

UC San Diego

UC San Diego Electronic Theses and Dissertations

Title

Individual Differences in the Format of Mental Representations and Their Cognitive Consequences

Permalink

<https://escholarship.org/uc/item/4v68s2jc>

ISBN

9798255421664

Author

Chkhaidze, Ana

Publication Date

2026-06-25

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA SAN DIEGO

Individual Differences in the Format of Mental Representations and Their Cognitive Consequences

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy

in

Cognitive Science

by

Ana Chkhaidze

Committee in charge:

Professor Lera Boroditsky, Co-Chair
Professor Anastasia Kiyonaga, Co-Chair
Professor Seana Coulson
Professor Marta Kutas
Professor John Serences

2026

Copyright

Ana Chkhaidze, 2026

All rights reserved.

The Dissertation of Ana Chkhaidze is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2026

DEDICATION

To Coco, who will hopefully never have to read this dissertation.

EPIGRAPH

An ‘inner process’ stands in need of outward criteria.

— *Ludwig Wittgenstein, Philosophical Investigations, §580*

TABLE OF CONTENTS

Dissertation Approval Page	iii
Dedication	iv
Epigraph	v
Table of Contents	vi
List of Figures	ix
List of Tables	xi
Acknowledgements	xii
Vita	xv
Abstract of the Dissertation	xvi
Introduction	I
o.1 What Are Thoughts Made Of?	I
o.2 Sensorimotor Theories of Cognition	2
o.3 Individual Differences in Mental Representations	3
o.4 The Puzzle of Absent Imagery	4
o.5 Approaching the Puzzle	5
Chapter I Visual and Verbal Thought Tendencies Shape Working Memory Under Distrac- tion	8
I.1 Introduction	8
I.2 Research Transparency Statement	10
I.2.1 General Disclosures	10
I.2.2 Study Disclosures	10
I.3 Methods	11
I.3.1 Participants	11
I.3.2 Materials and Measures	12
I.3.3 Experimental Task Procedure	15
I.3.4 Statistical Analysis	18
I.4 Results	19
I.4.1 Model-Derived Memorability Predicts WM Performance	19
I.4.2 Individual Imagery Tendencies Predict WM Error Patterns	21
I.4.3 Imagery Profiles Matter Most for WM Content in the Back of the Mind	23
I.4.4 The Effects of Visual Input During WM Depend on Stimulus Memorability and Imagery Profile	24
I.4.5 Visual Imagery Amplifies Distractor Costs for Weakly Memorable Items	27

1.5	Discussion	29
1.5.1	Memorability Shapes WM Performance and Its Vulnerability to Visual Input	29
1.5.2	Different WM Strategies Carry Different Costs	30
1.5.3	Rethinking Imagery in WM	30
1.5.4	Limitations and Future Directions	31
1.5.5	Concluding Remarks	32
Chapter 2	Visual Imagery Modulates Neural, but Not Behavioral, Signatures of Mental Simulation During Language Comprehension	33
2.1	Introduction	33
2.2	Experiment 1: Behavioral Study	36
2.2.1	Methods	36
2.2.2	Results	39
2.3	Experiment 2: ERP Study	40
2.3.1	Methods	40
2.3.2	Results	45
2.4	Discussion	50
2.4.1	Neural Underpinnings of the Match Effect	50
2.4.2	Visual Imagery as a Stress Test for Perceptual Simulation	52
2.4.3	Limitations and Future Directions	53
2.4.4	Concluding Remarks	54
Chapter 3	From Dots to Faces: Individual Differences in Visual Imagery Capacity Predict the Content of Ganzflicker-induced Hallucinations	55
3.1	Introduction	55
3.1.1	The Layered Model of Imagery	57
3.1.2	Individual Differences in Imagery and the Induction of Visual Hallucinations	59
3.1.3	The Present Study	62
3.2	Materials and Methods	64
3.2.1	Participants	64
3.2.2	Procedure	65
3.2.3	Topic Modeling	65
3.2.4	Sensorimotor Content Analysis	74
3.3	Results	77
3.3.1	Unsupervised Topic Modeling Identifies Systematic Themes Across the Imagery Spectrum	77
3.3.2	Human-Assessed Sensorimotor Content of Descriptions Varies With Imagery Vividness	81
3.4	Discussion	86
3.4.1	Phenomenological Profile of Ganzflicker-Induced Hallucinations	86
3.4.2	Hallucinated Content Maps Onto the Visual Processing Hierarchy Across the Imagery Spectrum	88
3.4.3	Sensorimotor Content of Hallucination Descriptions May Reflect Multisensory Simulation	90

3.4.4	Limitations	91
3.4.5	Conclusion	93
Chapter 4	Some Bottlenecks in Imagery Research	95
4.1	The Promise and Puzzle of Visual Imagery Diversity	95
4.2	Bottleneck 1: Visual Imagery Is Ill-Defined	96
4.3	Bottleneck 2: Mismatch Between Imagery Measures and Cognitive Tasks	98
4.4	Bottleneck 3: Beyond Representational Absence in Aphantasia	100
4.5	Bottleneck 4: Similar Outcomes, Different Routes	101
4.6	Concluding Remarks	103
Conclusion		104
4.7	Summary of Contributions	104
4.8	Limitations and Future Directions	106
4.9	Closing Remarks	107
Appendix A	Chapter 1	108
A.1	Supplementary Tables	108
A.2	VVIQ and Attentional Priority	118
A.3	Three-way Interaction Between Imagery, Memorability, and Target-Distractor Similarity	119
Appendix B	Chapter 2	121
B.1	Distribution of Visual Imagery Vividness Scores in Experiment 1	121
B.2	Picture-Locked LMM Effects Across the Scalp	122
B.3	Behavioral Model Tables	123
B.4	ERP Model Tables	124
Appendix C	Chapter 3	127
C.1	Example Hallucination Descriptions	127
C.2	Hierarchical Structure of Topics	129
C.3	Lasso Regression Coefficients for Continuous Vividness Prediction	130
C.4	Imagery Group-Level Topic Expression	131
C.5	Lancaster Sensorimotor Norm Dimensions	132
C.6	Robustness of Sensorimotor Effects	135
C.7	Ganzflicker Experience and Imagery Vividness	136
References		138

LIST OF FIGURES

Figure 1.1.	Task schematic and study variables	17
Figure 1.2.	Model-derived memorability scores predict WM performance	20
Figure 1.3.	Individual differences in imagery predict distinct WM performance patterns . .	22
Figure 1.4.	Visual similarity interacts with target memorability and imagery profile	26
Figure 1.5.	Visual imagery vividness amplifies the interaction between target memorability and target-distractor similarity	28
Figure 2.1.	Sentence-picture verification task and behavioral results	38
Figure 2.2.	Distribution of imagery questionnaire scores in the EEG sample	41
Figure 2.3.	Behavioral and neural responses to sentence-picture verification in Experiment 2	46
Figure 2.4.	ERP signatures of the match effect and its modulation by visual imagery	49
Figure 2.5.	Alternative accounts of how visual imagery may relate to perceptual simulation during language comprehension	53
Figure 3.1.	Ganzflicker-induced hallucination drawings from a pilot participant	56
Figure 3.2.	Layered model of visual imagery across the visual hierarchy	58
Figure 3.3.	Ganzflicker-induced simple and complex hallucinations	61
Figure 3.4.	Ganzflicker task and analytical approach	63
Figure 3.5.	Topic modeling and analysis pipeline	66
Figure 3.6.	Topic representations and hierarchical relationships among hallucinations . . .	71
Figure 3.7.	Hallucination content patterns distinguishing weak from moderate and strong imagery groups	79
Figure 3.8.	Lancaster sensorimotor analysis pipeline and results	82
Figure A.1.	WM performance as a function of imagery vividness and attentional priority .	118
Figure A.2.	Visual imagery vividness and target-distractor similarity effects across memorability levels	119

Figure A.3.	Visual imagery quality moderates the interaction between stimulus memorability and target-distractor similarity	120
Figure B.1.	Distribution of visual imagery vividness (VVIQ) scores in Experiment 1	121
Figure B.2.	Picture-locked LMM effects across the scalp	122
Figure C.1.	Hierarchical relationships among topics	129
Figure C.2.	Lasso regression coefficients for continuous visual vividness prediction	130
Figure C.3.	Imagery group-level topic expression	131
Figure C.4.	Distributions of Lancaster Sensorimotor Norm dimensions across hallucination descriptions	132
Figure C.5.	Individual sensorimotor dimensions by visual imagery vividness	134
Figure C.6.	Sensorimotor effects remain unchanged after controlling for gender	135
Figure C.7.	Altered consciousness states during Ganzflicker are associated with stronger visual imagery	136
Figure C.8.	Participants with stronger visual imagery report more positive emotional evaluations of the Ganzflicker experience	137

LIST OF TABLES

Table 3.1.	Representative hallucination descriptions by topic	67
Table 3.2.	Hallucination content mapped onto the visual processing hierarchy	89
Table A.1.	Response time model results for IRQ predictors	109
Table A.2.	Accuracy model results for IRQ predictors	110
Table A.3.	Error-type model results for IRQ predictors	111
Table A.4.	Response time model results for OSIVQ predictors	112
Table A.5.	Accuracy model results for OSIVQ predictors	113
Table A.6.	Error-type model results for OSIVQ predictors	114
Table A.7.	Response time model results for VVIQ predictors	115
Table A.8.	Accuracy model results for VVIQ predictors	116
Table A.9.	Error-type model results for VVIQ predictors	117
Table B.1.	Reaction time model results for Experiment 2	123
Table B.2.	Accuracy model results for Experiment 2	123
Table B.3.	Mixed-effects model output for picture-locked N170 voltage	125
Table B.4.	Mixed-effects model output for picture-locked N400 voltage	126
Table C.1.	Descriptive statistics for Lancaster Sensorimotor Norm dimensions	133

ACKNOWLEDGEMENTS

I have so many people to thank, for so many things, that if I were allowed to, this would be the longest section of my dissertation.

First, Anastasia, I want to thank you for patiently watching me take the long way around every sensible piece of advice you gave me, and for never once saying, “I told you so.” I am forever impressed by the rare way you combine intellectual rigor with creative freedom: asking questions that are sprawling and ambitious in the best way, and then insisting on answering them with psychophysics-level precision. This probably explains why so many of my papers are still almost ready.

