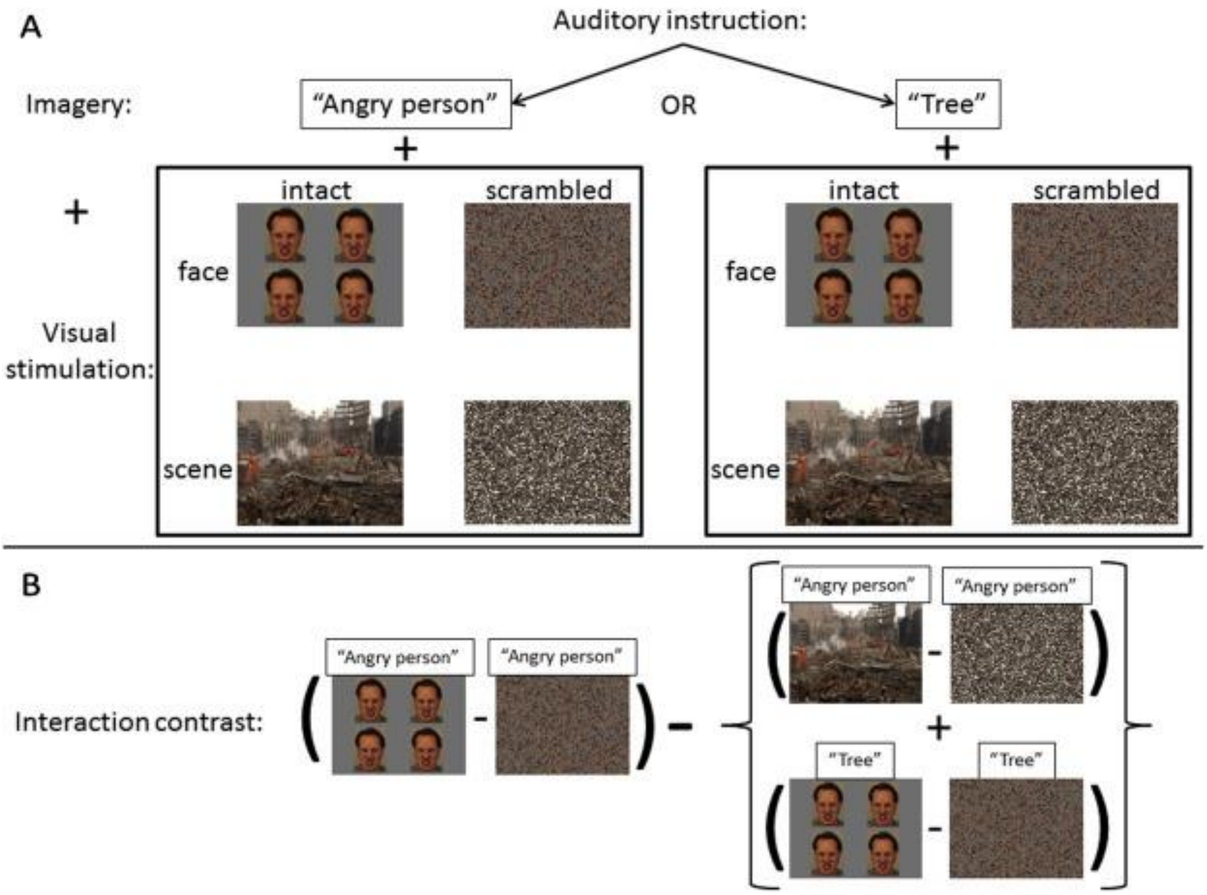


*Figure 2. Schematic overview of de Gelder et al.'s experiment. A shows a schematic overview of the design. T.N. was presented with an auditory instruction (angry person/tree), along with different visual stimuli (face/scene; scrambled/unscrambled). B shows a schematic presentation of the interaction contrast, which was used to test other hypotheses (see footnote 15). Adapted from de Gelder et al. (2015).*

I will not report on all the results here, as only some are relevant to the present argument.[15] The results which are interesting here are the ones that pertain to which brain areas were more activated during imagery, and the researchers calculated results for two different contrasts: active imagery (of tree/person) vs. imagery of darkness, and imagery of an angry person vs. imagery of

[15] The researchers were also interested in testing further hypotheses about residual vision and blindsight, hence the visual stimuli of person/scene and scrambled/unscrambled were included. They were also interested in whether imagery can boost unconscious visual perception, and found evidence to that extent. For these results, see: de Gelder et al. (2015).

a tree. For the first, the left hemisphere activation consisted of fronto-parietal regions, while the right hemisphere activation also included a medial anterior occipital region partly overlapping the inferior part of the parieto-occipital sulcus. In the second case, the brain areas that were more active during imagery of an angry person than a tree were motor and somatosensory areas, and as well as in temporal and occipital areas bilaterally. Further, T.N. verbally reported that he was confident that he experienced visual imagery during both of these conditions, but could not perceive any of the visual stimulus presented. He also reported that he had drawn on autobiographical memories to produce the imagery. Comparisons between T.N. and controls showed imagery activated frontal, parietal and occipital brain regions in all of them. Since V1 was completely dysfunctional in T.N., the results suggest that visual imagery does not rely on V1 activation, whereas conscious vision does.

### *3.3 Consequences for the Perceptual View*

Where does this leave us? Taken together, activation in early visual areas is unlikely to be necessary for visual imagery, as there is both correlational and causal evidence against the hypothesis. Moreover, not only is the implementational claim falsified, but this is also a serious blow against the positing of a visual buffer in the architecture, since such a buffer should not allow for a double dissociation between vision and visual imagery, regardless of where it is implemented. For Kosslyn's theory, it means that the visual buffer can at most be necessary for vision. Note, however, that the evidence does not *exclude* the possibility of any involvement of V1 or early visual processing in *certain cases* of visual imagery, and this is compatible with much of the data amassed by Kosslyn over the years (Kosslyn, Thompson and Ganis, 2006).

One could further use this evidence as an argument against Kosslyn's format claim. Pictorial theories rely on both behavioural evidence and neuroimaging evidence for the claim that visual imagery is pictorial. Behavioural evidence famously concern results from reaction times on mental rotation and scanning tasks (Shepard and Metzler, 1971; Kosslyn, Ball and Reiser, 1978), which Kosslyn claims are best explained by positing that the format of imagery is pictorial, but this has been widely disputed.[16] The neuroimaging data arguably makes for a stronger case in the following

[16] Behavioural evidence for (ii) comes mainly from mental rotation tasks and scanning experiments. In mental rotations tasks, participants are shown a 3D shape, and then asked how this would look were it to be rotated (Shepard and Metzler, 1971). They normally see four additional pictures, one of which correctly represents the rotated first shape. These experiments have found that the further the angle of the shape has

way. Visual imagery is displayed on the visual buffer, realised in V1. Because of V1's retinotopic organisation, it is likely that visual imagery also is pictorial. That is, the fact that V1 is spatially organised makes plausible the claim that the format of visual imagery is also organised spatially (Kosslyn *et al.*, 2001). But if V1 activation is not necessary for visual imagery, this puts pressure on this inference about the format, as it now does not seem plausible that visual imagery inherits its format from the spatial layout of the visual buffer. So, if V1 is not involved in the production of visual imagery, whence the pictorial format?

Of course, this argument does not show that visual imagery is *not* pictorial. It is possible that visual imagery might be realised in some other brain area which might also be organised in such a way as to suggest that visual imagery is pictorial, but until there is evidence to show so, the burden is on the pictorial theorist to argue the case. But since the format of imagery will not be important to the overall argument I make in this thesis, I will leave this issue aside here, and return to the cognitive architecture.

## 4. Towards a New View

The correlational and dissociative evidence deal a serious blow to the Perceptual View, and especially to Kosslyn's theory. It moreover puts pressure on other philosophical views held by Currie and Ravenscroft, Goldman, and Nanay, which rely heavily on the claim about common neural substrates for vision and visual imagery for their theories. This puts us in a position where looking for alternatives is warranted. Here, I will briefly recount the neurological story which is proposed by the same researchers who argue against the Perceptual View. This hints towards the cognitive architecture which I will develop in chapter 2, which claims that imagination and memory share a common architecture, and make the additional implementational claim that there are also many overlapping neural substrates.

---

to be rotated, the slower participants are to pick out the correct rotated shape (reaction time increases). Pictorialists explain this result by appealing to the format of the image: reaction time increases because the represented image also needs to be rotated further. The story is similar in mental scanning experiments, where participants commit a simple map to memory, and are then asked to locate various places on the map. Again, if two places are further apart, reaction time increases (Kosslyn, Ball and Reiser, 1978). An overview and discussion of these experiments can be found in Pylyshyn (2002). Pylyshyn has heavily disputed this evidence, and has proposed, as well as tested, alternative hypotheses, arguing that there are potential confounds (Pylyshyn, 1973, 1979, 1981, 2003a, 2003b). The dispute has not been settled, so whether behavioural evidence supports the pictorial view is unclear.

First, neuroscientists have suggested a new implementational story. Spagna et al. (Spagna *et al.*, 2021) as well as Bartolomeo (Bartolomeo, 2002a, 2002b, 2008; Bartolomeo *et al.*, 2020) suggest a network of brain areas which are involved in visual imagery. Noticeably, these do not necessarily include early visual areas[17], but rather they suggest that that a core network of underwriting visual imagery builds upon a region in the FG4 area of the left fusiform gyrus (Bonino *et al.*, 2015) – they label this the 'Fusiform Imagery Node' – together with regions in the ventral temporal cortex (Mahon and Caramazza, 2011). Areas in the left fusiform gyrus seem to be heavily involved in visual imagery, to the extent that patients experience deficits in visual imagery when this area is damaged, such as the architect who lost his ability to visualise following a lesion in the area (Bartolomeo, 2002b; Thorudottir *et al.*, 2020). They also suggest that attentional control is needed to sustain an episode of visual imagery, and we can see that attentional control networks in the fronto-parietal networks and the cingulo-opercular network have been implicated in multiple studies (Corbetta, 1998; Dosenbach *et al.*, 2008; Rossi *et al.*, 2009; Sheffield *et al.*, 2015; Xuan *et al.*, 2016). Moreover, networks known be active when processing semantic information, such as the anterior regions of the left temporal lobe, are also active during visual imagery. This indicates that semantic memory is drawn on when visualising. Finally, they also suggest that the medial temporal lobe is active during visual imagery, and it might be responsible for recombining elements when creating visual imagery. Recombining elements, as we will see in chapter 2, is also necessary for retrieving episodic memories.

Though this is at present only a sketch of which brain areas are involved in visual imagery, rather than a cognitive architecture, it already represents a significant shift from the Perceptual View. In particular, it no longer conceives of the generation of visual imagery as a bottom-up process, which engages the earliest areas of visual processing in order to build up a full image, similarly to how a representation is built up following visual stimulus. Instead, what is suggested is that higher level cognitive areas are heavily implicated.

Shifting the focus to developing a new theory of the cognitive architecture of visual imagery has many benefits. Firstly, it fosters a shift focus from *visual* imagery to *mental* imagery in general,

---

[17] Though they allow that these could be active in certain cases, like if the subject imagines a simple line. They suggest further research should be done to establish whether early visual processing is involved in such cases (Spagna *et al.*, 2021).

as it does not specify that imagery is built up relying on low-level visual areas. This is welcome, since other kinds of imagery have largely fallen on the wayside in this research, and domain-general processes that might support all kinds of imagery are not well understood. Secondly, as mentioned, recent evidence points to close links between memory and imagination, something which is not very prominent on Kosslyn's model, even though long-term memory is stated as an input into visual memory. Along with other researchers (Schacter and Addis, 2007, 2020), I believe that memory plays a bigger role in imagination and mental imagery than what it has previously been credited for. Thirdly, despite sketching out an argument against the neuroimaging evidence for the pictorial format in this chapter, a cognitive architecture would do well to stay neutral as to the format of mental imagery. Given that the format of mental imagery is extremely hard to infer both from behavioural and neuroimaging studies, pronouncing on the format of mental imagery to me seems premature.

## 5. Conclusion

The aim of this chapter has been to cast doubt on the Perceptual View of the architecture of visual imagery, in order to pave the way for exploring a new theory. I focused particularly on Kosslyn's theory, as this is the most popular theory of the architecture and implementation of visual imagery, both in neuroscience and philosophy. I argued that this theory is problematically committed to (i) visual imagery relying on the same neural substrates as early visual processing. Using correlational and causational evidence, I argued that (i) is false. Particularly, I appealed to evidence which showed a dissociation between visual imagery and processing in early visual areas, demonstrating that visual imagery production is possible even when V1 is completely dysfunctional. Finally, I briefly examined the alternative neurological picture proposed by Kosslyn's opponents, and suggested that this implicates memory processes. However, at present, this is only a story about which brain areas are responsible for the creation of imagery, rather than an architecture which specifies the different functions of the brain areas. As my interest lies in developing a cognitive architecture, this is not sufficient, though it is certainly informative. In the next chapter, I will thus take this one step further and develop an architecture, which respects the implementation claims of the alternative view proposed by neuroscientists.

# Chapter 2

# Towards a Cognitive Architecture of Memory and Imagination

## 0. Introduction

In the last chapter, I discussed the Perceptual View of visual imagery, according to which visual imagery relies on the same architecture, and is implemented in the same neural substrates, as vision. On the basis of my arguments against this view, I recommended starting to build an alternative theory which does not rely so heavily on the architecture of vision. This chapter will start constructing the alternative. As we will see, the theory that I will advocate holds that there is a shared architecture between memory and *mental* imagery in general. In doing so, this chapter also shifts the focus from visual imagery to mental imagery, and I argue that memory and imagery generation are underscored by the same system across the board – the so-called *episodic system*.

To better understand the remit of my proposal, it is important to distinguish between two types of questions concerning memory and imagination. One type of question is metaphysical. For example, how similar is a mental state of imagining to a mental state of remembering? Philosophers have discussed this issue extensively. Some philosophers suggest that these are continuous with each other (Michaelian, 2011; Perrin and Michaelian, 2017), whereas others suggest that they are distinct states (Debus, 2014; Fernandez, 2017; Robins, 2020). For example, Debus (2014) argues that they are distinct mental kinds, as a remembering involves remembering a *particular* event, whereas imagining involves imagining a *general* event. Another metaphysical question concerns what it takes for something to be a memory or an imagining, where many philosophers hold that for something to be a *genuine* memory, it requires the right causal connection to the past event, whereas there is no such causal connection that needs to hold in imagination (Martin and Deutscher, 1966; Bernecker, 2010; Michaelian and Robins, 2018; Michaelian, 2020).

In this thesis, I am *not* concerned with these metaphysical questions. Whatever the answer to these metaphysical questions, there are separate questions relating to the cognitive architecture of memory and imagination, and these are the questions I am concerned with: how are imaginings and rememberings *generated*? In particular, is there is a unified system underlying the formation of both imaginings and rememberings? Further, are these processes realised in the same neural areas? Recently, both philosophers and cognitive scientists have defended the view that there is an episodic system, which underwrites both memory and imagination (Hassabis, Kumaran and Maguire, 2007; Schacter and Addis, 2007, 2020; Perrin and Michaelian, 2017).[1] In this thesis, I endorse this proposal, and the present chapter will lay the foundation for this view, to which I further add original contributions in chapter 3 in order to make it more explanatorily powerful.

The aim of this chapter is to argue that both memory and imagination involve generative processes. This suggests that they rely on the same cognitive architecture, and I will outline a specific theory – the Constructive Episodic Simulation Hypothesis (CESH) – which specifies these processes and their functions, as well as where they are likely to be implemented on a neural level. To achieve this, I proceed as follows. In §1, I start by giving some preliminary distinctions which are necessary to structure the discussion on memory. These cover distinctions between long-term memory and short-term memory, as well as episodic and semantic memory. In §2, I start from the computational level and discuss two general views of the function of memory. The Preservative View[2] hold that the function of the memory system is to *preserve* content in representations called *memory traces*, and that the architecture of memory supports this by encoding, storing, and retrieving these. As the function of imagination is uncontroversially accepted to be to *generate* content, rather than to *preserve* content, memory and imagination are seen as having distinct functions implemented in distinct systems on this view. On the other hand, on the Generative View of memory, whilst memories are encoded, and often said to be stored in memory traces, retrieval processes are taken to *re*construct content, rather than simply retrieving it. Moreover, these are signs of an optimally functioning episodic system, as the function of the system is to generate rather than preserve content. I suggest that the Generative View is plausible, and in §3, I move to the algorithmic level and argue for a particular generative architecture on empirical basis. More specifically, I present evidence for CESH, according to which many of the same retrieval processes are involved in memory and imagination (Schacter and Addis, 2007, 2020). I elaborate on CESH's proposal that episodic memory and imagination employ the same

[1] This is not to say that these arguments are not interrelated. In fact, philosophers argue from evidence about the cognitive architecture to the metaphysical argument (Michaelian, 2011), but it is less clear if one could infer anything about the cognitive architecture from the metaphysics.

[2] The Preservative View is sometimes also called the 'Archival View' and Generative View is called the 'Constructive View'. To the best of my knowledge, this is just a terminological difference.

constructive processes (the Construction Claim) and are implemented in largely the same neural areas (the Implementation Claim), and offer empirical support for both of these claims. But there are also certain questions which CESH is silent on, such as how exactly the retrieval processes operate and how memory traces are stored. These are questions which I will answer in the next chapter, when I develop CESH into CESH+ with these added modifications.

## 1. Preliminary Distinctions

In this thesis, I will primarily be concerned with how so-called episodic rememberings and episodic imaginings are constructed. That is, rememberings and imaginings which are accompanied by imagery. To understand this discussion, it is useful to see how episodic memory fits into the wider structure of memory.

The research on memory is one of the most prospering fields in psychology, and has been so since the very early days of psychology (Ebbinghaus, [1885] 1913). By now, the study of memory has been divided up into many different sub-fields, and different memory sub-systems and brain areas are posited to be responsible for different functions. The most high-level distinction is between long-term memory and short-term memory (see Figure 3). Long-term memory has the capacity to store a large number of items for a long time, whereas short-term memory stores a small number of items for a short amount of time (Atkinson and Shiffrin, 1968). Empirical research indicates that these are indeed two different *systems* with different storages, where items in short-term memory can be moved into long-term memory storage.[3] For example, an agent can learn a telephone number, which is held in short-term memory and could be repeated back at an experimenter. If repeated enough times, the item is transferred into long-term storage. There are plenty of models of short-term memory, which posit internal mechanisms as well as how many items can be stored, probably the most famous of which is Baddeley and Hitch's Working Memory Model (Baddeley and Hitch, 1974; Baddeley, 2000) (see Figure 4). This model posits different

[3] However, some argue that short-term memory is just *reactivated* long-term memory, and that there is in fact only one system (Cowan, 1988, 1999; Oberauer, 2002, 2009). There are strong arguments against this view. For example, Norris (2017) argues that short-term memory and long-term memory have different architectural constraints, and short-term memory as reactivation of long-term memory cannot account for features we see in short-term memory, such as the ordering of items. Along with Norris and others, I believe that short-term memory is not the reactivation of long-term memory, though no arguments in my thesis depend on this claim.

storages for different kinds of information, such as visuospatial information and auditory information.

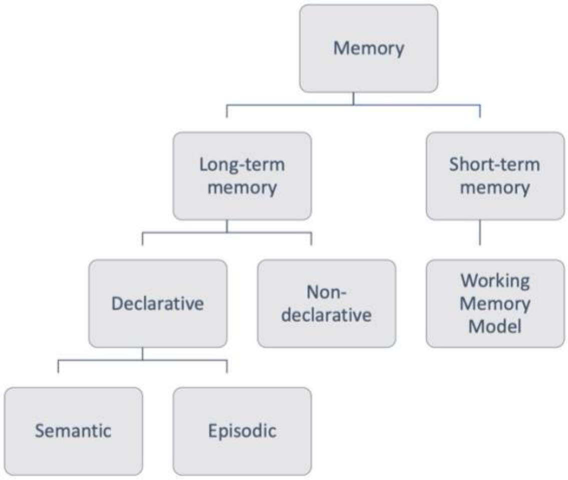


*Figure 3. Memory structures.*

Long-term memory, which I will focus on in this thesis, is further divided into declarative (explicit) and non-declarative (implicit) memory. Declarative memory is generally taken to be memory that can be consciously accessed by an agent, such as what their date of birth is, or what the last hotel they stayed in looked like. Non-declarative memory, on the other hand, is sometimes said to be 'unconscious memory'; it can generally be accessed by showing, but not by saying. For example, an agent can remember procedures, such as how to tie one's shoelaces or how to ride a bike, but they are not necessarily able to say exactly how they accomplish these tasks.

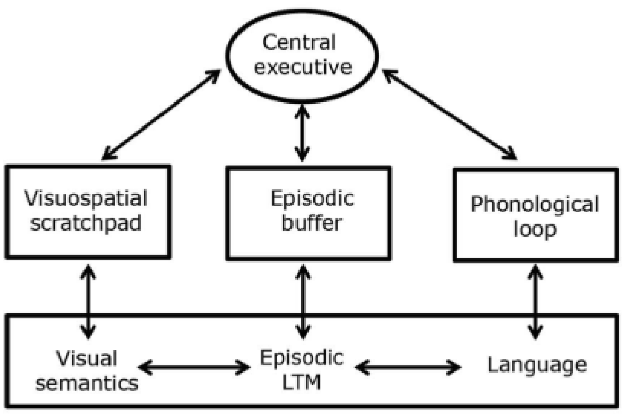


*Figure 4. Baddeley and Hitch's Working Memory Model with the 'episodic buffer', added in 2000 (adapted from Baddeley (2000).*

Finally, there are sub-divisions within declarative memory, and these are the ones I am concerned with in the thesis. Within declarative memory, we find episodic memory and semantic memory. This distinction goes back to Tulving (Tulving, 1972, 1983), who characterised episodic memory as memory for personal events, often accompanied by a sense of reliving (autonoetic consciousness). The information in episodic memory pertains to information concerning the who-what-where-when.[4] For example, an agent can episodically remember their last birthday party, including where it was and who was present, as well as having a sense of reliving this experience when they recall it (e.g., by experiencing vivid visual and auditory imagery). By contrast, semantic memory is memory for facts, including narration and references to other events. For example, an agent might semantically remember that their last birthday party was on the 22nd of January, and the names on the list of invitations.

The semantic/episodic distinction has been seen as a robust one, with research in the two fields forming different traditions and research programmes. But many recent findings point to that, although these are architecturally distinct kinds of memory, there are many points of convergence (Renoult *et al.*, 2019). In §3, I will further discuss both the distinctiveness and the points of convergence in more detail. A small caveat is in place before moving on. Though I am mainly concerned with episodic rememberings/imaginings, as we will see, semantic information is still vital for constructing these. Hence, even if the semantic and episodic processes can be double dissociated, this does not entail that semantic information does not normally go into constructing an episodic remembering/imagining. With preliminary distinctions out of the way, I will now focus my discussion on the generation of episodic memory and its relationship to imagination, as this is what is at stake in recent debates.

[4] Episodic memory has also been investigated in non-human animal cognition. As autonoetic consciousness is a feature which is only verbally reported on by humans, and it has been seen as a necessary component of episodic memory, the research on non-human animals has largely focused on whether they have something *approximating* human episodic memory. So far, there is ample evidence indicating that animals at least posit episodic-*like* memory, where this is taken to encode information such as who-what-where-when (Clayton and Dickinson, 1998; Suddendorf and Corballis, 2007).

## 2. The Preservative View and the Generative View of Episodic Memory

In this section, I will discuss two views of episodic memory, which are found in both philosophy and cognitive science, and differ on what function they take episodic memory to have. From considering the computational function of memory, I will then discuss which system enables this function in §3. The Preservative View is sometimes called a 'replay' view, as it holds that memory is similar to an event being recorded as it is experienced, and then 'replayed' during recall (Schacter and Addis, 2007). On the Preservative View, the function of memory is to encode events, more or less as they happened, into a memory trace – a representation of the event. This is archived, and if nothing malfunctions, an agent should then be able to retrieve that the memory trace, and 'relive' the experience. If this is the case, and this is indeed the function of episodic memory, it follows that any time this does not happen smoothly (if there are memory errors), it signals a *mal*functioning of the system.

The Generative View disagree with memory recall as 'replaying'. Here, an experienced event is encoded and stored in some way, but has to be re-assembled during recalled. The Generative View locates episodic memory within a larger *episodic system*, and argue that the function of this system is to *generate* representations of events. Though many agree that there is some kind of memory trace (for example, see: Michaelian, 2011), they disagree about how this is retrieved. In contrast to Preservationists, they hold that processes operate to generate and transform the representational content at the stage of retrieval. Here, a memory error when retrieving content does not actually signal a malfunctioning of the system, but rather, as we will see, it signals that the system functions as it optimally should. There is thus no simple notion of 'replaying' on this view, as retrieval involves reassembling content, rather than just retrieving and 'replaying' it. In this section, I find the suggestion that memory is generative as more plausible, and I suggest that we should consider the possible architecture with this function in mind. This is a crucial steppingstone for developing my cognitive architecture of memory and imagination, since the Generative View of episodic memory highlight the close architectural and functional relations between memory and imagination.

Before getting to the disagreements, there is one point which Preservationists and Generationists generally agree one, namely, that there has to be some kind of memory trace. That is, once an event is experienced, an encoding process forms a memory trace, which contains content reflecting the event. There are major questions concerning how to characterise a memory trace.

For example, there is an implementational question about where exactly memory traces are stored, where they could either be *local* or *distributed*. A local picture was preferred by early theorists, like the Preservationists Martin and Deutscher (1966), who argued that a memory trace is stored locally (i.e., all in one place). Consider an event such as a birthday party. Storing the representation of the birthday party *locally* means storing all the details of it in one place – in a particular box in the mind, if you will. This fits well with a picture on which a retrieval process is simply searching for a memory trace to retrieve it. As all details are stored in one place, there is no risk of them getting jumbled and causing a memory error. Though content cannot get jumbled, it can still decay, and some believe that this causes forgetting (Robins, 2016a).[5] But many, especially Generativists, prefer a distributed picture nowadays, whereby bits of traces are distributed around a neural network (Michaelian, 2011). These views have a connectionist flavour and, as such, the bits of traces are conceived of as different weights in a neural network. For example, the memory trace of a birthday party might not be all stored in the same place, but rather <cake> could be stored in one node, and <seeing friends> in another (Sutton, 2007).[6] To use the box analogy again, the content of the trace is stored in different boxes in the mind, and a retrieval process would need to search all these boxes and retrieve the right content in order to retrieve the memory trace. On this kind of distributed view of memory traces, recalling a memory consists in the network of nodes being re-activated, whereas on the local view, only the one memory trace needs to be reactivated (see Robins (2016b) for a discussion of local vs. distributed traces). As we will see in chapter 3, I will endorse a distributed view of memory traces, but I will deny that these are implemented in a connectionist network, and instead suggest a view compatible with classical computationalism. Now, let us get to the main disagreement between Preservationists and Generationists.

### *2.1 The Preservative View*

According to the Preservative View, memory is a preservative capacity that stores discrete representations of particular past events. The function of memory is to preserve these

[5] A caveat is in place here. The debate about trace decay seems to be limited to whether trace decay could be a cause of forgetting in *short*-term memory (and even this is debated, see e.g., Neath and Nairne (1995) and Berman, Jonides and Lewis (2009)), whereas the consensus seems to be that it cannot be a cause of forgetting in *long*-term memory. Since I focus on long-term memory in this thesis, trace decay seems of limited importance.

[6] This is probably even too crude, and the bits of traces might even be smaller than this, but the examples only serve to illustrate the point.

representations. Most accounts hold that it does so by creating a link to the past in the form of a memory trace (Martin and Deutscher, 1966; Bernecker, 2010). When an agent experiences an event, such as going to a birthday party, this event is encoded into a representation in the form of a memory trace. The memory trace holds the content of the memory, such as a visual content of the what the birthday cake looked like and an gustatory content of what the cake tasted like. This memory trace is stored and its content preserved, until the agent retrieves it. Retrieving a memory trace can be thought of as activating a search process, whereby some search mechanism rummages through the agents' memory traces, until it finds the right one, which is then promptly retrieved.[7]

This stands in stark contrast to imagination, which is widely agreed to be a generative process, whose function is to generate new representations by recombining content (Schacter and Addis, 2007, 2020; Suddendorf and Corballis, 2007; De Brigard, 2014).[8] In contrast to remembering, imagining is not a processes which works by simply retrieving content stored in memory traces. If so, we would find that we could not imagine anything we had not yet experienced, but this has been disconfirmed by showing that previously experienced content can be flexibly recombined in imagination (Addis *et al.*, 2009). So, if one holds a Preservative View of memory, but believes that imagination is generative, it does not seem likely that the same system could subserve both of these functions. Most likely, a separate system would need to be responsible for imagination, where the function of this system would be something more akin to generating novel content, rather than preserving content. So, if one holds the Preservative View of memory, it is likely that one needs to develop a distinct architecture for imagination, where the outputs from memory only feed into an imagination system, but the systems underscoring imagination and memory are distinct.

I do not find Preservative Views persuasive, because of their problem in accounting for memory errors, such as misrememberings. Since Preservationists hold that the function of memory is to

---

[7] Of course, the search mechanism could search in an orderly way. It could for example go through traces in the order they were encoded, or by 'looking at' the most recent trace first (in computer science, these ways of searching are referred to as 'queue searching' and 'stack searching'). The logistics of the search mechanism need not detain us here as these arguments do not hinge on it, but a thorough account of how content is searched for and selected for retrieval is provided in chapter 3.

[8] Imagination has long since been conceived of as a process which takes building blocks from the content of memory, and recombines them in novel ways into imaginings. For an early philosophical account, see: Hume (1738 [1975]).

preserve content in memory traces, it follows that a misremembering should be characterised as a *mal*function of the system. However, as pointed out by Robins (2016a), this explanation is unsatisfactory. Let me explain. Robins describes how, in the Deese-Roediger-McDermott (DRM) paradigm, participants study a list of related or unrelated words, and are later presented with a list of either the same or different words. They are asked to indicate which words they remember from earlier, and it has been found that participants in fact misremember words in a highly systematic way (Roediger and McDermott, 1995; for an overview, see: Gallo, 2013). In particular, when presented with a word they have not previously seen but which is systematically related to a word they have seen (e.g., 'nurse', if they have seen 'doctor'), they report remembering it. Now, there are four different explanations available for Preservationists for these experimental results, which Robins discusses. Firstly, one could put the effect down to participants guessing what was on the original list. But, as pointed out by Robins, this possibility has been controlled for extensively, as experiments often require participants to indicate whether they 'remember' or 'know' (Roediger and McDermott, 1995; Dewhurst and Farrand, 2004). They are instructed to answer that they 'remember' an answer if they can recall specific, vivid details about the item, and to respond that they 'know' if they are guessing. But participants often claim to 'remember' items that were not on the original list, even citing details about the item (Dewhurst and Farrand, 2004). This goes against the interpretation that they are simply guessing.[9] Secondly, a Preservationist could argue that the DRM effect is a results of a malfunction in the search process responsible for retrieving content. Again, this seems unlikely as we should then expect an overall decrease in recognition performance, rather than a selective decrease on particular items which are related to the original items presented. Thirdly, the DRM effect might be explained as an effect of trace decay. Preservationists allow that not all information has to get encoded into a memory trace, and that traces can decay over time, resulting in forgetting. However, what we see in DRM effects is not easily characterised as forgetting, as it does not involve *under*-reporting on content, as one would expect from forgetting. Instead, it involved *over*-reporting on content, and generating details about items that were not present. This is not easily reconcilable with trace decay. Finally, Preservationists could claim that the DRM effect is simply the result of noise and that it reflects a random variance in the retrieval process. Unfortunately, this too does not seem likely, since the DRM effect shows a very systematic error, and the noise explanation cannot

[9] Further, Robins rightly argues that if participants were simply guessing, we should expect the rate of recognition to be indistinguishable from chance. But participants are more likely to recognise items that were on the list, and less likely to recognise items that were not on the list – this reflects at least a partial reliance on memory.

account for this systematicity. Hence, Robins concludes, and I agree, that being unable to explain the DRM effect is a serious limitation for the Preservative View.[10]

The problem of accounting for memory errors highlights a problem with the posited function of memory, as it entails that our memory system works sub-optimally *most of the time*. How so? One might think that memory errors are uncommon, but in fact we frequently exhibit various memory errors, such as false recognition and the retrospective bias. False recognition can arise from misleading post-event suggestions, leading the changes in the reported memory content (Roediger and McDermott, 1995), for example if a leading question is used. The retrospective bias can be seen in cases where memory details are updated to conform to the agent's current beliefs (Levine, 1997). For example, in light of hearing who won the election, a person might claim that they always believed that that candidate would win, despite their prior conviction.[11] A more extreme memory error is confabulation, which has been widely studied since Loftus and Palmer's pioneering study in 1974 (Loftus and Palmer, 1974), where subjects were induced with a false memory by leading questions. Subjects observed a video of a car crash, and even though there was no broken glass at the scene, many subjects reported seeing broken glass when queried on it. Another famous example of confabulation are the Hyman et al. studies (1995; 1996), where experimenters induced false memories of childhood experiences in subjects.

The point is that memory errors are widespread, and occur both in everyday and extraordinary cases. A Preservationist would have to say that the memory system malfunctions in *all* of these cases. But this seems deeply undesirable from a theoretical perspective. A generally accepted tenet in cognitive science is that we should assume that systems function optimally, unless otherwise proven (Norris and Cutler, 2021). Here, having assumed the function of memory is to preserve content, we end up claiming that it malfunctions most of the time. Now, the Preservationist might just bite the bullet and claim that it already has been proven that the system function sub-

[10] Robins also argues that Generativists also have problems accounting for the DRM effect. However, the reason for Generationists' problem is not to do with their posited function of memory, but rather with the limits of the architecture. That is, it is not a computational level problem. Since my focus here is on function, I will not go over the objection to the Generative View. Robins suggest a hybrid way forward, whereby one takes a preservative approach to the retention of information, but a constructive approach to retrieval. As we will see in chapter 3, I am also sympathetic to this kind of view; it is similar to what I am suggesting for CESH+, where I will argue that there are indeed memory traces, but that content is flexibly recombined upon retrieval.

[11] For a good overview of different kinds of memory errors, see: Hyman and Loftus (1998).

optimally. But I think this is an undesirable conclusion, and that we can avoid it if we instead reconsider the function of memory. Doing so has led to new ideas about the function of memory, which brings me to the Generative View.

### *2.2 The Generative View*

The Generative View of memory has been popular in psychology for some time (Bartlett, 1932; Schacter, Norman and Koutstaal, 1998; Schacter and Slotnick, 2004), and are getting increasingly popular in philosophy (Sutton, 2007; Michaelian, 2011; De Brigard, 2014). According to this view, the function of episodic memory is not to preserve content encoded in representations. Instead, the function of memory is to provide a database from which representations of *potential future scenarios* can be constructed (Schacter, Addis and Buckner, 2007; Suddendorf and Corballis, 2007; Suddendorf, Addis and Corballis, 2011). In fact, the memory system is reconceptualised into a more general *episodic system*, which functions to construct representations of both the past and the future (De Brigard, 2014). This is suggested to be evolutionarily advantageous to organisms as it allows flexible planning for the future, which increases chances of survival (Suddendorf and Corballis, 2007). Indeed, a system that produces representations that help an organism plan for the future seems more useful than a system that simply preserves past event representations.

Now, if the function of the system is not preservation, it also makes little sense that the various processes involved in memory should be designed to optimise for preservation. Taking the Generativist View, whereby the function of episodic memory is to generate content, we should expect them to instead be optimised for *generation*. But this does not mean that Generative theorists throw out the idea of encoding, storage, and retrieval; they do not. But contrary to Preservative Views, all these processes can see content added, modified, or subtracted, all whilst working optimally. To further elaborate on the idea of generative memory, I will focus on the details of a particular representative philosophical account, namely, Michaelian's account (Michaelian, 2011, 2016; Perrin and Michaelian, 2017).

Because of the shift in function, Michaelian's account differs from the Preservative View in the processes involved in both encoding and retrieval, such that multiple processes can affect the encoding and retrieval of a memory trace, and in that this trace is conceived of as gist-like. Let me explain these claims in turn. Firstly, multiple processes affect the encoding of content into a memory trace. These include selection, abstraction, interpretation, and integration processes. The selection and abstraction processes both *reduce* the content that goes into a memory trace compared to the experience as lived. Consider encoding a memory of a birthday party. The selection process selects certain features to be encoded into a memory trace, such as the features that might be most useful for an agent to remember (Roediger and McDermott, 1995; Koriat, Goldsmith and Pansky, 2000). Hence, even if the agent saw exactly how many glasses there were on the table, this information might not be selected for encoding. The abstraction process likewise abstracts information and encodes content that is more general, rather than more specific, and so particular details of the birthday party might be omitted. Instead, other details might be *added* by the interpretation process, which interprets the information in light of the agent's prior beliefs and memories (Intraub, Bender and Mangels, 1992). Finally, the information needs to be integrated by the integration process into a holistic representation (Rhodes, 1996).[12]

Content can also be modified at other stages in the process, such as during consolidation, retrieval, and reconsolidation. The process of consolidation serves to stabilise the content, and it can take up to several years to do so (Ambrogi Lorenzini *et al.*, 1999). During this time, the content is malleable and can change or be lost entirely. Content can also be added when a memory trace is retrieved, due to the trace being reconstructed upon retrieval. Indeed, the cases of false recognition, the retrospective bias, and confabulation discussed above are all examples of content being added during retrieval. After the retrieval of content, the memory trace needs to be reconsolidated, and content can be altered here too.[13]

Secondly, all of these processes contribute to a memory trace being *gist*-like, such that the memory trace represents only some details or the general gist of an event. This is not a new idea in memory research; it dates back to Bartlett's (1932) Schema Theory, which claims that we store information in general schemas. According to this, instead of storing precise details of an event (or person/place/etc.), we store more general details of what an event is typically like. An oft-mentioned example is the so-called 'office schema', which is a representation of items that are generally present in a traditional office (desk, chair, books, ruler, etc.). If memory traces indeed stored as schemas or gists, we should expect that items that are commensurate with this

---

[12] This does not entail that the representation is locally stored. It is compatible with a distributed implementation of the representation.