Lera, thank you for doing the kind of research that made me want to do cognitive science in the first place. Thank you for bringing me to UCSD, for being such a kind advisor, and for pushing me to think bigger, deeper and more boldly.

Seana, you have been my advisor throughout these years, whether you wanted to be or not. Thank you for teaching us the best class in the world and for poisoning my intellectual taste with imagery. You may or may not be partly responsible for the jobs I fail to find because of it. Also, thank you for teaching me how to work with EEG data. I still can’t believe I can somewhat do that.

Marta, thank you for somehow thinking more highly of me than I probably deserved, and for telling me so at exactly the moments when I needed to hear it. Thank you also for noticing what was happening in Georgia, and for checking in. I am also deeply grateful to John and Sean for their guidance, generosity, and thoughtful feedback on my work.

I want to thank everyone in the Coulson and Kiyonaga labs. Holly, for the deep conversations and after-hours dark humor; Pria, for incredible generosity and for teaching me how to collect EEG data; Sihan and Ruby, for doing brilliant work so effortlessly that it was both inspiring and mildly intimidating; Janna and Vivien, for joining the lab just as I was leaving and making leaving feel even sadder than it already was. And Lana, for the nine-hour data-collection days, but long before that, and far more than that, for being my friend in sickness and in health, through this PhD and long after it.

It has been a great privilege to be part of this historic department: a rare intellectual home, full of brilliant people who have shared their wisdom so generously and without reservation. I have also

been lucky to have some of the best co-authors and collaborators. In particular, thank you to Reshanne Reeder, from whom I have learned so much about visual imagery and hallucinations, and perhaps even more about what good collaboration can be.

To EuroTrashHaus — Oisín, Felix, and Cameron — thank you for making San Diego feel less like San Diego. To be fair, we lived in La Mesa. Felix, thank you for teaching me how to drive and for doing so with the most unreasonable car, and for so much more that made me feel safe and at home. Oisín, for the late-night conversations about everything and nothing, and for poetry nights. Cameron, for possibly making me a slightly worse person, but definitely a funnier one. Thank you to Tornike, Parla, Pam, Sean, Harshada and other San Diego friends for being exactly the kind of clever and caring people one hopes to find in graduate school. To BB — Anita, Nina, Sophie, and Teko — for the never-ending chat notifications, and for lifelong laughter and love. And Tamara, thank you for going through life with me, even from 9,000 kilometers away. There are many more friends, in Tbilisi, Berlin, and all over the world who carried me through this PhD in ways big and small: with conversations, distractions, voice messages, gossip, advice, and the kind of care that made distance feel less absurd. I cannot name everyone here, but I am grateful for all of you.

Thank you most of all to my family: for unconditional love and patience as I wandered from class to class, field to field, and eventually country to country. Thank you for trusting me through all the detours, and for making everywhere I went feel connected to home.

Finally, Benni, thank you for loving me even when this dissertation took up more of our life than it should have; for so graciously rearranging so much of your life around mine; and, most of all, for Coco.

Chapter 1, in full, is currently being prepared for submission for publication of the material as Visual and verbal thought tendencies shape working memory under distraction. Chkhaidze, A., Gaspariani, L., Piriyaikitakonkij, M., Pappas, I., Ballard, I. C., & Kiyonaga, A. The dissertation author was the primary investigator and author of this material.

Chapter 2, in full, is currently being prepared for submission for publication of the material as Visual imagery modulates neural, but not behavioral, signatures of mental simulation during language

comprehension. Chkhaidze, A., Boroditsky, L., Trott, S., Coulson, S., & Kiyonaga, A. The dissertation author was the primary investigator and author of this material.

Chapter 3, in full, is a reprint of the material as it appears in Chkhaidze, A., Reeder, R. R., Gag, C., Kiyonaga, A., & Coulson, S. (2026). From dots to faces: individual differences in visual imagery capacity predict the content of Ganzflicker-induced hallucinations. *Neuroscience of Consciousness*, 2026(1), niago16. The dissertation author was the primary investigator and author of this material.

Chapter 4, in full, is currently being prepared for submission for publication of the material as Some bottlenecks in imagery research. Chkhaidze, A., & Kiyonaga, A. The dissertation author was the primary investigator and author of this material.

VITA

2015	Bachelor of Science in Social Sciences, Free University of Tbilisi
2017	Visiting Scholar, Gladstone Institute of Neurological Disease, San Francisco
2017	Master of Science in Biology, Agricultural University of Georgia
2018	Visiting Lecturer in Neuroscience, Free University of Tbilisi
2019	Master of Science in Neuroscience, University of Nevada, Reno
2019	Data Scientist, Leavingstone
2020	Visiting Lecturer in Neuroscience, Ilia State University
2026	Doctor of Philosophy in Cognitive Science, University of California San Diego

ABSTRACT OF THE DISSERTATION

Individual Differences in the Format of Mental Representations and Their Cognitive Consequences

by

Ana Chkhaidze

Doctor of Philosophy in Cognitive Science

University of California San Diego, 2026

Professor Lera Boroditsky, Co-Chair
Professor Anastasia Kiyonaga, Co-Chair

People vary greatly in how they experience their thoughts. Some say they think in vivid mental images, whereas others say they think in words. These internal representations play a central role in cognition, supporting memory, language, planning, and imagination. This raises the possibility that diversity in how people internally represent the world produces diversity in how cognition unfolds across minds. In this dissertation, I examine this possibility across three cases: when internal representations are maintained from visual input, constructed from language, or generated with little external constraint on their content.

First, I examine visual working memory, when people hold visual information in mind across a

short delay. Using a task with naturalistic images, I show that people with different imagery profiles may rely on different representational routes even when performing the same task with the same stimuli. These routes differ in what kind of information they preserve, how well they support simultaneous maintenance of multiple items, and whether similar visual input during maintenance interferes with or reinforces the memory trace.

Second, I turn to language comprehension, when people internally represent meaning conveyed through a non-visual medium. Embodied theories of meaning suggest that understanding language can involve recruiting sensory and motor systems. In a task previously used to index perceptual simulation during comprehension, I find that imagery vividness does not reliably modulate overt behavior. Neural responses, however, suggest that similar behavior may reflect different processing routes. Participants across the imagery spectrum show sensitivity to the task manipulation in a later window linked to meaning processing, whereas stronger visual imagers also show sensitivity in an earlier window linked to visual-perceptual processing. These findings suggest that the extent to which language comprehension recruits perceptual simulation may depend on imagery vividness.

Third, I explore internally generated visual experience through Ganzflicker-induced hallucinations, where visual content can emerge freely rather than being specified by a particular image or sentence. Analyzing free-text descriptions from over four thousand participants, I identify systematic differences in the kinds of content people report experiencing across the imagery spectrum, from simpler visual phenomena such as colors, flashes, and geometric patterns to complex objects, faces, scenes, and richer sensorimotor content.

Finally, I synthesize these findings into a broader framework for imagery research. I argue that mixed findings in the literature may reflect bottlenecks in how imagery is defined, measured, and linked to cognitive tasks, and propose some concrete ways forward. Together, these studies suggest that imagery diversity is not merely a curiosity about private experience. It is a tool for revealing how different minds can reach similar cognitive outcomes through different internal routes.

Introduction

0.1 What Are Thoughts Made Of?

As humans, we have a remarkable ability to guide behavior based on internally generated representations rather than reacting solely to external environments. We remember things that are long gone, plan futures we have not yet experienced, and reason about ideas with no physical referent. This capacity extends cognition beyond the immediate sensory world to the world of ideas and concepts, to what we call higher order cognition. Given their fundamental role in human cognition and behavior, uncovering the nature of these mental representations has been a central pursuit for centuries across philosophy (Dennett, 1969; Hume, 2000; Locke, 1948; Wittgenstein, 1953), neuroscience (Baker et al., 2022; Clark & Toribio, 1994; Kragel et al., 2018; Mirski & Bickhard, 2019), and cognitive science (Chalmers, 1993; Kosslyn, 1980; Pylyshyn, 1979; Shepard & Cooper, 1986). So, what are thoughts made of?

Some have argued that mental representations are depictive; that they iconically re-present the world and our experiences, recruiting neural resources used in perception and action (Kosslyn, 1994; Pearson & Kosslyn, 2015; Pearson, 2019). Others have maintained that information is stored in a descriptive, language-like format, and that the perceptual quality of imagery is epiphenomenal to the underlying representations (Pylyshyn, 1973, 1981, 2002). These debates have been difficult to resolve because internal representations are not directly observable.

Over the past several decades, neuroimaging has narrowed this gap. Studies have shown that visual imagery and visual perception recruit overlapping neural systems, including category-selective regions and early visual cortex (Ganis et al., 2004; O'Craven & Kanwisher, 2000; Slotnick et al., 2005).

The contents of imagery can be decoded from visual areas, and disrupting the visual cortex can impair imagery-based performance (Harrison & Tong, 2009; Naselaris et al., 2015; Stokes et al., 2009; Kosslyn et al., 1999). Beyond visual imagery, networks involved in action planning and execution are recruited during motor imagery, including premotor-parietal and subcortical systems (Hardwick et al., 2018; O’Shea, 2022), while inner speech and auditory-verbal imagery engage partially overlapping speech-production and speech-perceptual systems (Barber et al., 2021; Pratts et al., 2023; Rollo et al., 2026). Together, this evidence suggests that internally generated representations can draw on modality-specific neural systems used in perception and action, supporting depictive accounts of mental representations. This depictive format may play a functional role in cognition more broadly. By preserving aspects of how absent content would be perceived or acted upon, it allows the mind to maintain, transform, and infer information that is not directly available through perception.

0.2 Sensorimotor Theories of Cognition

Sensorimotor theories of cognition build on this possibility. They propose that cognition is grounded in the systems through which we perceive the world, act within it, and simulate experience in its absence (Barsalou, 1999, 2008; Glenberg, 1997; Wilson, 2002). Such accounts span cognitive processes including spatial navigation (Guariglia & Pizzamiglio, 2007), learning and memory (Barsalou, 1999; Kosslyn, 1980), planning the future (Schacter et al., 2012; Schacter & Madore, 2016), and visual and verbal working memory (Buchsbaum & D’Esposito, 2019; Christophel et al., 2017; Scimeca et al., 2018).

In visual working memory, for instance, peripheral visual systems (Dong & Kiyonaga, 2024; Zokaei et al., 2019) and early sensory regions (Ester et al., 2009; Rademaker et al., 2019; Serences, 2016; Serences et al., 2009) are modulated by the content of maintained information. Sensory recruitment during working memory is also evident behaviorally: perceived and maintained visual stimuli can compete with one another (Kiyonaga & Egnér, 2014; Rademaker et al., 2015), and working memory content can bias perception in favor of matching stimuli in the scene (Gayet et al., 2013; Pan et al., 2014).

Sensorimotor grounding has also been proposed as a mechanism for language comprehension. On this view, understanding language involves, at least in some cases, the partial activation of perceptual and motor systems relevant to the meanings being conveyed (Pulvermüller, 2005; Pulvermüller & Fadiga, 2010; Zwaan, 2014). For example, words and sentences can activate perceptual and motor brain regions associated with their referents, and these simulations can be fine-grained enough to elicit language-induced motion aftereffects, suggesting overlap between internally generated representations and visual percepts at the level of direction-sensitive neural populations (Dils & Boroditsky, 2010). Across different cognitive domains, sensorimotor theories show that cognition is broadly shaped by the systems through which we perceive and act on the world. Yet how, and how readily, individuals recruit these systems may vary.

0.3 Individual Differences in Mental Representations

People differ greatly in how they experience and describe their thoughts. Some say they hear their thoughts in their mind's ear (Roebuck & Lupyan, 2020). Others report seeing them as images in the mind's eye. At the extremes, people with aphantasia report little or no voluntary visual imagery, while people with hyperphantasia report imagery as vivid as actual perception (Zeman et al., 2020). These differences are not restricted to introspective reports. People with aphantasia show reduced or absent imagery-based priming in binocular rivalry (Keogh & Pearson, 2018, 2024), and reduced pupillary light response to imagined bright and dark stimuli (Kay et al., 2022). These findings suggest that imagery reports track measurable behavioral and physiological differences.

This variation has driven a wave of recent research asking how cognition unfolds for people who differ in their access to internal representations (Muraki et al., 2023; Zeman, 2024; I. Phillips, 2025). If visual, verbal, and sensorimotor representations support cognition, what happens when people differ in how readily, how vividly, or how flexibly they recruit these representational formats? These questions motivated my dissertation work. My starting prediction was naive and straightforward. If a cognitive task requires visual mental representation, then stronger imagers should perform better

than weaker imagers. A task from visual working memory, language comprehension, or perception could be paired with a visual imagery questionnaire, and individual differences in imagery would help explain why people sometimes perform very differently on the same cognitive task. My first attempts at this approach did not work as expected: imagery differences did not consistently predict performance success.

0.4 The Puzzle of Absent Imagery

The broader imagery literature has shown a similar pattern. People with weak visual imagery or aphantasia often perform similarly to typical imagers on tasks that seem likely to depend on visual imagery, including visual working memory, visual attention, and behavioral measures of language-induced perceptual simulation (Jacobs et al., 2018; Keogh et al., 2021; Muraki & Pexman, 2021; Pecher, van Dantzig, Zwaan, & Zeelenberg, 2009; Pounder et al., 2022; Weber et al., 2024; Zwaan & Pecher, 2012). This pattern has been described as the “puzzle of aphantasia” (I. Phillips, 2025): how can people who report no voluntary imagery perform tasks that appear to depend on visual representation?

One explanation might be that imagery differences are real but cognitively peripheral: a phenomenological and physiological trait without broader functional consequences. However, imagery-related effects are not completely absent in cognition. Rather, the existing literature suggests that these effects are selective. First, they may depend on whether relevant content is externally available or must be generated internally. For instance, people with aphantasia show reduced physiological fear responses to frightening scenarios conveyed through language, but not to frightening pictures. This suggests that emotional processing may differ across the imagery spectrum especially when affective meaning must be constructed internally rather than directly perceived (Wicken et al., 2021).

Second, imagery differences may depend on what kind of information needs to be represented. When asked to recall a visual scene from memory, people with weak imagery reproduce fewer object details while spatial memory remains intact (Bainbridge et al., 2021). Similarly, memory traces for simple visual features, such as orientation or color, may be successfully maintained over short delays in

weak imagers (Knight et al., 2026; Weber et al., 2024), whereas richer forms of internally generated experience appear more affected: people with aphantasia report diminished ability to re-experience past events and simulate possible futures (Dawes et al., 2022). These findings suggest that imagery differences may not affect all forms of representation equally. They may be more consequential for rich, object-based, episodic, or scene-like representations than for simpler features or spatial structure.

Third, imagery differences may depend on how cognition is measured. In many tasks, overall accuracy or reaction time may look similar across imagery groups, but this does not necessarily mean that people are solving the task in the same way. In tasks where people with aphantasia perform similarly to typical imagers, they still report less use of visual imagery and greater reliance on alternative strategies, such as semantic or spatial strategies (Keogh et al., 2021; Reeder, Pounder, et al., 2024). Thus, preserved performance may sometimes reflect compensation or alternative representational routes rather than the absence of imagery-related cognitive differences.

0.5 Approaching the Puzzle

In this dissertation, I pursue two empirical questions that tackle the complex ways in which imagery diversity manifests in cognitive diversity, and then develop a conceptual framework that situates this work within broader challenges in the field. In the first question, I ask whether people with different imagery profiles may be recruiting different strategies to achieve comparable performance. In the second, I ask whether we can glimpse the internal representations themselves to characterize qualitative differences in what people across the imagery spectrum “see.” Building on these empirical findings, I propose a set of bottlenecks that may partly explain persistent inconsistencies in imagery research, and suggest ways forward.

In Chapter 1, I take up the first question in the context of visual working memory, when participants are asked to maintain visual information in mind for short delays. I ask whether subjective tendencies to think in words versus pictures can produce unique memory signatures and change how mental content interacts with ongoing sensory input. In a large behavioral experiment, I tracked

accuracy, reaction times, and the types of errors people made in working memory. These errors provided an indirect window into how information may have been represented. For example, recalling a different giraffe instead of the target giraffe may suggest that the participant preserved the semantic category or label of the item, while failing to maintain the fine-grained visual details needed to identify the exact exemplar. Although overall performance was comparable across individuals with different imagery tendencies, the way they erred was not. The error patterns were consistent with the interpretation that stronger verbal thinking tendencies support more semantic memory traces, whereas stronger visual imagery supports more fine-grained visual traces. Further, when perception competed with working memory, sensory input had diverging effects depending on individual imagery profile: verbal thinkers benefited from similar visual input during the delay, whereas visual thinkers were impaired by it. Together, these findings suggest that similar working memory performance can be supported by different representational strategies, with consequences for how memory withstands interference from the perceptual world.

In Chapter 2, I extend the first question to language comprehension, when individuals are presumably simulating meaning conveyed through language. I ask whether an individual's visual imagery vividness shapes how readers process the perceptual content of language. In a preregistered behavioral study and an EEG follow-up using sentence-picture verification, participants read sentences and judged whether subsequent pictures matched their content. People typically respond faster when the picture's shape matches the shape implied by the sentence (for example, an eagle with outstretched wings after "The eagle flew across the sky"). This is called a match effect. I found that behaviorally, imagery vividness did not reliably predict the size of the match effect. However, imagery vividness modulated neural responses to shape mismatch, specifically in an early posterior time window associated with visual processing. Together, these findings provide a neural complement to the behavioral results in Chapter 1: people may be arriving at similar overt performance through different underlying mechanisms. For language comprehension theories, this means that the extent to which people rely on embodied representations may depend on their imagery profile.

In Chapter 3, I take up the second question. I ask whether the content of internally generated

experience can reveal qualitative differences in what people across the imagery spectrum "see." To do this, I turned to Ganzflicker, a paradigm in which sustained exposure to a rhythmically flickering visual field produces hallucination-like visual experiences. I drew on free-text descriptions of these experiences from over four thousand participants, treating the content and structure of these descriptions as a signature of the internal representations being generated. Using topic modeling and natural language processing, I characterized what people report experiencing during these hallucinations. Stronger imagers described more complex, naturalistic content, including faces, objects, and scenes, whereas weaker imagers described simpler patterns, flashes, and geometric forms. This chapter extends the study of imagery phenomenology beyond the question of how vivid internal experience feels to ask what that experience contains and how its content differs across the imagery spectrum.

In Chapter 4, I step back from the empirical work and develop a theoretical framework. I argue that recurring inconsistencies in imagery research reflect conceptual and methodological bottlenecks in how the field has defined, measured, and inferred the role of imagery in cognition. The chapter reviews these bottlenecks and proposes concrete responses, including multidimensional measures of imagery and paradigms that probe strategy and content rather than only overt behavioral success. This chapter brings together the empirical lessons of this dissertation and recent developments in the broader imagery literature into a framework for studying imagery diversity and its role in cognition.

This dissertation also makes a broader claim: imagery diversity is not merely a curiosity about private experience, but a testbed for cognitive theories. The field has moved from asking whether thoughts are depictive or descriptive to asking how cognition unfolds when people differ in their access to, reliance on, or experience of these representational formats. This shift makes new theoretical questions visible. When the same task can be solved through different cognitive routes, theories must specify how different representational formats are recruited, substituted, or combined to support cognition. Diversity in how people think is, in this sense, both a topic in its own right and a tool for understanding cognition.

Chapter 1

Visual and Verbal Thought Tendencies Shape Working Memory Under Distraction

1.1 Introduction

Working memory (WM) allows us to keep information accessible over short delays while we think, plan, and act. This information might be maintained in different formats, like a visual depiction or verbal label (Baddeley, 1983). For instance, when rehearsing where you parked your car, you might picture the exact spot and surrounding landmarks, or keep a mental note like “level 3, near the elevator.” Both representational strategies can be effective, and it is unclear whether there are behavioral advantages to either (Pereira Seabra et al., 2025; Souza et al., 2021). Some individuals may default to one tendency or another (Pearson & Keogh, 2019), but the format of WM representations may also depend on the information being remembered or the situational context. Here, we examine how features of the memory content, individual, and task may jointly influence how – and how well – WM information is maintained.