[13] For a comprehensive discussion on memory consolidation and reconsolidation, see Albertini (2013). Interestingly, Albertini et al. suggest that the process of reconsolidation are adaptive, as it can serve to strengthen a memory trace as well as incorporate new information into it.

schema/gist are more frequently remembered, whereas items which are not are more frequently forgotten. Indeed, this result has been obtained in a study which tested participants' memory accuracy for items present in offices (Brewer and Treyens, 1981). More recently, a number of empirical studies have also supported that this is indeed the way in which memory traces are encoded (Schacter and Addis, 2007). For example, results from DRM experiments (as discussed above), where a subject claims to recognise a word that has not been presented to them, but is thematically related to words they have seen, can be explained by subjects encoding the gist of the word, rather than the exact word. On the basis of these experiments, it has been argued that storing a gist-like representation is more economical than storing a complete representation, and as such, it seems preferable from an evolutionary perspective.[14]

Now, if the function of memory is to generate representations rather than preserve them, we can construct a view on which memory is significantly more similar to imagination. I think this is a viable option, and in the rest of the thesis, I will develop the architecture which I believe underwrites both memory and imagination. I call the system underwriting it, 'the episodic system'. Moving forward, I will offer arguments for this view, and discuss how both imaginings and remembering are constructed according to this view. In contrast to Michaelian, I give a more in-depth picture of the processes involved in retrieval.

## 3. The Constructive Episodic Simulation Hypothesis

With the function of the episodic system posited as the generation of representations of events, I will not move to the architectural level and suggest how this function is fulfilled. In this section, I will present empirical evidence for the cognitive architecture I develop in this thesis: the Constructive Episodic Simulation Hypothesis (CESH). This section will present the basic tenets of CESH, and I modify the model in the next chapter. The model follows in the general tradition of thinking that memory is constructive (Tulving, 1983; Suddendorf and Corballis, 1997; Dudai and

[14] One might worry about what counts as a genuine memory, as opposed to a misremembering or mere imagining, on the Generative View. Preservative Views also commonly hold that a memory trace must stand in the right causal connection to the real event in order to count as a genuine memory. Perhaps surprisingly, this is also something which advocates of the Generative View can get behind. For example, Michaelian (2011) suggest that the causal chain must go via a memory trace, but also through a properly functioning memory system, where this is cashed out in reliabilist terms. Since, on his view, a memory system which functions in a constructive way is a properly functioning memory system, such a memory system reliably produces genuine memories. The causal claim is thus compatible with the Generative View.

Carruthers, 2005; Buckner and Carroll, 2007), and the particular claims for CESH have been developed for some time by various researchers and most extensively by Schacter and Addis (Schacter and Addis, 2007, 2020; Schacter, Addis and Buckner, 2007; Addis *et al.*, 2009; Devitt, Addis and Schacter, 2017). CESH holds that the episodic system is responsible for generating memories and imaginings, both semantic and episodic. It posits various processes which are all, to different extents, involved in these generative processes. There are two claims that are essential for the model; (i) the Construction Claim, and (ii) the Implementation Claim. The first is architectural; it is a claim about the processes which support memory and imagination. The second is a claim about where these are implemented. This section presents empirical support for both of them, and lays the foundation for the substantial additions I make to the theory in chapter 3. As I will point out along the way, there are many details which are glossed over in the current version of CESH, and I contribute to filling in the gaps in the theory in chapter 3. This will make for a more robust defence both for CESH and the Generative View.

CESH claims that both episodic memory and imagination are constructive, and largely rely on the same processes. Stored information is retrieved and integrated into a fully-fledged representation containing all the relevant content. This is enabled by various processes which are involved both in retrieving a memory and in producing an imagining. These are *the semantic retrieval process, the episodic retrieval process,* and *the recombination process and reintegration process* (see Figure 3).[15]

Let me give an overview of the theory. Recall that there is a distinction between semantic and episodic memory. CESH has it that there is a semantic storage where semantic memories are stored, and an episodic storage, where episodic memories are stored.[16] To retrieve content from either of these, retrieval processes are recruited. A *semantic retrieval process* retrieves information from the semantic storage, and an *episodic retrieval process* retrieves information

---

[15] This is not intended as an exclusive list. It is likely that further research will discover more processes, and I argue for some in chapter 3. Note also that it is unclear from how CESH is stated whether the recombination and reintegration process should be conceived of as one process or two. As it does not matter to my arguments in this thesis, I will conceive of it as one process, and use 'the recombination process' for short.

[16] A possible alternative would be that there is one storage for all kinds of information, but different retrieval processes. Since the main tenets of CESH concern how information is retrieved and recombined, I do not see the possibility of a joint storage as a major threat to the theory. In my diagrams, I conceptualise them as different storages.

from the episodic storage. The semantic retrieval process is thus responsible for retrieving semantic content, such as the names of capital cities in Europe, and the episodic retrieval process is responsible for retrieving episodic information, such as visual information about what someone looks like.

The retrieval processes are also involved when generating imaginings. When we imagine, like when we remember, the content needs to be retrieved from somewhere, and the theory posits that it is retrieved from the same place by the same processes. But there are some differences is the relative reliance on the different retrieval processes; imaginings are thought to depend to a greater extent on the semantic retrieval process than do episodic memories, possibly because semantic information serves as a 'scaffold' for an imagining, which is then filled out with further episodic information (Irish *et al.*, 2012). The proposal for how imaginings are generated is hence that semantic schemas are retrieved, and supplemented with particular episodic details. For example, if imagining visiting a castle, a general schema of what a castle looks like would be retrieved, but the imagining could be filled out with particular details, such as details from one's own life about what clothes one would wear when visiting the castle. So, episodic imagination relies to a *high degree* on the semantic retrieval process, whereas episodic memory relies to a *low degree* on the semantic retrieval process. That is, episodic imagination typically recruits both semantic and episodic information using the episodic and semantic retrieval processes to a high degree, and then combines this information into an imagining. Episodic memory, on the other hand, recruits the episodic retrieval process to a high degree, but the semantic retrieval process to a low degree, and then recombines the information into an episodic memory.[17]

Despite some differences, both imaginings and rememberings require that the content that is retrieved is recombined. This is the responsibility of the (re)combination and (re)integration processes, which serve to (re)combine and (re)integrate information. It is not clear from the account whether these are actually to be seen as distinct processes, and if so, which operates on the content first. For this reason, I will conceptualise them as one process which both (re)combines and (re)integrates content for now (see Figure 5). What is clear is that this process takes its input from the retrieval processes, that is, it takes semantic and episodic content as input.

---

[17] Note that this is how it works in normal cases. In cases of semantic or episodic amnesia (Klein, Loftus and Kihlstrom, 2002; Rosenbaum *et al.*, 2005), subjects might rely wholly on one of the processes, as discussed in §3.1.1.

Depending on whether an imagining or remembering is created, the process needs to recombine elements, or combine elements for the first time. For example, if an agent remembers eating a birthday cake, visual information concerning the birthday cake might be integrated with gustatory information of what the cake tasted like. If imagining, on the other hand, information might be combined for the first time, such as if an agent imagines eating a soap flavoured birthday cake. Presumably, this is not something which has been experienced before, and information regarding the look of the cake and the flavour would need to be combined for the first time. Schacter and Addis hypothesise that the recombination process is more active in generating imaginings, as novel imaginings require binding content together for the first time, whereas for an episodic memory, relations between different bits of content need to be reinstated, which presumably is less effortful (Schacter and Addis, 2020).

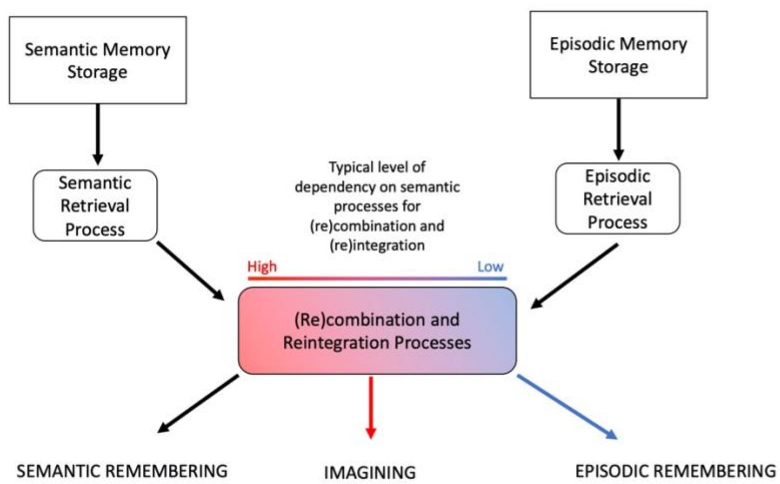


*Figure 5. A boxological representation of the Constructive Episodic Simulation Hypothesis* (Schacter and Addis, 2007, 2020).

### *3.1 (i) The Construction Claim*

In order to defend the theory, I will show how some crucial predictions it makes are carried out by data. Firstly, the theory predicts that there should *not* be a dissociation between episodic memory and episodic imagination, if they depend on the same processes. To show that this prediction is carried out, I consider (a) clinical evidence from patient populations. Secondly, the theory predicts a double dissociation between the episodic retrieval process and the semantic retrieval process, such that one can be impaired without the other being impaired. To show that this prediction is carried out, I also rely on the data in (a), but further consider (b) research on aging effects on memory, and (c) experimental data obtained in an *episodic specificity induction* (ESI) paradigm.

*3.1.1 (a) No Dissociation Between Episodic Memory and Imagination*

A major prediction of CESH is that impairments in episodic memory should be accompanied by impairments in episodic imagination, as both are hypothesised to depend on the episodic retrieval process. Various patient populations have been studied to see whether this is the case, and there are indeed patient populations which show impairments both with respect to episodic memory and imagining hypothetical future scenario[18] (for reviews, see: Schacter et al. 2007; Schacter et al. 2012; Ward 2016; Hallford et al. 2018). For example, two patients with episodic amnesia – K.C. and D.B. – have been found to be unable to both recall episodic memories *and* episodically imagine the future (Klein, Loftus and Kihlstrom, 2002; Rosenbaum *et al.*, 2005). K.C. suffered a motorbike accident, and was subsequently not able to recall any personal memories or encode new memories. He is not impaired with respect to other functions such as reasoning, perception, language, or empathy. He can correctly describe and navigate spatial layouts, such as that of the family's cottage, describe procedures, like how to perform the breast stroke, and he can execute skills, such as playing pool. Nevertheless, he retains no recollection of his personal past, and cannot imagine future possible events. For example, when asked to describe his future, K.C. responded that his mind was 'blank', and he was not able to give any answers about either the close future (tomorrow) or far future (next year). Another patient, D.B., suffered a cardiac arrest and had a total loss of episodic memory as a result. D.B. retained language skills and semantic memory, and could for example recite facts about public figures and events. Similarly to K.C., D.B. was found to be impaired with both respect to episodic memory and imagination, providing potentially confabulated answers to future questions (e.g., claiming that he was going to visit his mother this evening, even though she had been dead for two decades). These patients fail to both retrieve episodic memories *and* to produce future imaginings about personal events. CESH proposes that the explanation for this impairment is that episodic memory and episodic

[18] A hypothetical future scenario is a particular kind of imagining, which is taken to be realistic from the subject's point of view, such as imagining meeting their neighbour at the shops.

imagination both rely on the episodic retrieval process, such that when this is impaired, it negatively affects both recall and imagination.

But this data also supports CESH's claim that there are two separate retrieval processes, namely the semantic retrieval process and the episodic retrieval process. Both of the patients with episodic amnesia discussed above did not have any impairments when it came to recalling semantic details. For example, as mentioned, D.B. was able to recite semantic information about public figures and events (Klein, Loftus and Kihlstrom, 2002). K.C. was also able to recite semantic facts about himself, and performs at average on word recognition tests such as the WASI Vocabulary Subtest (Rosenbaum *et al.*, 2005). This indicates that both encoding and retrieval of semantic details were unimpaired in K.C., as he could both recite long-term facts from before the accident, as well as information learned after the accident. The fact that semantic memory is intact for both these individuals suggests that semantic information relies on a different retrieval process than does episodic information. Hence, this supports that there are two separate retrieval processes with different functions; retrieving semantic information *vs.* retrieving episodic information.

*3.1.2 (b) Memory and Aging*

The theory predicts that the semantic and episodic retrieval processes should doubly dissociate, and there are two further sources of evidence for this claim. Firstly, researchers have studied individuals with semantic dementia, and found that they have a selective impairment for words and concepts, though their episodic memory seems intact, such that they can episodically remember both long-term autobiographical events, and recently experienced events (Adlam, Patterson and Hodges, 2009; Mion *et al.*, 2010).[19] For example, Adlam et al. tested individuals with semantic dementia on their semantic knowledge of names and locations of famous buildings, such as the Eiffel Tower. Compared to controls, individuals with semantic dementia were impaired with respect to this task. However, they were not impaired when it came to reporting episodic information about a recent event. When asked questions about the previous day, such as what they were wearing, they did not score differently to controls. This is consistent with CESH's

[19] It is also possible that individuals with semantic dementia have impaired encoding or storage, as it has been shown that they perform differently on DRM tasks compared to an average population (Budson *et al.*, 2000, 2003). Specifically, in experiments, they are less susceptible to the false recognition effect than control participants. Schacter and Addis (2007) have argued that the best explanation for these results is that individuals with semantic dementia do not form a gist representation of the studied words, and are thus less likely to report recognising new words.

predictions, and it indicates a double dissociation between the episodic retrieval process and the semantic retrieval process.

Secondly, more evidence supporting the claim that there are two separate retrieval processes comes from studies showing that older adults produce fewer episodic details than do younger adults, but older adults instead produce more semantic details than do younger adults (Levine *et al.*, 2002; Addis, Wong and Schacter, 2008).[20] This natural decline in episodic details and rise in semantic details also suggest that the episodic retrieval process and semantic retrieval process can be dissociated, since a deterioration of the episodic retrieval process could explain these results. These studies often use pictures as cues, such as in a study by Gaesser et al. (2011). Here, younger and older adults were asked to either generate memories that were related to the cue (condition 1), or future imaginings that were related to the cue (condition 2). It was not specified whether they should try to generate episodic or semantic details. A third condition had participants simply describing the picture in as much detail as possible, without any acts of remembering of imagining. The details were scored as either semantic or episodic by blind coders, and the study found that older adults generated fewer episodic details and more semantic details than did younger participants. This result is explained by positing that there are two different kinds of retrieval processes – one episodic and one semantic – which dissociate, as it is then possible that one deteriorates independently of the other.

*3.1.3 (c) Experimental Data from an Episodic Specificity Induction (ESI) Paradigm*

This kind of result is not only obtained by studying aging, but can also be manufactured by clever experimental designs, lending further support to the claim that the episodic retrieval process is independent from the semantic retrieval process. An *episodic specificity induction* (ESI) can be used in experimental designs to test the dissociation (Madore, Gaesser and Schacter, 2014; Madore, Jing and Schacter, 2016). An ESI is an induction training in recalling episodic details of a recent experience that a participant receives before completing another task, and it was originally used to boost the amount of details that eyewitnesses could retrieve (Fisher and Geiselman, 1992). The expectation is that an ESI should increase the performance of a subsequent task only if the subsequent task relies on episodic retrieval processes, since it is designed to amplify only the

[20] 'Episodic details' are also referred to as 'internal details'; 'semantic details' are also referred to as 'external details'.

episodic retrieval process. Hence, if both imagination and episodic memory rely on episodic retrieval processes, performing an ESI before a memory or imagination task should boost the amount of details of the memory or imagining, but semantic details should not be boosted.

Madore, Gaesser, and Schacter (2014) used an ESI task to test whether the semantic retrieval process is independent from the episodic retrieval processes. In the first stage of their study, participants viewed a video of people performing various tasks in the kitchen. In the experimental condition they then received an ESI that directed them to recall the video in as much detail as possible, but in the control condition they received a control induction in which they were supposed to provide their general impressions of the video. In the second stage of the study, participants viewed visual cues (pictures) and were asked to recall a related personal memory, imagine a related personal event, or describe the picture. Results found that following the ESI, both older and younger adults generated more episodic details in the memory and imagination tasks. However, the ESI as expected had no effect on the number of semantic details in these tasks.[21] We can thus see that the claim that there are two separate retrieval processes, one which handles episodic information and one which handles semantic information, is supported by data using an ESI paradigm.

Hence, I take it that both the predictions that CESH makes are well-confirmed by empirical evidence. This gives us an architecture where the semantic and episodic retrieval processes are distinct, but where the construction of episodic rememberings and imaginings are both dependent on the episodic retrieval process. Now, I will turn my attention to how these processes are implemented, and what neuroscientific evidence indicates about the cognitive architecture.

### *3.2 (ii) The Implementation Claim*

Schachter and Addis claim that a particular neural network is responsible for this generating both imaginings and rememberings. I will show that many of the neural substrates that support episodic memory also support imagination, but I will not go as far as to say that episodic memory and imagination rely on *all and only* the same neural substrates, or that they rely on the same neural substrates *to the same extent*.

[21] These results have also been replicated in a study using words instead of pictures as the cues for memory and imagination tasks (Madore, Jing and Schacter, 2016).

Supporting the Implementation Claim, a number of studies have indicated that the *core network*, a sub-part of the *default mode network*, is active both during episodic memory and episodic imagination. Various experimental designs have been used in these studies. For example, Addis, Wong, and Schacter's (2007) fMRI study had healthy participants imagine a future event or remember a past event in response to a word cue. When this event was in their mind, they pressed a button, and then elaborated on the event for 20 seconds. Other studies have involved tasks such as imagining hypothetical scenes (Axelrod, Rees and Bar, 2017), past events that never occurred (Addis *et al.*, 2009), or counterfactual events (De Brigard *et al.*, 2013). Further studies still asked participants to recombine details (people, places, objects) in imagination, hence controlling for the possibility that participants could simply use a past event and recast it to a future setting when imagining future scenarios (Addis *et al.*, 2009). These studies converge on that the areas that are active both when episodically remembering and imagining are the medial prefrontal cortex, lateral temporal and parietal cortices, posterior cingulate and retrosplenial cortices, and the medial temporal lobes (including the hippocampus) (Benoit and Schacter, 2015). These reflect the core network supporting both memory and imagination, such that both functions are implemented in the same neural areas.

But there are also differences in how neural areas are activated during imagination and memory. Firstly, a number of areas show *more activity* during imagination phases than memory phases, one of these areas being the hippocampus (Benoit and Schacter, 2015). Schacter and Addis argue that this is because episodic memory requires the reactivation and reintegration of existing relations between details (such as representations of people and places), whereas the construction of novel imagined events require the *formation* of new relations by recombining and integrating details retrieved from different episodic memories into a new event representation (Schacter and Addis, 2020).[22] Thus, even though both functions are implemented in the same neural areas, it indicates that some processes are more active for one function than for another.

Secondly, neuroscientific data also indicate that imagination and episodic memory demonstrate differentially activated neural areas, such as the anterior temporal lobe, which is thought to implement the semantic retrieval process (Schacter and Addis, 2007). Though the episodic and semantic retrieval processes dissociate, as argued above, they still typically work together to

[22] For a different account of how this process works, see: Hassabis et al. (2007).

produce rememberings and imaginings, which contain both episodic and semantic information. Evidence from semantic dementia patients suggest that imagination is more heavily reliant on the semantic retrieval process than what episodic memory is. Irish et al. (2012) showed that whilst being able to retrieve episodic memories with similar amounts of details as controls and recast past events in the future, semantic dementia patients are unable to imagine future novel events. The best explanation for these results is that they are impaired with respect to the semantic retrieval process, and that this process is more heavily involved in producing imaginings than episodic memories. This seems likely, as further studies support this proposal, reporting differential activity during imagination phrases in areas known to support semantic processes, such as the medial prefrontal and lateral temporal cortices (Addis, Wong and Schacter, 2007; Addis *et al.*, 2009). As suggested earlier, imagination's reliance on the semantic retrieval process could be due to imagination utilising semantic details as a 'scaffold'.

This echoes the same implementational picture as advocated by Spagna et al. (2021) and Bartolomeo (2002a), which I discussed in chapter 1. I remarked that Spagna et al. and Bartolomeo do not develop a cognitive architecture, but it seems to me that their implementational pictures overlaps significantly with the one suggested in this chapter. Recall that they suggested a network of brain areas involved in *visual* imagery, comprising of the left fusiform gyrus, the fronto-parietal networks, the cingulo-opercular network, anterior regions of the left temporal lobe, and the medial temporal lobe. They posited that semantic information is drawn on, as well as that content is recombined into visual imagery. Now, with CESH, we have a more exact picture of what processes are implemented in different brain areas, and we can further say that memory and imagination are largely implemented in the *same* neural areas, though there are also areas which show differential activity during imagination and episodic memory, indicating on the architectural level that imagination is more heavily reliant on the semantic retrieval process. In all, this is good evidence for positing the episodic system, implemented in the core network, which functions to construct representations of both the past and the future in terms of rememberings and imaginings.

### *3.3 Taking Stock*

I have defended CESH as a cognitive architecture which supports both memory and imagination, by providing empirical evidence for the Construction Claim and the Implementation Claim. The upshot of this is that there is one system – the episodic system – which functions to construct episodic simulations of the past and the future. This is in line with the Generative View of memory, such as Michaelian's account, and the empirical support for CESH gives us a further reason to endorse the Generative View.

But CESH only gives us so much, as the theory is sketchy when it comes to a number of issues. Firstly, CESH does not tell us how the retrieval processes are able to locate the information they need to retrieve. It does not tell us, for example, whether the retrieval processes simply searches through all possible content until it finds the right content to retrieve, or if there is a more sophisticated way in which a retrieval process could be directed to the right content, such as by addressable content. Secondly, it does not tell us how particular content is selected for retrieval. Some kind of process needs to be able to account for why certain content is selected over other content, especially if there are many possible contents that could fit with what an agent is intending to do. For example, if an agent intends to imagine going on holiday, what is to say that content of a beach holiday is selected over content of a mountain holiday? Thirdly, CESH does not tell us how fine-grained the retrieval processes are. There are many different kinds of episodic details, such as visual details or auditory details; are there different episodic retrieval processes that reflect this fine-grained content, or does the same episodic retrieval process function to retrieve all kinds of content? Fourthly, CESH is not committed to an implementational story about memory traces. Since how content is stored is a crucial part of any theory of memory, one might expect CESH to say something about this. Hence, in order to build up CESH as a more explanatorily powerful view, I dedicate the next chapter to answering these questions and modifying CESH accordingly. This gives me a new model, which I call 'CESH+'. In chapter 4 and 5, I will show that CESH+ is even more explanatorily powerful, and this gives us further reasons to think that the Generativists are right about the function of the episodic system.[23]

## 5. Conclusion

In this chapter, I have developed a new cognitive architecture of memory and imagination, which I claim underwrites both semantic and episodic memory – including mental imagery generation. After narrowing down the focus to the kind of memory I am interested in, I discussed the

[23] There are, of course, even further questions to answer, and CESH+ by no means represents a complete architecture. I focus on making these particular modifications because of the problems I am concerned with in chapters 4 and 5.

Preservative View and the Generative View of episodic memory, which crucially disagree about what the function of memory is. I argued that the Preservative View is committed to saying that our memory system malfunctions most of the time, and that this is undesirable, leading us towards the Generative View instead. Suggesting that the function of episodic memory that Generativists posit might be more viable for developing an architecture of memory and imagination, I then argued for the Constructive Episodic Simulation Hypothesis. This theory gives an architectural as well as implementational model of how rememberings and imaginings are generated. I supported the model by arguing for two claims: the Construction Claim and the Implementation Claim. Together, these claims indicate that the episodic system is responsible for generating both rememberings and imaginings, and that its processes are largely implemented in the same neural areas in the default mode network. The proposal also shares a significant overlap in the neural areas proposed to underlie visual imagery in chapter 1. But I indicated that there is more to say about the cognitive architecture than what CESH is currently committed to. Accordingly, I will give a more sophisticated take on the different retrieval processes, and argue that we should posit memory indices, distributed memory traces, and a content selection mechanism. This will give me CESH+, which the next chapter is dedicated to developing.

# Chapter 3

# The Constructive Episodic Simulation Hypothesis +

## 0. Introduction

In the last chapter, I proposed that both imagination and memory are generative, and I introduced the CESH architecture. The main tenet of CESH is clear enough: there is one single system underwriting the generation of both imagination and memory. But there are several places in which the theory is sketchy. In this chapter, I focus on two. The first is on the architectural level. Here, CESH does not detail any mechanisms to direct retrieval processes to where content is stored, thus leaving the question open of how the retrieval processes are able to retrieve content successfully. CESH also posits very course-grained retrieval processes, on the level of episodic/semantic information, but empirical evidence suggests that retrieval processes should be individuated at a finer grain. Secondly, CESH is also sketchy when it comes to how these retrieval processes operate. That is, CESH does not specify the step-by-step procedure which is followed when a retrieval process retrieves any particular content. This chapter seeks to rectify these shortcomings by modifying CESH into what I call 'CESH+'.

To a first approximation, CESH+ contains further architectural features, including memory indices, multiple fine-grained retrieval processes, and distributed memory traces. Moreover, CESH+ also includes a Bayesian ranking/selection mechanism[1], and details the procedure which is followed when content is selected to be retrieved to be recombined into a remembering/imagining. Not only do these proposals solve the problems mentioned above, but they have other beneficial consequences as well. As we will see in chapter 4, CESH+ is able to explain how imagination can be both controlled and improved, and thus support the argument that imagination is a skill. In chapter 5, I will show that the addition of multiple fine-grained retrieval processes enables the theory to explain the selective impairment with respect to different kinds of sensory details as seen in aphantasia. Overall, my contributions make CESH+ more explanatorily powerful than CESH, and in doing so, lends greater support to the Generative View.

[1] I use 'selection mechanism' for short throughout the thesis.

The chapter takes the following structure. In §1, I start by adding memory indices and multiple retrieval processes to CESH. The idea of a memory index is borrowed from computer science, where indices store the addresses of information. Similarly, I suggest that our cognitive architecture contains such indices. I also argue that we have reasons to think that there are multiple retrieval processes. In line with this, I suggest that CESH+ should include a distributed view of memory traces. Having completed my architectural additions to CESH, in §2 I switch focus to how the retrieval processes operate. In order to explain why certain content is retrieved instead of other content, we need an account of exactly how content is selected. Here, I posit a content selection mechanism, which operates using Bayesian generation. This is the most optimal way for this kind of mechanism to operate, and I give detailed toy examples of how content is selected. This kind of account is in line with other current accounts in cognitive science, which posit similar operations for other tasks (see for example: Norris, 2006; Gopnik and Tenenbaum, 2007). With CESH+ on the table, §3 then considers and fends off an objection to the proposal that the mechanism could operate using Bayesian probabilities, as some might think that it seems computationally intractable. In §4, I conclude that CESH+ lends further credence to the Generative View. This sets the stage for the coming two chapters, where I show that, with CESH+ in hand, we can make sense both of imagination as a skill which can be improved and controlled, as well as give a cognitive explanation of the condition aphantasia.

## 1. CESH+

CESH, as presented and argued for in chapter 2, is silent on a number of important questions. To begin with, though positing retrieval processes is a major tenet, the theory is silent on how exactly the right content can be retrieved. A general point about retrieving anything is that in order to retrieve it, one must know where to retrieve it *from*. But CESH does not posit a way for the retrieval processes to access the address of the location of information. For this reason, in §1.1, I develop the idea of memory indices, which is borrowed from computer science, and which has been employed in other theories of memory (Teyler and DiScenna 1986; Rudy and O'Reilly 2001; Rudy, Huff, and Matus-Amat 2004; Teyler and Rudy 2007; Langille and Gallistel 2020). Moreover, as there is empirical evidence suggesting that episodic content is stored in different places – such as the auditory cortex or the motor cortex –, I argue that we should in fact posit *multiple* retrieval processes, such that each retrieval functions to retrieve only one type of content. A further argument for this will be given in chapter 5, where I show that we need to posit such diverse retrieval processes in order to account for selective impairments in aphantasia. In §1.3, I then show that the fact that content is stored in different places suggest that my theory should be committed to a distributed view of memory traces – a question which CESH is also silent on. Finally, in §1.4, I raise a worry one might have as a consequence of discussions in chapter 1, where I argued against the necessary involvement of early visual processing in visual imagery. Since I am now arguing that there is indeed activation in the visual cortex when visual content is retrieved, it might seem as I am smuggling the Perceptual View in through the back door under a different name. I show that this is not correct as the cognitive architecture I posit is fundamentally different.

### *1.1 Memory Indices*

My first addition to CESH is *memory indices*. Consider the following example: an agent episodically remembers riding a horse at her old riding school. What happens in the agent's mind? The first step  is that the agent *intends* to remember riding a horse in her old riding school. On the basis of this intention, multiple commands must be issued. These are commands to retrieve particular elements needed to reconstruct the memory, such as a visual representation of a horse, and an olfactory representation of what horses smell like. Retrieving these is the responsibility of the episodic retrieval process. But in order to retrieve these elements, the retrieval process needs to know where to find them. That is, it needs to know the locations, or addresses, of these elements. But CESH is silent on how this works, which is why I am suggesting the addition of memory indices.

A memory index stores the address of an element, much like how the address of a person is stored in an official register. This is an idea borrowed from computer science, and it might need some initial motivation. When searching for something, say, the number of a person in a phonebook, there are many ways of doing it. One way of doing it would be to start going through the phonebook from start to finish, until you find the right person and can read off their number. But say that you are looking for 'John Smith', who is logged under 'S' for 'Smith'.[2] Using this kind of search algorithm would have you look through most of the phonebook before finding him – it goes

---

[2] I here make the assumption that the content is sorted, rather than unsorted. This might be a controversial assumption to make about memory systems, but since having sorted content dramatically improves efficiency for searching, and since nobody has argued otherwise (to the best of my knowledge), I will help myself to this assumption.

without saying that this is very inefficient. This is why a phonebook incorporates an index, allowing you to look up which page contains surnames beginning with 'S' (or sometimes even 'Sm', 'Smi', etc.). Looking at the index and finding out that 'S' starts on page 500, you can immediately go there, and thereafter start searching alphabetically, significantly increasing your efficiency. In computer science, the address which an index stores is often referred to as a 'pointer' as it points to the location where the information can be retrieved.

Many researchers working on how human memory is organised borrow this idea from computer science, often also citing efficiency as an argument, and have accordingly argued that that there are memory indices which store the addresses of memory elements (Teyler and DiScenna 1986; Rudy and O'Reilly 2001; Rudy, Huff, and Matus-Amat 2004; Teyler and Rudy 2007; Langille and Gallistel 2020). Moreover, given that there are different kinds of retrieval processes, there might also be different kinds of indices, storing addresses of the different kinds of information. Again, this makes sense if we think in terms of efficiency, since looking through an *index*, as well as the phonebook itself, also takes time. Suppose that the index of the phonebook not only incorporated pointers to page numbers of surnames that start with 'S', but also, say, pointers to a completely different kind of information, like pointers to page numbers alphabetically listing native flowers in a *Flora*. This would be a crazy system, which would hugely impact search time. Naturally, to design an efficient system, we put the index for the flora in the *Flora* book, not in the phonebook. It stores information related to a completely different kinds of things (flowers, not people's phone numbers), so there is no reason to put this index in the phonebook. Similarly, having a huge index containing all the pointers to all the different *kinds* of memory elements would be inefficient. So, instead, the suggestion is that there are different indices, each designated to have pointers to certain kinds of information. These are divided up along the familiar lines of episodic information and semantic information, so accordingly, there is an episodic index and a semantic index.

The *a priori* argument for positing memory indices is that computer memory is structured by having indices that store addresses of content. If we think that the mind is indeed a computer, and we take computer science as a guiding light to figuring out how the mind operates, we have a good *a priori* reason to think that the mind must implement indices too (Goode *et al.*, 2020; Langille and Gallistel, 2020). But there is also empirical evidence to support positing memory indices. On the implementation level, memory indices were introduced to explain the role of the hippocampus in memory, positing that different parts of the hippocampus implement indices that store the addresses of sensory information, taking advantage of intracellular structures to do so

(Moscovitch *et al.*, 2005; Goode *et al.*, 2020). The theory specifies the intrinsic organisation of the hippocampus, and its synaptic physiology as well as its anatomical relationship to other regions of the brain (Teyler and Rudy, 2007). One line of support for memory indices comes from experiments which have attempted to *prevent* the forming of an index (for further experiments and discussion, see: Goode *et al.*, 2020). The thought here is that if an index is not established, we should expect that an animal fails to perform an otherwise expected behaviour if the behaviour depends on the index. For example, an experiment by Tanaka et al. (2020) induced a fear response in mice by creating a shock-association in a certain environment. When this association has formed, it is expected that the animal will display the fear behaviour in this environment. However, when particular neurons (CA1 neurons) thought to be necessary for establishing the index, were 'silenced' using laser stimulation, memory retrieval was impaired, and areas in the amygdala and cortex were not able to be reactivated, leading to reduced fear behaviour. Some experiments have also been carried out on humans using fMRI imagining, and these have supported the predictions of the theory, such as that cued recall should trigger the reactivation of the memory index, which will then reactivate the entire pattern of neocortical activity related to the episode (Rudy and O'Reilly 2001; Rudy, Huff, and Matus-Amat 2004).[3] This lends credibility to the idea of memory indices, as one would expect that the relevant index is first activated, before activation is reinstated where the information is stored. Positing memory indices as part of the CESH+ architecture is thus a well-supported proposal, which takes into account considerations about efficiency raised in computer science.

### *1.2 Retrieval Processes and Memory Traces*

There is also empirical evidence suggesting that the retrieval processes are more sophisticated than suggested by Addis and Schacter (2007, 2020) when it comes to the kind of information they retrieve. There are good reasons to think that depending on what *kind of episodic information* is requested (i.e., visual, auditory, olfactory, etc.), a different episodic retrieval process is recruited to retrieve it. Therefore, whereas CESH posits *one* episodic retrieval process, I posit (at least) eight for CESH+: visual, auditory, gustatory, tactile, olfactory, and affective, as well as a spatial semantic

[3] Interestingly, the idea of a memory index is often developed as a connectionist idea (Teyler and DiScenna, 1986; Teyler and Rudy, 2007), despite being an idea originating in classical computationalism with the idea of addressable memory (Langille and Gallistel, 2020). I find it somewhat puzzling that the idea seems to have taken hold in connectionism more so than in classical computationalism, but the idea seems compatible with both views. See Teyler and Rudy (2007) for a review of recent evidence.

retrieval process and a spatial episodic retrieval process (Smith et al. 2004; Gottfried et al. 2004; Barrós-Loscertales et al. 2012) (see Figure 6). The first six might seem more intuitive, so let me explain and motivate these first, before elaborating on the spatial ones, which also require positing a new spatial index.

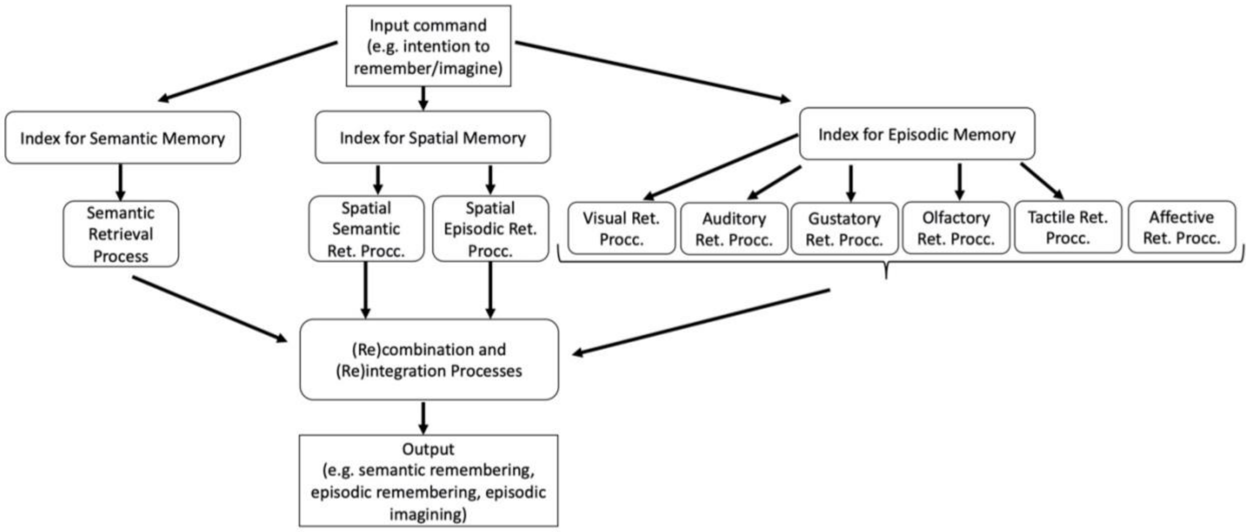


*Figure 6. A boxological representation of the Constructive Episodic Simulation Hypothesis+. 'Ret. Procc.' is short for 'retrieval process'. The depiction shows the memory indices which store addresses to memory content, the semantic, episodic, and spatial retrieval processes, which retrieve content, as well as the recombination process, which recombines content.*

A large body of independent research supports the existence of a different episodic retrieval processes dedicated to retrieving different sensory details. For example, recent studies indicate that brain regions involved in encoding an episodic memory are partially reactivated when that content is later remembered, and according to Danker and Anderson (2010), many PET and fMRI studies show the reactivation of sensory regions when retrieving an episodic memory. Studies have for example used an associative paradigm, where a word ('dog') is either coupled with hearing a sound (woof!) or a picture (of a dog) (Wheeler, Petersen and Buckner, 2000; Wheeler and Buckner, 2003, 2004; Wheeler *et al.*, 2006). Results from these studies show that upon seeing the word 'dog' again, activity in the visual association cortex is reinstantiated during retrieval of visual information (picture of dog), and activity in the auditory association cortex is reinstantiated during retrieval of auditory information (woof!).[4] Retrieval of olfactory memories has been studied in a similar way. In a study by Gottfried et al. (2004), objects were first presented to

[4] See Danker and Anderson (2010) for a discussion of these studies, as well as variations of the paradigm.

participants together with neutral, positive, or negative odours. Later, the previous objects were presented again, together with some new objects, and participants were asked to determine if they were presented with a new object or not. Results here showed that when participants were presented with a previously seen object, there was more activation in the primary olfactory cortex, indicating that the object was associated with the previous odour, and that this process underwrites odour retrieval. Unsurprisingly, similar patterns of reactivation have been found in the primary and secondary gustatory cortices when participants in another experiment read food-related words (Barrós-Loscertales *et al.*, 2012), and Smith et al. (2004) found that the fusiform gyrus was more active during both encoding and retrieval of emotional information, thus suggesting an affective retrieval process.[5] This evidence indicates that we should posit more sophisticated episodic retrieval processes than what was originally suggested by Schacter and Addis.[6]

The reader might worry that this is in fact not evidence for a multitude of retrieval processes, but simply evidence that information can be stored in different places. An alternative suggestion is that there is only one retrieval process, which accesses different types of information stored in different places. However, given a plausible functionalist construal of retrieval processes, the latter suggestion entails my proposal. Let me explain. Let's assume that a retrieval process is individuated functionally in terms of what domain of information it takes as input, and what domain of information it produces as output. It follows that if a process accesses information in a different domain than another (say, visual, as opposed to auditory), they should be categorised as separate processes. In other words, the visual retrieval process is the process that has the function of retrieving visual information from the visual domain, whereas the auditory retrieval process is the process that has the function of retrieving auditory information from the auditory domain.