Some things are easily held in mind, while others quickly decay. Objects that are more realistic (Brady et al., 2016; Brady & Störmer, 2024), familiar (Xie & Zhang, 2017, 2018), and nameable (Overkott & Souza, 2022) show better WM performance. This suggests that WM benefits from maintaining representations associated with existing knowledge or long-term traces. Long-term memory (LTM) research has also sought to capture a holistic ‘memorability’ property that reliably predicts performance (Bainbridge, 2019). Memorability can be estimated computationally from perceptual and semantic

features of an image (Khosla et al., 2015; Needell & Bainbridge, 2022), and is thought to reflect statistical distinctiveness of an item (Bainbridge et al., 2017). If WM benefits from stimulus familiarity, which may index stronger LTM traces, it may also benefit from properties that encourage LTM encoding. If so, we would expect ‘memorable’ stimuli to evoke more distinctive WM representations that are more immune to competition and decay. Such stimuli may imprint a stable trace at encoding, which survives in the background without effortful active maintenance, and therefore competes less with other demands for the focus of attention (Lewis-Peacock et al., 2012; Stokes, 2015).

Yet even for the same stimulus, not all minds remember alike. For instance, individuals report wide variability in the tendency to think in images versus words (Roebuck & Lupyan, 2020). Those with stronger visual imagery report relying more on visual representations during WM maintenance, while verbal thinkers report greater reliance on linguistic codes (Keogh et al., 2021; Reeder, Pounder, et al., 2024). However, it is unclear whether such subjective differences in thought format evince objective differences in behavior (Dong & Kiyonaga, 2024; Jacobs et al., 2018; Knight et al., 2026; Nedergaard & Lupyan, 2024; Pounder et al., 2022; Weber et al., 2024). Successful WM may be achievable through several cognitive avenues, so different underlying strategies might only reveal themselves in select performance measures. For instance, verbal thinkers may retain semantic WM category labels, but lose the visual detail needed to distinguish specific exemplars. This strategy would be efficient in many cases, but could cause confusions when WM content competes with other similar content in mind or in the environment.

That is, WM rarely unfolds in isolation. We often keep multiple goals in mind amid a stream of sensory input that may compete for attention or representational real estate (Adam et al., 2025). Visual vs. verbal formats may be more or less effective for simultaneously holding multiple goals, and may render the WM content differently susceptible to interference from competing sensory input. For instance, it may be infeasible to maintain several detailed visual codes in the focus of attention, whereas several verbal labels can be efficiently juggled (Lewis-Peacock et al., 2012; Souza & Oberauer, 2016; Souza & Skóra, 2017; van Moorselaar et al., 2014). Representational format may likewise determine how WM content interacts with sensory input (Zhang & Lewis-Peacock, 2023a, 2023b). Distraction from visual

input is commonly assumed to scale with feature similarity between memoranda and percepts—the more alike, the greater the interference—but similar sensory input can sometimes have little effect on WM, or even facilitate performance (Bettencourt & Xu, 2016; Lorenc et al., 2021; Rademaker et al., 2015, 2019; Xu, 2017). This discrepancy could arise if competing input disrupts fine-grained visual codes but leaves categorical codes relatively unaffected or even invigorated (Lorenc et al., 2018, 2021).

Here, we examine how stimulus properties and individual imagery differences shape WM performance under perceptual competition and varied attentional demands. We ask, for one, does model-derived stimulus memorability predict WM success and robustness to competition? Second, do individual differences in visual imagery and verbal thinking relate to different WM representational strategies? Third, do multiple tasks and individual factors interact to determine WM outcomes? Participants completed a delayed recognition task for real-world images, where task-irrelevant visual inputs appeared during WM maintenance. They then completed a battery of self-report imagery questionnaires. Considering the joint influence of stimulus, individual, and task properties may illuminate the conditions that determine what stays in mind and what slips away.

1.2 Research Transparency Statement

1.2.1 General Disclosures

Conflicts of interest: All authors declare no conflicts of interest. **Artificial intelligence:** AI-assisted tools, including ChatGPT and Claude, were used to revise limited portions of the manuscript for grammar and clarity. Claude was also used for coding assistance. All scientific content, analyses, interpretations, and final text were reviewed and approved by the authors. **Ethics:** The study was approved by the Office of IRB Administration at the authors’ institution.

1.2.2 Study Disclosures

Preregistration: The study sample size was preregistered, following simulation power analyses for mixed-effects models with a similar structure to those tested here. The WM task performance-

based exclusion criteria were also preregistered. The preregistration is available at <https://osf.io/32c6x>. However, the specific analyses reported here, involving stimulus memorability and self-report questionnaire data, were not preregistered. **Materials, data, and analysis scripts:** All study materials, data, and analysis scripts are publicly available at the OSF repository: https://osf.io/xqfke/overview?view_only=a742c594d69546c8887288d0b7bc6322.

1.3 Methods

We conducted a complementary analysis of a dataset that has been described in detail elsewhere (Ballard et al., 2026). The study was designed to examine the conditions under which visual WM for naturalistic images is susceptible to distraction from sensory input (Figure 1.1a). The task manipulated the behavioral relevance among multiple WM samples (attentional priority), as well as the image-level correlation between the WM sample and a task-irrelevant image that appeared during the delay (target-distractor visual similarity). After completing the WM task, participants completed a series of questionnaires that assessed multiple dimensions of both visual imagery and verbal thinking tendencies. This self-report data, which has not been previously analyzed, allow us to examine how subjectively-experienced thought formats relate to WM function. We first look at how model-derived stimulus memorability and individual differences in thought formats separately relate to multiple WM performance indices. We then test whether task context factors—attentional priority and task-irrelevant visual inputs—interact with these stimulus and individual characteristics to shape performance. Finally, we examine how all three factors jointly determine WM outcomes.

1.3.1 Participants

A total of 330 participants completed the original study online. Two pilot experiments were first collected to identify expected effect sizes for a priori power analyses. Simulation power analyses were conducted to determine the necessary sample size to detect interactions between attentional priority and target-distractor similarity. Those analyses yielded a conservative target sample size of $n = 136$. Two participant cohorts were then collected ($n = 157$ and $n = 173$) to meet that sample size in

each. Both cohorts completed the same task, but each with a different set of naturalistic images used as sample stimuli. Here, we combine the data across the two cohorts, because all critical task procedures were matched across them.

Participants were recruited through the SONA system and received course credit for participation. From the initial sample of 330, 16 participants were excluded for not completing the imagery questionnaires, and an additional 35 participants were excluded for failing at least two of the three attention checks that were embedded within the questionnaires. To those remaining 279 participants, we then applied the preregistered WM task performance-based exclusion criteria. If participants did not respond within the 3-second response window, the trial was treated as a timeout (accuracy coded as incorrect). Participants were excluded if their overall accuracy was below 40% (chance = 16.7%) or if they failed to respond on more than 30% of experimental trials. This resulted in the removal of an additional 33 participants, leaving us with a final sample of 246 participants for analysis (188 female, 40 male, 4 non-binary, 14 preferred not to say; mean age = 20.7 ± 2.2).

All participants provided informed consent in accordance with local ethics approval. Because the study was conducted online, there is some possibility of automated or AI-generated responses (Van der Stigchel et al., 2026). However, participants were recruited through the SONA system, requiring authenticated student accounts and access codes to enter the task, reducing the risk of paid bot infiltration. Embedded attention checks and performance-based exclusions targeted chance-level and internally inconsistent responses, behavioral signatures that current bots may not be able to easily mimic (Chetverikov, 2026).

1.3.2 Materials and Measures

The experimental stimuli consisted of high-resolution color photographs of real-world objects drawn from the THINGS database (Hebart et al., 2023). All images depicted common, nameable objects (e.g., household items, tools, animals), with the goal of assessing WM in the context of naturalistic stimuli. A total of 1,800 images were used across the two participant cohorts, with a set of 900 images for each cohort. In each cohort, 600 images served as WM samples and 300 as task-irrelevant sensory

inputs (i.e., ‘distractors’).

Image memorability was quantified using the ResMem computational model, a deep residual neural network that combines both perceptual and semantic features to predict human memory performance for visual images (Figure 1.1b) (Needell & Bainbridge, 2022). The model was trained on large-scale image memorability datasets (Goetschalckx & Wagemans, 2019; Khosla et al., 2015). The model outputs memorability scores ranging from 0 to 1, where higher scores indicate images that are more likely to be remembered consistently across individuals (Figure 1.1b, Figure 1.2a). For our implementation, we used the publicly available ResMem model (Needell & Bainbridge, 2022) to extract memorability scores for all images in the current stimulus set. These scores provide an objective assessment of stimulus memorability, independent of individual differences or experimental context. These scores were calculated post hoc, and were not incorporated into the task manipulation. This ensures that any observed effects reflect unbiased model predictions rather than stimulus selection bias.

Visual similarity between distractors and WM samples was quantified using feature representations from the VGG-16 convolutional neural network (Simonyan & Zisserman, 2015). VGG-16 is a 16-layer deep neural network architecture originally developed for image classification, but it has since been widely used in cognitive neuroscience to model human responses to visual objects and features (Bayet et al., 2020; De Cesare et al., 2021; German & Jacobs, 2020; Jacob et al., 2021; Shoham et al., 2024; Zohar et al., 2024). We focused on features from the third max-pooling layer (Pool3), which captures visual features such as edge orientations, spatial frequencies, and basic shape configurations while remaining relatively invariant to higher-level semantic content (Figure 1.1c). This layer has been argued to compute features roughly analogous to those represented in early visual areas V2-V4 and has been shown to effectively model human behavioral choices based on visual similarity (Piriyajitakonkij et al., 2025).

For our purposes, visual similarity between any two images was computed as the Pearson product-moment correlation coefficient between their respective VGG-16 Pool3 feature vectors, where higher scores indicate stronger similarity (Figure 1.1c, Figure 1.4d). See Piriyajitakonkij et al. (2025) for more VGG-16 feature extraction and similarity computation details. Each task trial incorporated a

unique stimulus triplet, comprising two WM samples and one distractor image. These stimulus triplets were grouped a priori by their VGG-16 correlations, to systematically control similarity between WM samples and distractors (Ballard et al., 2026).

Individual differences in visual imagery and verbal thinking were assessed using three established questionnaires: The Internal Representations Questionnaire (IRQ; Roebuck & Lupyan, 2020), The Object-Spatial Imagery and Verbal Questionnaire (OSIVQ; Blazhenkova & Kozhevnikov, 2009), and The Vividness of Visual Imagery Questionnaire (VVIQ; Marks, 1973) (Figure 1.1d, Figure 1.1e). Throughout the manuscript, we use ‘imagery profiles’ as a shorthand for individual differences in visual imagery and verbal thought.

The IRQ (Roebuck & Lupyan, 2020) measures individual propensity to spontaneously rely on visual versus verbal thought formats in daily cognitive activities. The questionnaire contains separate 10-item visual and 12-item verbal subscales. The visual subscale assesses the tendency to use visual imagery (e.g., "When I recall a conversation, I can visualize the person with whom I was speaking," "I prefer to read a map rather than have someone tell me directions"). The verbal subscale assesses propensity for verbal/linguistic processing (e.g., "I think in words more than in images," "When I think about a person, I usually think about his or her name before I think about how the person looks"). Participants rate each item on a 5-point Likert scale from 1 (strongly disagree) to 5 (strongly agree). In the results, we refer to IRQ scores as “propensity” scores.

The OSIVQ (Blazhenkova & Kozhevnikov, 2009) assesses the quality, preference, and ability for visual and verbal processing across three subscales (15 items each): object (e.g., "My mental images of different objects very much resemble the size, shape and colour of actual objects," "I can easily remember a great deal of visual details that someone else might never notice"), spatial ("I can easily imagine and mentally rotate three-dimensional geometric figures"), and verbal (e.g., "I think more in words than in images," "I enjoy doing work that requires the use of words"). Participants rate items on a 5-point scale from 1 (totally disagree) to 5 (totally agree). Here we focus on the object and verbal subscales, as these dimensions most directly map onto the representational formats of interest here. Object and spatial imagery are considered distinct (I. Phillips, 2025), and the task here emphasizes object features rather

than spatial relationships. For cohesion across analyses, we use “visual” to refer to the object subscale throughout the manuscript. In the results, we refer to OSIVQ scores as “quality” scores.

The VVIQ (Marks, 1973) assesses the subjective vividness of visual mental images across 16 scenarios. Participants are asked to imagine specific scenes (e.g., “The sun is rising above the horizon into a hazy sky” or “A relative or friend whom you frequently see, considering the exact contour of face, head, shoulders and body”). They then rate the vividness of their mental image on a 5-point scale from 1 (“No image at all, you only ‘know’ that you are thinking of the object”) to 5 (“Perfectly clear and as vivid as normal vision”). In the results, we refer to VVIQ scores as “vividness” scores.

1.3.3 Experimental Task Procedure

Participants performed a delayed recognition WM task, where task-irrelevant visual input appeared during the memory maintenance period (Figure 1.1a). Each trial began with a 1000ms central fixation cross, followed by the simultaneous presentation of two trial-unique sample images for 1000ms. The images appeared on opposite sides of central fixation, and participants were instructed to encode and remember both images for a subsequent forced choice test.

After a 500ms delay, a central cue appeared for 1000ms indicating which of the two images was more likely to be tested. This retro-cue was a partially filled horizontal bar, where the filled portion of the bar pointed to the cued item. Cues varied trial-by-trial as to their validity and reliability. On most trials the cued item was tested (valid), but on a subset of trials the uncued item was tested (invalid). The bars could be filled either 5/6 or 2/3 of the way, indicating different reliability levels, and participants were explicitly instructed about the meaning of different cue types. The retro-cue manipulation was meant to encourage participants to attentionally prioritize the item that was most likely to be tested. Thus, the WM item that was tested at the end of the trial could have been relatively prioritized or deprioritized during the delay, and we refer to this factor as “attentional priority.”

After distractor offset, there was an additional 750ms blank delay, followed by the recognition test display for 3sec. The display contained six images arranged in a circle, including the two originally presented WM samples plus four novel lure images. The lures were 2 other unique exemplars from

each of the WM sample categories (e.g., if a WM sample was a red cherry, a lure might be a yellow cherry). A central arrow pointing left or right indicated which spatial location from the encoding phase participants should retrieve (corresponding to either the originally cued or uncued location). Participants were instructed to click on the image that had appeared at that location during the encoding phase. We refer to the WM sample that was tested as the WM target. We refer to the visual similarity between the target and distractor as “target-distractor similarity.” The trial timed out if participants did not choose an image within 3 sec of the test display onset.

Sample and distractor images did not repeat across trials, and all attentional priority and target-distractor similarity conditions occurred in randomized order. Before the main task, participants completed a multi-stage practice with feedback. The initial phase consisted of a basic cued recognition task at a slowed pace (5 trials, 80% accuracy criterion), followed by retrocue introduction (5 trials), and finally full task practice including distractors and at the experimental timing (12 trials, 60% accuracy criterion). Failed phases were repeated until criteria were met. The main experiment consisted of 300 total trials distributed across 10 blocks.

To summarize, the experiment included factors of attentional priority (whether the tested WM item had been prioritized or deprioritized), cue reliability (how likely the prioritized item was to be tested), and target-distractor similarity (how correlated the distractor image was with the WM samples). However, the cue reliability manipulation is not germane to our current questions, so analyses collapse across that condition to maximize the number of trials in each condition of interest.

Following experimental task completion, participants completed the imagery questionnaires in randomized order. Three attention check questions were embedded throughout the questionnaires (e.g., “Most people have five legs” (rate from 1 (strongly disagree) to 5 (strongly agree))), and participants who failed two or more checks were excluded from analysis.

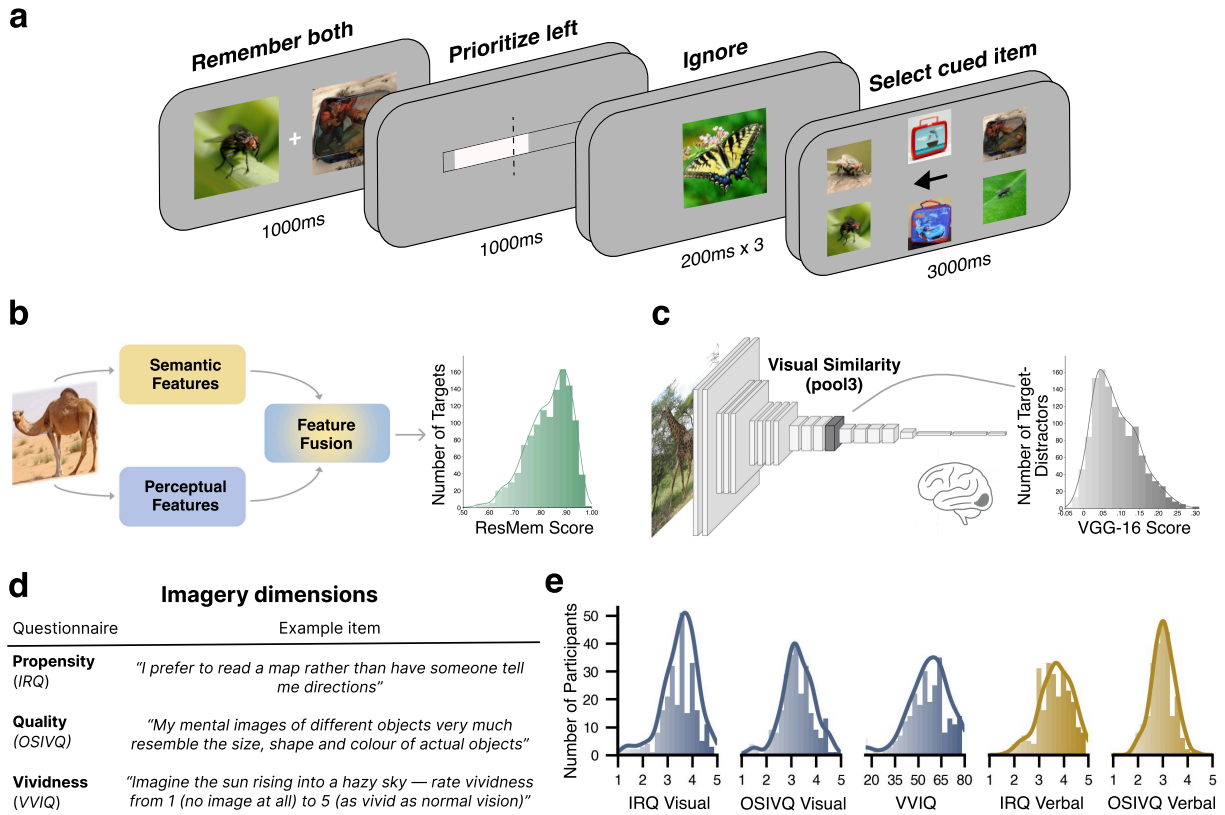


Figure 1.1. Task schematic and study variables. (a) Delayed recognition WM task trial sequence. Participants encoded two images, received a priority cue indicating which item was more likely to be tested, saw a task-irrelevant distractor during the delay period, and chose the tested WM item from a six-alternative array. (b) ResMem model architecture combines semantic features and perceptual features to predict image memorability. Histogram shows the distribution of ResMem scores for the WM target stimuli used here. (c) Features from the third max pooling layer of VGG-16 were used to quantify similarity between WM targets and distractors. Histogram shows the distribution of these similarity scores across target-distractor pairings. (d) Individual differences in visual and verbal thinking were assessed using three questionnaires: IRQ, OSIVQ, and VVIQ. The table shows example items from the three visual questionnaires. (e) Histograms show participant score distributions for each questionnaire subscale used here, with density curves overlaid.

1.3.4 Statistical Analysis

All analyses were conducted in *R* (Team, 2023) using mixed-effects modeling to account for the nested structure of trials within participants. We conducted separate analyses for three dependent variables: response time (RT; analyzed only for correct trials), choice accuracy (analyzed for all trials), and error types (analyzed only for incorrect trials). RTs were log-transformed prior to analysis.

For RT analysis, we used linear mixed-effects models implemented with the *lmer* function from the *lme4* package (Bates et al., 2015). For accuracy and error type analysis, we used generalized linear mixed-effects models (*glmer* function) with binomial link functions, modeling accuracy as correct/incorrect responses and error types as within-category/cross-category errors. All models used the "bobyqa" optimizer (Powell, 2009; Bates et al., 2015).