Evidence also points to that there are special retrieval processes dedicated to the retrieval of *spatial* information, which has been independently argued by other researchers, who claim that there are two spatial retrieval process – one semantic and one episodic – which are independent

[5] Further supporting studies are discussed in Danker and Anderson (2010).

[6] Moreover, certain people are impaired only with respect to retrieving a particular kind of information (e.g. visual), which suggests that there are indeed different processes that are recruited. For example, in a memory test using a drawing paradigm, people with aphantasia were shown to be impaired with respect to recalling visual details, but not spatial details (Bainbridge *et al.*, 2020). This will be discussed at length in chapter 5.

from all other retrieval processes (Moscovitch *et al.*, 2005; Rosenbaum *et al.*, 2005). The spatial episodic retrieval process retrieves allocentric and egocentric information about locations, including landmarks and typography, and supports re-experiencing the location. The spatial semantic retrieval process retrieves schematic representations of environments, and does not support re-experiencing the location. The existence of two dedicated spatial retrieval processes has been defended by Moscovitch et al. (2005) on the basis of double dissociation evidence. Two patients, K.C. and E.P. were tested on tasks related to semantic spatial information (distance judgements, proximity judgements, sequencing landmarks along routes, recognising gross features on world maps) and episodic spatial information (identifying smaller neighbourhood landmarks and smaller features on maps). Whilst they were not impaired on the former, they were severely impaired on the latter. This points to that schematic information as involved in the former task is retrieved differently to the more detailed information involved in the second task (Rosenbaum *et al.*, 2000).[7] The existence of spatial retrieval processes also suggest that there might be an index for spatial memory. Recall, storing information in separate indices is more efficient than having one bigger index. To the best of my knowledge, there is no empirical support for positing a further index for spatial memory, but these a prior considerations about efficiency give us some reason to think that there are indeed separate indices. For now, I will assume a spatial index into CESH+, though it is not a main claim of the theory. To reiterate, CESH+'s main claim about indices is that there has to be *at least* one index, but there are possibly multiple.

### *1.3 Memory Traces*

This picture of different retrieval process further suggests a distributed view of memory traces. Recall, memory traces are normally posited by Preservationists to make good on the idea of what is being retrieved when an agent recalls something. But Generativists need to posit memory traces too, because future and past representations cannot be generated *ex nihilo*, but have to be generated *from* something. The idea that generally does not appeal to Generativists is that memory traces are stored locally, that is, all the content being stored in one place, as this does not tally well with content being generated by flexible recombining content. Locally stored memory traces was the original idea proposal by Martin and Deutscher (1966), but a distributed view is preferred by many now (Sutton, 2007; Bernecker, 2010; Michaelian, 2011).

---

[7] See Moscovitch et al. (2005) for further evidence.

To clarify what is meant by a distributed view of memory traces, Robins (2016b) distinguishes between two different options. In particular, she distinguishes between distributed *memory traces* versus distributed *mental contents*. Firstly, mental content could be distributed in a connectionist network. Content such as *party* or *bicycl*e might activate different nodes in the network. But content lack a constituent structure. This means that there need not be any overlapping activation between similar concepts, such as BICLYCLE and TRICYCLE. This is because the nodes, à la connectionism, are not semantically evaluable as they do not store symbols representing content. Secondly, one can have a network of distributed *memory traces*. Here, memories are represented as patterns of connections between event features, such that a recent birthday party might be represented as <party><wine><birthday>. A similar event, such as a retirement party, might be represented as <party><wine><retirement>. The nodes in this network are semantically evaluable. However, there is no further implication of any distributed mental content.[8]

As I am not developing a connectionist view, my idea is most akin to distributed memory traces. That is, I believe that the data I have cited as showing that retrieval processes activate different neural regions suggest a distributed picture of the trace, such that visual content of <wine> might be stored in one place, and gustatory content of <wine> in another place. However, even though this is where the content is stored, I am also suggesting that there is a higher level on which a bundle of addresses to all the content of one memory, such as 'birthday party', are stored, namely, the index. The index stores the pointes to all of the addresses of a memory trace, and the trace itself is distributed in these locations.

My idea for CESH+ is thus that memory traces are indeed distributed, but this does not entail a connectionist picture. This is because the implementational details differ. Recently, Gallistel (2020) has argued that a connectionist picture of memory is untenable, as strengths between nodes could not code for content, and it is more likely that content is coded for on an intra-molecular level, which tallies better with a classical computationalist conception of the mind. In

[8] Robins (2016b) raises the distinction between the distribution of memory traces and the distribution of mental content in order to object to the Causal Theory of Memory. In particular, she argues that distributed memory traces are incompatible with the Causal Theory of Memory as defended by Bernecker (2010) and Michaelian (2011). This is because distributed memory traces do not provide way to track the causal history of memories for particular events, which causes a problem in distinguishing between remembering and relearning (something which Martin and Deutscher (1966) saw as crucial). This problem does not arise for my account, as it concerns the architecture rather than metaphysics of memory.

the connectionist tradition, it is common to think that the memory trace resides on the neural level. Specifically, connectionists have a synaptic theory of memory, where the strength between synapses in a neural network implements a memory trace. But classical computationalist have raised worries about this proposal, since it seems like an associative bond between neurons is not able to encode symbolic information. Recently, Gallistel (2020) has suggested an alternative, namely that memory is implemented on the molecular level inside the cell. Gallistel claims that there are information-carrying molecules inside neurons, and his argument draws on recent work on the intracellular structure of the Purkinje cell (Johansson *et al.*, 2014, 2015). This proposal has the further advantage that information-bearing molecules can be easily generated and destroyed, without much energy, whereas synaptic associations are comparatively costly. This matters because however memory is implemented, we need a lot of it, and it needs to be generated quickly as an agent learn new information. I will tentatively endorse this view about the implementation, though I recognise that it still has many questions to answer, such as how the address is implemented differently to the content stored in a molecule. Still, I believe it is more fitting for the architecture I am proposing. It is important to notice that my main argument for the architecture of CESH+ does not hang on either memory traces being stored in a distributed way, or memory being implemented intra-molecularly. One could develop CESH+ along different lines on these questions if one so wished.

### *1.4 Differences between CESH+ and the Perceptual View*

Having argued for the modifications, let me finish this discussion by comparing the new account to the Perceptual View discussed in chapter 1. At this point, the reader might worry that my account collapses into the kind of Perceptualist and Simulationist accounts I argued against (Kosslyn, 1994; Currie and Ravenscroft, 2002; Goldman, 2006; Nanay, 2016). After all, I said before that visual imagery is not dependent on the same neural substrates as vision, contrary to Kosslyn's conviction, but in this chapter, I am suggesting that both episodic remembering and imagining – per definition accompanied by imagery – re-activates sensory areas, such as areas in the visual cortex (Wheeler, Petersen and Buckner, 2000; Wheeler and Buckner, 2003, 2004; Wheeler *et al.*, 2006). This might look like smuggling the Perceptual View in by the back door, and still making them dependent on sensory areas.

Here is how I answer this challenge. My proposal is significantly different from Perceptualism/Simulationism because the *processes* I argue are instantiated in the sensory neural areas are significantly different from those proposed by Kosslyn. Recall, Kosslyn posits a particular cognitive architecture whereby both vision and visual imagery are realised in the visual buffer. It is this visual buffer that is supposedly implemented in the primary visual cortex. As such, Kosslyn's theory makes the prediction that these areas should be activated both during vision and visual imagery. But this is very different from the prediction CESH+ makes. CESH+ does not posit a visual buffer, and as a consequence, does not predict that specific areas in the primary visual cortex ought to be activated both during vision and visual imagery. Hence, the empirical evidence I discussed in chapter 1, where I showed that there is a double dissociation between vision and visual imagery, is only a problem for Kosslyn's theory, but not for CESH+.

It can also be put like this: the Perceptual View and CESH+ claim that the function of visual areas for imagery are widely different, where the Perceptual View claim they implement a buffer, whereas CESH+ claims they reflect content retrieval. There should thus be no reason to think that these theories are the same, despite some of the same neural regions being relevant for implementation. I will now turn to a very different question, namely, the question of how the retrieval process operates.

## 2. Bayesian Inference and Bayesian Generation

Another question which CESH does not answer is how the retrieval processes operate. This is crucial if we are to say why certain content is retrieved in favour of other content. For example, if an agent intends to imagine a party, what is to say that they will imagine a birthday party rather than a retirement party? The answer clearly lies in how content is being selected, but so far, CESH does not specify this as it says nothing about the procedures that mechanisms in the architecture employ. Here, I will both posit a selection mechanism and detail how it operates. This will be of major importance for chapter 4, where I argue that imagination is a skill, and improving imagination amounts to selecting more apt content for imaginings.

My suggestion is that the selection mechanism works by Bayesian generation. The idea that selection mechanisms operate according to Bayesian principles is common in cognitive science, partly because using Bayesian probabilities is the optimal way of making selections in light of many available options. That is, it is most efficient. Though it might not always be true that the mind works optimally, we should start from this assumption and test the hypothesis using experimental paradigms (Norris and Cutler, 2021). Results might then prove us wrong, and we might need to modify the theory. But before testing the hypothesis, we cannot know the particular ways in which the mind might work sub-optimally, so we should not build such assumptions into our first theory.

There are generally two kinds of Bayesian accounts: Bayesian inference accounts and Bayesian generation accounts, which are used in a vast amount of theories. These include Predictive Processing accounts of perception (Clark 2013; Kirchhoff 2018; Hohwy 2020; Williams 2021), as well as theories about causal reasoning and counterfactual thought (Walker and Gopnik, 2013), language acquisition (Jusczyk, 2003), and word recognition (Norris, 2006). For example, Walker and Gopnik argue that children are rational Bayesian learners of causal models, and that Bayesian inference is actually *implemented* in a cognitive mechanism. As a consequence, this theory can *explain* behaviour, because the model attempts to mirror what goes on in the mind. Their appeal to Bayesian inference is thus more than a mere modelling of behaviour, which would not have this explanatory power.[9] As I will show, Bayesian generation suits my purpose better here, but my claim is otherwise very similar to Walker and Gopnik's, as I hypothesise that the Bayesian generation is indeed implemented in this cognitive mechanism. This means that the calculations I describe on the following pages are analogous to ones carried out by the mechanism. The upshot is that my account is also an attempt to explain behaviour, rather than model it.

For my account to make sense, going over how Bayesian inference works before describing Bayesian generation will be helpful. The general idea of Bayesian inference is that a particular hypothesis can be tested by calculating how likely it is to be true, given the data an agent has observed. For example, if I observe my friend Sally coughing, how likely is it that she is coughing because she has a cold? The data I am observing here is Sally's coughing, and the hypothesis I am interested in testing is whether she is coughing because she has a cold. There are several hypotheses available, and we will focus on three different ones here: Sally could be coughing because she has a cold, she could be coughing because she has heartburn, or she could be coughing because she has lung cancer.[10]

To calculate which hypothesis is the most likely, we need to set values for *likelihoods* and *priors*. These are supposed to capture what we know about the case. The likelihoods capture how likely

[9] Jones and Love (2011) discuss this and other ways of using Bayes in theory and modelling.
[10] This example is adapted from Perfors et al. (2011).

it is that coughing is a symptom of any of the diseases. Since both colds and lung cancer cause coughing, whereas heartburn does not, the likelihoods favour these options. That is, it is common to cough when one has a cold or lung cancer, but not when one has heartburn. These will thus receive higher values and heartburn a lower one (see Table 1*)*. The priors capture how likely it is that one has a cold, heartburn, or lung cancer, independently of any observed symptoms. The priors favour colds and heartburns over lung cancer, since it is more common to have a cold or a heartburn than to have lung cancer. We capture this intuition by assigning higher values to colds and heartburns, and a lower one to lung cancer. This can be formalised in the following way, where *P* is the probability, *d* is the data observed, and *h* is a specific hypothesis:

| Likelihoods | Priors |
|---|---|
| $P(d\|h_{heartburn}) = 0.1$ | $P(h_{heartburn}) = 0.4$ |
| $P(d\|h_{cold}) = 0.8$ | $P(h_{cold}) = 0.5$ |
| $P(d\|h_{cancer}) = 0.9$ | $P(h_{cancer}) = 0.1$ |

*Table 1. Likelihoods and priors for the Sally example* (Perfors *et al.*, 2011).

We can then calculate what we are interested in, which is the *posterior probability*, i.e. the probability that Sally is coughing *because* she has a cold. We do so using Bayes' formula:

$$P(h_{cold}|d) = \frac{P(d|h_{cold})\,P(h_{cold})}{P(d|h_{heartburn})P(heartburn) + P(d|h_{cancer})P(h_{cancer}) + P(d|h_{cold})P(h_{cold})}$$

This gives us:

$$P(h_{cold}|d) = \frac{(0.8)(0.5)}{(0.1)(0.4)+(0.9)(0.1)+(0.8)(0.5)} = 0.754$$

Hence, the hypothesis that Sally is coughing because she has a cold has a probability of 0.75. We can then perform the same calculation for the other hypotheses, and see which one comes out as the most likely.

*2.1 Bayesian Generation for Content Selection*

With this in place, let me now elaborate on my account of the cognitive mechanism of control in imagination. Remember, when an imagining is generated, it draws on memory content in semantic and episodic memory. Given that we have a vast number of different memory representations, there is also a vast array of imaginings that could be generated, and a cognitive mechanism needs to have a structured way of *selecting* content from one representation over other content. We need to find out how this is accomplished to account for how imagination can be controlled with respect to the cognitive mechanism involved. But it does not look like a Bayesian inference account is apt for this task, as it is unclear what would count as a hypothesis or a piece of evidence when generating imaginings. Instead, a generative account, which does not invoke these notions, is more suitable. This kind of account generally allows us to calculate the probability of two events co-occurring, by utilising the priors and likelihoods. For example, we can find out the probability of Sally having a cold *and* Sally coughing. Assume that P(Cough|Cold) = 0.8, as it is very common to cough when one has a cold. This gives us:

$$P(\text{Cough}, \text{Cold}) = P(\text{Cough}|\text{Cold})\ \text{P}(\text{Cold})$$

Plugging in the numbers:

$$P(\text{Cough}, \text{Cold}) = \text{P}(0.8)\ \text{P}(0.5) = 0.4$$

What we have found here is the joint probability distribution; the probability that two events co-occur. In other words, the probability that one event occurs, given that another event occurs. This seems more apt for the task at hand, as a generative account could give us the probability for an imagining/remembering to occur, given the contents of an agent's memory representations.[11]

But an agent is bound to have more than one memory content, so the mechanism needs to calculate the joint probability distribution for multiple contents, and then select a few which are forwarded to the recombination process. I suggest that the contents of the memory representations assigned the highest probabilities are selected to be recombined. Note thus that the selection mechanism both ranks *and* selects content. But before we get to this point, we need

[11] See Williams (2021) for a discussion of generative Bayesian accounts in imagination.

to assign priors and likelihoods just like in the case above. Let us use a simplified example to illustrate this.

An agent, Matilda, tries to accurately imagine what the weather will be like in Croatia in spring, because she wants to decide whether to go there. This is her imaginative project, $I_{Holiday}$. We are interested in what content is selected for this imagining, given the probabilities that the selection mechanism assigns to the content. For ease of example, say that Matilda only has four memory representations and that these are all mostly episodic: visiting Croatia in the hot summer; a person in Croatia saying it often rains there in spring; visiting a rainy and flooded Venice in spring; and visiting Leeds for shopping in the summer. To make things easier, let us use these categories as shorthand: 1) *Croatia*; 2) *Conversation*; 3) *Venice*; 4) *Leeds*. Each of these memory representations is comprised of different elements, such as information relating to who-what-where-when, all of which are candidates for content that could be retrieved and recombined into a novel imagining. For example, *Conversation* contains both episodic information about a person and semantic information about the weather.

How should the priors be assigned in this case? Recall that in the example of Sally, priors were assigned on the basis of how likely an event was to occur independently of any other event. Could it be the case that priors here too are determined by episodes' independent probabilities of occurring? This does not seem likely, as the mind does not have *access* to these probabilities. As mentioned, I am giving a realist account of the selection mechanism, which aims to explain behaviour, whereby the Bayesian selection mechanism is actually implemented in the mind, and which aims to explain behaviour. For this to be viable, the sub-personal mechanism needs to have access to the probabilities that it uses in calculations. But it does not seem like sub-personal mechanisms have access to these independent probabilities. Instead, I suggest that prior probabilities are determined by something that the sub-personal mechanism does have access to: the emotional intensity of stored episodes. It should be noted that this is a tentative suggestion, and not a main claim of the account. There are many other proposals which could be developed, such as that prior probabilities are set by the recency, frequency of recall, or strength of connection between elements (Talarico, Labar and Rubin, 2004). I will develop the proposal along the lines of emotional intensity now to illustrate the theory, but I am open to the possibility that priors are set in a different way.

This hypothesis predicts that emotional episodes (negative or positive) are more readily retrieved in greater detail than neutral episodes (Brown and Kulik, 1977; Talarico, Labar and Rubin, 2004). *Prima facie*, the emotional intensity of episodes seems highly relevant for remembering/imagining, and this suggestion tallies well with evolutionary theories, which suggest that emotionally intense episodes are more useful for an animal to remember/imagine, than neutral ones (Suddendorf and Corballis, 2007). To take an extreme but illustrative example, if a pre-historic human is planning a trip to the waterhole, it holds a greater payoff for them to more readily retrieve a memory of a fearful episodic of being attacked at the waterhole, than a neutral one of having a quiet drink, even if the neutral episodes have happened more frequently. More readily recalling intense episodes drastically increases the chances of survival, and it can thus be more useful to more easily retrieve emotionally intense episodes.

Now, going on holiday might not be *that* emotionally intense, but it still has a significant degree of emotional intensity. Hence, both *Croatia* and *Venice* should be assigned high priors, reflecting that they are highly likely to be recalled and used in imaginings. Particular conversations about the weather are not emotionally intense, so *Conversation* should be assigned a very low prior. Similarly, visiting Leeds for shopping is not likely to be very emotionally intense, so this should also be assigned a low value.

What about the likelihoods? These should capture how likely it is that a particular imagining is generated, given the content of a memory representation. Now, we should assign *Croatia* a high likelihood, since it is likely that a memory of Croatia in the summer will be drawn on when trying to imagine the weather there in spring. *Conversation,* though having a low prior, should be assigned a high likelihood, since it concerns the weather in Croatia – the very thing the imaginer is trying to imagine. *Venice* is also fairly likely to be drawn on, since it is a memory of a place geographically close to Croatia and it is a memory of springtime. On the other hand, *Leeds* seems very unlikely to be drawn on, since the location is nowhere near Croatia, and the memory concerns summer rather than spring.[12] Below, the *priors* and *likelihoods* are assigned to reflect this:

[12] It might be helpful to think of the agent as 'priming' certain memory representations here. When priming a memory, a subject is presented with a certain cue, and this makes it more likely that they recall memories that are related to the cue (e.g. mouse-cheese) (McNamara, 2005).

| *Likelihoods* | *Priors* |
|---|---|
| $P(i_{Holiday} \mid m_{Croatia}) = 0.9$ | $P(m_{Croatia}) = 0.9$ |
| $P(i_{Holiday} \mid m_{Conversation}) = 0.8$ | $P(m_{Conversation}) = 0.3$ |
| $P(i_{Holiday} \mid m_{Venice}) = 0.2$ | $P(m_{Venice}) = 0.8$ |
| $P(i_{Holiday} \mid m_{Leeds}) = 0.2$ | $P(m_{Leeds}) = 0.3$ |

*Table 2. Likelihoods and priors for the Croatia example. I use the notation 'm' for 'memory' and 'i' for 'imagining'.*

What we are trying to find out, then, is how likely it is that a particular imagining is generated, given the memory representation: $P(i_{Holiday}, m_{Croatia})$. In other words, we are trying to find out the joint probability distribution. We calculate this in the following way:

$$P(i_{Holiday}, m_{Croatia}) = P(i_{Holiday} | m_{Croatia})\, P(m_{Croatia})$$

This gives us:

$$P(i_{Holiday}, m_{Croatia}) = (0.9)(0.9) = 0.81$$

From this, we can see that the content of the memory representation *Croatia* is assigned a high probability, making it highly likely to be used in generating the imaginative project. From this alone, it looks likely that the Matilda would produce an imagining of a sunny Croatia, but, remember, we have only considered content in the memory representation *Croatia* so far. Since we are interested in how likely it is that the agent will imagine a certain scenario (rainy spring in Croatia) given *all* her memory representations, we need to also assess how likely the content from other memory representations are to be drawn on; I will leave these calculations out here, but the results are as follows:

| *Croatia* | *Conversation* | *Venice* | *Leeds* |
|---|---|---|---|
| $P(i_{Holiday}, m_{Croatia})$ | $P(i_{Holiday}, m_{Conversation})$ | $P(i_{Holiday}, m_{Venice})$ | $P(i_{Holiday}, m_{Leeds})$ |
| 0.81 | 0.24 | 0.16 | 0.06 |

*Table 3. The joint probability distribution for the Croatia example.*

What we have done above is analogous to the computations carried out by the selection mechanism. But the selection mechanism needs to do more than just calculate the probabilities of various memory contents' being used in imaginings; it needs to actually *select* the content too. Just like there is a structured way of calculating probabilities, there needs to be a structured way of selecting content based on these probabilities. A plausible assumption is that only the top-ranking content gets included in the imagining/remembering, since otherwise all content would be included to some degree, no matter their joint probability distribution. It will suffice for my demonstration here to assume that the three top ranking are selected. In that case, the contents of *Croatia, Conversation*, and *Venice* will be selected to be recombined into an imagining. Since these together contain the vital semantic information about rain in the spring, an episodic representation of rain, and the episodic representation of the location Croatia, Matilda generates an accurate imagining of it raining in the spring in Croatia.[13]

The reader might at this point worry that this is a 'just so' story, which does not make any precise predictions, and lacks any empirical support. Without predictions and support, characterising the selection process as Bayesian might itself seem unfounded. This would be a considerable problem for the account, but luckily, it is more than a 'just so' story. Here is an example of a prediction of the account: if emotional intensity determines priors, emotionally intense episodes should be more likely to be recalled/imagined than neutral episodes. There are two telling instances of this. Firstly, patients with PTSD are more likely to recall/imagine negative emotionally intense episodes, than neutral episodes (Pearson *et al.*, 2015; Pearson, 2019). Secondly, multiple studies have found that average populations often suffer from an *intensity bias* when imagining/recalling episodes, whereby episodes are imagined/recalled as overly emotionally intense (Wilson, Meyers and Gilbert, 2003; Gilbert and Wilson, 2007; Ebert, Gilbert and Wilson, 2009). Both of these findings are predicted and can be explained by appealing to priors being set by emotional intensity. Note that this is not to say that we *cannot* imagine neutral episodes; of course we can. Even though

[13] Interestingly, there are actually two different probabilities at play here. The probability of Matilda generating an accurate imagining is an *objective* probability, whereas the probabilities assigned by the selection mechanism to content are *subjective* probabilities. Subjective probabilities are probabilities represent*ed* by a vehicle, whereas objective probabilities are probabilities of represent*ing* something. Like most Bayesian theories in cognitive science, the account I have given is concerned with subjective probability. However, if we are interested in modelling behaviour, we can also infer an objective probability from the subjective probabilities, as the subjective probabilities assigned by the selection mechanism bears on the probability that Matilda generates an accurate imagining. See Hájek (2019) for a discussion.

an episode could have a low prior due to low emotional intensity, it could still be selected due to high likelihood (recalling cereal at the breakfast table would have a low prior due to not being emotionally intense, but in a context where an agent's intention is to remember their breakfast table, the likelihood would rise significantly). The claim is rather that, *ceteris paribus*, emotional episodes are more likely to be chosen over non-emotional ones, due to emotional intensity determining priors. Hence, I take it that this is a plausible story of how the procedures are carried out by the selection mechanism. I will now consider one objection.

## 3. Is a Bayesian Selection Mechanism Computationally Tractable?

There might be a general worry about the tractability of my Bayesian account. The reader might think that the account looks computationally intractable, since for any imagining/remembering, the cognitive mechanism would need to assign a prior and likelihood to every single one of thousands of contents in memory representations. This simply looks like too big a task and it does not seem plausible that this is what goes on sub-personally. I have three responses to this.

Firstly, on my proposal, only likelihoods would need to be re-assigned for every imagining depending on the agent's intention, as priors are set by the emotional intensity and do not change with a change in intention. This halves the amount of values that need to be assigned. But one might think that this is still problematic, as it might seem unrealistic that the selection mechanism is able to quickly assign a large quantity of likelihoods to contents. I do not share this worry. Whichever mechanisms operate sub-personally must do so at an incredible speed to produce behaviour, often within a fraction of a second. My suggestion is that the selection mechanism operates using Bayesian generation, but *whichever* mechanism one posits must operate at an incredible speed. Consider a very simple selection mechanism, which could select memory content and is based on linear search. Assume for the sake of the demonstration that it is simply searching through memory content which has been ordered from earliest encoded to most recently encoded. This might be more like the kind of mechanism that someone would find computationally tractable, as well as *prima facie* plausible, since it fits neatly with our picture of sub-personal mechanisms as searching for and retrieving content. If this search mechanism is tasked with searching for some specific content to be retrieved, the best case scenario is if the content it is looking for is the first one in the list (that is, the oldest content). But at worst, it has to search through the whole list in order to find the content (the most recent content). This means that the search time increases linearly with how much content is stored. This is not ideal for efficiency, and would on average be slower than the mechanism I propose. The general point is that whichever kind of mechanism one develops, there is always the possibility that is needs to search through/rank all of the content, because there is always a possibility that the content that needs to be retrieved is the last one looked at. This possibility does not go away by positing a simpler, but less optimal, mechanism. So, for anyone pursuing this line of objection, a simple mechanism should seem as computationally intractable as the one I am positing. But clearly there has to be *some* mechanism or other operating, and the one I am suggesting has the benefit of being optimised for efficiency.

Moreover, a Bayesian mechanism does not have to perform precise calculations; calculations could be approximate. Bayesian accounts can be simplified and approximation methods can be used, such as Markov-Chain Monte Carlo (Hastings, 1970), or Gibbs sampling (Geman and Geman, 1984), and this has been successful in other models.[14] Other methods include variational approximation methods, which are widely used in machine learning and cognitive science (for example, see: Smith (2021)).[15] Finally, another reason to think that this actually is a computationally tractable account is that Bayesian models very similar to this have already been developed in computer science for generating novel virtual worlds from a limited set of 'memory representations' (Fisher *et al.*, 2012; Davies, 2020). I thus take it that this is a computationally tractable account of content generation for imagination and memory.

## 4. Conclusion

In this chapter, I proposed CESH+ as a modification of CESH, as CESH is sketchy in a number of places. My first contribution was to add memory indices, which store the addresses of the locations of memory content. This is an idea adapted from computer science, and it can explain how the retrieval process knows where to go to retrieve a certain memory content. I then proposed that there are also multiple episodic retrieval processes, as well as spatial retrieval processes. The function of these is to retrieve different types of memory content, such as visual memory content or auditory memory content. I argued that we should posit these, as there is evidence that retrieving particular sensory memories reactivates sensory areas. As we will see, this modification will pay off in chapter 5, where I argue that selective retrieval processes can explain some of the

---

[14] See Jones and Love (2011) and Shultz (2007) for a discussion of further methods and examples.
[15] This is similar to the linguistic cues discussed by Jones and Wilkinson (2020).

impairments that we see in aphantasia. Furthermore, I more tentatively suggested that CESH+ fits well with a distributed picture of memory traces, where different neural areas store different features of memory traces. I further suggested that these traces are implemented intra-molecularly, rather than in a connectionist network. Finally, I added a Bayesian selection mechanism, intended to explain how certain content is retrieved over other. In particular, I argued that content is ranked and selected by a mechanism which operates using Bayesian generation. Memory content is thus assigned likelihoods and priors when an agent intends to imagine/remember something, and depending on the values, it can be selected for recombination. This particular idea will be vital for chapter 4, where I argue that the selection mechanism can account for how imagination can be improved and controlled.

Before moving on to that task, it is worth noting that though CESH+ is an improvement on CESH, the cognitive architecture is by no means complete. There are still many more questions to answer, such as how the recombination process operates and how the indices are organised. But what I have done here is sufficient to shed light on the issues with which I will concern myself in this thesis, namely imagination as a skill and a cognitive explanation of aphantasia. The ability to successfully apply CESH+ to elucidate these issues gives us further reasons to continue on the project of expanding CESH+ into a full architecture in future research.

# Chapter 4

# Imagination as a Skill

## 0. Introduction

In the previous chapters, I put forward a novel architecture of memory and imagination, namely, CESH+. In this chapter, I show that CESH+ sheds considerable light on the issue of whether imagination is a skill.

Other mental actions, such as performing mental arithmetic, have recently been argued to be skills (Buskell, 2015; DeKeyser, 2015), and I will show that imagination is no different, so the first aim of this chapter is to establish that imagining is in fact a skill. To do so, I will argue it meets two hallmarks of skill: *improvability by practice* and *control*.[1] These hallmarks are not intended as necessary and sufficient conditions, but rather as indicators of that something is a skill, and they have been widely agreed upon by both philosophers and psychologists (Anderson, 1982; Dreyfus, 2002; Stanley and Krakauer, 2013; Fridland, 2014; Wu, 2016).

My claim, however, faces an 'how-possibly' challenge. For imagination to be a skill, imagination should be improvable and controllable, but how is this possible? We need a cognitive architecture that enables this. I argue that the Bayesian selection mechanism I posited in chapter 3 can do the job. In particular, I tentatively suggest that the likelihood of content being selected for recombination can change as the result of an agent's *feeling of error* about a produced imagining, though this thesis is in need of empirical testing. Moreover, this kind of account ties together control and improvability, such that the functioning of the control mechanism can be invoked in an explanation of how imagination improves.  The fact that CESH+ is able to explain how imagination can be improved importantly lends further support to the architecture, giving us another reason to endorse it.

[1] I use 'imagination' and 'the capacity to imagine' interchangeably in sentences like 'imagination is a skill' and 'the capacity to imagine is a skill'.

I proceed as follows. §1 – 5 argue *that* imagination is a skill. In §1, I provide reasons for thinking that *improvability by practice* and *control* are hallmarks of something's being a skill. §2 discusses Kind's (2020a) recent arguments that imagining is a skill, but argues that her case is not empirically well-supported and can therefore not establish the conclusion. To show that imagination is improvable by practice, §3 draws on experimental evidence from the study of mental rotation and auditory imagery tasks. In §4, I provide a theoretical account of controlled imagination, where I argue that controlling our imagination amounts to constraining it in certain ways. In §5, I provide empirical data from developmental psychology supporting that imaginings are in fact controlled in the proposed way, and I conclude that imagination is a skill. In §6, I consider the issue of *how* this could be — in particular, I discuss how we can make sense of imaginings being controlled and improvable on the sub-personal level. Drawing upon CESH+, I tentatively suggest that imaginings are often accompanied by epistemic feelings which play a causal role in changing the likelihood that certain memory content is chosen to be recombined into an imagining. Specifically, I suggest that a feeling of error about an imagining reliably indicates that the imagining is indeed inaccurate, and that this serves to change the likelihoods such that a more accurate imagining can be produced. I conclude that not only is imagination a skill, but also that CESH+ can give us a sub-personal account of how it works.

## 1. The Hallmarks of Skill

There is widespread agreement among philosophers and psychologists that the two hallmarks of something's being a skill are: a) skills can be improved through practice (Anderson, 1982; Stanley and Williamson, 2001; Dreyfus, 2007; Fridland, 2014), and b) skills involve control (Dreyfus, 2002; Stanley and Krakauer, 2013; Fridland, 2014; Wu, 2016).[2] Hence, a viable way to find out whether something is a skill is to check whether it can indeed be improved by practising, and whether it can be controlled.

Whether skills improve through practice is a well-tested hypothesis within psychology, as it has been tested since the 1960's (for an early example, see: Fitts (1964)), culminating with Anderson's

[2] Philosophical debates about skill are often tied up in the Know-How Debate, which concerns what kind of knowledge we have when we know how to perform a skill, and particularly whether know-*how* is a subset of know-*that*. The answer to this question does not matter for my purposes here. For an overview and in-depth discussion of the Know-How Debate, see: Pavese (2021)

influential work on ACT-R theory (Anderson, 1982, 1993, 2007). Anderson lays out three stages in skill acquisition[3] - a declarative stage, a procedural stage, and an automatic stage – where each level signifies an improvement in executing the skill. Learning the skill of driving a car is often used as an example to illustrate this. At the declarative stage, an agent has to explicitly think about each step in, say, changing gears ('depress the clutch, grab the gear stick, move the gear stick down…'). This is often an arduous task which requires a lot of executive resources, and is not executed smoothly. At the procedural stage, the execution becomes smoother, and fewer executive resources are needed. For example, instead of thinking about each step involved in changing gears, these steps can be 'chunked' into larger units such as, 'put the car in neutral gear' and 'put the car in gear 4'. Finally, at the automatic stage, this action can be performed smoothly and with minimal executive recourses involved. Importantly, Anderson argues that an agent moves to the second and third stages by *practising* the skill and, to the best of my knowledge, everybody nowadays accepts that practising is vital for skill improvement.[4]

But in order to see whether something can be *improved* by practising, we need to specify a notion of what it means to be good at something – otherwise it would not be possible to compare an agent's performance before and after practising. How good one is at something normally depends on the goal we evaluate it against. For example, it is difficult to evaluate how good someone is at running without knowing if the goal is to run 100 metres or a marathon. Endurance, for example, matters to how good one is at running if the goal is to run a marathon, but matters less if the goal is to run a 100 metre race. A good marathon runner does thus not necessarily make a good sprinter. If we pick endurance as a measure of goodness for running a marathon, we can then compare a runner's performance with respect to endurance before and after practising. Similarly, if we pick time as a measure of goodness for sprinting 100 metres, we can measure performance before and after practising with respect to how fast they run.

For imagining, we also need a standard of measurement in order to see if someone improves by practising, and the same general point applies here: we should evaluate improvement with respect to goal. For example, when using imagination to plan for the future, we could say that someone is good at imagining if they imagine the future as *accurately* as possible (Suddendorf and Corballis,

[3] In psychology, the study of 'skill acquisition' incorporates both *acquiring* a skill and *improving* that skill.
[4] For an overview of the literature in psychology and a summary of studies on improving with practise, see: DeKeyser (2015). Studies on skills include performing arithmetic, motor skills, and second language acquisition.

2007). This seems relevant as they are trying to accurately imagine a future state of affairs. If, on the other hand, the goal is to create a fictional story, an accurate imagining does not seem like a better imagining. Instead, we might want to consider how fantastical the imagining is, or how detailed it is. For this reason, I will not specify a general notion of *goodness* with respect to imagination, but since all the cases I consider have to do with attempting to imagine accurately, I use accuracy as a standard of measurement in this chapter. Call this kind of imagination 'practical imagination'.

The reader might have an additional worry about accuracy: namely, how can we test imaginings for their accuracy? This is a warranted worry, as we do not carry out many of the things we imagine about the future. For example, someone might imagine that if they went to Spain on holiday, they would feel happy. However, they might end up never going to Spain, and hence we cannot tell whether they would have been happy or not. But psychological experiments using between-group design can help with get an answer to this. For example, in the research on affective forecasting, it is common to ask one group of subjects to imagine how they would feel in the future if certain events took place/did not take place. But instead of having these subjects then experience the event, they are instead compared to another group, which have already experienced the event, in order to see if the imagining of the first group is accurate. This gives us a statistical likelihood of the subjects' imagining being accurate, and it has for example been using in testing the accuracy of imagining how happy one would feel getting tenure (Gilbert *et al.*, 1998).[5] So, there is a way of measuring accuracy of imaginings, even if the imagined state of affairs is never experienced by imaginer.

Secondly, skills also involve control (Dreyfus, 2002; Stanley and Krakauer, 2013; Fridland, 2014; Wu, 2016). Fridland makes a particularly good case for this when she describes how an expert can intervene in their actions because they have greater control than a novice, and they can also choose to make voluntary errors. For example, an expert gymnast is skilled partly because she can control her every movement; if something is going slightly wrong, she can intervene and correct this. She also has the option of making voluntary errors, e.g. to demonstrate to novices how *not* to perform an action. Fridland also offers an account for how we should understand control of bodily actions

[5] Other experiments use an in-group design to test accuracy of imaginings of happiness for exam results (Levine *et al.*, 2012, Experiment 3), election results (Levine *et al.*, 2012, Experiement 1), and social aptitude test results (Gilbert, Meyers and Wilson, 2008).

on in terms of strategic control, selective control[6], and motor control. Strategic control is a personal level phenomenon, and it can be consciously exercised by an agent when planning and setting out strategies for their skilled action. This control does not automate, but is intentionally exercised by an agent. For example, footballer could set out a particular strategy during a game. Selective control is responsible for selecting the relevant features in an environment that a skilled agent should gather information about and respond to, given her goals, plans, and strategies (Wu, 2011, 2016). In contrast to strategic control, this kind of control does automate (Wu, 2011), such that the football player, once an expert, will automatically pay attention to the right features of her opponent, given her strategy (Yarbus, 1967; Wu, 2011). Third, motor control is constituted by automatised motor routines learned through training, such that an agent who is able to hold their limbs in exactly the right positions is one that is exercising motor control. Of course, motor control is not involved in skilled mental actions, like imagining (Buskell, 2015).