All continuous predictors were standardized (z-scored) prior to model fitting to facilitate interpretation of interaction effects and improve model convergence. The predictors of primary interest included stimulus memorability (ResMem scores), target-distractor similarity (VGG-16 correlations), attentional priority (prioritized vs. deprioritized), and individual differences in imagery profiles as measured by the three questionnaires. We fitted separate models for each imagery questionnaire (propensity (IRQ), quality (OSIVQ), and vividness (VVIQ)) to examine how different dimensions of imagery relate to WM performance.

Our primary theoretical interest was in the interactions between stimulus properties, individual differences, and task-context factors. Because interaction effects require all component terms to be estimated simultaneously, we fitted comprehensive models that simultaneously included all main effects, theoretically interesting two-way interactions, and the critical three-way interaction between all three factors. For each questionnaire, we specified models with the following structure. We use the term *imagery* here as shorthand for all visual and verbal thought tendency factors:

$$\begin{aligned}
Y \sim & \text{ResMem Scores} + \text{Target-Distractor Similarity} + \text{Attentional Priority} + \text{Imagery} \\
& + \text{Imagery} \times \text{ResMem Scores} + \text{Imagery} \times \text{Target-Distractor Similarity} \\
& + \text{Imagery} \times \text{Attentional Priority} + \text{ResMem Scores} \times \text{Target-Distractor Similarity} \\
& + \text{ResMem Scores} \times \text{Attentional Priority} + \text{Target-Distractor Similarity} \times \text{Attentional Priority} \\
& + \text{Imagery} \times \text{ResMem Scores} \times \text{Target-Distractor Similarity} \\
& + (1 \mid \text{Participant}) + \varepsilon
\end{aligned}$$

Where Y represents the dependent variable (log RT, accuracy, or error type), *Imagery* represents the relevant imagery profile measure(s) for each questionnaire (visual and verbal subscales for propensity (IRQ) and quality (OSIVQ) models; visual score for vividness (VVIQ) model), and $(1 \mid \text{Participant})$ represents random participant intercepts.

1.4 Results

1.4.1 Model-Derived Memorability Predicts WM Performance

We first examined whether ResMem model-derived stimulus memorability would relate to WM performance. This model has been tested in the LTM context (Cao et al., 2025; Roberts et al., 2023), but it is unclear whether the features that drive durable long-term encoding also benefit short-term maintenance.

We found that memorability showed strong main effects for all three WM performance measures (Figure 1.2). Items with higher ResMem scores elicited faster RTs (Figure 1.2b; $\beta = -0.02$, 95% CI $[-0.03, -0.02]$, $SE = 0.001$, $t = -18.58$, $p < .001$) and more accurate choices (Figure 1.2c; $OR = 1.17$, 95% CI $[1.15, 1.19]$, $SE = 0.01$, $z = 17.20$, $p < .001$). Therefore, this independently derived memorability property supports memory on the order of seconds.

To probe the underlying processes that might drive these memorability effects, we also examined

patterns of WM errors. In this task, incorrect choices could come from the same semantic category as the target sample (within-category errors: e.g., choosing a different giraffe when the correct response was another giraffe) or from a different category (cross-category errors: e.g., choosing a car when the target was a giraffe; Figure 1.3a). Within-category errors suggest memory preserved the correct semantic label but lacked the visual detail to distinguish exemplars. Cross-category errors suggest complete forgetting or confusion with the untested WM sample.

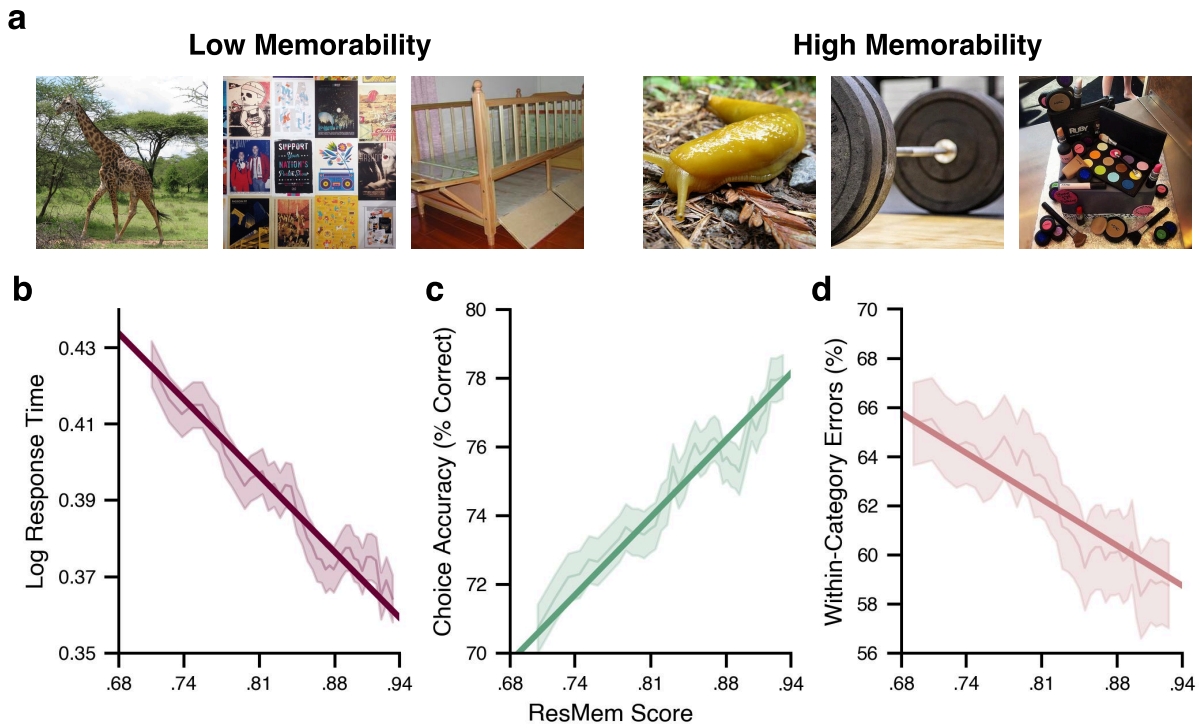


Figure 1.2. Model-derived memorability scores predict WM performance. (a) Example experimental stimuli, depicting low memorability (left) and high memorability (right) items as determined by the ResMem model. (b) Log-transformed response times (RT) as a function of ResMem score. RT reflects the time from test display onset to mouse click; the Log RTs correspond to a raw RT change from 1625ms at lower ResMem scores to 1500ms at higher ResMem scores. (c) WM choice accuracy (% correct) as a function of ResMem score. (d) Within-category error rates (%) as a function of ResMem score. Within-category errors reflect incorrect choices to lures from the same semantic category as the target. For all plots, jagged lines represent the rolling average of the performance measures computed using sliding windows spanning 20% of observations and moved in 2% steps along the x-axis. Solid lines represent linear regression fits, and shaded bands represent 95% confidence intervals. All effects $p < .001$.

We found that lower memorability items elicited relatively more within-category errors than higher memorability items, which elicited more cross-category errors ($OR = 0.92$, 95% CI [0.89, 0.96], $SE = 0.02$, $z = -4.31$, $p < .001$; Figure 1.2d). In other words, less memorable items produced more errors overall, but those errors tended to preserve semantic content (i.e., be from the same category as the target). This is consistent with the possibility that forgettable items leave fragile WM traces that need to be actively maintained and are susceptible to interference from same-category lures. More memorable items, by contrast, were remembered better, but when errors occurred, they were more likely to miss the mark completely (vs. forgettable items). This suggests that these distinctive traces were less likely to be confused with same-category lures, but may sometimes be forgotten through attentional lapses at encoding or swapped with other content in mind.

Model-derived memorability is therefore a strong predictor of WM performance, shaping not only how well items are retained, but also specific ways that memory can fail.

1.4.2 Individual Imagery Tendencies Predict WM Error Patterns

The results above are consistent with the idea that within-category errors stem from representations with coarser or more fragile visual information (vs. rich perceptual detail). If that is the case, then individuals with stronger verbal thinking tendencies — who may rely more on semantic labeling strategies — should show more of this type of error.

Indeed, within-category errors increased with verbal propensity (IRQ) (Figure 1.3b, right; $OR = 1.14$, 95% CI [1.02, 1.27], $SE = 0.06$, $z = 2.32$, $p = .02$) and decreased with visual propensity (IRQ) (Figure 1.3b, left; $OR = 0.89$, 95% CI [0.79, 0.99], $SE = 0.05$, $z = -2.15$, $p = .03$). When incorrect, strong verbal thinkers were likely to choose items from the same category, which would occur if they used semantic coding strategies that sacrifice visual detail. Strong visual imagers showed the opposite pattern, with a greater share of their errors falling across categories compared to those with weak imagery. This is consistent with detailed visual representations that either succeed in preserving item-specific information (when correct) or fail due to attentional lapses, confusion, or unsuccessful encoding (when incorrect). These effects were specific to imagery propensity

(IRQ), while neither quality (OSIVQ) nor vividness (VVIQ) showed reliable effects on error patterns (Tables A.6 and A.9). And while error patterns varied with imagery profiles, overall performance was roughly comparable across the imagery spectrum (see Tables A.1–A.2, A.4–A.5, and A.7–A.8 for RT and accuracy statistics).

These results highlight that individual variability in visual and verbal thinking may manifest in the maintenance strategies people employ during WM, but that there can be multiple routes to successful performance.

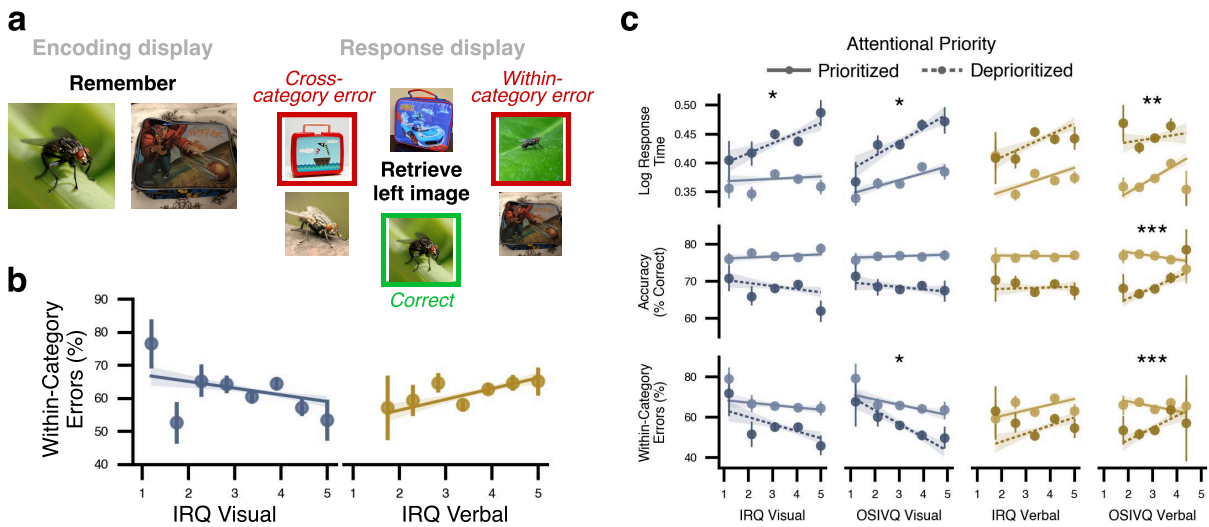


Figure 1.3. Individual differences in imagery predict distinct WM performance patterns. (a) Example trial showing different error types: Participants could make within-category errors (choosing wrong exemplar from correct category, e.g., different fly when target was a specific fly) or cross-category errors (choosing item from wrong category entirely, e.g., choosing a lunchbox when target was fly). (b) Within-category error rates (%) as a function of IRQ visual (left) and verbal (right) imagery scores. (c) WM performance measures as a function of imagery scores, separated by attentional priority of the tested WM item (solid lines = prioritized trials, dashed lines = deprioritized trials). Top row shows log-transformed response times, middle row shows accuracy (% correct), bottom row shows within-category error rates (%). Blue panels show visual imagery measures, and yellow panels show verbal thinking measures. (b–c) For all plots, imagery scores were binned into five equal-width bins for visualization only; points represent mean values within each bin, error bars represent 95% confidence intervals, and lines represent linear regression fits with shaded 95% confidence bands. Asterisks indicate significant interactions between imagery scores and attentional priority (* $p < .05$, ** $p < .01$, *** $p < .001$).

1.4.3 Imagery Profiles Matter Most for WM Content in the Back of the Mind

These individual imagery tendencies may also shape how attention is distributed over multiple goals in mind. For instance, vivid visual representations may occupy the focus of attention, limiting how many items can be maintained in detail. Categorical codes, though less precise, may be more sustainable across multiple items. If so, visual thinkers should show greater costs when deprioritized content is tested, while verbal thinkers should show more similar performance between prioritized and deprioritized content, and therefore smaller effects of attentional priority.

Visual imagery and attentional priority. All visual imagery measures interacted with attentional priority for RT. As imagery scores increased, so did the magnitude of the attentional priority effect. This was the case for imagery propensity (IRQ; $\beta = -0.003$, 95% CI $[-0.006, -0.001]$, $SE = 0.002$, $t = -2.28$, $p = .02$), quality (OSIVQ; $\beta = -0.003$, 95% CI $[-0.006, -0.001]$, $SE = 0.001$, $t = -2.35$, $p = .02$), and vividness (VVIQ; $\beta = -0.003$, 95% CI $[-0.006, -0.001]$, $SE = 0.001$, $t = -2.12$, $p = .03$; Figure 1.3c, top; Figure A.1, left). Specifically, responses for deprioritized WM items got progressively slower with increasing imagery strength, suggesting that strong visual imagers may struggle with WM content held in the back of the mind (vs. weaker imagers). Visual thinkers may rely on fine-grained visual traces that are harder to distribute across multiple items in WM. Accuracy and error patterns showed similar but less consistent trends across questionnaires (see Tables A.2, A.5, A.6, A.8, and A.9; Figure 1.3c; Figure A.1).

Verbal thinking and attentional priority. Verbal thinking quality (OSIVQ) interacted with attentional priority across all performance metrics (Figure 1.3c), but showed the opposite pattern from visual imagery. As verbal thinking scores increased, the magnitude of the attentional priority effect diminished. RTs ($\beta = 0.004$, 95% CI $[0.001, 0.006]$, $SE = 0.001$, $t = 2.83$, $p = .005$), accuracy ($OR = 0.95$, 95% CI $[0.94, 0.97]$, $SE = 0.009$, $z = -5.21$, $p < .001$), and error types ($OR = 0.93$, 95% CI $[0.90, 0.96]$, $SE = 0.02$, $z = -4.19$, $p < .001$) showed virtually no attentional priority effect at the highest levels of verbal thinking. Verbal thinkers may rely on categorical representations

that can be spread across multiple items more readily than detailed visual traces. However, verbal propensity (IRQ) did not show this effect (Tables A.1–A.3).

In sum, individuals who default to visual imagery show a greater cost to WM content that is maintained in the background. Those who default to verbal strategies instead show less of a cost to deprioritized WM content, but also less of a benefit to prioritized WM content.

1.4.4 The Effects of Visual Input During WM Depend on Stimulus Memorability and Imagery Profile

A prior report on this dataset showed that WM is impacted by the visual similarity of concurrent visual inputs (Ballard et al., 2026). Here, we also found a main effect of target-distractor similarity on RT ($\beta = -0.003$, 95% CI $[-0.006, -0.001]$, $SE = 0.001$, $t = -2.26$, $p = .02$; Table A.1), where greater visual similarity led to faster responses. Having shown that stimulus memorability and imagery tendencies may shape the strength and format of WM representations, we next asked whether these differences alter how WM interacts with concurrent visual input.

Stimulus memorability matters most when visual input is similar to WM content. If memorability shapes the stability of WM traces and the degree to which they require active maintenance, visually similar input (*distractors*) may have different consequences for high- and low-memorable items.

For all metrics, we found an interaction whereby the influence of memorability increased with increasing target-distractor similarity. For RT, as similarity increased, performance for highly memorable items improved, while performance for weakly memorable items worsened (Figure 1.4a; $\beta = -0.004$, 95% CI $[-0.007, -0.002]$, $SE = 0.001$, $t = -3.59$, $p < .001$). A similar pattern emerged for accuracy (Figure 1.4b; $OR = 1.05$, 95% CI $[1.03, 1.07]$, $SE = 0.009$, $z = 5.51$, $p < .001$). This finding is consistent with similar visual input reinvigorating stable traces of memorable items, but disrupting fragile traces of forgettable items.

Error patterns for low- and high-memorable items also diverged with increasing similarity. Specifically, within-category errors increased, and cross-category errors decreased, for low-memorable items as distractors became more similar to the memory target (Figure 1.4c; $OR = 0.94$, 95% CI

[0.92, 0.98], $SE = 0.02$, $z = -3.24$, $p = .001$). This could occur if similar visual input biases or permeates weakly memorable items, causing more confusions with visually similar lures. Intriguingly, across all metrics, memorability effects largely vanished when distractors were most dissimilar from WM content, suggesting that the properties underlying memorability might matter most when WM faces competing visual input.

Visual input helps or hinders WM depending on imagery profile. If visual imagers maintain detailed visual representations while verbal thinkers rely on linguistic codes, similar visual input may interfere with the former while leaving the latter unaffected or augmented.

We found that visual imagery propensity (IRQ) interacted with target-distractor similarity for RT ($\beta = 0.004$, 95% CI [0.001, 0.007], $SE = 0.002$, $t = 2.53$, $p = .01$). For relatively stronger imagers, responses became slower with increasing target-distractor similarity, whereas relatively weaker imagers showed the reverse (Figure 1.4e, left). This suggests that similar visual input may interfere when representations are highly visual (i.e., in strong imagers), but may boost coarse or muted representations (i.e., in weak imagers). Verbal thinking propensity (IRQ) showed an opposing pattern ($\beta = -0.004$, 95% CI [-0.007, -0.001], $SE = 0.001$, $t = -2.53$, $p = .01$). Stronger verbal thinkers showed increasingly fast responses with increasing target-distractor similarity, whereas weaker verbal thinkers remained relatively flat (Figure 1.4e, right). This suggests that similar visual input might complement existing semantic representations to improve memory. These IRQ interactions were specific to RT ($ps > .15$ for accuracy and error type; Tables A.2–A.3), and neither imagery quality (OSIVQ) nor vividness (VVIQ) showed interactions with target-distractor similarity (all $ps > .09$; Tables A.4–A.9).

These opposing patterns between visual and verbal thinking suggest that thought formats shape how WM content interacts with sensory input. Visual imagers show vulnerability to competing visual information, while verbal thinkers may rely less on detailed visual representations and benefit from the relevant input.