Hence, I take it that skills are characterised both by improvability by practice and control. It is worth noting that I am not claiming that improvability by practise nor control are necessary or sufficient conditions on something's being a skill – they might be, but my argument need not be that strong. All I need is for them to be reliable indicators of something's being a skill, which is a claim that is generally agreed upon. Then, if we want to find out whether something is a skill, meeting these hallmarks is a good indicator. What I do in the rest of this chapter is argue that imagining indeed meets these hallmarks, and therefore is best thought of as a skill.

## 2. Kind's Case for Imagination as a Skill

With the hallmarks in place, let us turn to a recent argument for imagination being a skill. We often praise the imaginative abilities of inventors, writers, and the like, and we might have the natural intuition that some are better at imagining than others. To demonstrate this point, Kind (2020b) considers three people with incredible imaginative abilities — in particular, with extraordinary visual imagination – namely, the engineer Nikola Tesla, the scientist Temple Grandin, and the origami folder Satoshi Kamiya. Tesla described how he used visual imagination to work out, improve, change, and operate his inventions; Grandin was able to use her imagination to visualise how cattle would react when entering a conveyor track in a slaughtering plant in order to design a new process where cattle would not panic; and Kamiya has said that he uses his visual

[6] Fridland calls this 'selective, top-down, automatic attention'.

imagination to figure out how to fold origami, where he visualises the piece completed and then sees in his mind's eye how it unfolds. Since these people all seem to be very *good* at imagining, Kind concludes that we should take them to be *skilled* imaginers.[7] But it is not so clear that they are indeed skilled imaginers, since it has not been shown that they meet the two hallmarks of skill. In particular, I will show that it is unclear that they have improved by *practice*.

The problem with the cases of Tesla, Grandin, and Kamiya is that there are two possible explanations for their exceptional performance, and we cannot tell these two apart: (i) they are exceptionally good at imagining because they *improved* their imagination by practising; (ii) they were simply born with extraordinary imaginative abilities. In another recent paper, Kind acknowledges this issue, which she labels 'the nativist objection' (Kind, 2020a). Specifically, this objection states that the fact that somebody is a very good imaginer does not provide evidence for the claim that imagination is a skill, since that person might simply have an innate capacity to imagine well, rather than having developed this capacity through practice. To address this objection, Kind turns to evidence from sports psychology. She considers how *visualisation techniques* are used in sports, where athletes imagine themselves achieving the desired outcome (e.g. spinning a ball). Since athletes *practice* visualising and, purportedly on the basis of this, improve their sport performances, Kind infers that practising visualising improves visual imagination, and this in turns improves the sport performance. Hence, it appears that imagination meets the first hallmark of skill.

The crux of the matter is whether this inference is justified: do we have good reasons to think that by practising visualising, athletes become better at visual imagination? Unfortunately for Kind, sports psychology cannot provide an empirical basis for this conclusion, for the following reason: empirical research into athletes' use of visual imagery does not actually *test* whether athletes improve at visualising.[8] Instead, researchers are interested in whether athletes improve their athletic skill, and questionnaires such as the Sport Imagery Questionnaire are designed to test this improvement, rather than to test their potential visualisation improvement (Hall et al. 1998; Munroe et al. 2000; Beauchamp et al. 2002; Williams and Cumming 2014). Though some studies report on how good athletes are at visualising (often cashed out in terms of how vivid, rather than

---

[7] See Kind (2016) for a longer discussion of Grandin and Tesla.

[8] For a general discussion on visual imagery in sports, see: Munzert and Lorey (2013) and Guillot (2020). For a meta-analysis, see: Driskell et al. (1994).

accurate, their imaginings are), there is still no comparison between visualisation ability before and after practising (Baddeley and Andrade, 2000).

However, even if improvement in visualisation is not directly measured, one could potentially argue that there is evidence for imagination's improving with practice, since the *sports performance* is improved. One could claim that the best explanation for why athletes perform better at their sport is because their capacity to imagine has improved with practice, and this improvement directly affects their performance. Unfortunately, even this conclusion is hard to establish, since it is notoriously difficult to tease out whether and to what extent visualising improves sport performance because of the many confounding factors. For example, athletes are often in physical training *as well as* using visualising techniques, so visualising cannot easily be singled out as the sole factor that contributes to performance improvement, and some research has also suggested that visualisation techniques do in fact not improve performance directly, but rather work to boost confidence (Munzert and Lorey, 2013), which in turn makes for a better performance. For these reasons, Kind's inference is not as straightforward as it first seems, and sports psychology cannot currently give us a conclusive answer as to whether imagination is improvable by practice. Luckily, there are other areas of research which directly test whether imagination is improvable by practice, making them more suitable for my purposes here.

## 3. Improving Imagination by Practising

Whether imagination is improvable by practice has been extensively tested using other tasks, and I here discuss mental rotation tasks and auditory imagery tasks. As mentioned briefly in chapter 1, mental rotation tasks are tasks where a subject is presented with a picture of a 2D or 3D object, next to four other pictures of very similar-looking objects at various angles (see Figure 7). Their task is to determine which of the four shapes matches the first shape, if the first shape were to be rotated. In order to solve the task, the subjects are asked to mentally rotate the original shape to 'see' if it matches the others. Subjects are put through practice trials where they get to practise on a large number of different shapes before they go through the test-sessions. In both practice trials and test-sessions, subjects are instructed to press one button if they think the shapes are the same, and another one if they think that the shapes are different. Reaction times and accuracy are measured on both occasions. If the subject is able to provide a correct answer more quickly in the test-session than in the practice trial, they are taken to have improved. This is an important measure, since being skilled is positively correlated with performing the task quicker (Anderson, 1982). This experimental design, and slight variations of it, has repeatedly been used and the results robustly indicate that subjects can improve their capacity for mental rotation (Shepard and Metzler, 1971; Kosslyn, 1994; Habacha, Molinaro and Dosseville, 2014).[9]

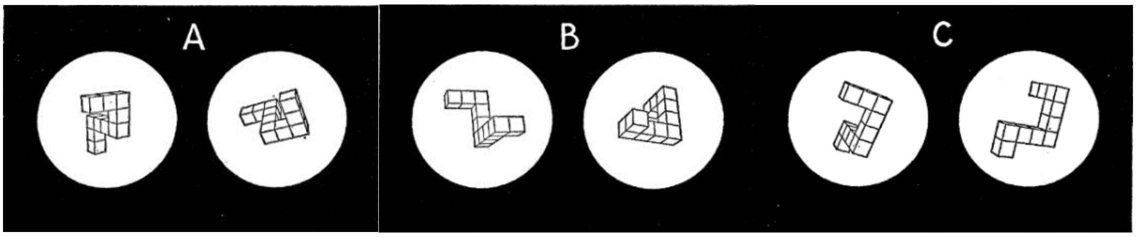


*Figure 7. Figures used in a mental rotation task* (Shepard and Metzler, 1971).

One might worry about confounding factors here too. For example, these results could be due to subjects using techniques other than mental rotation, such as retrieving the right answer from memory.[10] But there are strong reasons to think that this is not the case. Provost et al. (2013) observed that participants could potentially solve the mental rotation task in either of these ways, so to disentangle these two hypotheses, they introduced a further EEG measurement into the rotation experiment. More precisely, they measured the EEG signal Rotational Related Negativity (RRN) during task performance, since this positively correlates with mentally rotating shapes, but not with retrieving a shape from memory. RRN signals were present during both practice and test sessions, thus indicating that subjects are indeed mentally rotating the shapes on both occasions. At the same time, subjects reduced their reaction time by an average of 800ms from practice to test sessions, demonstrating significant improvement through practice.

These experiments show that *visualising* can be practised such that subjects can improve at it. However, my argument in this chapter concerns imagination *in general,* not just mental rotation through visualising, so one possible objection is that because visualisation is just one kind of imagination, we cannot draw such a general conclusion from these studies. As discussed in the Introduction, we ordinarily distinguish among different kinds of imagination, such as visual imagination, auditory imagination, gustatory imagination, and so on. It looks possible that visual

[9] More studies can be found in Kosslyn (1994) and Kosslyn et al. (2006). For a review, see: Zacks (2008).
[10] Though this is a common objection, it is perhaps hard to see how it could have teeth given the architecture I propose whereby the same system underlies memory and imagination. Because of its prevalence, I discuss it anyway.

imagination is improvable, whereas auditory imagination might not be, and hence my conclusion over-generalises.

Now, to the best of my knowledge, researchers have not directly investigated whether imagination in other modalities can improve through practice. Still, I think that we can make a case for it at least for auditory imagination. Research on auditory imagination has repeatedly shown that musicians are better than novices at generating accurate auditory imagery (Aleman *et al.*, 2000; Halpern *et al.*, 2004; Bishop, Bailes and Dean, 2013; Jakubowski, Farrugia and Stewart, 2016; Jakubowski, 2020). For example, Aleman et al. (2000) tested 15 musically trained subjects and 20 novices on two auditory tasks. Musically trained subjects have at least two years of formal music training and activity played an instrument. Firstly, they administered a musical imagery task adapted from Halpern (1988) whereby they were supposed to mentally compare pitches of notes corresponding to lyrics taken from familiar songs. In the experimental condition, participants would see a lyric from a familiar song on a screen, with two highlighted words. They were to judge whether the second word was higher in pitch than the first one. For the control condition of the same tasks, participants also heard the song, and were thus not required to mentally compare the pitches. They were instructed to respond as fast as possible by pressing a button. Another task adapted from Mehta et al. (1992) that was administered was a non-musical auditory imagery task. In the experimental condition of this task, a triad of descriptions of sounds were presented, such as 'crying baby', 'laughing baby', and 'meowing cat'. Using auditory imagery, participants were asked to indicate the deviant item in terms of acoustic characteristics. In the control condition, the sounds were actually presented.[11] Results showed that musically trained participants had more correct responses than novices for both the musical and non-musical tasks. Though performance within a subject was not measured in this study, the best explanation for the results is that musically trained subjects improved their mental imagery compared to novices.[12] Given that there was only one testing occasion, the study does not suffer from the potential confounds as sports psychology, where other kinds of training could have given rise to improvement. Further research

[11] A visual task was also administrated, with an experimental and control conditions, where participants were supposed to indicate the deviant item out of a triad of three (descriptions of) visual items. The task was adapted from Mehta et al. (1992). Musically trained participants did not perform differently to novices on this task. Administrating this task controlled for the possibility that musically trained participants might be better than novices at imagery *tout court*.

[12] A possible alternative explanation is the following: musically trained subjects already had better auditory imagery abilities. There might thus be innate differences between subjects. To my knowledge, this possibility has not yet been controlled for.

should investigate exactly how to improve musical imagery. Presumably, there might be various ways of doing this, where practising an instrument might be one, but possibly, a subject could also improve their ability by continuously performing auditory imagery tasks as well.

In all, I take it that these studies show that imagination in general is improvable through practice. Though not all areas of imagination have been tested, the ones that have robustly indicate improvement, and we have no reason to think that the other areas cannot be improved. Imagination thus meets the first hallmark of being a skill.

## 4. Control in Imagination

Things are looking good for imagination so far. I have shown that there is robust empirical data supporting the thesis that imagination is indeed improvable by practice, but that is not enough to conclude that imagination is a skill, since skills also involve control (Dreyfus, 2002; Stanley and Krakauer, 2013; Fridland, 2014; Wu, 2016; Kind, 2020a). For example, a runner can control the speed they are running at, and a chess player can control their play by setting a strategy for a game. But it is less clear what it means to control one's imagination. What I will do next is specify a plausible theoretical framework for what controlling imagination means, where I again draw on Kind's research. In particular, I draw on another paper of Kind's, where she argues that we can constrain our imagination by employing the *Reality Constraint* and *Change Constraint* (Kind, 2016), and I will argue that controlling imagination plausibly amounts to employing these constraints. Thereafter, I use empirical evidence from research in developmental psychology to show that these are indeed ways that imagination is controlled such that when these constraints are employed, more accurate imaginings are produced.[13]

[13] Kind also rightly acknowledges that imagination can be controlled, but does not treat this at length and offers no empirical support (Kind, 2020a, p. 339).

### *4.1 A Theoretical Framework for Control*

Kind argues that we can constrain our imagination by using the Reality Constraint and Change Constraint (Kind, 2016), and I will argue that controlling imagination plausibly amounts to employing these constraints.[14]

Consider the following example. There are many cases where our goal is to imagine things accurately, such as when I moved into my narrow terraced house and tried to accurately imagine if my bedframe would fit up the stairs. But note that imagination is voluntary, and it is not bound by how things actually are (Balcerak Jackson, 2018). That is, I seem to be able to imagine whatever I want irrespectively of the state of the world. Given this, how can we ensure that it is accurate? Kind argues that it can be so only if we constrain it by employing the Reality Constraint and the Change Constraint.

According to the Reality Constraint, an imagining should be constrained by the world as it actually is. For example, I should imagine the bedframe and the stairs as being the sizes they actually are. According to the Change Constraint, when we imagine a change in the world, we need to be guided by the consequences of that change. When I imagine the bedframe going up the stairs, I should imagine the consequences of that state of affairs, namely, that the bedframe ends up upstairs. I should not imagine that it suddenly vanishes or that it miraculously shrinks. Kind's reason for thinking that controlling one's practical imagination amounts to constraining it in this way comes from a consideration about machines as *ideal imaginers*. She conceives of a science fiction scenario where machines are tasked with accurately predicting the future to defend humanity from an alien attack, and because of their logical way of imagining, they can do so perfectly. In particular, the machines are able to accurately represent a state of affairs, and then proceed to change the state of affairs in a logical way. This means that the machines' prediction about the future is bound to be accurate.[15] In contrast, the imagination of humans does not proceed in this kind of way, and

[14] Ideas of constraining imagination have also been developed elsewhere, see Williamson (2016). There might also be further constraints, but my goal here is not to list all of them, but rather sketch out a framework.

[15] There are reasons to think that this inference is not correct, and that there is still room for error in these machines' future predictions. Presumably, these machines must work with *probabilistic* representations about the future, so even if they correctly conceive of the most likely future scenario, there is a non-zero chance that that scenario will not happen. So, it is not the case that whatever the machines predict will happen necessarily does happen and their imagination might be a bit less ideal than what Kind presumes, but this does not impact my argument about controlling one's imagination. To be clear, we should not

this is why, according to Kind, we can make mistakes when imagining the future. What we need to do to imagine the future correctly is to be more like the machines and constrain our imagination.

One might wonder whether these constraints are indeed two separate constraints, or whether one could be cashed out in terms of the other. It might seem as though the Change Constraint just adds a temporal aspect to the Reality Constraint, as imagining the bed ending up upstairs when I take it up the stairs is also a way of imagining a state of affairs changing *realistically*. But note that there are cases where these constraints come apart, which suggest that one does not reduce to the other. I can choose not to employ the Reality Constraint and imagine something unrealistic, but still employ the Change Constraint to imagine things changing as they normally would. Imagining an alligator driving a car is an example of this, where this is clearly unrealistic, but changes in the world remain normal (the alligator turning the steering wheel turns the car's wheels, etc.).[16] In the next section I come back to this issue and provide further evidence that they are different, as the ability to use them also develops at different times in children.

This gives us a plausible theoretical picture of how imagination could be controlled: an act of imagining is controlled if the imaginer constrains it by employing the Reality Constraint and the Change Constraint. On this basis, we can assess whether imagination is a skill by considering whether there are actual cases where imagination *is* controlled in this way. At this point, a caveat is in place. I am at this point only interested in subjects that are attempting to imagine accurately. It is to this extent that I am suggesting that the above constraints be employed. But of course, there might be many cases where imagination could be controlled in different ways. For example, an author trying to imagine a fantasy world is *prima facie* all the more skilled, the less she employs the Reality Constraint. Presumably, there are other constraints that she might want to employ to

---

assume that the right outcomes is always required for something to count as controlled, it should be enough that the right outcomes is *reliably* achieved. For an interesting discussion about reliability and success, see: Small (2017).

[16] Kind also offers reasons to think these constraints are indeed different, by arguing that employing them can malfunction in different ways (Kind, 2016). Drawing on the example of machines as ideal imaginers, she claims that one machine might be faulty in a way such that instead of representing 500 missiles in a defence facility, it represents 50 missiles – this is a violating of the Reality Constraint. Another machine might correctly represent the amount of missiles, but wrongly infer that updating the defence system entails that the defence shields can now withstand fewer missile blasts than they actually would be able to withstand – this violates the Change Constraint.

control her imagining. This should not surprise us as it directly parallels how other skills are controlled; a marathon runner wants to control her running in a way that is different from how a sprinter wants to control his running as they are trying to achieve different goals. Hence, what I am proposing here should not be generalised to how all acts of imagining should be controlled in order to be skilled. Determining what the constraints are that should be employed when the goal is not an accurate imagining is a project for a different day.

## 5. Empirical Research on Control in Imagination

So, do we actually employ these constraints? To answer this question, I will turn to developmental psychology. There are two reasons to focus on research on children rather than adults. Firstly, and trivially, there is no research testing how adults constrain their imagination (to the best of my knowledge). Secondly, developmental studies can tell us when children start being able to employ these constraints, and as such, it can inform whether these constraints are indeed different, as seeing if they develop independently would give us a good indication that they are different. To make this argument, I will firstly briefly go over when imagination in general seems to emerge in children, before tackling evidence for children employing the Reality Constraint and Change Constraint respectively.

Recall, I called imagining accurately for planning purposes 'practical imagination'. There are two questions pertaining to when practical imagination develops in children. In order for children to be able to use practical imagination, it is necessary that their general imagination capacity has developed. Based on empirical research, I will argue that children have a minimal capacity to imagine as early as 7 months, as can be shown by violation-of-expectation studies (Onishi and Baillargeon, 2005; Scott and Baillargeon, 2017), and motor planning studies (Claxton, Keen and McCarty, 2003).[17] These studies show that infants are able to take their representational processes offline, which I will take as a minimal requirement for possessing imagination and as a precursor for *practical imagination* (Suddendorf and Redshaw, 2013). Then, I will suggest that the ability to apply relevant constraints to imagination for the use of practical imagination emerges slightly

[17] Suddendorf and Redshaw (2013) have made convincing arguments that more capacities need to mature in order for children to *intentionally* imagine the future. However, since, as I will show, infants at 7 months are surprised when their expectations are violated, they must minimally be capably of offline processing at this age, even if this does not allow them to intentionally imagine future scenarios.

later in life – through the ages three to five, depending on the constraint. Thus, by the age of five, it is safe to say that the competence to use practical imagination has matured in children.

### *5.1 Offline Processing*

As argued in chapter 2, and also observed by Suddendorf and Redshaw (2013), the activation of a lot of disparate mental processes are necessary for an agent to be able to imagine. For example, retrieval processes need to retrieve content from memory, and recombination processes are needed in order to flexibly recombine elements into novel representations. Suddendorf and Redshaw further suggest that one needs to be able to entertain times other than the present, and be able to entertain one's own future states as separate from one's current states. Another key component of imagination is offline processing, that is, being able to represent states, such as belief states or emotional states, that one currently is not in. This is the component I focus in this section, where I argue that non-verbal studies, such as violation-of-expectation studies, indicate that at least a minimal capacity for offline processing is present in infants as young as seven months.

False belief tasks test offline processing by seeing if a subject can attribute a belief to an agent that is different to the belief they themselves currently hold (Onishi and Baillargeon, 2005; Scott and Baillargeon, 2017). For example, a child and an assistant see an experimenter hide a marble under one of two boxes. The assistant then leaves the room, and the experimenter moves the marble to the other box. The child sees this and forms a true belief about where the marble is, but the assistant, who did not see the marble being moved, holds a false belief. Then the assistant re-enters, and the child is asked where they will look for the marble. If children are capable only of online processing, it is likely that they will wrongly answer that the assistant will look where the marble *actually* is, since this is the belief the child themselves hold. If, on the other hand, they pass the task by indicating the location where the assistant last saw the marble, it can be inferred that they are capable of offline processing.

Interestingly, studies using a non-verbal violating-of-expectation paradigm suggest that offline processing is present already in 7-month-olds, but since infants of that age are not able to give verbal feedback, looking times have been used as an indicator of what the infant is capable of. Particularly, infants look at an action *for a longer time* if they are surprised, and researchers exploit this, hypothesising that infants will look for longer if an agent act incongruously with their belief (Clements and Perner, 1994). In keeping with this hypothesis, studies have shown that when the assistant, who has not witnessed the re-hiding of an object, reaches to the place where the object actually is, infants look for a significantly longer time than when the assistant reaches for the place where they last saw the object (Onishi and Baillargeon, 2005; Scott and Baillargeon, 2017) (see Figure 8) This is best explained by infants already possessing the capacity for offline processing, such that they attribute a false belief to the agent, and are surprised when the agent does not act in accordance with this belief.

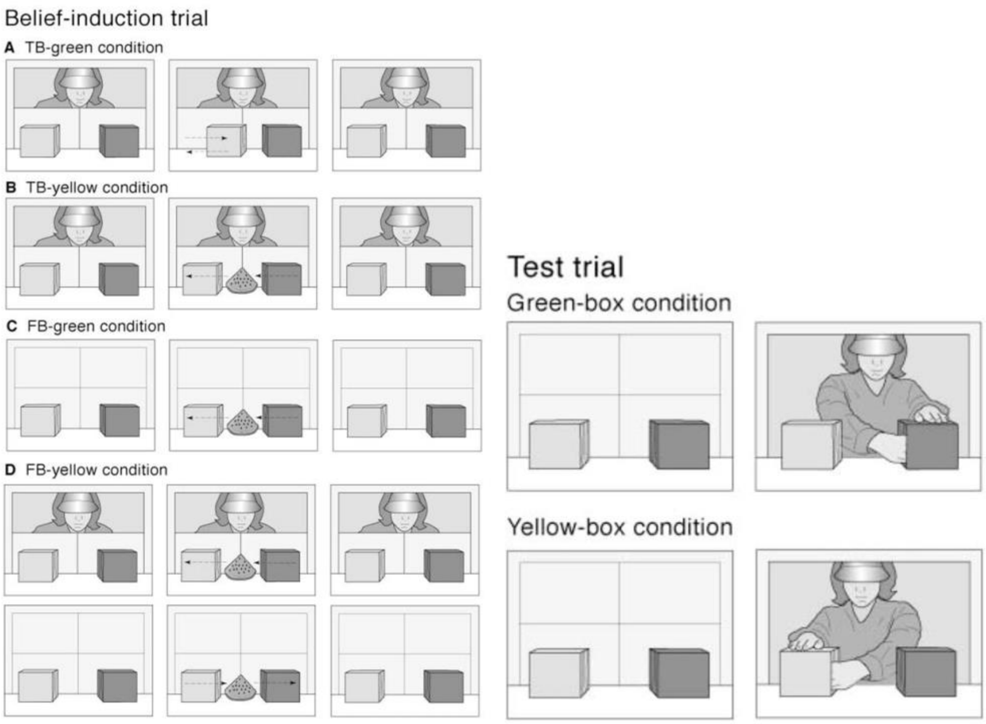


*Figure 8. Trials from Onishi and Baillargeon (2005) false belief experiment with 15-month old infants.*

Further support for offline processing in infants comes from motor planning studies (Claxton, Keen and McCarty, 2003). People approach objects at different speeds depending on what they are intending to do with the object, such as throwing it or placing it somewhere carefully. That is, depending on the future goal, agents plan their motor behaviour differently. Claxton et al. tested infants at the age of 10½ months, and they were found to exhibit the same behaviour. The study showed that they changed the speed at which they were approaching a ball, depending on whether they were going to place it into a narrow tube, or throw it into a tub in front of them. Given that we know that adults do this depending on what future goal they entertain, the best explanation for the same behaviour in infants is that they are also able to entertain a future goal. Again, this is a sign of offline processing, since infants would need to represent a non-current states of affairs.

Hence, we have good evidence that very young infants already possess offline processing, so there is a possibility that infants this age could also constrain their imagination, though this has not yet been tested.[18] Instead, experiments on constraining imagination have been conducted with slightly older children, and I will discuss these findings before coming back to a tentative suggestion about the case of infants.

### *5.2 The Reality Constraint*

One reason why children, rather than infants, have been used in this research is because it involves *verbal* paradigms.[19] Within this paradigm, there is evidence that children of the age of three are able to apply the Reality Constraint when imagining future scenarios. In a study by Atance and O'Neill (2005), three-year-olds were asked to imagine that they were going on a trip with their parents, and to select three items out of eight to bring with them. The items they could choose from were juice, raisins, sunglasses, Band-Aids, a book, a teddy bear, a telephone, and money. The children were then asked what they wanted to bring with them and why. Now, if the children were able to control their imagination using the Reality Constraint, we should expect that they decide to bring items with them that they might realistically need, such as Band-Aids, in case they tripped and fell. This could be reflected in the verbal answers they gave to why they decided to bring a certain item. If they were not able to constrain their imagination using the Reality Constraint, we should expect that they chose items that made less sense, and possibly also that their explanations for these were less realistic. Though note that it is possible that they might be able to apply to Reality Constraint, but without giving verbal evidence for this.

---

[18] It is worth noting that alternative explanations have been forwarded where it is argued that a theory of mind is not necessary for succeeding in these tasks (Perner and Ruffman, 2005; Heyes, 2014). However, my argument does not hinge on infants being capable of offline processing. The studies I offer in support of children using constraints in imagination concern older children (ages 3 – 5), and so my conclusion is also limited to these ages. A more tentative suggestion is that if infants are capable of offline processing, it might be the case that they could also employ constraints in imagination. This remains to be tested.

[19] There is ample evidence that children engage in imagination, with the emergence of pretend play around two years (Singer and Singer, 1990; McCune, 1995; Nielsen and Dissanayake, 2004), making up and engaging with fictional stories from early pre-school years (Applebee, 1978; Appleyard, 1990; Engel, 1995; Walker, Gopnik and Ganea, 2015), and sometimes having imaginary companions from the age of three (Taylor, 2001).

How did the children fare? Juice, money, and a phone were all about 20% likely to be chosen to be brought by the children, whilst the rest of the items were about 10% likely to be chosen, so this does not reveal any one item to be preferred. Of greater interest is that in 37% of the cases, children referenced the imagined future as a *reason* for why they had chosen a particular item. For example, they could say that they had chosen juice because 'when I am thirsty, I *will* drink it', or the Band-Aids '*in case* someone has an owie' (hurts themselves).[20] This study gives us evidence that children applied the Reality Constraint, since it shows that they tried to realistically imagine their future needs, as indicated by the items chosen and the explanations given. However, it should be noted that only 37% of the children demonstrated the use of the Reality Constraint.

But a limitation of this study is that is that language might be a confounding factor. It is possible that the three year olds were able to apply the Reality Constraint, but they failed to verbally express this (e.g. stating 'I don't know' as a reason for why they brought something). Another limitation is that all the items are realistic items for a child to bring on a trip; though it might not be realistic for an adult to bring a teddy bear, this is definitely something that is realistic for a child who might want comfort or something to play with. It is thus possible that the children were not really applying the Reality Constraint, but that they were instead constrained by the experimental design regarding what items they could imagine to bring. To better bring out whether children can employ the Reality Constraint, it would be useful to contrast realistic items with unrealistic items. Luckily, another study has been conducted which addresses this issue, though it still suffers the same limitation about language.

Atance and Meltzof (2005) conducted a similar study, testing children's ability to plan for the future. Here, children of three, four, and five years were tested. They were to imagine themselves in various outdoors locations (such as a desert) and asked what item they wanted to bring with them from a set of three. For example, in the desert-case, they could choose from soap, a mirror, or sunglasses. If children applied the Reality Constraint, we should expect them to choose sunglasses, since it is realistic to imagine that one might need sunglasses in the desert, whereas a mirror or soap are unlikely to be needed. In fact, this is what children chose. All age groups

[20] Notice that the second sentence-structure also references the uncertainty of the future, demonstrating that children are aware that the future is not set like the past. The primary purpose of this study was to investigate the children's linguistic abilities, in particular if they spoke about the *uncertainty* of the future ('I will bring Band-Aids because someone *might* get hurt'), but it also indirectly tested whether children are able to apply the Reality Constraint.

performed significantly above chance, with four-and five-year-olds performing significantly better than three-year-olds. Four-and five-year-olds also explained their choice with reference to future needs about two thirds of the time.

This replicates the results of the previous study, but it also tells us something new, as children at the age of three are able to employ the Reality Constraint, but children at the age of four do so more consistently. However, language could again be a confounding factor, and the improved performance in children at the age of four might reflect a better grasp of language rather than an improved ability to employ the Reality Constraint. Hence, three-year-olds might be able to exercise the Reality Constraint even if they are not able to express their reasons for their choice.

But a non-verbal study speaks against this interpretation.[21] Suddendorf and Busby (2005) tested three-, four-, and five-year-olds. They were first introduced to an empty room, in which the only item was a puzzle frame without any puzzle pieces. Then they were led into another room to play unrelated games for five minutes, before being presented with four items of which they had to choose one to bring back to the first room. One of these items were the puzzle pieces, which they could use in the first room. Results showed that four- and five-year-olds were more likely than the control group (which had not been introduced to the first room) to choose to bring the puzzle pieces, whereas three-year-olds were not. Nevertheless, though this study diminishes the possibility of language being a confounding factor as children were not required to provide verbal answers, it is still worth noting that instructions were given verbally. Hence, it is not possible to entirely exclude language as a confound, as children could have misunderstood instructions. They might for example not have understood that they would only return to the puzzle-room once, or that they could not at a later point bring back further items. Future research should attempt to utilise a completely non-verbal paradigm to exclude language as a confounding factor.

Despite these shortcomings, taken together, these studies indicate that children are able to employ the Reality Constraint at the age of four, though it is uncertain whether they can do so even earlier. For my purposes in this chapter, it demonstrates that employing the Reality Constraint is indeed a viable way of partially cashing out control in imagination.

[21] This is a non-verbal study since verbal *answers* were not required by children, but note that instructions to the children were still given verbally, meaning that children needed to have a developed linguistic capacity. It thus does not parallel the non-verbal false belief tasks discussed above, which require no linguistic competence at all.

### *5.3 The Change Constraint*

To the best of my knowledge, only one study tested whether children are able to apply the Change Constraint. McColgan and McCormack (2008) tested three-, four-, and five-years olds on whether they were able to place an item in the right place, given how a sequence of events were supposed to unfold. For example, children were asked to place a camera on a path along which a doll was going to walk. It was explained that the doll wanted to take a picture of some zoo animals which appeared further down the path, and the interesting question was whether children would place the camera before or after the zoo animals. If children were able to apply the Change Constraint, they should imagine that the world changes in different ways depending on whether they place the camera before or after the zoo animals. If they place it before, the doll would be able to get the camera and take the picture. If they place it after, she would not be able to obtain the camera and could not take the picture. Interestingly, only five-year-olds reliably passed this task, with both four- and three-year-olds failing.

This result indicates that five-year-olds are able to apply the Change Constraint, whereas three- and four-year-olds are not. However, it is also likely that language is a confounding factor here, especially as the instructions for this task were more complicated than instructions for the earlier tasks. Though children were asked if they understood the instructions, it is still possible that language processing got in the way of younger participants' being able to correctly complete the task.

### *5.4 Competence or Performance?*

This survey of studies indicates that cashing out control in terms of particular constraints that can be employed by the agent does enjoy empirical support. Hence, we now both have a plausible theoretical framework of what it means to control one's imagination, and we can also see that people can indeed control their imagination in these particular ways in order to obtain accurate imaginings. Further, these results indicate that there is an asymmetry between the development of the Reality Constraint and the Change Constraint, whereby ability to employ the Reality Constraint develops earlier than the ability to employ the Change Constraint. This points towards a dissociation between the two constraints, such that one could be competent at employing one but not the other.

Before moving on, a final note about these studies is in place. It is unclear why children below a certain age seem to fail to employ these constraints. Drawing a parallel with the research on false belief is informative, and two possible options have been discussed. Either, it could be due to their lacking the competence, that is, it has not yet developed (Perner, 1991; Sabbagh *et al.*, 2009; Gopnik and Wellman, 2012; Wellman, 2014). Or, it could be due to children underperforming because of factors, such as processing load or misunderstanding what the experimenter wants, despite actually having the competence (Leslie, 2000; Baillargeon, Scott and He, 2010; Carruthers, 2013; Helming, Strickland and Jacob, 2016). Why children fail tasks has been a debate in the false belief literature for some time now, especially since the view that children could not pass false belief tasks before the age of five was overturned by research showing that infants did pass non-verbal version of the false belief task. This starkly speaks in favour of the performance explanation, whereby children of three – and even infants – have the competence to attribute false beliefs, but they fail to do so in experimental conditions because of some other factors. Given that a similar age pattern has emerged for employing the Reality Constraint and the Change Constraint in imagination, with four- and five-year-olds more reliably passing than three-year-olds, and it is possible that this pattern could have a similar explanation. Perhaps it is the case that three-year-old can employ both constraints, but they fail to do so, say, because of processing load. Even more interesting is the prospect of finding out whether infants could pass a non-verbal task of employing the constraints. This should be tested in future research.

### *5.5 Is Imagining a Skill?*

Summing up, we now have good evidence to think that imagination can indeed be controlled in particular ways, and that greater control leads to a more accurate imagining. In particular, an agent can employ the Reality Constraint and Change Constraint in order to obtain an accurate imagining. We further have reasons to think that agents indeed can employ these constraints, and that the ability to employ them develops during early childhood. A question remains about whether infants can also employ these constraints, and given the research on false belief attribution, one might think that this is plausible.

In total, the empirical evidence cited provides support for my argument that imagination is indeed a skill. I have now not only shown that imagination can be improved by practice, but also that it can be controlled, meaning that imagination meets the two hallmarks of skill identified in the literature. As these are good indicators that something is a skill, I conclude that imagination is indeed a skill.

## 6. Improving and Controlling Imagination via Epistemic Feelings

But the account now faces a how-possibly challenge. A common theme from both philosophy and psychology when it comes to giving sub-personal accounts of control is to focus on how control is implemented in a *selection mechanism*, and how the functioning of this mechanism can lead to an improved skill. For example, in her work on bodily actions, Fridland (2014) argues that control involves a mechanism that selects the relevant features of the environment. For example, a batsman in cricket focuses their attention on the relevant features of the bowler's delivery to determine which shot to play. When this selection mechanism selects better attentional foci, the skill is improved. Note here that this links together control and improvability, such that the mechanism responsible for sub-personal control can be improved, resulting in an overall improved skill. This, of course, gives rise to the following puzzle: *how* can control be implemented at the sub-personal level in the case of imagination, and how does this enable improvement?

An interesting answer to the first question comes from Pezzulo and Castelfranchi (2009), who developed an account of control for *mental actions* which they apply to imagination specifically. They propose that controlled imagination involves a selection mechanism selecting relevant representations from memory to form imaginings, which are then compared to the agent's goal. When imagining something, you typically have a goal with that imagining, for example to imagine something accurately in order to solve a problem. Consider their example of a person whose goal is putting some beers in the fridge, and who is trying to solve the task of how to open the fridge door with their hands full, using imagination. Pezzulo and Castelfranchi argue that control is exercised here as follows: the selection mechanism produces an imagining by selecting a relevant content from memory, which is then judged for its aptness. For example, the person can imagine opening the fridge door with their foot, and compare this to their goal representation to find out if it is satisfied. If the answer is 'no', they can use this feedback to form a new, improved imagining.

I think that Pezzulo and Castelfranchi are on the right track, but their proposal only gets us so far. A particular problem with their account is that they do not specify *how* the control mechanism can select new content to form an improved imagining. In the remainder of this chapter, I argue that the Bayesian selection mechanism developed in chapter 3 provides a better account of control as selection when it comes to imagination, and that positing it can explain how imagination can improve, as it can tell us *how* different content is selected to be recombined for imaginings, and hence how more apt content can be selected in order to produce more accurate imaginings. This means that my account can make sense of the connection between control and improvability, something which is rightly stressed by Fridland, but which is lost on Pezzulo and Castelfranchi's account.

Recall that the Bayesian selection mechanism assigns likelihoods and priors to content, determining how likely it is to be retrieved for recombination. The priors, I suggested, are set depending on the emotional intensity of a particular content, such that something with higher emotional intensity is assigned a higher prior. Likelihoods, on the other hand, are assigned depending on the agent's intention. That is, if an agent intends to imagine a birthday party, content relevant to this are assigned higher likelihoods. Here, I will make an additional suggestion to this proposal: better imaginings are generated as the result of likelihoods changing, and likelihoods change because of an agent's so-called epistemic feelings. Let me elaborate on this.

I suggest that imaginings are accompanied by *epistemic feelings,* and that these play a causal role in changing the likelihoods of content being selected for recombination. The particular feelings that I will focus on are the and the *feeling of error* and the *feeling of certainty*. I suggest that if the agent's imagining is accompanied by a feeling of error, the agent is likely to decide to revise the imagining, with the likelihood of the memory content that caused the error being readjusted, so that a different imagining is produced the next time. I also tentatively suggest that if an agent's imagining is accompanied by a feeling of certainty, the agent believes the imagining to be realistic, and likelihoods are not adjusted, since an optimal imagining might already have been produced. Before getting further into how this works, let me introduce epistemic feelings properly.

### *6.1 Epistemic Feelings*

Epistemic feelings are feelings about an agent's own mental capacities or mental processes (Koriat, 1993; Winkielman *et al.*, 2003; Dokic, 2013; Proust, 2013; Arango-Muñoz, 2014; Michaelian and Arango-Muñoz, 2014). These feelings have affective phenomenology which can be positive or negative, depending on the particular feeling. The study of epistemic feelings is a new line of research in philosophy, and has so far received a lot of attention from philosophers such as Proust, Arango-Muñoz and Michaelian (Dokic, 2013; Proust, 2013; Arango-Muñoz, 2014;

Michaelian and Arango-Muñoz, 2014; Fernandez Cruz, Arango-Muñoz and Volz, 2016). Though epistemic feelings such as the feeling of certainty have recently attracted attention in philosophy, they have been commonly investigated and thought to be about other mental states in psychology for some time. For example, it is common practice to ask eyewitnesses about their confidence about their memory (Loftus and Palmer, 1974), and participants in imagination studies are often asked to rate their confidence in the vividness they report of imagery (Keogh and Pearson, 2018). Many different kinds of epistemic feelings have been discussed, including the feeling of forgetting, the feeling of knowing, the feeling of certainty, and the feeling of error (for an overview, see: Arango-Muñoz, 2014). The ones that are applicable to imagination are the feeling of error and the feelings of certainty, so I will illustrate the general argument for the existence of epistemic feelings using these, before I move on to their function and applicability to imagination.

Arango-Muñoz and Michaelian (2014) illustrate epistemic feelings by giving an example of Santiago losing his passport when returning from a holiday. When arriving at the airport, he discovers that he does not have his passport, even though he is confident that he had seen it the night before as he was packing. He finds it puzzling that he does not have it with him, since he is confident in his memory of having packed it. Later, he receives a phone call from the hotel that he stayed in, saying that the passport has been found in his room. It seems like he had accidentally knocked it off the table where he kept his hand luggage when he left. The epistemic feeling in this case is thus a *feeling of certainty.* Another epistemic feeling is the *feeling of error*. This could for example arise when a person has made a calculation that they suspect is wrong, or after a faulty reasoning process. We can consider the case of Santiago again. In the above example, it could have been the case that the Santiago's memory was accompanied by a feeling of error, rather than a feeling of certainty. If this were the case, he would not have trusted his memory, and instead not come to believe that he had not packed his passport. Epistemic feelings are metacognitive feelings, that is, feelings *about* another mental state or process. In the case of Santiago, the feelings of certainty and error are feelings about another mental process – the memory process.

To my knowledge, it has not been disputed *that* we have these kinds of feelings. What is disputed, however, is what the function of epistemic feelings is.[22] It has been suggested that the function of

[22] It might be more apt to talk about the function*s* of epistemic feeling*s*, as it is not clear that they all have the same function, or that we have identified a natural kind. I am open to the possibility that different epistemic feelings have different functions.

epistemic feelings is to evaluate a subject's cognitive process for success of failure.[23] It is commonly argued that epistemic feelings can fulfil this role by being sensitive to cues or heuristics of cognitive processes (Winkielman *et al.*, 2003; Schwartz and Metcalfe, 2011; Proust, 2013) For example, it has been suggested that they might be sensitive to cue familiarity or fluency.[24] An agent who has a feeling of certainty about retrieving a memory might thus not have this feeling in virtue of the retrieved content, but rather in virtue of the fluency of the retrieval process. Exactly what epistemic feelings are sensitive to is not something which I will get deeper into here, as I am only interested in sketching a potential role for epistemic feelings in improving imagination. As my main focus is on the feeling of error, and this has been suggested to be sensitive to the fluency of a process, I will likewise subscribe to this idea for now (Fernandez Cruz, Arango-Muñoz and Volz, 2016).

### *6.2 Epistemic Feelings and Imagination*

Here is how I think epistemic feelings help improve imagination. Suppose that Sally imagines going to Peru on holiday in order to decide whether to do so. She imagines seeing all the beautiful natural sites in the mountains, and she imagines having a great time there. Her imagining is accompanied by the epistemic feeling of confidence. Now picture Matilda who imagines the same thing about herself. However, her imagining is accompanied by the feeling of error. It does not seem right to her that she will have a great time up in the Peruvian mountains, however beautiful they may be. Having realised this, she might start to think about *why* her imagining seems inaccurate to her, and realise that she has suffered from altitude sickness in high places before. This realisation might prompt her to produce a different imagining where she is not having such a great time in the mountains. This hopefully seems like a fairly intuitive suggestion.

Here is the more controversial suggestion. These feelings are significant to *the improvement of our capacity to imagine* because they can affect the likelihoods of content being retrieved. Recall, a particular imagining is produced as a result of the agent's intention, where this intention sets

[23] This evaluative proposal might seem like a strange suggestion for some epistemic feelings, like the 'tip of the tongue' feelings. However, my argument here does not require that all epistemic feelings are evaluative – only that the feelings of certainty and the feeling of error are – nor does it require that all phenomena that have been suggested to be epistemic feelings actually turn out to actually be so.

[24] There is a live debate about what epistemic feelings are sensitive to. For views advocating heuristics such as fluency, see Winkielman et al. (2003) and Duke et al. (2014). For an argument against this, see Greely (2021).

likelihoods for different memory contents being used to produce an imagining. My suggestion is that epistemic feelings play a causal role in changing these likelihoods, and thereby make the agent produce a new different imagining. So, what happens on the sub-personal level? Take the example of Matilda, who suffers from altitude sickness. As posited, Matilda has a memory of having altitude sickness when in the mountains. However, her first imagining does not incorporate this, due to the likelihood of that memory content being initially ranked as low.[25] But she has a feeling of error about this imagining. This feeling of error can causally affect the likelihood of content being selected, and change the likelihood of the content of altitude sickness, such that it now becomes very likely to be used to produce this imagining. Hence, in forming the new imagining, if the likelihood of the altitude sickness content is high enough, Matilda might produce an imagining of experiencing altitude sickness in the mountains. Her new imagining is thus more accurate than the old one, and we can say that her imagination has improved with respect to her goal.

But there is an obvious worry with this proposal, namely that epistemic feelings might not be reliable. If so, we have no reason to think that epistemic feelings could *improve*, so much as just *change*, what imagining is being produced. For example, if the feeling of error is only correct 50% of the time, a feeling of error should not be taken as a sign that someone is actually incorrect, as the feeling is not accurate above chance level. Whether epistemic feelings are reliable *tout court* is an empirical question that largely remains to be tested. Luckily, there is some empirical evidence for the feelings of error reliably indicating errors actually having been committed. For this reason, I only commit my account to feelings of error being able to reliably indicate that an inaccurate imagining has been produced, whilst I do not commit my account to that a feeling of certainty indicates that an accurate imagining has been produced. Let us have a closer look at the feeling of error.

The feeling of error was investigated in an experiment by Fernandez Cruz et al. (2016), which focused on whether the feeling of error reliably indicates committing a mathematical error. In the study, participants were asked to solve a so-called number bisection task (NBT) where they judged whether the number in the middle of a triplet was the mean of the numbers on either side. For example, the stimulus might be '07_16_25', and the participant was supposed to judge whether

[25] There are a number of reasons which can explain why the memory content is initially ranked as low. For example, she might have been primed by beautiful photos of the mountains, or recently heard testimony from people who really enjoyed visiting Peru. The recency or intensity of these memories could make the likelihood of these memories much higher than the likelihood of the memory of altitude sickness.

16 is the mean of 7 and 25. The stimulus was presented on a computer screen for 2.5 seconds. After solving this task, participants were asked whether they thought they had made an error ('Do you think you have committed an error?') and given 2 seconds to indicate a response. If they answered that they did not think they had committed an error, they were asked to rate their confidence about their answer to the NBT on a Likert-scale from 1-6. Results found that there was a significant positive correlation between the feeling of error and being incorrect (see Figure 9).

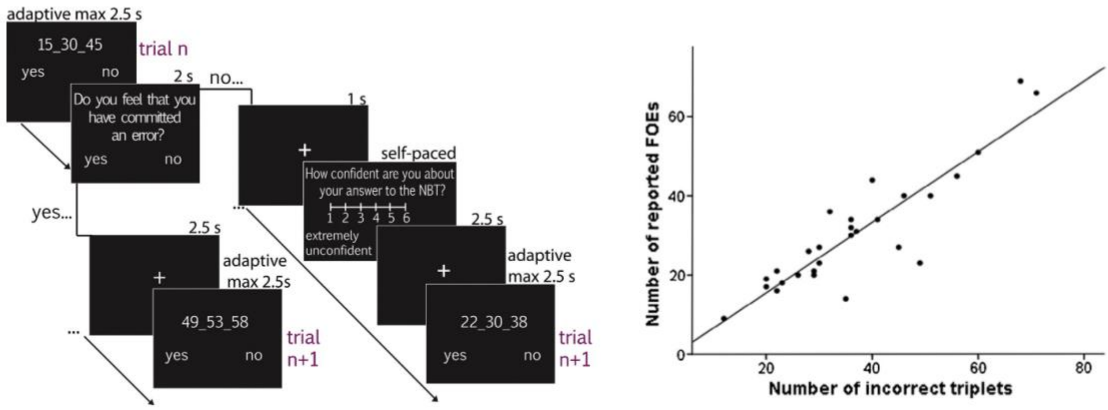


*Figure 9. Experimental design and results from Fernandez Cruz* (2016). *The results show the Pearson correlation of the number of incorrect answered triplets and reported feelings of error (r = 0.896, p < 0.001, n = 28). Each point represents the averaged number of FOEs and incorrect triplets, per participant, over the 190 trials.*

Though no study, to my knowledge, has tested whether people have a feeling of error about imaginings, this result lends plausibility to the idea that the feeling of error reliably indicates actually being wrong. If so, it is possible that the feeling of error could reliably indicate being wrong about the accuracy of imaginings too. Notice though, that the converse does not follow: we cannot infer from the results of this study that the *lack of* a feeling of error indicates being correct. We can thus not say that being confident about being *correct* about the answer to a mathematical problem or the accuracy of an imagining indicates that a subject is correct, nor can we necessarily infer that the mere lack of a feeling of error indicates correctness. To be clear, these results only indicate that the feeling of error reliably indicates being wrong, and as such, my suggestion is only that the feeling of error could indicate being wrong about the accuracy of imaginings too.[26]

[26] Some older studies, such as Loftus and Palmer's (1974) eyewitness study, suggest that other epistemic feelings, such as the feeling of certainty, might not be reliable. But it should be noted that this study used leading questions, which are likely to have affected participants' response and level of certainty about the correctness of the response.

The more tenuous suggestion is that the feelings of confidence might indicate being accurate. However, as there is no empirical evidence supporting this claim, no argument can be made. This is not detrimental for my account. Recall, this account is supposed to tell a story about how imagination can be improved. It is pivotal that something like a feeling of error plays a causal role in changing the likelihoods such that an agent can produce a new imagining that is different to the previous one, and that the likelihoods are changed in the right kind of way such as to produce a more accurate imagining. This constitutes improvement. But it is not crucial for this argument that the agent then feels confident about the accurate imagining, as I do not suggest that this plays a role in improvement.

The hypothesis concerning the feeling of error needs to be tested. A design similar to Fernandez Cruz et al.'s could be used for this purpose. But since I am interested in improvement, it would also be beneficial to see whether an agent produces a *better* imagining following a feeling of error. The idea that this would be the case is initially plausible, as having some kind of feedback loop to adjust likelihoods has proven useful in improving other Bayesian systems, as in machine learning, and it therefore looks viable as a way of improving imagination too (Gopnik and Tenenbaum, 2007). This account can thus also account for improvement sub-personally by appealing to the likelihoods being changed, and hence a new imagining being generated. Hence, CESH+ illuminates the account of imagination as a skill, since it can give an account of control as selection being implemented in the Bayesian selection mechanism, and tie this together with improvability.

### *6.3 The Possibility of Inaccurate Epistemic Feelings*

Here is a worry about my proposal that the feeling of error is causally effective in improving imagination. One might think that an agent might have the *wrong* epistemic feeling about their imagining, thus hindering improvement. That is, Matilda might have an epistemic feeling of certainty about her inaccurate imagining of enjoying the mountains, but to improve her imagination, she would need to have a feeling of error. If this is the case, she is unlikely to produce a new imagining and thus cannot improve her imagination. This is certainly a problem for how to *actually* improve one's imagination, but it misses the mark as an objection to my account. What I have argued here is that the feelings of error can serve as a way to improve one's imagination, but this does not entail that an agent always experiences this feeling in the right circumstances. In fact, there are many examples of ways of improving other skills, where agents fail to have the right feeling or attitude. For example, practising scales improves guitar playing, but since many think that practising scales is boring, they do not do it. But the fact that agents do not practise scales is not an objection to guitar playing improving through practising scales. An agent not practising scales is compatible with this. Similarly, an agent not having the feeling of error about an inaccurate imaging is compatible with my account. All I need for my account is that it is statistically more likely that an agent experiences the feeling of error when an error has indeed been committed, as shown by Fernandez Cruz et al., but this does not entail that there will not be cases where an agent does not have a feeling of error, despite being wrong. A case of an agent having the wrong feeling is therefore compatible with my account.

Another more pressing objection is that even though it has not been tested whether people have a reliable feeling of error about inaccurate imaginings, many imaginings themselves do seem inaccurate. As mentioned, studies from affective forecasting highlight how people are in fact systematically wrong when imagining the intensity of their future emotions (Wilson and Gilbert, 2000; Wilson *et al.*, 2000; Wilson, Meyers and Gilbert, 2003; Gilbert and Wilson, 2007). Since these inaccurate imaginings are nonetheless reported by participants in the studies, and we have no reason to think they thought they had committed an error, one might think that a large number of the population does not have feelings of error about inaccurate imaginings. This is more worrying, since it would be a problem for my account if it is statistically unlikely that people in general have a feeling of error about inaccurate imaginings.

There are a few things to say about this. Firstly, it is worth pointing out that these studies are conducted in low-stakes scenarios where participants are *not* asked to revise their imagining, nor are they asked to rate the confidence in their imagining or whether they had a feeling of error. It is thus possible that most participants' imaginings are accompanied by a feeling of error such that they have a low confidence in it, but that they report on the imagining anyway as the experimental condition dictates it. Further, it is also possible that participants perform poorly on this because they are in a low-stakes scenario, and not motivated to *really* try to imagine accurately. Studies investigating affective forecasting have focused on imagined future feelings in relation to football wins/losses, and these imaginings are inconsequential to the participants, as they do not, for example, involve any decision-making on the part of the agent (Wilson *et al.*, 2000). Even studies that sound like they deal with high-stakes cases such as getting a job, are in fact fairly low-stakes, since the job in question is a one-off job to test products and look at advertisements, paying $25 (Gilbert *et al.*, 1998, study 6). These studies do not currently pose a threat to my account, as they do not rule out that participants actually had a feeling of error despite reporting on an inaccurate imagining, and they are often low-stakes, meaning that participants potentially are not trying to control their imagination properly. Going forward, it would be interesting to also include a question about the feelings of error in affective forecasting experiments, such that we can find out whether it is the case that participants systematically get the intensity wrong, but also have a feeling of error about their imagining. This could inform further studies, which could investigate whether having a feeling of error is efficacious in revising one's imagining to improve imagination.

## 7. Conclusion

In this chapter, I identified *improvability by practice*, and *control* as two hallmarks of skill, and I argued that imagination meets both. Specifically, I showed that Kind's recent argument for the thesis that imagination is a skill is not well-supported by empirical evidence, as there could be confounding variables in the evidence Kind uses from sports psychology. To make a stronger case for the improvability of imagination, I instead provided evidence from mental rotation and auditory imagery studies which more clearly show that subjects improve with practice. Next, I raised the question of whether imagination can be controlled. As it was unclear what it means to control an imagining, I suggested that we understand control in terms of employing certain constraints to imagination. In the case of practical imagination, these are the Reality Constraint and Change Constraint. I then provided empirical evidence from developmental psychology, demonstrating that children at a young age have the capacity to employ these constraints, thereby controlling their imagination. I also raised the possibility that testing infants in non-verbal paradigms might show that control can be exercised at an even earlier age, paralleling how false-belief attribution has arguably been demonstrated in infants in a non-verbal paradigm. Since imagination is both improvable by practice, and can be controlled, this indicates that imagination is indeed a skill.

I finally tied this proposal together with CESH+, and showed how this cognitive architecture can shed light on how exactly imagination can be controlled and improved. My proposal is that control should be cashed out as selection, and that the selection mechanism, which is responsible for selecting and ranking memory content to be recombined into imaginings, can improve as the result of an agent's feeling of error about the retrieval process. This changes the likelihoods of certain content to be retrieved, enabling other content to be selected for recombination instead, hence potentially producing a more accurate imagining. Importantly, this does not only provide a more exact account of how imagination can improve, but it also provides further support for CESH+ as the model can illuminate the thesis that imagination is a skill. Because of this, we now have further reason to endorse both CESH+ and the Generative View. In the next chapter, I will similarly offer further support for these accounts, by showing that CESH+ can is the only model which can adequately give a cognitive explanation of aphantasia.

# Chapter 5

# Aphantasia: In Search of a Theory

## 0. Introduction

This chapter will use CESH+ to give a cognitive theory of aphantasia – a condition where people, amongst other things, are impaired with respect to generating voluntary visual imagery. I show that we can use CESH+ to see which processes malfunction and hence give rise to this condition. As we will see, this marks CESH+ out as the only theory which can currently explain all impairments in aphantasia, and hence, as giving the best explanation of the condition. Doing so indicates further that CESH+ is a plausible theory of the cognitive architecture of memory and imagination, and it favours the Generative View of memory.

Until recently, it has been commonplace to assume that everybody has the capacity to voluntarily generate mental imagery. But since 2015, an increasing number of people who are *unable* to do so have been identified – this condition has become known as *congenital aphantasia*.[1] Despite the great attention it has received from researchers (Greenberg and Knowlton, 2014; Zeman, Dewar and Della Sala, 2015; Fulford *et al.*, 2018; Jacobs, Schwarzkopf and Silvanto, 2018; Keogh and Pearson, 2018, 2021; Pearson, 2019; Bainbridge *et al.*, 2020; Dawes *et al.*, 2020; Zeman *et al.*, 2020; Milton *et al.*, 2021; Nanay, 2021) and media, we still do not know much about this condition.[2] Not only have very few explanatory theories of aphantasia been proposed (Pearson, 2019; Nanay, 2021), but it even remains unclear which cluster of impairments characterise the condition in the first place.

---

[1] Galton (1883) first documented the condition in 1883, but no modern research was conducted until 2010. A distinction is made between acquired aphantasia (Zeman *et al.*, 2010) and congenital (lifelong) aphantasia (Zeman, Dewar and Della Sala, 2015; Fulford *et al.*, 2018; Milton *et al.*, 2020; Zeman *et al.*, 2020). I limit my discussion to congenital aphantasia, since acquired aphantasia is extremely rare and could be accompanied by further cognitive impairments, and I use the shorthand 'aphantasia'.

[2] See for example an article in the Scientific American (https://www.scientificamerican.com/article/when-the-minds-eye-is-blind1/) and an article by the BBC (https://www.bbc.co.uk/news/health-34039054). Both accessed on 7/1/2021.

For example, some claim aphantasia primarily involves a *visual* imagery impairment, selectively impairing the generation of visual imagery (Greenberg and Knowlton, 2014; Fulford *et al.*, 2018; Keogh and Pearson, 2018; Bainbridge *et al.*, 2020; Zeman *et al.*, 2020; Milton *et al.*, 2021). Others claim that there are further impairments associated with the condition, which affect other forms of imagery too, such as auditory and olfactory imagery, as well as other impairments related to episodic memory (Zeman, Dewar and Della Sala, 2015; Jacobs, Schwarzkopf and Silvanto, 2018; Pearson, 2019; Dawes *et al.*, 2020; Nanay, 2021). There is also disagreement about whether aphantasia only affects the production of *voluntary* imagery, as when intentionally imagining (Zeman, Dewar and Della Sala, 2015; Jacobs, Schwarzkopf and Silvanto, 2018; Keogh and Pearson, 2018; Pearson, 2019; Bainbridge *et al.*, 2020; Dawes *et al.*, 2020), or if it also affects *involuntary* imagery, such as imagery generated when dreaming (Greenberg and Knowlton, 2014; Fulford *et al.*, 2018; Zeman *et al.*, 2020; Milton *et al.*, 2021; Nanay, 2021). Most importantly, it remains unclear whether aphantasia is a condition resulting from a malfunction in a system producing visual imagery (Pearson, 2019), or if it results from a malfunction in a different system.

The lack of significant progress towards a theory of aphantasia, I contend, is the result of a piecemeal approach: so far, different researchers have focused on different subsets of data, but there has been no overarching project of drawing all the available data together into a theory of aphantasia. This has hampered the possibility of giving an *explanation* of the impairments as resulting from a malfunctioning of a cognitive system. In cognitive science, there is a well-established tradition of providing theories of cognitive architecture that describe how the mind is structured by positing internal mechanisms, and to explain impairments with respect to the malfunctioning of these mechanisms (Fodor and Pylyshyn 1988; Newell 1990; Anderson 1993; Nichols and Stich 2004). In this chapter, I seek to provide a better understanding of aphantasia by offering such a cognitive explanation of the condition.

I thus aim to do three things in this chapter. First, after illustrating the current confusion of tongues in aphantasia research in §1, I examine the data from recent studies on aphantasia and show that they cluster neatly into six robust data points in §2. Accordingly, I propose that a theory of aphantasia ought to explain the following:

> (1) the impairment in generating voluntary visual imagery;
>
> (2) the impairment in generating mental imagery with respect to different sensory systems;

(3) the differential impairment in producing voluntary imagery and involuntary imagery;

(4) the impairment in recalling episodic memory details;

(5) the impairment with respect to content generation for both atemporal events and future events;

(6) the retained ability to solve spatial imagery tasks and score averagely on spatial imagery questionnaires.[3]

Secondly, in §3, I discuss two recent accounts of aphantasia, namely, Nanay's (2021) account involving unconscious imagery, and Pearson's (2019) account based on the cognitive architecture of visual imagery, and I show that neither of them can explain (1) - (6). Both of these views are versions of the Perceptual View, whereby visual imagery production is dependent on early visual processing, introduced in chapter 1. Finally, in §4, I put forward a novel theory of aphantasia. This theory is based on the CESH+ architecture forwarded in previous chapters, and I show that the cluster of impairments in aphantasia can be explained by the malfunctioning of different episodic retrieval processes, making aphantasia an *episodic system condition*. This represents a major step forward in the research on aphantasia, and it lends significant support to CESH+, as it is the only theory which can cogently explain all impairments.

## 1. Definitions of Congenital Aphantasia

Let us begin by taking a look at what definitions of 'aphantasia' are currently used in the literature (see Table 4):

[3] These points are not necessary and sufficient conditions for having aphantasia, but rather characterise the condition.

| | Authors (year) | Definition of Aphantasia |
|---|---|---|
| 1. | Greenberg and Knowlton (2014) | total congenital absence of visual imagery |
| 2. | Zeman et al. (2015) | reduced or absent voluntary imagery |
| 3. | Keogh and Pearson (2018) | inability to create visual images in one's mind |
| 4. | Fulford et al. (2018) | lifelong absence of visualisation |
| 5. | Jacobs et al. (2018) | the congenital inability to experience voluntary mental imagery |
| 6. | Pearson (2019) | lack of the ability to voluntarily form mental images |
| 7. | Milton et al. (2021) | lifelong lack of visual imagery |
| 8. | Zeman et al. (2020) | lifelong absence of mind's eye |
| 9. | Dawes et al. (2020) | absence of voluntarily generated internal visual representations |
| 10 | Bainbridge et al. (2020) | inability to create voluntary visual mental images |
| 11. | Nanay (2021) | no conscious mental imagery |
| 12. | Keogh and Pearson (2021) | inability to visualise |

*Table 4. Definitions of aphantasia used in the literature.*

As should be clear, a first major point of disagreement is whether people with aphantasia are impaired with respect to *visual* imagery only. Recall that it is widely accepted that there are many kinds of imagery other than visual imagery, such as:

- Auditory imagery (Okada and Matsuoka, 1992; Halpern *et al.*, 2004; Zatorre, Halpern and Bouffard, 2010; Herholz, Halpern and Zatorre, 2012);
- Olfactory imagery (Moustafa Bensafi et al. 2003; Bensafi and Rouby 2007; Flohr et al. 2014);
- Gustatory imagery (Croijmans *et al.*, 2019);
- Motor imagery (Munzert and Lorey, 2013; Guillot, 2020);
- Affective imagery (Wilson and Gilbert, 2000; Gilbert and Wilson, 2007; Blackwell, 2020);
- Spatial imagery (Shepard and Metzler, 1971; Whitwell *et al.*, 2015);
- Temporal imagery (Viera and Nanay, 2020)

Stating that aphantasia is a condition where only visual imagery is impaired (as definitions 1, 3, 4, and 7 – 10, and 12 do) implies that aphantasics could perhaps generate all other kinds of mental imagery. This conflicts with what is stated in definitions 2, 5, 6, and 11, which use the all-encompassing term 'mental imagery'. It thus appears that there is no consensus about whether people with aphantasia are only (or primarily) impaired with respect to visual imagery, or if this impairment clusters with other mental imagery impairments.

Secondly, while it is common to make a distinction between the generation of voluntary and involuntary imagery (Dorsch, 2015; Jakubowski, 2020; Pearson, 2020b), the above definitions often do not specify which of these two abilities aphantasics supposedly lack. For example, definitions 1, 4, 7, 8, and 11 do not make this explicit, thus allowing for both involuntary and voluntary imagery to be affected, whilst definitions 2, 3, 4, 5, 6, 9, 10, and 12 explicitly state an impairment in only voluntary imagery. Again, we lack a precise description of the type of impairment involved in aphantasia.

A final problem is that all these definitions tacitly assume that aphantasia is mainly, if not exclusively, a problem of generating *imagery*. That is, they presuppose that the core impairment in aphantasia, if not the only impairment, is an impairment in producing imagery (visual or otherwise, voluntary or otherwise). This, as I will show, goes against a large body of data indicating that aphantasic subjects exhibit a *cluster* of cognitive impairments, which are not limited to impairments involving imagery. It would be a mistake to assume from the outset that these impairments are not central to aphantasia.

These three problems, I maintain, are symptomatic of a more serious issue: the research on aphantasia has so far been piecemeal, with each study providing a new definition based only on its own data. If we want to provide an adequate explanation of aphantasia, we ought to instead analyse *all* the available data. This is what I do next. As far as I know, this is the first time an exhaustive review of data on aphantasia is presented.

## 2. Empirical Data on Congenital Aphantasia

Below, I present the data from all studies on aphantasia to date. My review follows the common practice of operationalising aphantasia in terms of scoring below a certain threshold on the

Vividness of Visual Imagery Questionnaire (VVIQ) (Marks, 1973).[4] This questionnaire asks subjects to form a voluntary *visual* image, and aphantasia is thus operationalised in the literature in terms of an impairment in *voluntary visual imagery*. In turn, I consider data on: voluntary visual imagery (2.1), non-visual imagery (2.2), involuntary visual imagery (2.3), memory (2.4), atemporal and future imagination (2.5), spatial imagery (2.6), and prosopagnosia (2.7). I will argue that the findings related to the first five subsections are robust enough to indicate a cluster of impairments in aphantasia.

### *2.1 Voluntary Visual Imagery*

All studies on aphantasia have administered the VVIQ and established that subjects are impaired with respect to voluntary visual imagery (Greenberg and Knowlton, 2014; Zeman, Dewar and Della Sala, 2015; Fulford *et al.*, 2018; Jacobs, Schwarzkopf and Silvanto, 2018; Keogh and Pearson, 2018, 2021; Bainbridge *et al.*, 2020; Dawes *et al.*, 2020; Zeman *et al.*, 2020; Milton *et al.*, 2021). Recently, there have also been some experimental findings pointing in the same direction.

Three experiments (Keogh and Pearson, 2018, 2021, experiment 3 and 4) (n = 15, n = 10, n= 15, respectively) have used a binocular rivalry paradigm, showing that aphantasics demonstrate no priming effect following a visual imagery condition, whereas controls did.[5] As this paradigm is integral to an argument later in the chapter, an in-depth discussion is provided in §3.1.1. For now, it suffices to say that the three experiments provided support that aphantasics are impaired in generating voluntary visual imagery.

---

[4] The VVIQ asks subjects to form a voluntary *visual* image. The maximum score on the VVIQ is 80/80, and the minimum is 16/80, where a subject would have answered 'no image at all' on all questions. The threshold for counting as aphantasic varies. Some studies use 16/80 (Fulford et al., 2018; Zeman et al., 2015), other studies use ranges, such as 17-30 (Zeman, Dewar and Della Sala, 2015), or 16-23 (Zeman *et al.*, 2020; Milton *et al.*, 2021), or 16-25 (Bainbridge et al. 2020). Some studies also make further distinctions between groups of aphantasics – notably Zeman et al. (2015) distinguishes between subjects who score 16 ('no imagery'), and subjects who score between 17-30 ('limited imagery'), and Zeman et al. (2020) distinguishes between subjects who scored 16 ('extreme aphantasia), and subjects who scored between 17-23 ('moderate aphantasia'). Some studies do not report what operationalised definition was used (Greenberg and Knowlton, 2014; Keogh and Pearson, 2018; Dawes *et al.*, 2020). Keogh and Pearson (2021) used self-ascribed aphantasics, though the VVIQ was also administered.

[5] A forthcoming study by Wicken et al. (*forthcoming*) also uses the same paradigm, but as a peer-reviewed draft is not yet available, I do not discuss this data.

One study also carried out a further experiment on voluntary visual imagery (Keogh and Pearson, 2021, experiment 4). This experiment tested whether participants could form so-called attentional templates. These are templates based on visual imagery, which include spatial and object information, and they are thought to aid our attentional performance (Treisman, 2006; Battistoni, Stein and Peelen, 2017). Results showed that aphantasics showed no evidence of being able to form attentional templates, confirming their inability to form voluntary visual imagery.[6]

Based on the results from the VVIQ and these experimental results, we need to explain the following:

(1) the impairment in generating voluntary visual imagery.

### *2.2 Non-Visual Imagery*

In Zeman et al.'s (2015) study, 10/21 aphantasics reported that all their sensory systems were affected, such that they could not voluntarily produce mental imagery in any of them. These results were replicated in Zeman et al. (2020). This study had a total of 2000 participants, comprising of aphantasics and hyperphantasics, though the sizes of the subsets were not specified.[7] Results found that 54.2% of aphantasics reported that *all* their sensory systems were seriously affected. 'Extreme aphantasics' were also more likely than 'moderate aphantasics' to report all their sensory systems affected (see footnote 4 for operational definitions).

Dawes et al. (2020) reported similar results based on the Questionnaire upon Mental Imagery, in which participants are asked to rate the vividness and clarity of voluntary imagery in different sensory systems (visual, auditory, tactile, kinaesthetic, gustatory, olfactory, affective) on a 7-point scale for 35 items. Here, results showed that 26.2% reported a complete lack of imagery for all sensory systems, and 73.8% reported overall significantly reduced imagery in all sensory systems compared to controls, but still some degree of non-visual imagery.

---

[6] Despite this, aphantasics do not have any impairments with respect to feature-based attention (Keogh and Pearson 2021, experiment 3).

[7] Hyperphantasia is often studied together with aphantasia, and it is characterised by particularly vivid mental imagery as indicated by extremely high scores on the VVIQ (Zeman *et al.*, 2020; Milton *et al.*, 2021).

Thus, more than half of aphantasics report reduced mental imagery in all sensory systems; and up to 26.2% report a total absence of mental imagery in all sensory systems.[8] The second data point to be explained is thus:

> (2) the impairment in generating mental imagery with respect to different sensory systems.

*2.3 Involuntary Imagery*

Though aphantasics cannot form voluntary mental imagery as indicated by the VVIQ scores in all studies, a few studies have reported that they *can* form involuntary mental imagery. In particular, studies have asked about 'flashes of visual imagery' (Zeman, Dewar and Della Sala, 2015), daydreaming (Dawes *et al.*, 2020), or night-time dreams (Dawes *et al.*, 2020; Zeman *et al.*, 2020).[9]

In the first study of congenital aphantasia, Zeman et al. (2015) administered a set of questions to 21 aphantasics.[10] They found that 10/21 participants reported involuntary flashes of imagery, and 17/21 reported visual dreaming. Thus, about half of the participants reported one form of involuntary imagery (flashes) and more than half reported another form of involuntary imagery (dreaming).

This finding has been replicated in a larger study by Zeman et al. (2020), using the same questions. This study divided aphantasics into 'extreme aphantasics' and 'moderate aphantasics' (see footnote 4 for operational definitions). Here, 63.4% of all aphantasics reported dreaming, but 'extreme aphantasics' were significantly less likely to report this than 'moderate aphantasics'.

---

[8] We have reason to believe that this is an underestimation of how many people have impairments to non-visual imagery, since the data reported here is only taken from a subset of those who have a visual imagery impairment as indicated by the VVIQ. It would not capture, say, a person who is *only* impaired when it comes to olfactory imagery, and they would thus be excluded from the sample. However, we do have independent evidence that show that there are in fact people who have only an olfactory imagery impairment, without any impairment in visual imagery. This was demonstrated by Bensafi and Rouby (2007) who, based on the VVIQ, developed the Vividness of Olfactory Imagery Questionnaire (VOIQ). Testing subjects on both the VVIQ and the VOIQ, they founds some subjects who scored normally on the VVIQ, but below average on the VOIQ. Thus, we should note that our current data likely underestimates the cases of non-visual imagery impairments.

[9] See Whiteley (2020) for an article discussing aphantasia and the nature of dreaming.

[10] See Zeman et al.'s (2015) Supplementary Data for these questions.

Further, about 30% of all aphantasics reported brief flashes of visual imagery, again with a similar significant difference between 'extreme aphantasics' and 'moderate aphantasics'.

Finally, these findings were replicated by Dawes et al.'s (2020) study of 267 aphantasics. They used the Imaginal Process Inventory (IPI) with 24 items assessing the frequency of daydreams and night dreams on a 5-point agreement scale; as well as the Subjective Experiences Rating Scale (SERS) comprising of 39 questions assessing participants' night dreams on a 5-point scale (these questions concerned sensory, affective, and cognitive components, as well as spatial complexity, perspective, and lucidity). Aphantasics reported experiencing significantly fewer night dreams than control participants, and that the dreams were also of qualitative difference.[11] Aphantasics' dreams were impaired across all sensory aspects, and they reported a lower sense of awareness and control over their dreams, as well as a less clear dreamer-perspective, but they did not differ on within-dream cognition or spatial features of the dream. Interestingly, aphantasics scored significantly higher than controls on how much time they spend thinking during their dreams, potentially reflecting semanticised dream content. There was no significant difference between the frequency of daydreams between aphantasics and control participants, but a comparison with a second non-age matched control group did show a significant difference, such that aphantasics experienced significantly fewer daydreams than controls.

This indicates a third impairment. All aphantasics have an impairment with respect to voluntary imagery, but some of them *also* have an impairment with respect to involuntary imagery. So, the third data point a theory of aphantasia should be able to explain is:

> (3) the differential impairment in producing voluntary imagery and involuntary imagery.

[11] Our ability to introspect and report on dreams has been doubted (Schwitzgebel, 2002). However, even if one is sceptical of this ability, there is still a statistical difference between aphantasics and non-aphantasics that needs to be explained.

### *2.4 Memory*

A wide range of findings have been made relating to autobiographical memory, episodic memory, semantic memory, and working memory in aphantasia. I discuss the two first ones in turn.[12]

In two studies (Zeman, Dewar and Della Sala, 2015; Zeman, 2020), aphantasics were asked if they think their autobiographical memory is 'normal'. In the 2015 study, results showed that 14/21 aphantasics answered negatively, and in the 2020 study aphantasics reported having significantly worse memory than both the control groups. Interestingly, there was no in-group difference between 'moderate aphantasics' and 'extreme aphantasics'.

In Dawes et al.'s (2020) study with 267 aphantasics, two questionnaires were used to assess their episodic and semantic memory. The Episodic Memory Imagery Questionnaire (EMIQ) assessed the vividness of episodic memories, with items based on the VVIQ, and the Survey of Autobiographical Memory (SAM) assessed episodic, semantic, spatial memory. SAM contains questions about recalling specific details (such as what people were wearing), recalling facts, and one's perceived competence at spatial navigation. Aphantasics reported almost no ability to generate visual sensory details when recalling past events, and they scored significantly lower than controls when it came to providing details of episodic memories. For semantic memory, aphantasics scored significantly lower than control group 1, but not significantly lower than control group 2. However, aphantasics did not score significantly lower than controls on the spatial section of SAM.

Many of these findings are echoed by Milton et al.'s (2021) study of 69 aphantasics, which is the most extensive study of memory in aphantasics to date. Here, participants took the Logical Memory Test (immediate, and 30-minute delayed recall of a prose passage), the Rey-Osterrieth Complex Figure (copy a figure immediately, and after a 30-minute delay), the Warrington Recognition Memory Test (word and facial recognition), and the Autobiographical Interview (recall as much information as possible about an event). Results showed that there was a small significant difference on the Logical Memory Test, aphantasics performing slightly worse than controls. The Rey-Osterrieth Complex Figure Task and the Warrington Memory Test showed no

[12] The studies on working memory were conducted by Jacobs et al. (2018) and Greenberg and Knowlton (2014), with one and two participants respectively. Due to the small sample sizes of these experiments we cannot say whether the findings would generalise, and I will therefore not report the findings here.

significant differences. The interesting findings relate to the Autobiographical Interview. Here, details provided by participants were coded as *episodic details* (location, people, etc) or *semantic details* (information, narrative, etc), and results showed that aphantasics produced significantly fewer episodic details, but not significantly fewer semantic details, than controls.

A drawing paradigm has also been used to investigate how many details aphantasics (n = 61) can reproduce from memory (Bainbridge et al. 2020). This found that they produce significantly fewer than controls, and that these details are particularly to do with memory of objects, rather than spatial memory. Interestingly, this study also found that aphantasics had significantly fewer memory errors than controls, where this was not due to drawing fewer details than controls (this possibility was adjusted for).

From this discussion, we can see that another result that has been replicated across many studies is that aphantasics have a memory impairment (Zeman, Dewar and Della Sala, 2015; Dawes *et al.*, 2020; Zeman *et al.*, 2020; Milton *et al.*, 2021). Particularly, they produce fewer episodic details than controls when retrieving episodic memories, and they also report having problems recalling autobiographical memories. It seems like episodic memory is particularly affected, since they do not perform sub-optimally on many other memory tests, such as immediate or delayed recall tasks. Thus, this is the fourth data point that a theory of aphantasia should be able to explain:

(4) the impairment in recalling episodic memory details.

### *2.5 Atemporal and Future Imagination*

Atemporal and future imagination  relate to voluntarily imagining general events (e.g., going to the market) and future events (e.g., going to the market *tomorrow*) respectively (Rendell *et al.*, 2012). In Milton et al.'s (2021) study, aphantasics engaged in one future and one atemporal imaginative task. In the atemporal task, they were provided with three different scenarios which they were to elaborate on as much as possible (e.g., imagining standing in the middle of a bustling street market). In the future task, they were asked to imagine three possible future events (e.g., a possible Christmas event). They tried to describe these events in as much detail as possible, and the information was coded and scored for different components, including spatial reference, entity presence, sensory description, thought/emotion/action. This score also took into account the participant's rating of their sense of presence, as well as the researcher's score of how detailed a picture was evoked in the researcher’s mind when reading the aphantasic’s description. Results showed that aphantasics scored significantly lower than the control group on both tasks.

Similarly, Dawes et al. (2020) studied aphantasics’ ability to voluntarily imagine the future using SAM. Subjects rated their agreement on a 1-5 point scale for six statements such as: ‘When I imagine an event in the future, the event generates vivid mental images that are specific in time and place’. Results showed that aphantasics reported a near  inability to imagine future events in any sensory detail.

These findings further bolster the preliminary data from the VVIQ showing that aphantasics cannot voluntarily generate mental imagery. But it also goes further than the VVIQ data, in that it not only encompasses visual imagery, and it provides information on both atemporal and future imagery. Thus, the fifth data point our theory should be able to account for is:

> (5) the impairment with respect to content generation for both atemporal events and future events.

### *2.6 Spatial Imagery*

Some studies have investigated whether aphantasics’ ability to use spatial imagery is intact. For example, Zeman et al. (2020) used the Windows Task, asking aphantasics to mentally count how many windows they had in their house. They found that there was a significant difference in how different groups of aphantasics solved this task, where ‘extreme aphantasics’ were more likely than ‘moderate aphantasics’ to rely on semantic memory rather than a visualisation technique.

Further, Dawes et al. (2020) used The Object and Spatial Imagery Questionnaire, consisting of 25 items which participants rate on a 5-point agreement scale (e.g., ‘I am a good Tetris player’). Aphantasics had significantly lower scores than controls for object imagery, but not spatial imagery. This same questionnaire was used by Keogh and Pearson (2018) and Bainbridge et al. (2020), with similar findings. Interestingly, Keogh and Pearson also used a questionnaire about the spontaneous use of spatial imagery – the Spontaneous Use of Imagery Scale (SUIS) – and found that aphantasics did not perform differently from controls here. Similarly, aphantasics performed well on spatial imagery tests administrated by Milton et al. (2021), which used Manikin’s Test (a mental rotation task), the Curved Segments Test, and the Animal Tails Test. All three tests are designed to be solved using visualisation techniques, however, unexpectedly, aphantasics did not perform significantly worse than controls in any of them.

These results indicate that aphantasics lack visual imagery, whilst retaining spatial imagery. But we should be somewhat cautious in this conclusion, since an alternative explanation is that aphantasics could solve the spatial tasks using other strategies. Pylyshyn (1981) has famously argued that spatial rotation tasks can be solved using tacit knowledge, and Bainbridge (2020) has reported that aphantasics rely to a higher extent on semantic coping techniques, making alternative strategies a viable possibility. A limitation of the current research is thus that these alternative strategies have not been controlled for. Now, the overwhelming consensus is that aphantasics retain spatial imagery (Jacobs, Schwarzkopf and Silvanto, 2018; Keogh and Pearson, 2018; Bainbridge *et al.*, 2020; Dawes *et al.*, 2020; Zeman *et al.*, 2020; Milton *et al.*, 2021), but given the possibility of these alternative strategies, we ought to still be cautious. Thus, rather than claim that aphantasics retain *spatial imagery*, the data point should be neutral, leaving open the different possible explanations. Note too that this point does not reflect an impairment, but it is still something for which a theory needs to account:

> (6) the retained ability to solve spatial imagery tasks and score averagely on spatial imagery questionnaires.

### *2.7 Prosopagnosia*

Zeman et al. (2020) asked whether aphantasics had problems recognising faces, and found a gender difference, whereby males reported poor facial recognition only significantly more often than hyperphantasics, but females reported it significantly more often than hyperphantasics and controls. In contrast, Milton et al. (2021) found no gender difference when using the Prosopagnosia Index (PI), which showed that aphantasics reported being significantly worse than controls at recognising faces. However, aphantasics did not perform any worse on a 15-item test for recognising famous faces.

Despite these findings, it is too early to say that a theory of aphantasia should explain these data at this stage. Given that Zeman et al. (2020) did not use a validated questionnaire, the claim that aphantasics potentially have impaired facial recognition comes only from one study. Before we put the burden of explaining the data on a theory, we ought to confirm that this data indeed needs explaining and that it is not spurious, and we do so by replicating results using multiple questionnaires, experimental paradigms, different research teams, etc. It will be an interesting task for future research to further investigate this, but I will set the issue aside for now.

### *2.8 Taking Stock*

I have identified what a theory of aphantasia ought to explain, namely: (1) the impairment in generating voluntary visual imagery; (2) the impairment in generating mental imagery with respect to different sensory systems; (3) the differential impairment in producing voluntary imagery and involuntary imagery; (4) the impairment in recalling episodic memory details; (5) the impairment with respect to content generation for both atemporal events and future events; and (6) the retained ability to solve spatial imagery tasks and score averagely on spatial imagery questionnaires. But the wide range of impairments identified here raises the question of how aphantasia could be explained by a theory. For this, we need to shift gear into thinking about a theory of the cognitive architecture whose *malfunctioning* could explain the impairments in aphantasia. This provides the missing link between the observed impairments and what underlies them.

Recall that in cognitive science, a cognitive architecture is a theory which posits internal mechanisms that support functions. We need to consider how different mechanisms could malfunction, resulting in the cluster of impairments. What I will do for the remainder of this chapter is apply this line of reasoning to aphantasia, and I argue that (1) – (6) above are best explained in terms of a malfunctioning in a mechanism in the cognitive architecture underlying episodic memory and imagination (the episodic system), as presented in chapters 2 and 3. This, as we will see, goes against previous conceptions whereby the malfunction is thought to lie in the visual imagery system (Pearson, 2019).

## 3. Objections to Current Theories

The aim of this section is to examine extant cognitive theories of aphantasia. I consider Nanay's (2021) account of aphantasics as lacking conscious mental imagery, and Pearson's (2019) theory based on the visual/dorsal architecture of visual imagery. Neither of the accounts attempt to explain all the data points I identified. I therefore first identify which points they focus on, before evaluating the explanation and considering whether they could be extended to inclusively account for (1) – (6). I find that neither account can satisfactorily explain everything, and hence we are in need of a new theory.

## *3.1 Nanay's No Conscious Imagery Account*

### *3.1.1 The Account*

Nanay's view is based on Kosslyn's Perceptual View, discussed in chapter 1 (Nanay, 2015, 2016, 2018, 2021). Recall that Kosslyn's view posits that the same architecture underwrites visual imagery production and vision, and that the processes involved are implemented in the same neural areas, particularly, in the primary visual cortex. For now, I will ignore the objections I made to the Perceptual View in chapter 1, in order to see whether the view could be developed into an explanation of aphantasia.

Nanay (2021) argues that there is *un*conscious visual imagery. Though this may seem mysterious at first glance, it parallels the widely accepted claim that perception can be conscious (as when presented with a stimulus) or unconscious (as when presented with a masked stimulus) (Kentridge et al. 1999; Kouider and Dehaene 2007; Weizkrantz 2009). Similarly, according to Nanay, there is conscious and unconscious visual imagery. Moreover, he maintains that this unconscious visual imagery can be voluntarily or involuntarily generated, just like how a subject can voluntarily generate visual imagery of a holiday, or involuntarily have a traumatic visual flashback. On this basis, he puts forward the following account: aphantasics lack all forms of conscious visual imagery (voluntary and involuntary), but (some) aphantasics retain involuntary unconscious visual imagery. I first motivate his claim that some aphantasics have *unconscious* visual imagery, and then why he thinks that this spared imagery is also *involuntary*.

Nanay argues that some aphantasics have *unconscious* visual imagery to explain the performance of *one* aphantasic subject in an experiment by Jacobs et al. (2018).[13, 14] In the experiment, the

[13] This study comprised of 1 aphantasic individual, and 11 controls. Due to the low participant number, there is an obvious problem about whether the findings would generalise.

[14] The subject in Jacobs et al.'s study performed poorly on the VVIQ, and no other mental imagery questionnaires were administrated. We can thus only make claims about the subject's retained/absent *visual imagery*, but we have no information about any non-visual imagery. In what follows, I therefore use 'visual imagery' where Nanay uses 'mental imagery'. Ditto for the Keogh and Pearson (2018) experiment discussed in 3.1.2, which also only administered the VVIQ.

subject was shown a geometrical shape (e.g., a triangle), and was then either instructed to imagine the triangle (imagination condition) *or* was shown placeholders for the triangle (placeholder condition), before being shown a single dot and asked whether this was within the boundaries of the original shape. It was expected that the aphantasic subject would not be able to solve the task in the imagination condition, since this presumably requires visual imagery. Surprisingly, the subject did not perform differently from controls in *either* condition, and performed well above chance levels (around 90%). How can the results in the imagination condition be explained? Nanay argues that the best explanation is the following: controls used *conscious* visual imagery in the imagination condition, whereas the aphantasic subject used *un*conscious visual imagery.

However, Nanay is aware that this hypothesis faces a potent objection. Keogh and Pearson (2018) tested 15 aphantasics and found that aphantasics seem to have *no* visual imagery *at all* – neither conscious, nor unconscious (results have been replicated in Keogh and Pearson (2021)). This experiment used a binocular rivalry paradigm, where average subjects normally exhibit a priming effect after imagining a stimulus. Participants were sat in front of a screen, and instructed to imagine either a red horizontal Gabor patch or a green vertical Gabor patch, before being presented with a binocular rivalry test where the different Gabor patches were independently presented to each eye (see Figure 10). They were then asked whether the pictures appeared to be overlapping or not. In controls, having first imagined one of the Gabor patches primed the visual system to perceive this patch more strongly when the patches were presented simultaneously. However, no such priming effect was found in aphantasics. Nanay admits that this finding appears out of line with the predictions of his own account, since his account predicts that there should *still be a priming effect*. After all, if retaining unconscious visual imagery allowed the aphantasic in Jacobs et al.'s experiment to solve the task in the imagination condition, it would be strange if unconscious visual imagery did not give rise to a priming effect here.

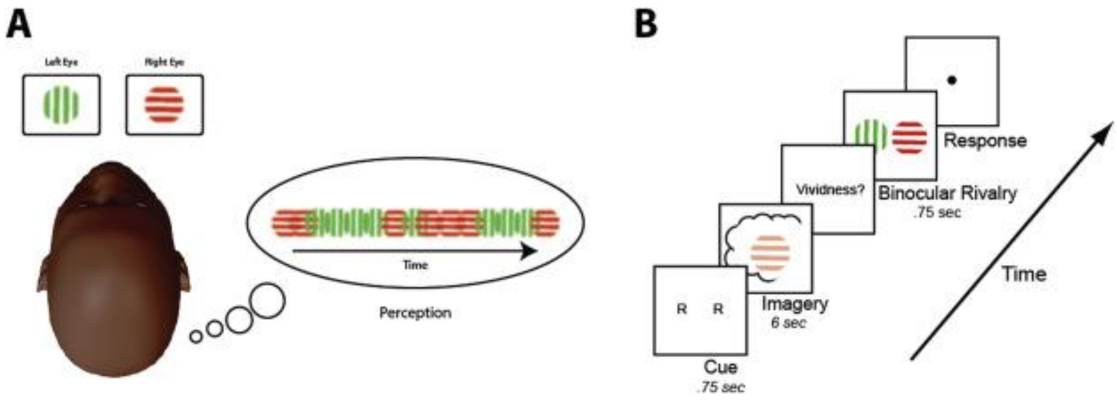


*Figure 10. Binocular rivalry paradigm and experimental timeline. Reproduced from Keogh and Pearson (2018).*

It is to rebut this objection that Nanay points to the distinction between *voluntary* unconscious visual imagery and in*voluntary* unconscious visual imagery. Keogh and Pearson's experiment involved the former as it was a voluntary task. Hence, Nanay argues that their finding is consistent with the claim that aphantasics have *involuntary* unconscious visual imagery, and this is how he arrives at his conclusion that some aphantasics retain involuntary unconscious visual imagery.

At a first glance, Nanay's account looks promising as a theory of aphantasia. To begin with, it can explain the impairment in voluntary visual imagery (1): aphantasics lack voluntary conscious visual imagery, and hence they report not experiencing any visual imagery on the VVIQ. Given that Nanay also holds that there are different kinds of mental imagery (Nanay, 2018), the account can also explain differential impairment across sensory systems (2), by positing differential impairments in different *kinds* of mental imagery. For example, it is possible on Nanay's account for a subject to have an impairment in conscious visual imagery, but no impairment in conscious auditory imagery. By the same accord, it should be able to explain the retention of spatial imagery (6), since this is also a kind of imagery, and it could just so happen that it is never impaired in aphantasics. Finally, since Nanay posits a distinction between voluntary and involuntary imagery, it could also account for the differential impairment in these and thus explain (3).

### *3.1.2 Problems for the Account*

Unfortunately, there are two serious problems with the account to which I now turn. Firstly, Nanay's attempt to avoid the objection from Keogh and Pearson (2018) leads to a contradiction in his own proposal; secondly, his theory cannot explain the episodic memory impairment (4) or the impairment in future/atemporal imagination (5).

Nanay explains Keogh and Pearson's finding by hypothesising that aphantasics lack voluntary unconscious visual imagery, but retain involuntary unconscious visual imagery. This interpretation, however, undermines his own explanation of Jacobs et al.'s experiment in terms of unconscious imagery, and by doing so, undermines the account. How so? We need to first ask whether *voluntary* or *involuntary* unconscious visual imagery was involved in Jacobs et al. task. As the subject was *instructed* to imagine something, it was quite clearly a voluntary task. In light of this, Nanay should say that the aphantasic subject used *voluntary* unconscious visual imagery to solve the task – it would make no sense to claim that the subject used involuntary imagery in a voluntary task. But this is inconsistent with interpreting Keogh and Pearson's finding as aphantasics *lacking* this very type of unconscious imagery.

It cannot be the case that aphantasics both retain and do not retain voluntary unconscious visual imagery – this is a contradiction. Now, Nanay could either stand by the explanation of Jacobs et al.'s finding, or stand by Keogh and Pearson's explanation of their finding. Choose the former, and his account would predict the opposite of what was found by Keogh and Pearson, rendering his account disconfirmed by the data. Choose the latter, and he would now lack support for the very claim that aphantasics retain unconscious visual imagery in the first place, as there is now no viable way of positing unconscious visual imagery to explain the Jacobs et al. finding. Thus, either route undermines the account.

Moreover, even if one thought that the hypothesis that aphantasics retain involuntary unconscious visual imagery were backed up by data, this theory still struggles to account for other impairments. Particularly, it cannot explain why aphantasics have problems with recalling episodic memory details (4) or imagining future and atemporal events (5). This is because the account offers no connection between mental imagery and the episodic processes involved in episodic memory and episodic imagination. Potentially, one could claim that aphantasics are only impaired with respect to recalling/imagining *conscious* imagery of past or future events, and perhaps they still retain unconscious imagery here too. However, this is an unsupported claim, and it moreover wrongly predicts that aphantasics could recall *no* conscious details at all, although we already know that aphantasics can recall some episodic details (see section 2.4). The account fails to explain that finding. Nanay's account thus fails both on its own terms and in accounting for the whole set of data concerning aphantasia. Let us now consider whether Pearson's proposal fares better.

## *3.2 Ventral and Dorsal Streams of Visual Imagery*

### *3.2.1 The account*

Pearson also endorses Kosslyn's Perceptual View, but the way in which he tackles aphantasia is different to Nanay. Pearson (2019) focuses on accounting for (1) and (6) – the impairment in voluntary visual imagery, and the retained ability to solve spatial tasks. His proposal starts from the by-now classical distinction between the *ventral* and *dorsal* pathways of vision (Goodale and Milner, 1992): the first one (also called the 'what' stream) provides information about what an object looks like; the second one (also called the 'where' stream) provides information about where an object is spatially located. Importantly, these pathways can dissociate, as can be seen in the patient DF (Servos and Goodale 1995; Milner and Goodale, 1995/2006; Whitwell et al. 2015), who has been found to be unable to report on what objects look like, but nevertheless is able to interact with these objects in a normal way (though see: Carey, Harvey, and Milner (1996); Dijkerman, Milner, and Carey (1998); Marotta, Behrmann, and Goodale (1997); Mon-Williams et al. (2001); and Hesse, Ball, and Schenk (2012)).

Pearson applies this distinction to the cognitive architecture for visual imagery: he claims that there is both ventral and dorsal *visual imagery*, and that these two types of visual imagery can also dissociate.[15] On this basis, Pearson argues that, in aphantasics, the ventral pathway is damaged, but the dorsal pathway is unimpaired. This is capable of explaining data points (1) and (6) at once, by spatial imagery produced by the dorsal pathway being retained, enabling aphantasics to solve mental rotation tasks, but visual imagery produced by the ventral pathway being damaged, leading to no generation of voluntary visual imagery.

Given that we have independent evidence that these streams dissociate, Pearson's account looks promising. But one might think that it over-generates by predicting that aphantasics should also lack *vision*, given the impairment in the ventral stream. But in fact, it does not, since Pearson also maintains that there is a distinction between the processing of *external* information (a tree) and

[15] Pearson uses the phrase 'mental imagery' throughout his article though he only discusses *visual* imagery. To avoid confusion, I use the more appropriate 'visual imagery' here.

the processing of *internal* information (a mental representation of a tree) in the ventral stream, such that these can also dissociate.[16] Hence, Pearson maintains that people with aphantasia only have a damaged ventral stream when it comes to internal processing.

By tweaking Pearson's account a little, we could extend its explanatory benefits even further. For example, the differential impairment in voluntary and involuntary imagery (3) could be explained by adding a distinction between top-down and bottom-up processing to the model. *Top-down* processing involves the process being triggered by a subject's intention, *whereas bottom-up* processing is triggered in the absence of intention. When I voluntarily imagine something, the process is triggered top-down by my intention, but when I experience visual imagery in dreams, the process is triggered bottom-up in the absence of an intention. If Pearson appeals to this distinction, he could explain why some aphantasics experience involuntary imagery whereas others do not: both groups are impaired with respect to internal top-down processing in the ventral stream, but the ones who experience involuntary imagery retain bottom-up processing.

Moreover, the theory could also explain (4) – i.e., the impairment in episodic memory. Pearson holds that visual imagery is produced by the ventral stream and it enables other functions, such as mind-wandering and episodic memory (see Figure 11). Therefore, if aphantasics have a ventral stream impairment, and the ventral stream underwrites episodic memory and mind wandering, we should expect to see an impairment there too.[17] Presumably, this is not an exhaustive list of functions that visual imagery supports, and Pearson could hold that visual imagery could also enable atemporal and future imagination too (5). It thus looks like this account explain the majority of the data points.

[16] Pearson refers to external processing as 'bottom-up' processing, and internal processing as 'top-down' processing. I have chosen different terminology to avoid confusion with the bottom-up/top-down distinction I make later.

[17] Note that Pearson does not hold that mental imagery is *necessary* for many of these functions, and that other coping strategies could account for these functions. He only claims that mental imagery normally underwrites them.

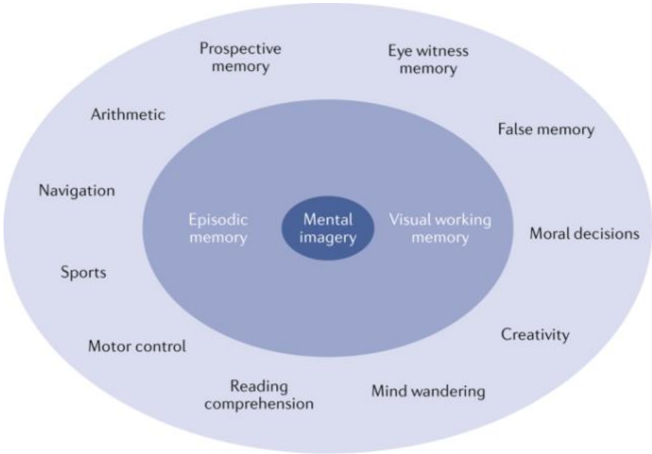


*Figure 11. Graphical depiction of the cognitive processes related to mental imagery in non-aphantasic individuals. Reproduced from Pearson (2019).*

### *3.2.2 Problems for the Account*

Despite the capacity to account for (1), (3), (4), (5) and (6), Pearson's narrow focus on the cognitive architecture of *visual* imagery leaves him with insufficient elements to explain the whole set of data on aphantasia. In particular, it seems practically impossible to explain impairments in *non*-visual imagery (2) in terms of impairments to visual imagery. It is not only the case that (2) cannot be *directly* explained by appealing to the mechanism involved in generating visual imagery, but it is also hard to see why an impairment in visual imagery could *indirectly* explain such impairments. That is to say, it looks unlikely that the generation of non-visual imagery would be dependent on the generation of visual imagery. This seems unlikely as we know, for example, that the system where visual imagery is realised is distinct from the system where olfactory imagery is realised (Flohr *et al.*, 2014; Winlove *et al.*, 2018). Pearson's proposal thus leaves unexplained the crucial point concerning why aphantasics are impaired with respect to different kinds of mental imagery.

This shortcoming of Pearson's model should not come as a surprise, since he characterises aphantasia as a *visual imagery* condition from the start. This is the key mistake, and we ought to revise our starting point. If we are looking for an account of aphantasia in which *all* its impairments are the result of a common faulty mechanism, then we need to move away from characterising aphantasia as a *visual* imagery condition. In fact, I propose that we ought to move away from characterising aphantasia as an *imagery* condition altogether, since some of the data we are trying to explain has nothing to do with *imagery*. My theory does exactly that.

## 4. A New Theory

I maintain that CESH+ can better explain (1) – (6). Recall that CESH+ posits memory indices, several different retrieval processes, a selection mechanism, and a recombination/reintegration process (see Figure 12).

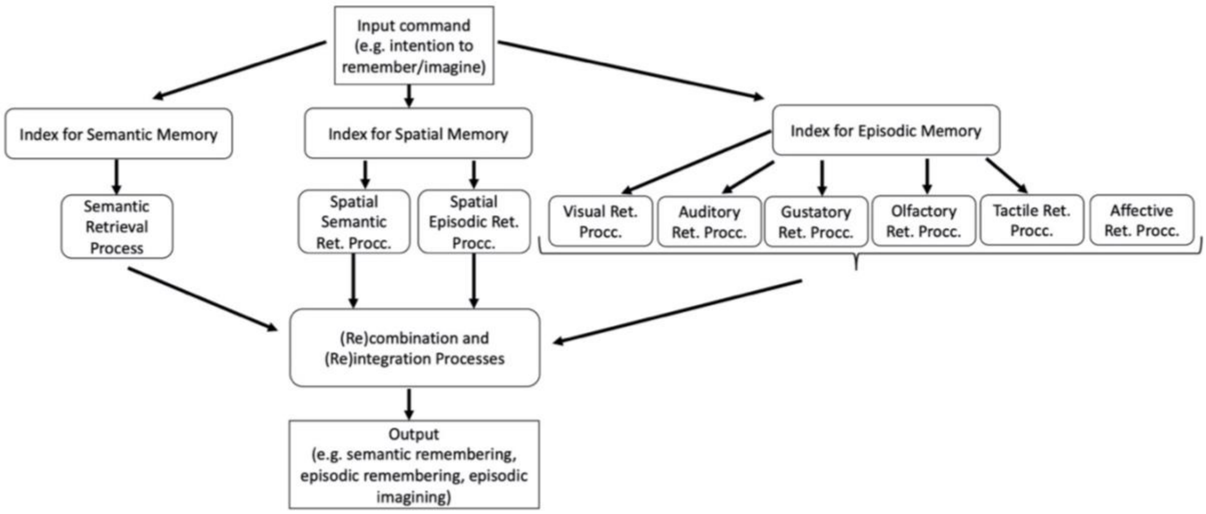


*Figure 12. A boxological depiction of the Constructive Episodic Simulation Hypothesis+, as argued for in chapters 2 and 3. The boxological depiction shows the memory indices which store addresses to memory content, the semantic, episodic, and spatial retrieval processes, which retrieve content, as well as the recombination process, which recombines content.*

I start with (1): why can aphantasics not voluntarily generate visual imagery? To explain this, we need to look at the mechanisms that generate voluntary imagery. According to the architecture at hand, generating voluntary imagery involves a subject's intention to trigger commands to retrieve elements from storage, the addresses of which are provided by the relevant index. When a subject is unable to voluntarily generate mental imagery, the top-down command fails to trigger the relevant retrieval process. That is, a command is issued, but the relevant episodic retrieval processes are not activated. This in turns means that no elements can be retrieved, and there is nothing to forward to the (re)combination process to recombine, resulting in no experience of visual imagery.

What exactly goes wrong here? We are not yet in a position to know exactly why the retrieval processes are not activated. There are three possibilities: either there is a problem with the memory index itself, or with the retrieval processes downstream from the memory index, or with the recombination process. The last option is unlikely as we know that the recombination process is also vital to recombining elements when forming semantic imaginings/rememberings, and we know that semantic memory is not impaired in aphantasics (Bainbridge *et al.*, 2020; Milton *et al.*, 2021). So we are left with two viable options. fMRI imaging could shed some light on this by telling us whether hippocampal areas are activated as normal as this is where the index for episodic memory is realised (Moscovitch *et al.*, 2005). If they are, it would indicate that the memory index works as normal, and hence it is more likely that aphantasics have a particular problem with the retrieval processes. In fact, fMRI has already shown that visual areas are abnormally activated in aphantasics, lending support to the second option, but this evidence is not conclusive (Fulford *et al.*, 2018).

Secondly, there are also aphantasics who cannot *involuntarily* generate mental imagery (3), where no intention is involved. I claimed earlier that the command to activate retrieval processes via a memory index can also be triggered in a bottom-up way, that is, in the *absence* of a subject's intention. This is how involuntary imagery is normally triggered. Hence, it follows that in a subject who cannot involuntarily generate mental imagery, the commands which are issued in the absence of an intention also fail to activate the episodic retrieval processes. Again, if the retrieval processes are not activated, there can be no information forwarded to the (re)combination process, and hence no experience of involuntary imagery. Considerably less research has been done on involuntary imagery, and we are not yet in a position to locate the exact problem – it is possible that there is a problem with the memory index, or with the retrieval processes.

Interestingly, this account makes a novel prediction with respect to voluntary and involuntary impairments. As we have seen from the data, some subjects are impaired with respect to both the top-down and bottom-up processing, resulting in no voluntary imagery *and* no involuntary imagery. But not all subjects lack both voluntary and involuntary imagery – many retain involuntary imagery. This points to a dissociation between these two processes, where one can be retained in the absence of the other. This cognitive architecture predicts that we should also find subjects who retain voluntary imagery, but lack involuntary imagery (see Figure 13). This intriguing prediction remains to be tested.

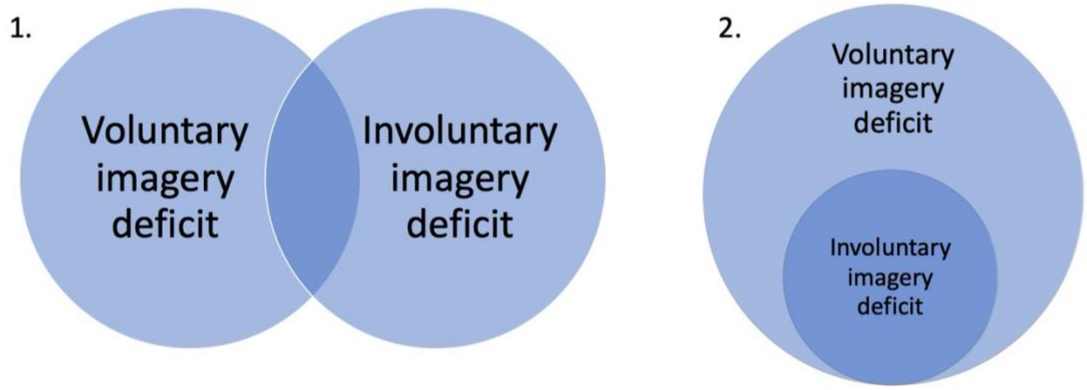


*Figure 13. Venn diagrams depicting possible relations of the voluntary imagery impairment and the involuntary imagery impairment in aphantasia. Alternative 1 allows the impairments to fully dissociate, whereas alternative 2 only allows for a partial dissociation whereby a subject could only have the voluntary imagery impairment without the involuntary imagery impairment, but not vice versa. My account supports the alternative depicted in 1.*

Thirdly, this theory is well-equipped to explain (4) and (5) – the impairments in retrieving episodic memory details and generating future/atemporal imaginings. In fact, these impairments should not come as a surprise given the architecture of the episodic system. If there is a general problem with activating the episodic retrieval process, we should expect this to affect all these aspects. Put simply, if the activation of the episodic retrieval processes is impaired, we should expect to see fewer details reported in episodic remembering, as well as in future/atemporal imaginings, since the output depends on the episodic retrieval processes. But note that the output also depends on other processes, such as the semantic retrieval process, which is not impaired. We know that the semantic retrieval process also contributes to the output of episodic memories and episodic imaginings (Schacter and Addis, 2020), and so this account predicts that aphantasics should rely more heavily on these than what other people do, resulting in some memory details being retrieved. Sensory details could be stored in semantic memory as semanticised content which has been rehearsed (Bainbridge *et al.*, 2020), though retrieving these is not accompanied by the sense of reliving that episodic memories are, as suggested by Greenberg and Knowlton (2014). Thus, my account can explain how aphantasics can still recall episodes in less detail by using different coping strategies, and it predicts that we should find that aphantasics rely more heavily on semantic memory.

Fourthly, we ought to account for why aphantasics can be differentially impaired across sensory systems (2), which I noted that Pearson's theory had trouble providing an answer to. In contrast, positing different retrieval processes can easily explain why it is the case that a person could be impaired with respect to one kind of sensory imagery but not another. The retrieval processes operate independently from each other, so it is possible for one to be impaired whilst others are not. For example, when a person is impaired with respect to visual imagery, the retrieval process that is responsible for retrieving visual information is impaired, whilst the other ones are not. That is, when a command is issued to activate the visual retrieval process, this fails, whereas commands to activate other retrieval processes succeed. We should expect neurological data to bear this out, by showing differential activity in the visual cortex when a person with a visual imagery impairment tries to visually imagine, compared to when an average person visually imagines. This is indeed what has been found (Fulford *et al.*, 2018). Neurological activation for other impairments, such as auditory or olfactory impairments, are yet to be tested, but we should expect similar results of differential activity there. The modified CESH+ is thus able to explain the data that Pearson's struggled with.

Finally, my theory can also account for the fact that aphantasics still score highly on spatial imagery questionnaires (6), and are able to solve tasks involving spatial imagery.[18] There are two possible explanations for these results, and further research needs to adjudicate between them. Recall that there is a semantic spatial retrieval process and an episodic spatial retrieval process. One possibility is that aphantasics retain the functionality of *both* of these processes, even though the episodic retrieval processes are impaired. Another possibility is that at least one of the spatial retrieval processes always remains functional. It is likely that only the spatial semantic retrieval process needs to work in order to solve spatial imagery tasks and navigate, so this would be sufficient to produce the results discussed in section 2.6 (Moscovitch *et al.*, 2005). If this is the case, it is unlikely that the subjects could experience conscious spatial mental imagery. Currently, we do not have data which can adjudicate between these explanations, as no experiments focusing on spatial imagery have been conducted. However, critically, my theory has the resources to explain both possibilities.

[18] One might be sceptical of spatial imagery as a *sui generis* kind of mental imagery, and think instead that it is just a kind of visual imagery. This suggestion entails that a person who lacks visual imagery also lacks spatial imagery, and we would need a new explanation for how spatial tasks can be solved by aphantasics. However, it looks unlikely that spatial imagery is a kind of visual imagery. Though there is no conclusive evidence, it has been shown that visual imagery activates brain areas which are not activated during spatial imagery (Trojano *et al.*, 2002; Ganis, Thompson and Kosslyn, 2004; Bonino *et al.*, 2015), and that people who are congenitally blind (presumably not able to generate visual imagery) can solve mental rotation tasks (Marmor and Zaback, 1976). This suggests that spatial imagery is indeed a *sui generis* kind.

We have now seen how this cognitive architecture is able to explain (1) – (6). Let me highlight two particularly noteworthy points to finish. The first point is that aphantasia is best characterised as an *episodic condition*, rather than a *mental imagery condition*. Though earlier accounts of aphantasia have characterised the condition as a (visual) imagery condition, the data on aphantasics does not in fact tally with this interpretation. We have no reason to think that the inability to form voluntary visual imagery should take precedence over the other impairments in defining the condition, even though the condition was first identified in this way. As we have learned more about the condition, we have seen that it is characterised by a cluster of impairments, of which one is the inability to form voluntary visual imagery. But as I have shown here, aphantasia cannot be a visual imagery condition as argued by Pearson, but it is instead a condition which can be wholly explained by the cognitive architecture of the episodic system. From this we can gather that aphantasia is not first and foremost a visual imagery condition.

Secondly, going forward, we ought to develop a new sampling method for aphantasia to reflect this insight. Given what we now know of aphantasia, we can see that the VVIQ focuses too narrowly on visual mental imagery. In fact, using it will treacherously skew our research sample towards people with a visual imagery impairment, and completely leave other aphantasics out of the sample, which would be a serious concern. We ought to prioritise developing a new method which focuses on various aspects of the condition, such as the generation of voluntary and involuntary imagery, the generation of mental imagery with respect to different sensory systems, and the generation of episodic memory details. With such a method, we will be able to investigate the condition further and test the predictions of the theory of aphantasia I have developed here.

## 5. Conclusion

In this chapter, I have laid the groundwork for a theory of aphantasia. I have argued that aphantasia is a condition which results from a malfunction in the episodic system. To argue my case, I considered currently available data on aphantasia, and identified six data points for which a theory ought to be able to provide a cognitive explanation. Examining Nanay's and Pearson's accounts, I found that these were unable to do so satisfactorily, and I therefore developed a new theory. My theory is not only able to provide a cognitive explanation of the current data, but it also makes interesting testable predictions. For example, it predicts that there should be people with only the involuntary imagery impairment; that people with aphantasia rely more heavily on semantic memory as a coping strategy; and that there should be aphantasics who are only selectively impaired with respect to non-visual imagery. Our next goal should be to test the predictions of this theory. The research on aphantasia is still in its infancy and there are many avenues left to explore, but I believe that this theory based on CESH+ can guide us in the right direction.

The fact that CESH+ provides the best explanation for aphantasia greatly speaks in favour of the theory. In particular, it speaks in favour of some of the modifications I added in chapter 3, such as the multiple episodic retrieval processes. We can now see that, in contrast to CESH, which only posits one episodic retrieval process and no spatial retrieval processes, CESH+ has an advantage as it can explain the selective impairments in aphantasia.

# Conclusion

The aim of this thesis has been to convince you that episodic memory and sensory imagination are products of the same system, the episodic system, and that the best way of modelling this system is provided by CESH+. I motivated CESH+ by first casting doubt on the received view of visual imagery production – the Perceptual View – since a major prediction of the view is falsified by recent neuroscientific data. This paved the way for developing an architecture of all kinds of sensory imagery and episodic memory. I then considered CESH as the architecture underlying imagination and memory, and argued that such an architecture is well-supported, since both capacities are optimised for content generation, and plausibly involve an episodic retrieval process, a semantic retrieval process, and a recombination process. But I also noted that CESH is sketchy in a number of places: it does not explain how stored content is selected, retrieved, and recombined to form a remembering or imagining. This provided the motivation for developing CESH+, which contains memory indices, several new retrieval processes, as well as a Bayesian selection mechanism. I supported these modifications by discussing a vast amount of empirical data from psychology, cognitive neuroscience, and computer science. To further demonstrate the explanatory power of CESH+, I showed that it can both provide a sub-personal account of how imagination can be controlled and improved, hence supporting the claim that imagination is a skill. Moreover, I argued that it is the only theory which can explain all the impairments associated with aphantasia. On this basis, I conclude that CESH+ is a promising architecture detailing the processes involved in producing episodic memory and sensory imagination.

At the same time, I do not wish to pretend that this is all there is to say about the cognitive architecture of memory and imagination. Building a cognitive architecture is a slow and arduous process, and the predictions of the proposed architecture need to be constantly tested empirically. I believe that CESH+ is fruitful also in this respect, since it generates interesting predictions and thereby opens avenues for new research. To demonstrate this, I will leave you with the skeletons of three experimental designs that can be used to shed further light on the architecture of the episodic system.

### *1.1 Experiment 1 – Testing the Perceptual View and CESH+ using an fMRI paradigm*

The Perceptual View and CESH+ make different predictions as to which processes are responsible for the voluntary visual impairment in aphantasia. I suggest that aphantasics (operational definition: minimal score on the VVIQ, minimal score on the visual imagery part of the QMI/Plymouth Sensory Imagery Questionnaire, average score on non-visual imagery parts of the QMI/Plymouth Sensory Imagery Questionnaire) and controls (operational definition: average score on VVIQ/QMI/Plymouth Sensory Imagery Questionnaire) undertake five fMRI tasks to shed light on which processes are impaired in aphantasics. The five proposed tasks are as follows:

1. *Visual imagery task*. This task should instruct participants to form voluntary visual imagery, e.g., by instructing them to 'visualise a beach', 'imagine seeing an apple', etc.
2. *Perception task*. This should consist of presenting visual stimuli to participants corresponding to the visual imagery content, e.g., a picture of a beach, a picture of an apple, etc.
3. *Semantic retrieval task*. This should consist of semantic retrieval tasks, such as recalling the names of famous capitals in Europe, or recalling one's date of birth, etc.
4. *Episodic retrieval task*. This should consist of recalling visual imagery of past events, e.g., instructing participants to 'picture your last birthday party', or 'visually recall what your house looks like'.
5. *Control/rest task*.

The Perceptual View suggests aphantasics have an impairment is early visual processing, and therefore predicts that early visual areas (e.g., V1) should show decreased activity in aphantasics performing a visual imagery task compared to a perception task. In a control group, we should instead expect similar activation in early visual areas for the imagery task compared to the perception task. CESH+, on the other hand, suggests that aphantasics have an impairment in the visual retrieval process *or* in the episodic memory index. This predicts that aphantasics, compared to controls, should show decreased activity in hippocampal areas and/or areas in the visual cortex (though not early visual areas) when performing the visual imagery task and the episodic retrieval task. Here is what we would expect to find if the impairment is in the episodic index, rather than in the visual episodic retrieval process. We should see a differential activation in the part of the hippocampus storing addresses to episodic details, compared to the areas storing semantic details (Moscovitch *et al.*, 2005), when comparing results from the episodic retrieval task to results from the semantic retrieval tasks between aphantasics and controls. If the impairment is instead in the visual retrieval process, we should not expect such differential results, and instead expect to see differential activation in the visual cortex when comparing the imagery task between the aphantasia group and the control group.

### *1.2 Experiment 2 – Are erroneous imaginings accompanied by a feeling of error, and can this aid improvement?*

Fernandez Cruz et al. (2016) designed a study which tested the accuracy of the feeling of error in subjects performing mental calculations. They found that the feeling of error reliably tracked making an error in the calculation. My study proposes to test whether a similar feeling of error accompanies inaccurate imaginings, and further, whether a feeling of error negatively correlates over time with improvement on the task. The experiment would use a mental rotation task, which is a task where subjects have previously demonstrated improvement over time (Provost *et al.*, 2013). Participants are presented with a shape, and asked which of four other shapes this would correspond to if rotated (Shepard and Metzler, 1971). Reaction times for identifying the correct shape have been measured, and shown to decrease as a subject improves at a task. Here, I suggest that once a subject has identified the shape they think the first shape cenorresponds to in a trial, they should indicate whether they think they are correct in their response (i.e., whether they have a feeling of error about it), and indicate their certainty about committing an error on a 9-point Likert scale. The account which I have developed in this thesis predicts that not only should subjects improve at the mental rotation task over time, but it also predicts that a feeling of error should positively correlate with a wrong answer, and that the frequency of feelings of error should decrease as a function of the subject improving at the task. These results would lend indirect support to my hypothesis that the Bayesian selection mechanism is affected by the feeling of error, such that the likelihoods of content is changed, and a new, more accurate, imagining is produced.

### *1.3 Experiment 3 – Can aphantasics improve their ability to recall/imagine more details?*

Aphantasics recall and produce fewer episodic details when imagining/remembering compared to controls (Bainbridge *et al.*, 2020; Milton *et al.*, 2021). Experiments on aging have shown that administering an Episodic Specificity Induction (ESI) task can boost the amount of details produced, at least when recalling past episodes (Madore, Gaesser and Schacter, 2014; Madore, Jing and Schacter, 2019). In this experiment, I propose to test three groups: extreme aphantasics (operational definition: lowest score on all sensory part of the QMI/Plymouth Sensory Imagery

Questionnaire), moderate aphantasics (operational definition: a lower than average score on all sensory parts of the QMI/Plymouth Sensory Imagery Questionnaire), and controls (operational definition: average score on all sensory part of the QMI/Plymouth Sensory Imagery Questionnaire). There should be a 2 x 3 factorial design. First, participants are shown a video, and they either receive an ESI task, or complete an alternative task. In the ESI condition, they are asked about details of this video, whereas in the alternative task condition, they are asked about their general impression of the video. Second, either participants are asked to recall an event, or imagine an event, or describe a visual stimulus in front on them (e.g., a photo). This study could tell us whether aphantasics could be 'trained' to recall/imagine more details than they first. In particular, it would be interesting to see whether there are any differential results between extreme aphantasics compared to moderate aphantasics, such that only moderate aphantasics' memory/imagination are boosted by the ESI. This could potentially indicate that 'extreme aphantasia' and 'moderate aphantasia' are in fact not conditions along a spectrum, as is commonly assumed (Zeman *et al.*, 2020). If imagination is a skill, as I have argued in this thesis, we should expect some normal variation in this skill, such that some people are better than others, just as some people are better than others at football or mental arithmetic. But being poor at imagining might not necessarily indicate a condition due to a malfunction in a cognitive system; it could just indicate a lack of practice. However, an *inability to improve* is more likely to indicate a malfunction in a cognitive system. Hence, if extreme aphantasics are unable to improve in the ESI condition, this could indicate that they indeed have the condition aphantasia, and if moderate aphantasics are able to improve following an ESI, this might just indicate a natural variation in the skill of imagining. If so, so-called 'moderate aphantasics' should potentially be excluded from the future study of the condition aphantasia, as data from this group would skew results.

# Bibliography

Addis, D. R. *et al.* (2009) ‘Constructive episodic simulation of the future and the past: Distinct subsystems of a core brain network mediate imagining and remembering’, *Neuropsychologia*, 47(11), pp. 2222–2238. doi: 10.1016/j.neuropsychologia.2008.10.026.

Addis, D. R. (2018) ‘Are episodic memories special? On the sameness of remembered and imagined event simulation’, *Journal of the Royal Society of New Zealand*. Taylor & Francis, 48(2–3), pp. 64–88. doi: 10.1080/03036758.2018.1439071.

Addis, D. R., Roberts, R. P. and Schacter, D. L. (2011) ‘Age-related neural changes in autobiographical remembering and imagining’, *Neuropsychologia*. Elsevier Ltd, 49(13), pp. 3656–3669. doi: 10.1016/j.neuropsychologia.2011.09.021.

Addis, D. R. and Schacter, D. L. (2008) ‘Constructive episodic simulation: Temporal distance and detail of past and future events modulate hippocampal engagement’, *Hippocampus*, 18(2), pp. 227–237. doi: 10.1002/hipo.20405.

Addis, D. R., Wong, A. T. and Schacter, D. L. (2007) ‘Remembering the past and imagining the future: Common and distinct neural substrates during event construction and elaboration’, *Neuropsychologia*, 45(7), pp. 1363–1377. doi: 10.1016/j.neuropsychologia.2006.10.016.

Addis, D. R., Wong, A. T. and Schacter, D. L. (2008) ‘Age-Related Changes in the Episodic Simulation of Future’, *Psychological Science*, 19(1), pp. 33–41.

Adlam, A. L. R., Patterson, K. and Hodges, J. R. (2009) ‘“I remember it as if it were yesterday”: Memory for recent events in patients with semantic dementia’, *Neuropsychologia*, 47(5), pp. 1344–1351. doi: 10.1016/j.neuropsychologia.2009.01.029.

Alberini, C. M. (ed.) (2013) *Memory Reconsolidation*. San Diego: Elsevier Science & Technology.

Aleman, A. *et al.* (2000) ‘Music training and mental imagery ability’, *Neuropsychologia*, 38(12), pp. 1664–1668. doi: 10.1016/S0028-3932(00)00079-8.

Ambrogi Lorenzini, C. G. *et al.* (1999) ‘Neural topography and chronology of memory consolidation: A review of functional inactivation findings’, *Neurobiology of Learning and Memory*. doi: 10.1006/nlme.1998.3865.

Anderson, J. R. (1982) ‘Acquisition of Cognitive Skill’, *Readings in Cognitive Science: A Perspective from Psychology and Artificial Intelligence*, 89(4), pp. 362–380. doi: 10.1016/B978-1-4832-1446-7.50032-7.

Anderson, J. R. (1993) *Rules of the mind*. Hillsdate, NJ: Lawrence Erlbaum.

Anderson, J. R. (2007) *How can the human mind occur in the physical universe?* New York, NY: Oxford University Press.

Applebee, A. N. (1978) *The child's concept of story: Ages two to seventeen*. Chicago: University of Chicago Press.

Appleyard, J. A. (1990) *Becoming a reader: The experience of fiction from childhood to adulthood*. Cambridge: Cambridge University Press.

Arango-Muñoz, S. (2014) 'The nature of epistemic feelings', *Philosophical Psychology*, 27(2), pp. 193–211. doi: 10.1080/09515089.2012.732002.

Arshamian, A. *et al.* (2008) 'Sniff your way to clarity: The case of olfactory imagery', *Chemosensory Perception*, 1(4), pp. 242–246. doi: 10.1007/s12078-008-9035-z.

Atance, C. M. and O'Neill, D. K. (2005) 'Preschoolers' talk about future situations', *First Language*, 25(1), pp. 5–18. doi: 10.1177/0142723705045678.

Atance, C. and Meltzoff, A. (2005) 'My future self: Young children's ability to anticipate and explain future states', *Cognitive Development*, 20, pp. 1-341–361. doi: 10.1038/jid.2014.371.

Atkinson, R. C. and Shiffrin, R. M. (1968) 'Human Memory: A Proposed System and its Control Processes', *Psychology of Learning and Motivation - Advances in Research and Theory*. doi: 10.1016/S0079-7421(08)60422-3.

Axelrod, V., Rees, G. and Bar, M. (2017) 'The default network and the combination of cognitive processes that mediate self-generated thought', *Nature Human Behaviour*. Springer US, 1(12), pp. 896–910. doi: 10.1038/s41562-017-0244-9.

Baddeley, A. (2000) 'The episodic buffer: A new component of working memory?', *Trends in Cognitive Sciences*. doi: 10.1016/S1364-6613(00)01538-2.

Baddeley, A. D. and Andrade, J. (2000) 'Working memory and the vividness of imagery', *Journal of Experimental Psychology: General*, 129(1), pp. 126–145. doi: 10.1037/0096-3445.129.1.126.

Baddeley, A. D. and Hitch, G. J. (1974) 'Working memory', in Bower, G. H. (ed.) *The psychology of learning and motivation*. New York, NY: Academic Press. doi: http://dx.doi.org/10.1016/S0079-7421(08)60452-1.

Baillargeon, R., Scott, R. M. and He, Z. (2010) 'False-belief understanding in infants', *Trends in Cognitive Sciences*. doi: 10.1016/j.tics.2009.12.006.

Bainbridge, W. A. *et al.* (2020) 'Quantifying Aphantasia through drawing: Those without visual imagery show deficits in object but not spatial memory', *Cortex*. Elsevier Ltd. doi: 10.1016/j.cortex.2020.11.014.

Balcerak Jackson, M. (2016) 'On the Epistemic Value of Imagining, Supposing, and Conceiving', in Kind, A. and Kung, P. (eds) *Knowledge Through Imagination*. Oxford University Press.

Balcerak Jackson, M. (2018) 'Justification by Imagination', in Macpherson, F. and Dorsch, F. (eds) *Perceptual Imagination and Perceptual Memory*. Oxford University Press.

Barlassina, L. and Gordon, R. M. (2017) 'Folk Psychology as Mental Simulation', *The Stanford Encyclopedia of Philosophy*. Available at: https://plato.stanford.edu/archives/sum2017/entries/folkpsych-simulation/.

Barrós-Loscertales, A. *et al.* (2012) 'Reading salt activates gustatory brain regions: FMRI evidence for semantic grounding in a novel sensory modality', *Cerebral Cortex*, 22(11), pp. 2554–2563. doi: 10.1093/cercor/bhr324.

Bartlett, F. C. (1932) *Remembering: A Study in Experimental and Social Psychology*. Cambridge University Press.

Bartolomeo, P. *et al.* (1998) 'Multiple-domain dissociation between impaired visual perception and preserved mental imagery in a patient with bilateral extrastriate lesions', *Neuropsychologia*, 36(3), pp. 239–249. doi: 10.1016/S0028-3932(97)00103-6.

Bartolomeo, P. (2002a) 'The relationship between visual perception and visual mental imagery: A reappraisal of the neuropsychological evidence', *Cortex*, 38(3), pp. 357–378. doi: 10.1016/S0010-9452(08)70665-8.

Bartolomeo, P. (2002b) 'The relationship between visual perception and visual mental imagery', *Neuroscience*, pp. 357–378.

Bartolomeo, P. (2008) 'The neural correlates of visual mental imagery: An ongoing debate', *Cortex*, 44(2), pp. 107–108. doi: 10.1016/j.cortex.2006.07.001.

Bartolomeo, P. *et al.* (2020) 'Assessing the causal role of early visual areas in visual mental imagery', *Nature Reviews Neuroscience*, 21(9), p. 517. doi: 10.1038/s41583-020-0348-5.

Bartolomeo, P., Bachoud-Lévi, A. C. and Thiebaut de Schotten, M. (2014) 'The anatomy of cerebral achromatopsia: A reappraisal and comparison of two case reports', *Cortex*, 56, pp. 138–144. doi: 10.1016/j.cortex.2013.01.013.

Basso, A., Bisiach, E. and Luzzatti, C. (1980) 'Loss of Mental Imagery : a Case Study Introduction', *Neuropsychologia*, 18, pp. 435–442.

Battistoni, E., Stein, T. and Peelen, M. (2017) 'Preparatory Attention in Visual Cortex', *Annals of the New York Academy of Sciences*, 1396, pp. 92–107.

Beauchamp, M. R., Bray, S. R. and Albinson, J. G. (2002) 'Pre-competition imagery, self-efficacy and performance in collegiate golfers', *Journal of Sports Sciences*, 20(9), pp. 697–705. doi: 10.1080/026404102320219400.

Beck, J. (2018) 'Analog mental representation', *Wiley Interdisciplinary Reviews: Cognitive Science*, 9(6), pp. 1–10. doi: 10.1002/wcs.1479.

Beck, J. (2019) ‘Perception is analog: The argument from Weber’s law’, *Journal of Philosophy*, 116(6), pp. 319–349. doi: 10.5840/jphil2019116621.

Benoit, R. G. and Schacter, D. L. (2015) ‘Specifying the core network supporting episodic simulation and episodic memory by activation likelihood estimation’, *Neuropsychologia*. Elsevier, 75, pp. 450–457. doi: 10.1016/j.neuropsychologia.2015.06.034.

Bensafi, M. *et al.* (2003) ‘Olfactomotor activity during imagery mimics that during perception’, *Nature Neuroscience*, 6(11), pp. 1142–1144. doi: 10.1038/nn1145.

Bensafi, M. and Rouby, C. (2007) ‘Individual differences in odor imaging ability reflect differences in olfactory and emotional perception’, *Chemical Senses*, 32(3), pp. 237–244. doi: 10.1093/chemse/bjl051.

Bergmann, T. O. and Hartwigsen, G. (2020) ‘Inferring causality from noninvasive brain stimulation in cognitive neuroscience’, *Journal of Cognitive Neuroscience*. doi: 10.1162/jocn_a_01591.

Berman, M. G., Jonides, J. and Lewis, R. L. (2009) ‘In Search of Decay in Verbal Short-Term Memory’, *Journal of Experimental Psychology: Learning Memory and Cognition*. doi: 10.1037/a0014873.

Bernecker, S. (2010) *Memory: A Philosophical Study*. Oxford: Oxford University Press.

Bértolo, H. *et al.* (2003) ‘Visual dream content, graphical representation and EEG alpha activity in congenitally blind subjects’, *Cognitive Brain Research*, 15(3), pp. 277–284. doi: 10.1016/S0926-6410(02)00199-4.

Bishop, L., Bailes, F. and Dean, R. T. (2013) ‘Musical Expertise and the Ability to Imagine Loudness’, *PLoS ONE*, 8(2). doi: 10.1371/journal.pone.0056052.

Blackwell, S. E. (2020) ‘Emotional Mental Imagery’, in Abraham, A. (ed.) *The Cambridge Handbook of the Imagination*. Cambridge University Press, pp. 241–257. doi: 10.1017/9781108580298.016.

Block, N. (1981) *Imagery*. Cambridge, MA: MIT Press.

Block, N. (1983) ‘Mental Pictures and Cognitive Science’, 92(4), pp. 499–541.

Boly, M. *et al.* (2007) ‘When thoughts become action: An fMRI paradigm to study volitional brain activity in non-communicative brain injured patients’, *NeuroImage*. doi: 10.1016/j.neuroimage.2007.02.047.

Bonino, D. *et al.* (2015) ‘Spatial imagery relies on a sensory independent, though sensory sensitive, functional organization within the parietal cortex: A fMRI study of angle discrimination in sighted and congenitally blind individuals’, *Neuropsychologia*. Elsevier, 68, pp. 59–70. doi: 10.1016/j.neuropsychologia.2015.01.004.

Brewer, W. F. and Treyens, J. C. (1981) 'Role of schemata in memory for places', *Cognitive Psychology*, 13(2), pp. 207–230. doi: 10.1016/0010-0285(81)90008-6.

Bridge, H. *et al.* (2012) 'Vivid visual mental imagery in the absence of the primary visual cortex', *Journal of Neurology*, 259(6), pp. 1062–1070. doi: 10.1007/s00415-011-6299-z.

De Brigard, F. *et al.* (2013) 'Remembering what could have happened: Neural correlates of episodic counterfactual thinking', *Neuropsychologia*. Elsevier, 51(12), pp. 2401–2414. doi: 10.1016/j.neuropsychologia.2013.01.015.

De Brigard, F. (2014) 'Is memory for remembering? Recollection as a form of episodic hypothetical thinking', *Synthese*, 191(2), pp. 155–185. doi: 10.1007/s11229-013-0247-7.

Brown, R. and Kulik, J. (1977) 'Flashbulb Memories', *Cognition*, 5, pp. 73–99. doi: 10.1016/B978-0-08-097086-8.51035-6.

Buckner, R. L. and Carroll, D. C. (2007) 'Self-projection and the brain', *Trends in Cognitive Sciences*. doi: 10.1016/j.tics.2006.11.004.

Budson, A. E. *et al.* (2000) 'When false recognition is unopposed by true recognition: Gist-based memory distortion in Alzheimer's disease', *Neuropsychology*. doi: 10.1037/0894-4105.14.2.277.

Budson, A. E. *et al.* (2003) 'Semantic versus phonological false recognition in aging and Alzheimer's disease', *Brain and Cognition*. doi: 10.1016/S0278-2626(03)00030-7.

Buskell, A. (2015) 'How to be skilful: opportunistic robustness and normative sensitivity', *Synthese*. Springer Netherlands, 192(5), pp. 1445–1466. doi: 10.1007/s11229-014-0634-8.

Camp, E. (2007) 'Thinking with Maps', *Philosophical Perspectives*, (21).

Carey, D. P., Harvey, M. and Milner, A. D. (1996) 'Visuomotor sensitivity for shape and orientation in a patient with visual form agnosia', *Neuropsychologia*. doi: 10.1016/0028-3932(95)00169-7.

Carruthers, P. (2006) *The Architecture of the Mind*. Oxford University Press.

Carruthers, P. (2013) 'Mindreading in infancy', *Mind and Language*. doi: 10.1111/mila.12014.

Cavedon-Taylor, D. (fothcoming a) 'Mental Imagery: Pulling the Plug on Perceptualism', *Philosophical Studies*.

Cavedon-Taylor, D. (forthcoming b) 'Untying the Knot: Imagination, Perception and their Neural Substrates', *Synthese*.

Clark, A. (2013) 'Whatever next? Predictive brains, situated agents, and the future of cognitive science', *Behavioral and Brain Sciences*, 36(3), pp. 181–204. doi: 10.1017/S0140525X12000477.

Claxton, L. J., Keen, R. and McCarty, M. E. (2003) 'Evidence of motor planning in infant reaching behavior', *Psychological Science*, 14(4), pp. 354–356. doi: 10.1111/1467-9280.24421.

Clayton, N. and Dickinson, A. (1998) 'Episodic-like memory during cache recovery by scrub jays', *Nature*, 395(6699), pp. 272–274. doi: 10.1038/255243a0.

Clements, W. A. and Perner, J. (1994) ‘Implicit understanding of belief’, *Cognitive Development*, 9(4), pp. 377–395. doi: 10.1016/0885-2014(94)90012-4.

Cooper, J. M. *et al.* (2011) ‘The effect of hippocampal damage in children on recalling the past and imagining new experiences’, *Neuropsychologia*. Elsevier Ltd, 49(7), pp. 1843–1850. doi: 10.1016/j.neuropsychologia.2011.03.008.

Corbetta, M. (1998) ‘Frontoparietal cortical networks for directing attention and the eye to visual locations: Identical, independent, or overlapping neural systems?’, *Proceedings of the National Academy of Sciences of the United States of America*. doi: 10.1073/pnas.95.3.831.

Cowan, N. (1988) ‘Evolving Conceptions of Memory Storage, Selective Attention, and Their Mutual Constraints Within the Human Information-Processing System’, *Psychological Bulletin*. doi: 10.1037/0033-2909.104.2.163.

Cowan, N. (1999) ‘An embedded-processes model of working memory’, *Models of working memory: Mechanisms of active maintenance and executive control.*

Croijmans, I. *et al.* (2019) ‘Measuring the multisensory imagery of wine: The Vividness of Wine Imagery Questionnaire’, *Multisensory Research*, 23(3).

Currie, G. (1995) ‘Visual Imagery as the Simulation of Vision’, *Mind and Language*, 10(1–2), pp. 25–44.

Currie, G. and Ravenscroft, I. (2002) *Recreative Minds: Imagination in Philosophy and Psychology*. Oxford University Press.

Danker, J. and Anderson, J. R. (2010) ‘The ghosts of brain states past: Remembering reactivates the brain regions engaged during encoding’, *Psychological Bulletin*, 136(1), pp. 87–102. doi: 10.1037/a0017937.The.

Davies, J. (2020) ‘Artificial Intelligence and Imagination’, in Abraham, A. (ed.) *The Cambridge Handbook of the Imagination*. Cambridge University Press (Cambridge Handbooks in Psychology), pp. 162–172. doi: 10.1017/9781108580298.011.

Dawes, A. J. *et al.* (2020) ‘A cognitive profile of multi-sensory imagery, memory and dreaming in aphantasia’, *Scientific Reports*. Springer US, 10(1), pp. 1–10. doi: 10.1038/s41598-020-65705-7.

Debus, D. (2014) ‘“Mental Time Travel”: Remembering the Past, Imagining the Future, and the Particularity of Events’, *Review of Philosophy and Psychology*, 5(3), pp. 333–350. doi: 10.1007/s13164-014-0182-7.

DeKeyser, R. (2015) ‘Skill Acquitision Theory’, in VanPatten, B. and Williams, J. (eds) *Theories in Second Language Acquisition*. Second Edi. New York: Taylor & Francis.

Devitt, A. L., Addis, D. R. and Schacter, D. L. (2017) ‘Episodic and semantic content of memory and imagination: A multilevel analysis’, *Memory and Cognition*, 45(7), pp. 1078–1094. doi:

10.3758/s13421-017-0716-1.

Dewhurst, S. A. and Farrand, P. (2004) 'Investigating the phenomenological characteristics of false recognition for categorised words', *European Journal of Cognitive Psychology*. doi: 10.1080/09541440340000088.

Dijkerman, H. C., Milner, A. D. and Carey, D. P. (1998) 'Grasping Spatial Relationships: Failure to Demonstrate Allocentric Visual Coding in a Patient with Visual Form Agnosia', *Consciousness and Cognition*. doi: 10.1006/ccog.1998.0365.

Dokic, J. (2013) 'Seeds of self-knowledge: Noetic feelings and metacognition', in *Foundations of Metacognition*. doi: 10.1093/acprof:oso/9780199646739.003.0020.

Dorsch, F. (2015) 'Focused Daydreaming and Mind-Wandering', *Review of Philosophy and Psychology*, 6(4), pp. 791–813. doi: 10.1007/s13164-014-0221-4.

Dosenbach, N. U. F. *et al.* (2008) 'A dual-networks architecture of top-down control', *Trends in Cognitive Sciences*. doi: 10.1016/j.tics.2008.01.001.

Dretske, F. (1981) *Knowlegde and the Flow of Information*. Cambridge, MA: MIT Press.

Dreyfus, H. L. (2002) 'Intelligence without representation – Merleau-Ponty's critique of mental representation', *Phenomenology and the Cognitive Sciences*, 1.

Dreyfus, H. L. (2007) 'The return of the myth of the mental', *Inquiry: An Interdisciplinary Journal of Philosophy*. Taylor & Francis, 50(4), pp. 352–365. doi: 10.1080/00201740701489245.

Driskell, J. E., Copper, C. and Moran, A. (1994) 'Does mental practice enhance performance?', *Journal of Applied Psychology*, 79(4), pp. 481–492. doi: 10.1037//0021-9010.79.4.481.

Dudai, Y. and Carruthers, M. (2005) 'The Janus face of Mnemosyne', *Nature*. doi: 10.1038/434567a.

Duke, D., Fiacconi, C. M. and Köhler, S. (2014) 'Parallel effects of processing fluency and positive affect on familiarity-based recognition decisions for faces', *Frontiers in Psychology*. doi: 10.3389/fpsyg.2014.00328.

Ebbinghaus, H. (1888) 'Memory: A Contribution to Experimental Psychology', in Ruger, H. A. and Bussenius, C. E. (eds). New York: Teachers College, Columbia University.

Ebert, J. E. J., Gilbert, D. T. and Wilson, T. D. (2009) 'Forecasting and Backcasting: Predicting the Impact of Events on the Future', *Journal of Consumer Research*, 36(3), pp. 353–366. doi: 10.1086/598793.

Engel, S. (1995) *The stories children tell: Making sense of the narratives of childhood*. New York: Henry Holt and Company.

Farah, M. J. (1984) 'The neurological basis of mental imagery: A componential analysis', *Cognition*, 18(1–3), pp. 245–272. doi: 10.1016/0010-0277(84)90026-X.

Farah, M. J., Levine, D. N. and Calvanio, R. (1988) ‘A case study of mental imagery deficit’, *Brain and Cognition*, 8(2), pp. 147–164. doi: 10.1016/0278-2626(88)90046-2.

Fernandez Cruz, A. L., Arango-Muñoz, S. and Volz, K. G. (2016) ‘Oops, scratch that! Monitoring one’s own errors during mental calculation’, *Cognition*. Elsevier B.V., 146, pp. 110–120. doi: 10.1016/j.cognition.2015.09.005.

Fernandez, J. (2017) ‘The intentional objects of memory’, in Bernecker, S. and Michaelian, K. (eds) *RoutledgeHandbook of Philosophy of Memory*.

Fisher, M. *et al.* (2012) ‘Example-based synthesis of 3D object arrangements’, *ACM Transactions on Graphics*.

Fisher, R. P. and Geiselman, R. E. (1992) *Memory-Enhancing Techniques for Investigative Interviewing: The Cognitive Interview*. Springfield, IL: Charles C. Thomas Books.

Fitts, P. M. (1964) ‘Perceptual-Motor Skill Learning’, in *Categories of Human Learning*. Elsevier Inc., p. Categ. Hum. Learn.

Flohr, E. L. R. *et al.* (2014) ‘The fate of the inner nose: Odor imagery in patients with olfactory loss’, *Neuroscience*, 268, pp. 118–127. doi: 10.1016/j.neuroscience.2014.03.018.

Fodor, J. A. (1975) *The Language of Thought*. Cambridge, MA: Harvard University Press.

Fodor, J. A. (2007) *Revenge of The Given*, *Contemporary Debates in Philosophy of Mind*. Edited by B. P. McLaughlin and J. D. Cohen. Blackwell.

Fodor, J. A. and Pylyshyn, Z. W. (1988) ‘Connectionism and cognitive architecture: A critical analysis’, *Cognition*, 28(1–2), pp. 3–71. doi: 10.1016/0010-0277(88)90031-5.

Fridland, E. (2014) ‘They’ve lost control: Reflections on skill’, *Synthese*, 191(12), pp. 2729–2750. doi: 10.1007/s11229-014-0411-8.

Fulford, J. *et al.* (2018) ‘The neural correlates of visual imagery vividness – An fMRI study and literature review’, *Cortex*. Elsevier Ltd, 105, pp. 26–40. doi: 10.1016/j.cortex.2017.09.014.

Gaesser, B. *et al.* (2011) ‘Characterizing Age-Related Changes in Remembering the Past and Imagining the Future’, *Psychology and Aging*, 26(1), pp. 80–84. doi: 10.1037/a0021054.

Gallistel, C. R. (2020) ‘The physical basis of memory’, *Cognition*. Elsevier B.V., (August), p. 104533. doi: 10.1016/j.cognition.2020.104533.

Gallo, D. A. (2013) *Associative illusions of memory: False memory research in DRM and related tasks*, *Associative Illusions of Memory: False Memory Research in DRM and Related Tasks*. doi: 10.4324/9780203782934.

Galton, E. (1883) *Inquiries into human faculty and its development*. London: Dent.

Ganis, G., Thompson, W. L. and Kosslyn, S. M. (2004) ‘Brain areas underlying visual mental imagery and visual perception: An fMRI study’, *Cognitive Brain Research*. doi:

10.1016/j.cogbrainres.2004.02.012.
de Gelder, B. *et al.* (2008) 'Intact navigation skills after bilateral loss of striate cortex', *Current Biology*. doi: 10.1016/j.cub.2008.11.002.
de Gelder, B. *et al.* (2015) 'Visual imagery influences brain responses to visual stimulation in bilateral cortical blindness', *Cortex*. Elsevier Ltd, 72, pp. 15–26. doi: 10.1016/j.cortex.2014.11.009.
Geman, S. and Geman, D. (1984) 'Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6(6), pp. 721–741. doi: 10.1109/TPAMI.1984.4767596.
Gilbert, D. T. *et al.* (1998) 'Immune Neglect: A Source of Durability Bias in Affective Forecasting', *Journal of Personality and Social Psychology*, 75(3), pp. 617–638. doi: 10.1037/0022-3514.75.3.617.
Gilbert, D. T., Meyers, J. and Wilson, T. D. (2008) 'Lessons From the Past: Do People Learn From Experience That Emotional Reactions Are Short-Lived? Timothy', 27(12), pp. 1648–1661. doi: 10.1177/01461672012712008.
Gilbert, D. T. and Wilson, T. D. (2007) 'Prospection: Experiencing the future', *Science*, 317(5843), pp. 1351–1354. doi: 10.1126/science.1144161.
Goldman, A. (2006) *Simulating Minds: The Philosophy, Psychology, and Neuroscience of Mindreading*. Oxford University Press USA.
Goodale, M. A. and Milner, A. D. (1992) 'Separate visual pathways for perception and action', *Trends in Neurosciences*. doi: 10.1016/0166-2236(92)90344-8.
Goode, T. D. *et al.* (2020) 'An Integrated Index: Engrams, Place Cells, and Hippocampal Memory', *Neuron*. Elsevier Inc., 107(5), pp. 805–820. doi: 10.1016/j.neuron.2020.07.011.
Gopnik, A. and Tenenbaum, J. B. (2007) 'Bayesian networks, Bayesian learning and cognitive development', *Developmental Science*, 10(3), pp. 281–287. doi: 10.1111/j.1467-7687.2007.00584.x.
Gopnik, A. and Wellman, H. M. (2012) 'Reconstructing constructivism: Causal models, Bayesian learning mechanisms, and the theory theory', *Psychological Bulletin*. doi: 10.1037/a0028044.
Gottfried, J. *et al.* (2004) 'Remembrance of odors past: Human olfactory cortex in cross-modal recognition memory', *Neuron*, 42(4), pp. 687–695. doi: 10.1016/S0896-6273(04)00270-3.
Greely, N. (2021) 'Epistemic feelings, metacognition, and the Lima problem', *Synthese*. Springer Netherlands, (February). doi: 10.1007/s11229-021-03094-8.
Greenberg, D. L. and Knowlton, B. J. (2014) 'The role of visual imagery in autobiographical memory', *Memory and Cognition*, 42(6), pp. 922–934. doi: 10.3758/s13421-014-0402-5.
Guillot, A. (2020) 'Neurophysiological Foundations and Practical Applications of Motor Imagery', in Abraham, A. (ed.) *The Cambridge Handbook of the Imagination*. Cambridge University Press (Cambridge Handbooks in Psychology), pp. 207–226. doi: 10.1017/9781108580298.014.

Habacha, H., Molinaro, C. and Dosseville, F. (2014) 'Effects of gender, imagery ability, and sports practice on the performance of a mental rotation task', *American Journal of Psychology*, 127(3), pp. 313–323. doi: 10.5406/amerjpsyc.127.3.0313.

Hájek, A. (2019) 'Interpretations of Probability', in Zalta, E. N. (ed.) *The Stanford Encyclopedia of Philosophy*.

Hall, C. R. *et al.* (1998) 'Imagery use by athletes: Development of the Sport Imagery Questionnaire.', *International Journal of Sport Psychology*. Italy: Edizioni Luigi Pozzi, 29(1), pp. 73–89.

Hallford, D. J. *et al.* (2018) 'Psychopathology and episodic future thinking: A systematic review and meta-analysis of specificity and episodic detail', *Behaviour Research and Therapy*. Elsevier, 102(September 2017), pp. 42–51. doi: 10.1016/j.brat.2018.01.003.

Halpern, A. R. (1988) 'Mental Scanning in Auditory Imagery for Songs', *Journal of Experimental Psychology: Learning, Memory, and Cognition*. doi: 10.1037/0278-7393.14.3.434.

Halpern, A. R. *et al.* (2004) 'Behavioral and neural correlates of perceived and imagined musical timbre', *Neuropsychologia*, 42(9), pp. 1281–1292. doi: 10.1016/j.neuropsychologia.2003.12.017.

Handy, T. C. *et al.* (2004) 'Visual imagery and memory: Do retrieval strategies affect what the mind's eye sees?', *European Journal of Cognitive Psychology*. doi: 10.1080/09541440340000457.

Hassabis, D., Kumaran, D. and Maguire, E. A. (2007) 'Using imagination to understand the neural basis of episodic memory', *Journal of Neuroscience*, 27(52), pp. 14365–14374. doi: 10.1523/JNEUROSCI.4549-07.2007.

Hastings, W. K. (1970) 'Monte Carlo Sampling Methods Using Markov Chains and Their Applications', *Biometrika*, 57(1), pp. 97–109.

Helming, K. A., Strickland, B. and Jacob, P. (2016) 'Solving the Puzzle about Early Belief-Ascription', *Mind and Language*. doi: 10.1111/mila.12114.

Herholz, S. C., Halpern, A. R. and Zatorre, R. J. (2012) 'Neuronal correlates of perception, imagery, and memory for familiar tunes', *Journal of Cognitive Neuroscience*, 24(6), pp. 1382–1397. doi: 10.1162/jocn_a_00216.

Hesse, C., Ball, K. and Schenk, T. (2012) 'Visuomotor performance based on peripheral vision is impaired in the visual form agnostic patient DF', *Neuropsychologia*. doi: 10.1016/j.neuropsychologia.2011.11.002.

Heyes, C. (2014) 'False belief in infancy: a fresh look', *Developmental science*, 17(5), pp. 647–659. doi: 10.1111/desc.12148.

Hohwy, J. (2020) 'New directions in predictive processing', *Mind and Language*, 35(2), pp. 209–223. doi: 10.1111/mila.12281.

Hume, D. (1739) *A Treatise of Human Nature*. Edited by A. L. Selby-Bigge. Oxford: Clarendon Press.

Hyman, I. E., Husband, T. H. and Billings, F. J. (1995) 'False memories of childhood experiences', *Applied Cognitive Psychology*. doi: 10.1002/acp.2350090302.

Hyman, I. E. and Loftus, E. F. (1998) 'Errors in autobiographical memory', *Clinical Psychology Review*, 18(8), pp. 933–947. doi: 10.1016/S0272-7358(98)00041-5.

Hyman, I. E. and Pentland, J. (1996) 'The role of mental imagery in the creation of false childhood memories', *Journal of Memory and Language*. doi: 10.1006/jmla.1996.0006.

Intraub, H., Bender, R. S. and Mangels, J. A. (1992) 'Looking at Pictures But Remembering Scenes', *Journal of Experimental Psychology: Learning, Memory, and Cognition*. doi: 10.1037/0278-7393.18.1.180.

Irish, M. *et al.* (2012) 'Considering the role of semantic memory in episodic future thinking: Evidence from semantic dementia', *Brain*, 135(7), pp. 2178–2191. doi: 10.1093/brain/aws119.

Irish, M. and Piguet, O. (2013) 'The pivotal role of semantic memory in remembering the past and imagining the future', *Frontiers in Behavioral Neuroscience*. doi: 10.3389/fnbeh.2013.00027.

Jacobs, C., Schwarzkopf, D. S. and Silvanto, J. (2018) 'Visual working memory performance in aphantasia', *Cortex*. Elsevier Ltd, 105, pp. 61–73. doi: 10.1016/j.cortex.2017.10.014.

Jakubowski, K. (2020) 'Musical Imagery', in Abraham, A. (ed.) *The Cambridge Handbook of the Imagination*. Cambridge University Press (Cambridge Handbooks in Psychology), pp. 187–206. doi: 10.1017/9781108580298.013.

Jakubowski, K., Farrugia, N. and Stewart, L. (2016) 'Probing imagined tempo for music: Effects of motor engagement and musical experience', *Psychology of Music*, 44(6), pp. 1274–1288. doi: 10.1177/0305735615625791.

Jeannerod, M. (1994) 'The representing brain: Neural correlates of motor intention and imagery.', *Behavioral and Brain Sciences*. United Kingdom: Cambridge University Press, 17(2), pp. 187–245. doi: 10.1017/S0140525X00034026.

Johansson, F. *et al.* (2014) 'Memory trace and timing mechanism localized to cerebellar Purkinje cells', *Proceedings of the National Academy of Sciences of the United States of America*. doi: 10.1073/pnas.1415371111.

Johansson, F. *et al.* (2015) 'Activation of a Temporal Memory in Purkinje Cells by the mGluR7 Receptor', *Cell Reports*. doi: 10.1016/j.celrep.2015.10.047.

Jones, M. and Love, B. C. (2011) 'Bayesian fundamentalism or enlightenment? on the explanatory status and theoretical contributions of bayesian models of cognition', *Behavioral and Brain Sciences*, 34(4), pp. 169–188. doi: 10.1017/S0140525X10003134.

Jones, M. and Wilkinson, S. (2020) ‘From Prediction to Imagination’, in Abraham, A. (ed.) *The Cambridge Handbook of the Imagination*. Cambridge University Press.

Jusczyk, P. W. (2003) ‘Chunking language input to find patterns.’, in *Early category and concept development: Making sense of the blooming, buzzing confusion.* New York, NY, US: Oxford University Press, pp. 27–49.

Kentridge, R. W., Heywood, C. A. and Weiskrantz, L. (1999) ‘Attention without awareness in blindsight’, *Proceedings of the Royal Society B: Biological Sciences*. doi: 10.1098/rspb.1999.0850.

Keogh, R. and Pearson, J. (2018) ‘The blind mind: No sensory visual imagery in aphantasia’, *Cortex*. Elsevier Ltd, 105(2015), pp. 53–60. doi: 10.1016/j.cortex.2017.10.012.

Keogh, R. and Pearson, J. (2021) ‘Attention driven phantom vision: Measuring the sensory strength of attentional templates and their relation to visual mental imagery and aphantasia: Measuring attentional templates’, *Philosophical Transactions of the Royal Society B: Biological Sciences*, 376(1817). doi: 10.1098/rstb.2019.0688rstb20190688.

Kind, A. (2001) ‘Putting the Image Back in Imagination’, *International Phenomenological Society*, 62(1), pp. 85–109.

Kind, A. (2016) *Imagining Under Constraints*, *Knowledge Through Imagination*. doi: 10.1093/acprof:oso/9780198716808.003.0007.

Kind, A. (2020a) ‘The Skill of Imagination’, in Fridland, E. and Pavese, C. (eds) *Routledge Handbook of Philosophy of Skill and Expertise*, pp. 335–346.

Kind, A. (2020b) ‘What imagination teaches’, in Schwenkler, J. and Lambert, E. (eds) *Transformative Experience*. Oxford University Press. Available at: https://search.proquest.com/docview/1876853930?accountid=45451%5Cnhttp://linksource.ebsco.com/linking.aspx?sid=ProQuest+Dissertations+%26+Theses+Global&fmt=dissertation&genre=dissertations+%26+theses&issn=&volume=&issue=&date=2016-01-01&spage=&title=What+.

Kirchhoff, M. D. (2018) ‘Predictive processing, perceiving and imagining: Is to perceive to imagine, or something close to it?’, *Philosophical Studies*. Springer Netherlands, 175(3), pp. 751–767. doi: 10.1007/s11098-017-0891-8.

Klein, S. B., Loftus, J. and Kihlstrom, J. F. (2002) ‘Memory and temporal experience: The effects of episodic memory loss on an amnesic patient’s ability to remember the past and imagine the future’, *Social Cognition*, 20(5), pp. 353–379. doi: 10.1521/soco.20.5.353.21125.

Koriat, A. (1993) ‘How do we know that we know? The accessibility model of the feeling of knowing’, *Psychological Review*. doi: 10.1037/0033-295X.100.4.609.

Koriat, A., Goldsmith, M. and Pansky, A. (2000) ‘Toward a psychology of memory accuracy.’, *Annual review of psychology*. doi: 10.1146/annurev.psych.51.1.481.

Kosslyn, S. M. (1973) 'Scanning Visual Images: Some Structural Implications', *Perception and Psychophysics*, 14, pp. 90–94.

Kosslyn, S. M. (1980) *Image and mind*. Cambridge, MA: Harvard University Press.

Kosslyn, S. M. (1983) *Ghosts in the Mind's machine: Creating and using images in the brain*. New York: Norton.

Kosslyn, S. M. *et al.* (1993) 'Visual mental imagery activates topographically organized visual cortex: PET investigations', *Journal of Cognitive Neuroscience*. doi: 10.1162/jocn.1993.5.3.263.

Kosslyn, S. M. (1994) *Image and Brain: The Resolution of the Imagery Debate*. MIT Press.

Kosslyn, S. M. *et al.* (1999) 'The role of area 17 in visual imagery: Convergent evidence from PET and rTMS', *Science*. doi: 10.1126/science.284.5411.167.

Kosslyn, S. M. *et al.* (2001) 'Neural foundations of imagery', *Nature Reviews Neuroscience*, 2(9), pp. 635–642.

Kosslyn, S. M., Ball, T. M. and Reiser, B. J. (1978) 'Visual images preserve metric spatial information: Evidence from studies of image scanning', *Journal of Experimental Psychology: Human Perception and Performance*, 4(1), pp. 47–60. doi: 10.1037/0096-1523.4.1.47.

Kosslyn, S. M., Behrmann, M. and Jeannerod, M. (1995) 'The cognitive neuroscience of mental imagery', *Neuropsychologia*, 33(11), pp. 1335–1344. doi: 10.1016/0028-3932(95)00067-D.

Kosslyn, S. M. and Shwartz, S. P. (1977) 'A simulation of visual imagery', *Cognitive Science*, 1(3), pp. 265–295. doi: 10.1016/S0364-0213(77)80020-7.

Kosslyn, S. M., Thompson, W. and Ganis, G. (2006) *The Case for Mental Imagery*. Oxford University Press.

Kouider, S. and Dehaene, S. (2007) 'Levels of processing during non-conscious perception: A critical review of visual masking', in *Philosophical Transactions of the Royal Society B: Biological Sciences*. doi: 10.1098/rstb.2007.2093.

Kung, P. (2010) 'Imagining as a guide to possibility', *Philosophy and Phenomenological Research*, 81(3), pp. 620–663. doi: 10.1111/j.1933-1592.2010.00377.x.

Langille, J. J. and Gallistel, C. R. (2020) 'Locating the engram: Should we look for plastic synapses or information-storing molecules?', *Neurobiology of Learning and Memory*. Elsevier, 169(June 2019), p. 107164. doi: 10.1016/j.nlm.2020.107164.

Leslie, A. (2000) '"Theory of Mind" as a Mechanism of Selective Attention', *The new cognitive neurosciences*.

Levine, B. *et al.* (2002) 'Aging and autobiographical memory: Dissociating episodic from semantic retrieval', *Psychology and Aging*, 17(4), pp. 677–689. doi: 10.1037/0882-7974.17.4.677.

Levine, L. J. (1997) 'Reconstructing Memory for Emotions', *Journal of Experimental Psychology:*

*General*. doi: 10.1037/0096-3445.126.2.165.

Levine, L. J. *et al.* (2012) 'Accuracy and artifact: Reexamining the intensity bias in affective forecasting', *Journal of Personality and Social Psychology*, 103(4), pp. 584–605.

Lewis, D. (1971) 'Analogue and digital', *Noûs*, 5, pp. 321–327.

Lewis, D. (1983) 'Mad Pain and Martian Pain', *Philosophical Papers*, pp. 122–130.

Loftus, E. F. and Palmer, J. C. (1974) 'Eyewitness Testimony', *Journal of Verbal Learning and Verbal Behavio*, (13), pp. 585–589.

Logie, R. H. *et al.* (2011) 'Low and high imagers activate networks differentially in mental rotation', *Neuropsychologia*. doi: 10.1016/j.neuropsychologia.2011.07.011.

Madore, K. P., Gaesser, B. and Schacter, D. L. (2014) 'Constructive episodic simulation: Dissociable effects of a specificity induction on remembering, imagining, and describing in young and older adults', *Journal of Experimental Psychology: Learning Memory and Cognition*, 40(3), pp. 609–622. doi: 10.1037/a0034885.

Madore, K. P., Jing, H. G. and Schacter, D. L. (2016) 'Divergent creative thinking in young and older adults: Extending the effects of an episodic specificity induction', *Memory and Cognition*. Memory & Cognition, 44(6), pp. 974–988. doi: 10.3758/s13421-016-0605-z.

Madore, K. P., Jing, H. G. and Schacter, D. L. (2019) 'Episodic specificity induction and scene construction: Evidence for an event construction account', *Consciousness and Cognition*, 68(August 2018), pp. 1–11. doi: 10.1016/j.concog.2018.12.001.

Mahon, B. Z. and Caramazza, A. (2011) 'What drives the organization of object knowledge in the brain?', *Trends in Cognitive Sciences*. doi: 10.1016/j.tics.2011.01.004.

Marks, D. F. (1973) 'Visual Imagery Differences in the Recall of Pictures', *British Journal of Psychology*, 64(1), pp. 17–24. doi: 10.1111/j.2044-8295.1973.tb01322.x.

Marmor, G. S. and Zaback, L. A. (1976) 'Mental rotation by the blind: Does mental rotation depend on visual imagery?', *Journal of Experimental Psychology: Human Perception and Performance*, 2(4), pp. 515–521. doi: 10.1037/0096-1523.2.4.515.

Marotta, J. J., Behrmann, M. and Goodale, M. A. (1997) 'The removal of binocular cues disrupts the calibration of grasping in patients with visual form agnosia', *Experimental Brain Research*. doi: 10.1007/PL00005731.

Marr, D. (1982) *Vision: A Computational Approach*. San Fransisco: Freeman & Co.

Martin, C. B. and Deutscher, M. (1966) 'Remembering', *The Philosophical Review*. [Duke University Press, Philosophical Review], 75(2), pp. 161–196. doi: 10.2307/2183082.

Martin, M. G. F. (2002) 'The transparency of experience', *Mind and Language*, 17(4), pp. 376–425. doi: 10.1111/1468-0017.00205.

McColgan, K. L. and McCormack, T. (2008) ‘Searching and Planning : Young Children’s Reasoning about Past and Future Event Sequences’, *Child Development*, 79(5), pp. 1477–1497.

McCune, L. (1995) ‘A Normative Study of Representational Play at the Transition to Language’, *Developmental Psychology*, 31(2), pp. 198–206. doi: 10.1037/0012-1649.31.2.198.

McNamara, T. P. (2005) *Semantic Priming: Perspectives from memory and word recognition*. New York: Psychology Press Taylor and Francis Group.

Mehta, Z., Newcombe, F. and De Haan, E. (1992) ‘Selective loss of imagery in a case of visual agnosia’, *Neuropsychologia*. doi: 10.1016/0028-3932(92)90069-X.

Michaelian, K. (2011) ‘Generative memory’, *Philosophical Psychology*, 24(3), pp. 323–342. doi: 10.1080/09515089.2011.559623.

Michaelian, K. (2016) *Mental time travel: Episodic memory and our knowledge of the personal past*. Cambridge, MA: MIT Press.

Michaelian, K. (2020) ‘Confabulating as Unreliable Imagining: In Defence of the Simulationist Account of Unsuccessful Remembering’, *Topoi*. Springer Netherlands, 39(1), pp. 133–148. doi: 10.1007/s11245-018-9591-z.

Michaelian, K. and Arango-Muñoz, S. (2014) ‘Epistemic feelings, epistemic emotions: Review and introduction to the focus section’, *Philosophical inquiries*, 2(1), pp. 97–122.

Michaelian, K. and Robins, S. K. (2018) ‘Beyond the causal theory? Fifty years after Martin and Deutscher’, *New Directions in the Philosophy of Memory*, pp. 13–32. doi: 10.4324/9781315159591.

Milner, A. D. and Goodale, M. A. (1995) *The visual brain in action*. 2nd edn. Oxford: Oxford University Press.

Milton, F. *et al.* (2020) ‘Behavioral and neural signatures of visual imagery vividness extremes: aphantasia vs. hyperphantasia’, *PsyArXiv*. Available at: https://doi.org/10.31234/osf.io/j2zpn.

Milton, F. *et al.* (2021) ‘Behavioral and Neural Signatures of Visual Imagery Vividness Extremes: Aphantasia versus Hyperphantasia’, *Cerebral Cortex Communications*, 2(2), pp. 1–15. doi: 10.1093/texcom/tgab035.

Mion, M. *et al.* (2010) ‘What the left and right anterior fusiform gyri tell us about semantic memory’, *Brain*, 133(11), pp. 3256–3268. doi: 10.1093/brain/awq272.

Mon-Williams, M. *et al.* (2001) ‘Monocular and binocular distance cues: Insights from visual form agnosia I (of III)’, *Experimental Brain Research*. doi: 10.1007/s002210000657.

Moro, V. *et al.* (2008) ‘Selective deficit of mental visual imagery with intact primary visual cortex and visual perception’, *Cortex*, 44(2), pp. 109–118. doi: 10.1016/j.cortex.2006.06.004.

Moscovitch, M. *et al.* (2005) ‘Functional neuroanatomy of remote episodic, semantic and spatial memory: A unified account based on multiple trace theory’, *Journal of Anatomy*, 207(1), pp. 35–

66. doi: 10.1111/j.1469-7580.2005.00421.x.

Munroe, K. J. *et al.* (2000) 'The four Ws of imagery use: Where, when, why, and what', *Sport Psychologist*, 14(2), pp. 119–137. doi: 10.1123/tsp.14.2.119.

Munzert, J. and Lorey, B. (2013) 'Motor and visual imagery in sports', *Multisensory Imagery*, 9781461458, pp. 319–341. doi: 10.1007/978-1-4614-5879-1_17.

Nanay, B. (2010) 'Perception and imagination: Amodal perception as mental imagery', *Philosophical Studies*, 150(2), pp. 239–254. doi: 10.1007/s11098-009-9407-5.

Nanay, B. (2015) 'Perceptual content and the content of mental imagery', *Philosophical Studies*. Springer Netherlands, 172(7), pp. 1723–1736. doi: 10.1007/s11098-014-0392-y.

Nanay, B. (2016) 'Hallucination as mental imagery', *Journal of Consciousness Studies*, 23(7–8), pp. 65–81.

Nanay, B. (2018) 'Multimodal mental imagery', *Cortex*. Elsevier Ltd, 105, pp. 125–134. doi: 10.1016/j.cortex.2017.07.006.

Nanay, B. (2021) 'Unconscious Mental Imagery', *Philosophical Transactions of the Royal Society B*.

Naselaris, T. *et al.* (2015) 'A voxel-wise encoding model for early visual areas decodes mental images of remembered scenes', *NeuroImage*. Elsevier B.V., 105, pp. 215–228. doi: 10.1016/j.neuroimage.2014.10.018.

Neath, I. and Nairne, J. S. (1995) 'Word-length effects in immediate memory: Overwriting trace decay theory', *Psychonomic Bulletin & Review*. doi: 10.3758/BF03210981.

Newell, A. (1990) *Unified Theories of Cognition*. Cambridge, MA: Harvard University Press.

Nichols, S. and Stich, S. (2000) 'A cognitive theory of pretense', *Cognition*, 74(2), pp. 115–147. doi: 10.1016/S0010-0277(99)00070-0.

Nichols, S. and Stich, S. P. (2004) *Mindreading: An Integrated Account of Pretence, Self-Awareness, and Understanding Other Minds*, *Mindreading: An Integrated Account of Pretence, Self-Awareness, and Understanding Other Minds*. Cleardon Press. doi: 10.1093/0198236107.001.0001.

Nielsen, M. and Dissanayake, C. (2004) 'Pretend play, mirror self-recognition and imitation: A longitudinal investigation through the second year', *Infant Behavior and Development*, 27(3), pp. 342–365. doi: 10.1016/j.infbeh.2003.12.006.

Nigel, T. (2021) *Mental Imagery*, *The Stanford Encyclopedia of Philosophy*. Edited by E. N. Zalta. Available at: https://plato.stanford.edu/archives/spr2021/entries/mental-imagery.

Norris, D. (2006) 'The Bayesian reader: Explaining word recognition as an optimal bayesian decision process', *Psychological Review*, 113(2), pp. 327–357. doi: 10.1037/0033-295X.113.2.327.

Norris, D. (2017) 'Short-term memory and long-term memory are still different', *Psychological*

*Bulletin*, 143(9), pp. 992–1009. doi: 10.1037/bul0000108.

Norris, D. and Cutler, A. (2021) 'More why , less how : What we need from models of cognition'. Elsevier B.V., (March).

Oberauer, K. (2002) 'Access to Information in Working Memory: Exploring the Focus of Attention', *Journal of Experimental Psychology: Learning Memory and Cognition*. doi: 10.1037/0278-7393.28.3.411.

Oberauer, K. (2009) 'Chapter 2 Design for a Working Memory', *Psychology of Learning and Motivation - Advances in Research and Theory*. doi: 10.1016/S0079-7421(09)51002-X.

Okada, H. and Matsuoka, K. (1992) 'Effects of Auditory Imagery on the Detection of a Pure Tone in White Noise: Experimental Evidence of the Auditory Perky Effect', *Perceptual and Motor Skills*, 74(2), pp. 443–448.

Olivetti Belardinelli, M. *et al.* (2009) 'An fMRI investigation on image generation in different sensory modalities: The influence of vividness', *Acta Psychologica*. Elsevier B.V., 132(2), pp. 190–200. doi: 10.1016/j.actpsy.2009.06.009.

Onishi, K. K. and Baillargeon, R. (2005) 'Do 15-month-old infants understand false beliefs?', *Science*, 308(5719), pp. 255–258. doi: 10.1126/science.1107621.

Page, J. W., Duhamel, P. and Crognale, M. A. (2011) 'ERP evidence of visualization at early stages of visual processing', *Brain and Cognition*. doi: 10.1016/j.bandc.2010.11.001.

Park, T. (2019) 'Why successful performance in imagery tasks does not require the manipulation of mental imagery', *Avant*, 10(2), pp. 1–11. doi: 10.26913/avant.2019.02.02.

Parkin, B. L., Ekhtiari, H. and Walsh, V. F. (2015) 'Non-invasive Human Brain Stimulation in Cognitive Neuroscience: A Primer', *Neuron*. doi: 10.1016/j.neuron.2015.07.032.

Pavese, C. (2021) 'Knowledge How', in Edward N. Zalta (ed.) *The Stanford Encyclopedia of Philosophy*. doi: https://plato.stanford.edu/archives/sum2021/entries/knowledge-how.

Pearson, J. *et al.* (2015) 'Mental Imagery: Functional Mechanisms and Clinical Applications', *Trends in Cognitive Sciences*. Elsevier Ltd, 19(10), pp. 590–602. doi: 10.1016/j.tics.2015.08.003.

Pearson, J. (2019) 'The human imagination: the cognitive neuroscience of visual mental imagery', *Nature Reviews Neuroscience*. Springer US, 20(10), pp. 624–634. doi: 10.1038/s41583-019-0202-9.

Pearson, J. (2020a) 'Reply to: Assessing the causal role of early visual areas in visual mental imagery', *Nature Reviews Neuroscience*, 21(9), p. 517. doi: 10.1038/s41583-020-0348-5.

Pearson, J. (2020b) 'The Visual Imagination', in Abraham, A. (ed.) *The Cambridge Handbook of the Imagination*, pp. 175–186.

Pegna, A. J., Landis, T. and Khateb, A. (2008) 'Electrophysiological evidence for early non-conscious processing of fearful facial expressions', *International Journal of Psychophysiology*. doi: 10.1016/j.ijpsycho.2008.08.007.

Perfors, A. *et al.* (2011) 'A tutorial introduction to Bayesian models of cognitive development', *Cognition*. Elsevier B.V., 120(3), pp. 302–321. doi: 10.1016/j.cognition.2010.11.015.

Perner, J. (1991) *Understanding the Representational Mind*. Cambridge, MA: MIT Press.

Perner, J. and Ruffman, T. (2005) 'Infants' insight into the mind: How deep?', *Science*, 308(5719), pp. 214–216. doi: 10.1126/science.1111656.

Perrin, D. and Michaelian, K. (2017) 'Memory as mental time travel', *The Routledge Handbook of Philosophy of Memory*, (Debus 2014), pp. 228–239. doi: 10.4324/9781315687315.

Pezzulo, G. and Castelfranchi, C. (2009) 'Thinking as the control of imagination: A conceptual framework for goal-directed systems', *Psychological Research*, 73(4), pp. 559–577. doi: 10.1007/s00426-009-0237-z.

Proust, J. (2013) *The philosophy of metacognition*. Oxford University Press.

Provost, A. *et al.* (2013) 'Two routes to expertise in mental rotation', *Cognitive Science*, 37(7), pp. 1321–1342. doi: 10.1111/cogs.12042.

Putnam, H. (1967) 'Psychological predicates', in Capitan, W. H. and Merrill, D. D. (eds) *Art, Mind, and Religion*. University of Pittsburgh Press.

Pylyshyn, Z. W. (1973) 'What the mind's eye tells the mind's brain: A critique of mental imagery', *Psychological Bulletin*, 80(1), pp. 1–24. doi: 10.1037/h0034650.

Pylyshyn, Z. W. (1979) 'The rate of "mental rotation" of images: A test of a holistic analogue hypothesis', *Memory & Cognition*, 7(1), pp. 19–28. doi: 10.3758/BF03196930.

Pylyshyn, Z. W. (1981) 'The imagery debate: Analogue media versus tacit knowledge', *Psychological Review*, 88(1), pp. 16–45. doi: 10.1037/0033-295X.88.1.16.

Pylyshyn, Z. W. (1984) *Computation and Cognition: Toward a Foundation for Cognitive Science*. Cambridge, MA: MIT Press.

Pylyshyn, Z. W. (2002) 'Mental imagery: In search of a theory', *Behavioral and Brain Sciences*, 25(2), pp. 157–182. doi: 10.1017/S0140525X02000043.

Pylyshyn, Z. W. (2003a) 'Return of the mental image: Are there really pictures in the brain?', *Trends in Cognitive Sciences*, 7(3), pp. 113–118. doi: 10.1016/S1364-6613(03)00003-2.

Pylyshyn, Z. W. (2003b) *Seeing and Visualizing: It's Not What You Think*. Bradford.

Rendell, P. G. *et al.* (2012) 'Older adults have greater difficulty imagining future rather than atemporal experiences', *Psychology and Aging*. doi: 10.1037/a0029748.

Renoult, L. *et al.* (2019) 'From Knowing to Remembering: The Semantic–Episodic Distinction', *Trends in Cognitive Sciences*. Elsevier Inc., 23(12), pp. 1041–1057. doi: 10.1016/j.tics.2019.09.008.

Rhodes, G. (1996) *Superportraits: Caricatures and recognition*. Hove: Psychology Press.

Riddoch, M. J. (1990) 'Loss of visual imagery: A generation deficit', *Cognitive Neuropsychology*. doi: 10.1080/02643299008253444.

Robins, S. K. (2016a) 'Misremembering', *Philosophical Psychology*, 29(3), pp. 432–447. doi: 10.1080/09515089.2015.1113245.

Robins, S. K. (2016b) 'Representing the past: memory traces and the causal theory of memory', *Philosophical Studies*. Springer Netherlands, 173(11), pp. 2993–3013. doi: 10.1007/s11098-016-0647-x.

Robins, S. K. (2020) 'Defending Discontinuism, Naturally', *Review of Philosophy and Psychology*. Review of Philosophy and Psychology, 11(2), pp. 469–486. doi: 10.1007/s13164-020-00462-0.

Roediger, H. L. and McDermott, K. B. (1995) 'Creating False Memories: Remembering Words Not Presented in Lists', *Journal of Experimental Psychology: Learning, Memory, and Cognition*. doi: 10.1037/0278-7393.21.4.803.

Rosenbaum, R. S. *et al.* (2000) 'Remote spatial memory in an amnesic person with extensive bilateral hippocampal lesions', *Nature Neuroscience*, 3(10), pp. 1044–1048. doi: 10.1038/79867.

Rosenbaum, R. S. *et al.* (2005) 'The case of K.C.: Contributions of a memory-impaired person to memory theory', *Neuropsychologia*, 43(7), pp. 989–1021. doi: 10.1016/j.neuropsychologia.2004.10.007.

Rossi, A. F. *et al.* (2009) 'The prefrontal cortex and the executive control of attention', in *Experimental Brain Research*. doi: 10.1007/s00221-008-1642-z.

Royet, J. P. *et al.* (2013) 'The impact of expertise in olfaction', *Frontiers in Psychology*, 4(DEC), pp. 1–11. doi: 10.3389/fpsyg.2013.00928.

Rudy, J. W., Huff, N. C. and Matus-Amat, P. (2004) 'Understanding contextual fear conditioning: Insights from a two-process model', in *Neuroscience and Biobehavioral Reviews*. doi: 10.1016/j.neubiorev.2004.09.004.

Rudy, J. W. and O'Reilly, R. C. (2001) 'Conjunctive representations, the hippocampus, and contextual fear conditioning', *Cognitive, Affective and Behavioral Neuroscience*. doi: 10.3758/CABN.1.1.66.

Sabbagh, M. A. *et al.* (2009) 'Neurodevelopmental correlates of theory of mind in preschool children', *Child Development*. doi: 10.1111/j.1467-8624.2009.01322.x.

Schacter, D. L. *et al.* (2012) 'The Future of Memory: Remembering, Imagining, and the Brain', *Neuron*. Elsevier Inc., 76(4), pp. 677–694. doi: 10.1016/j.neuron.2012.11.001.

Schacter, D. L. and Addis, D. R. (2007) 'The cognitive neuroscience of constructive memory: Remembering the past and imagining the future', *Philosophical Transactions of the Royal Society*

*B: Biological Sciences*, 362(1481), pp. 773–786. doi: 10.1098/rstb.2007.2087.

Schacter, D. L. and Addis, D. R. (2020) 'Memory and Imagination: Perspectives on Constructive Episodic Simulation', *The Cambridge Handbook of the Imagination*, pp. 111–131. doi: 10.5860/choice.50-4715.

Schacter, D. L., Addis, D. R. and Buckner, R. L. (2007) 'Remembering the past to imagine the future: The prospective brain', *Nature Reviews Neuroscience*, 8(9), pp. 657–661. doi: 10.1038/nrn2213.

Schacter, D. L., Norman, K. A. and Koutstaal, W. (1998) 'The cognitive neuroscience of constructive memory', *Annual Review of Psychology*. doi: 10.1146/annurev.psych.49.1.289.

Schacter, D. L. and Slotnick, S. D. (2004) 'The cognitive neuroscience of memory distortion', *Neuron*. doi: 10.1016/j.neuron.2004.08.017.

Schwartz, B. L. and Metcalfe, J. (2011) 'Tip-of-the-tongue (TOT) states: Retrieval, behavior, and experience', *Memory and Cognition*. doi: 10.3758/s13421-010-0066-8.

Schwitzgebel, E. (2002) 'How well do we know our own conscious experience? The case of visual imagery', *Journal of Consciousness Studies*, 9(5–6), pp. 35–53.

Scott, R. M. and Baillargeon, R. (2017) 'Early False-Belief Understanding', *Trends in Cognitive Sciences*. Elsevier Ltd, 21(4), pp. 237–249. doi: 10.1016/j.tics.2017.01.012.

Senden, M. *et al.* (2019) 'Reconstructing imagined letters from early visual cortex reveals tight topographic correspondence between visual mental imagery and perception', *Brain Structure and Function*. Springer Berlin Heidelberg, 224(3), pp. 1167–1183. doi: 10.1007/s00429-019-01828-6.

Servos, P. and Goodale, M. A. (1995) 'Preserved Visual Imagery in Visual Form Agnosia', 33(11), pp. 1383–1394.

Sheffield, J. M. *et al.* (2015) 'Fronto-parietal and cingulo-opercular network integrity and cognition in health and schizophrenia', *Neuropsychologia*. doi: 10.1016/j.neuropsychologia.2015.05.006.

Shepard, R. N. and Metzler, J. (1971) 'Mental rotation of three-dimensional objects', *Science*. US: American Assn for the Advancement of Science, pp. 701–703. doi: 10.1126/science.171.3972.701.

Shultz, T. R. (2007) 'The Bayesian revolution approaches psychological development', *Developmental Science*, 10(3), pp. 357–364. doi: 10.1111/j.1467-7687.2007.00588.x.

Singer, D. G. and Singer, J. L. (1990) *The house of make-believe: Children's play and the developing imagination*. Cambridge, MA: Harvard University Press.

Sirigu, A. and Duhamel, J. R. (2001) 'Motor and visual imagery as two complementary but neurally dissociable mental processes', *Journal of Cognitive Neuroscience*. doi: 10.1162/089892901753165827.

Slotnick, S. D., Thompson, W. L. and Kosslyn, S. M. (2005) 'Visual mental imagery induces retinotopically organized activation of early visual areas', *Cerebral Cortex*, 15(10), pp. 1570–1583. doi: 10.1093/cercor/bhi035.

Small, W. (2017) 'Agency and Practical Abilities', *Royal Institute of Philosophy Supplement*, 80(May), pp. 235–264. doi: 10.1017/s1358246117000133.

Smith, A. *et al.* (2004) 'fMRI correlates of the episodic retrieval of emotional contexts', *NeuroImage*, 22(2), pp. 868–878. doi: 10.1016/j.neuroimage.2004.01.049.

Smith, R. (2021) 'A Step-by-Step Tutorial on Active Inference and its Application to Empirical Data' doi: 10.31234/osf.io/b4jm6.

Soteriou, M. (2013) *The Mind's Construction*. Oxford University Press.

Spagna, A. *et al.* (2021) 'Visual mental imagery engages the left fusiform gyrus, but not the early visual cortex: A meta-analysis of neuroimaging evidence', *Neuroscience and Biobehavioral Reviews*. Elsevier Ltd, 122(December 2020), pp. 201–217. doi: 10.1016/j.neubiorev.2020.12.029.

Stanley, J. and Krakauer, J. W. (2013) 'Motor skill depends on knowledge of facts', *Frontiers in Human Neuroscience*, 7.

Stanley, J. and Williamson, T. (2001) 'Knowing How', *Journal of Philosophy*, 98(8), pp. 411–444.

Stanley, J. and Williamson, T. (2017) 'Skill', *Nous*, 51(4), pp. 713–726. doi: 10.1111/nous.12144.

Van den Stock, J. *et al.* (2014) 'Neural correlates of body and face perception following bilateral destruction of the primary visual cortices', *Frontiers in Behavioral Neuroscience*. doi: 10.3389/fnbeh.2014.00030.

Suddendorf, T., Addis, D. R. and Corballis, M. C. (2011) 'Mental Time Travel and the Shaping of the Human Mind', *Predictions in the Brain: Using Our Past to Generate a Future*, (February). doi: 10.1093/acprof:oso/9780195395518.003.0121.

Suddendorf, T. and Busby, J. (2005) 'Making decisions with the future in mind: Developmental and comparative identification of mental time travel', *Learning and Motivation*, 36(2 SPEC. ISS.), pp. 110–125. doi: 10.1016/j.lmot.2005.02.010.

Suddendorf, T. and Corballis, M. C. (1997) 'Mental time travel and the evolution of the human mind', *Genetic, Social, and General Psychology Monographs*.

Suddendorf, T. and Corballis, M. C. (2007) 'The evolution of foresight: What is mental time travel, and is it unique to humans?', *Behavioral and Brain Sciences*, 30(3), pp. 299–351. doi: 10.1017/S0140525X07001975.

Suddendorf, T. and Redshaw, J. (2013) 'The development of mental scenario building and episodic foresight', *Annals of the New York Academy of Sciences*, 1296(1), pp. 135–153. doi: 10.1111/nyas.12189.

Sutton, J. (2007) *Philosophy and memory traces: Descartes to connectionism*. Cambridge:

Cambridge University Press.

Talarico, J. M., Labar, K. S. and Rubin, D. C. (2004) ‘Emotional intensity predicts autobiographical memory experience’, *Memory and Cognition*, 32(7), pp. 1118–1132. doi: 10.3758/BF03196886.

Taylor, M. (2001) *Imaginary Companions and the Children Who Create Them*. Oxford University Press, Incorporated.

Teyler, T. J. and DiScenna, P. (1986) ‘The Hippocampal Memory Indexing Theory’, *Behavioral Neuroscience*. doi: 10.1037/0735-7044.100.2.147.

Teyler, T. J. and Rudy, J. W. (2007) ‘The Hippocampal Indexing Theory and Episodic Memory: Updating the Index’, *Hippocampus*, 17, pp. 1158–1169.

Thorudottir, S. *et al.* (2020) ‘The architect who lost the ability to imagine: The cerebral basis of visual imagery’, *Brain Sciences*, 10(2). doi: 10.3390/brainsci10020059.

Treisman, A. (2006) ‘How the deployment of attention determines what we see’, *Visual Cognition*, 14, pp. 411–443. doi: 10.4324/9780203783122.

Trojano, L. *et al.* (2002) ‘Coordinate and categorical judgements in spatial imagery. An fMRI study’, *Neuropsychologia*, 40(10), pp. 1666–1674. doi: 10.1016/S0028-3932(02)00021-0.

Tulving, E. (1972) ‘Chapter 10: Episodic and sematic memory’, in *Organisation of memory*.

Tulving, E. (1983) ‘Elements of Episodic Memory’, *Canadian Psychology*.

Tye, M. (1991) *The Imagery Debate*. Cambridge, MA: MIT Press.

Tye, M. (1995) *Ten Problems of Consciousness*. Cambridge, MA: MIT Press.

VanRullen, R. and Reddy, L. (2019) ‘Reconstructing faces from fMRI patterns using deep generative neural networks’, *Communications Biology*. doi: 10.1038/s42003-019-0438-y.

Viera, G. and Nanay, B. (2020) ‘Temporal Mental Imagery’, in Abraham, A. (ed.) *The Cambridge Handbook of the Imagination*. Cambridge University Press, pp. 227–240.

Walker, C. M. and Gopnik, A. (2013) ‘Causality and Imagination’, in Taylor, M. (ed.) *The Oxford Handbook of the Development of Imagination*.

Walker, C. M., Gopnik, A. and Ganea, P. A. (2015) ‘Learning to Learn From Stories: Children’s Developing Sensitivity to the Causal Structure of Fictional Worlds’, *Child Development*, 86(1), pp. 310–318. doi: 10.1111/cdev.12287.

Ward, A. M. (2016) ‘A critical evaluation of the validity of episodic future thinking: A clinical neuropsychology perspective’, *Neuropsychology*, 30(8), pp. 887–905. doi: 10.1037/neu0000274.

Weizkrantz, L. (2009) *Blindsight: A Case Study Spanning 35 Years and New Developments*. 2nd Editio. Oxford University Press.

Wellman, H. M. (2014) *Making Minds*. New York, NY: Oxford University Press.

Wheeler, M. E. *et al.* (2006) ‘Evidence for separate perceptual reactivation and search processes during remembering', *Cerebral Cortex*. doi: 10.1093/cercor/bhj037.

Wheeler, M. E. and Buckner, R. L. (2003) 'Functional dissociation among components of remembering: Control, perceived oldness, and content', *Journal of Neuroscience*. doi: 10.1523/jneurosci.23-09-03869.2003.

Wheeler, M. E. and Buckner, R. L. (2004) 'Functional-anatomic correlates of remembering and knowing', *NeuroImage*. doi: 10.1016/j.neuroimage.2003.11.001.

Wheeler, M. E., Petersen, S. E. and Buckner, R. L. (2000) 'Memory's echo: Vivid remembering reactivates sensory-specific cortex', *Proceedings of the National Academy of Sciences of the United States of America*, 97(20), pp. 11125–11129. doi: 10.1073/pnas.97.20.11125.

Whiteley, C. M. K. (2020) 'Aphantasia, imagination and dreaming', *Philosophical Studies*. Springer Netherlands, (August). doi: 10.1007/s11098-020-01526-8.

Whitwell, R. L. *et al.* (2015) 'Patient DF's visual brain in action: Visual feedforward control in visual form agnosia', *Vision Research*. Elsevier Ltd, 110(PB), pp. 265–276. doi: 10.1016/j.visres.2014.08.016.

Wicken, M., Keogh, R. and Pearson, J. (forthcoming) 'The Critical Role of Mental Imagery in Human Emotion: Insights from fear-based imagery and aphantasia', *Proceedings of Royal Society: B*.

Williams, D. (2021) 'Imaginative Constraints and Generative Models', *Australasian Journal of Philosophy*. Taylor & Francis, 99(1), pp. 68–82. doi: 10.1080/00048402.2020.1719523.

Williams, S. E. and Cumming, J. (2014) 'The Sport Imagery Ability Questionnaire Manual', *University of Birmingham*, (April), pp. 1–7. doi: 10.13140/RG.2.1.1608.6565.

Williamson, T. (2016) 'Knowing by Imagining', *Knowledge Through Imagination*, pp. 113–123. doi: 10.1093/acprof:oso/9780198716808.003.0005.

Wilson, T. D. *et al.* (2000) 'Focalism: A source of durability bias in affective forecasting', *Journal of Personality and Social Psychology*, 78(5), pp. 821–836. doi: 10.1037/0022-3514.78.5.821.

Wilson, T. D. and Gilbert, D. T. (2000) 'Affective Forecasting', *Advances in Experimental Social Psychology*, 35, pp. 345–411. doi: http://dx.doi.org/10.1016/j.pain.2011.02.015.

Wilson, T. D., Meyers, J. and Gilbert, D. T. (2003) '"How Happy Was I, Anyway?" A Retrospective Impact Bias', *Social Cognition*, 21(6), pp. 421–446. doi: 10.1521/soco.21.6.421.28688.

Winkielman, P. *et al.* (2003) *The Hedonic Marking of Processing Fluency: Implications for Evaluative Judgment The Psychology of Evaluation: Affective Processes in Cognition and Emotion*, *The Psychology of Evaluation: Affective Processes in Cognition and Emotion*.

Winlove, C. I. P. *et al.* (2018) 'The neural correlates of visual imagery: A co-ordinate-based meta-analysis', *Cortex*. Elsevier Ltd, 105, pp. 4–25. doi: 10.1016/j.cortex.2017.12.014.

Wu, W. (2011) ‘Attention as Selection for Action’, in Mole, C., Smithies, D., and Wu, W. (eds) *Attention: Philosophical and Psychological Essays*. Oxford University Press.

Wu, W. (2016) ‘Experts and Deviants: The Story of Agentive Control’, *Philosophy and Phenomenological Research*, 92(2).

Xuan, B. *et al.* (2016) ‘The activation of interactive attentional networks’, *NeuroImage*. doi: 10.1016/j.neuroimage.2016.01.017.

Yarbus, A. (1967) *Eye Movement and Vision*. New York: Plenum Press.

Young, B. (2019) ‘Olfactory imagery: is exactly what it smells like’, *Philosophical Studies*. Springer Netherlands. doi: 10.1007/s11098-019-01371-4.

Zacks, J. M. (2008) ‘Neuroimaging studies of mental rotation: A meta-analysis and review’, *Journal of Cognitive Neuroscience*, 20(1), pp. 1–19. doi: 10.1162/jocn.2008.20013.

Zatorre, R. J., Halpern, A. R. and Bouffard, M. (2010) ‘Mental reversal of imagined melodies: A role for the posterior parietal cortex’, *Journal of Cognitive Neuroscience*, 22(4), pp. 775–789. doi: 10.1162/jocn.2009.21239.

Zeman, A. *et al.* (2010) ‘Loss of imagery phenomenology with intact visuo-spatial task performance: A case of “blind imagination”’, *Neuropsychologia*, 48(1), pp. 145–155. doi: 10.1016/j.neuropsychologia.2009.08.024.

Zeman, A. (2020) ‘Phantasia - The psychological significance of lifelong visual imagery vividness extremes - Supplementary Material’.

Zeman, A. *et al.* (2020) ‘Phantasia – The psychological significance of lifelong visual imagery vividness extremes’, *Cortex*. Elsevier Ltd, 130, pp. 426–440. doi: 10.1016/j.cortex.2020.04.003.

Zeman, A., Dewar, M. and Della Sala, S. (2015) ‘Lives without imagery - Congenital aphantasia’, *Cortex*. Elsevier Srl., 73, pp. 378–380. doi: 10.1016/j.cortex.2015.05.019.