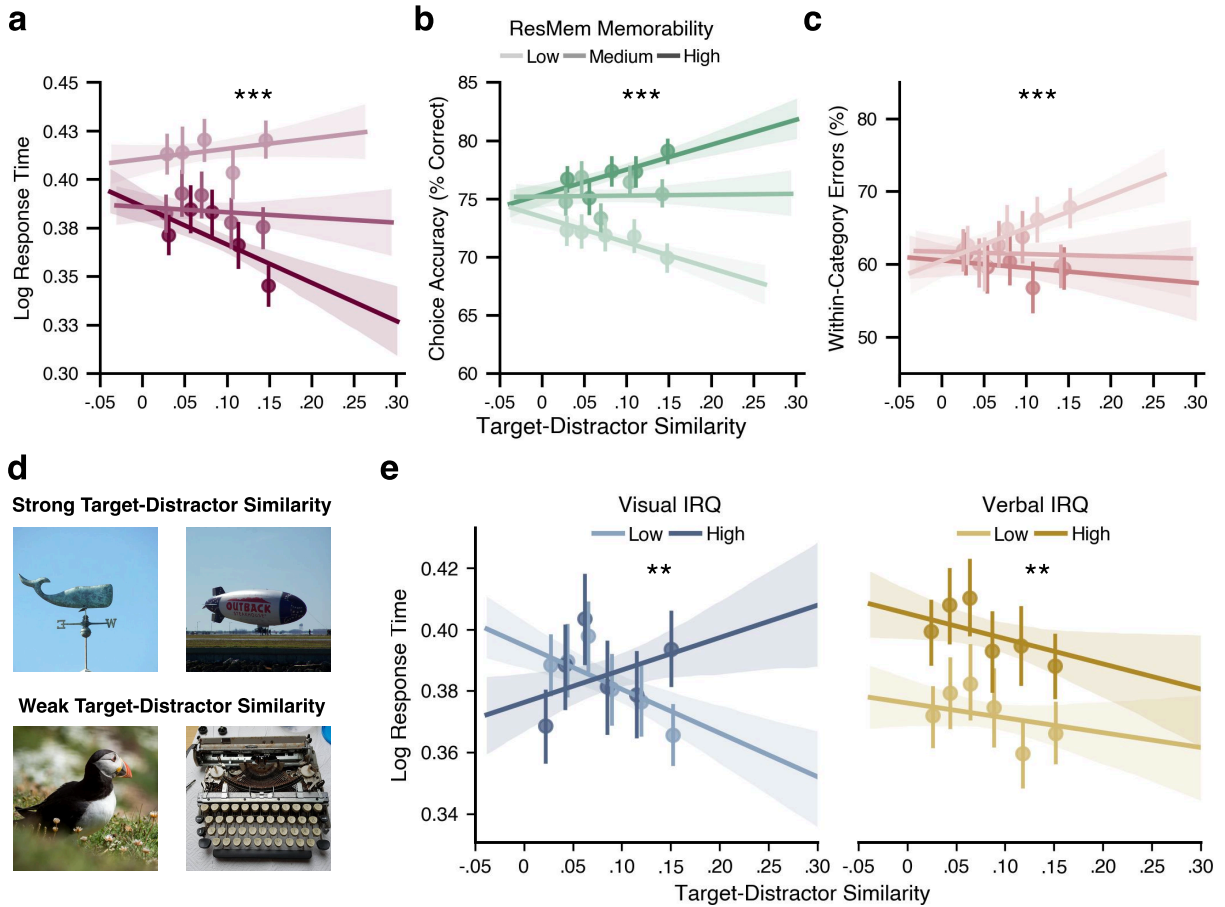


Figure 1.4. Visual similarity between WM targets and distractors interacts with target memorability and individuals' imagery profiles. (a–c) Interacting effects of stimulus memorability and target-distractor similarity on WM performance. For illustrative purposes, the data are split by memorability level and color saturation indicates memorability strength from low to high. (a) Log-transformed response times (RTs). For highly memorable items, the log RTs correspond to a raw RT change from approximately 1550ms at the lowest similarity level to approximately 1450ms at the highest similarity level. (b) Choice accuracy. (c) Within-category error rates. (d) Example stimulus pairs illustrating strong or weak target-distractor similarity, based on VGG-16 Pool3 feature correlations. (e) Interacting effects of visual imagery or verbal thinking and target-distractor similarity on response times. For illustrative purposes, data are median split into low and high imagery groups, with color saturation indicating imagery level from low to high. Across panels (a–c) and (e), solid lines represent linear regression fits with shaded bands showing 95% confidence intervals. On the x-axis, target-distractor similarity scores are binned into quantiles, where points represent mean values within each bin and error bars represent 95% confidence intervals. Asterisks indicate significant interaction effects in all plots ($*p < .05$, $**p < .01$, $***p < .001$).

1.4.5 Visual Imagery Amplifies Distractor Costs for Weakly Memorable Items

Finally, we asked whether stimulus memorability and imagery profile may jointly shape visual distractor effects on WM. If weakly memorable items require more effortful maintenance (as compared to highly memorable items), then their traces may be especially vulnerable to similar visual input in those who rely on a visual maintenance strategy.

Indeed, stronger visual imagery amplified the memorability $\times$ distractor similarity interaction. Imagery quality (OSIVQ) showed three-way interactions in RT ($\beta = -0.003$, 95% CI $[-0.005, -0.001]$, $SE = 0.001$, $t = -2.13$, $p = .03$; Figure A.3, top) and accuracy ($OR = 1.02$, 95% CI $[1, 1.04]$, $SE = 0.009$, $z = 2.00$, $p = .05$; Figure A.3, middle). Imagery vividness (VVIQ) showed a similar interaction for accuracy ($OR = 1.03$, 95% CI $[1.01, 1.05]$, $SE = 0.01$, $z = 2.63$, $p = .008$; Figure 1.5a). Visually similar input had its most disruptive effect on weakly memorable items among those who reported the strongest imagery, while those with weaker imagery were less affected (Figures 1.5a and A.3). No other three-way interactions were significant, and there were no reliable two-way interactions between imagery and memorability (see Tables A.1–A.9).

Imagery strength had little effect on memorable items, but stronger visual tendencies made weakly memorable items even more vulnerable to similar visual input. Across analyses, the impact of both imagery profile and stimulus memorability was most pronounced when WM was in competition with similar sensory input (Figure 1.5b).

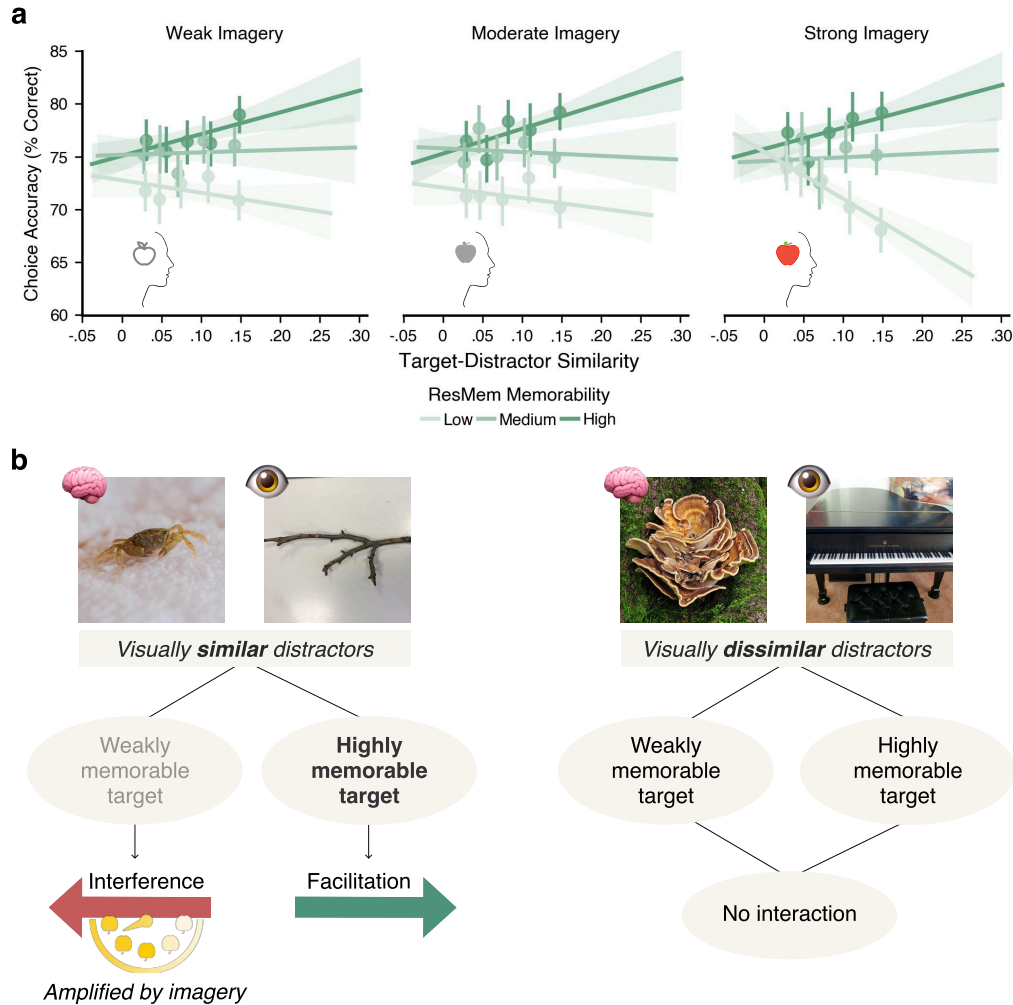


Figure 1.5. Visual imagery vividness amplifies the interaction between target memorability and target-distractor similarity. (a) Choice accuracy (% correct) as a function of target-distractor similarity, target memorability level (low, medium, high ResMem scores), and visual imagery vividness (VVIQ). Panels are separated by imagery vividness. For visualization, VVIQ scores were divided into three equal-sized groups (weak, moderate, strong), indicated by the apple icons below each panel. Within each panel, separate lines represent different stimulus memorability levels, with color saturation indicating binned ResMem scores from low (light green) to high (dark green). Solid lines represent linear regression fits with shaded bands showing 95% confidence intervals. On the x-axis, target-distractor similarity scores were binned into quantiles (for visualization only), with points representing mean values within each bin and error bars representing 95% confidence intervals. (b) Summary diagram of the three-way interaction. Each example pair shows the WM target on the left (brain icon) and the distractor on the right (eye icon). Similar distractors facilitated highly memorable targets but interfered with weakly memorable ones, with interference amplified in strong visual imagers. No effects emerged when distractors were visually dissimilar.

1.5 Discussion

We examined how stimulus memorability, imagery profiles, and task context influence WM success and patterns of failure. We found that individual differences in imagery tendencies can dictate how WM interacts with the task context. Stronger visual imagers showed greater costs to WM content maintained in the background (compared to verbal thinkers), and showed greater vulnerability to visually similar distracting input, especially for content that was intrinsically less memorable. Together, these findings highlight how WM is jointly shaped by interactions between stimulus, individual, and contextual factors.

1.5.1 Memorability Shapes WM Performance and Its Vulnerability to Visual Input

Model-derived memorability predicts better LTM (Needell & Bainbridge, 2022), and here we found that this benefit extends to WM. Items with higher memorability scores were recognized more quickly and accurately, corroborating prior findings from human behavior-derived memorability scores (Gillies et al., 2023; Ye et al., 2024).

We further found that memorable WM stimuli were resilient to visually similar distraction, and even benefited from it. Weakly memorable stimuli were instead more vulnerable to interference. Therefore, equally ‘similar’ distracting input can have opposite effects on stimuli with intrinsically memorable vs. forgettable properties. This may happen if memorable items imprint stronger traces that can persist with less active maintenance, such that overlapping input can bring them back into focus rather than compete with them. Forgettable items, by contrast, may evoke fragile traces that require constant activation to avoid decay, rendering them vulnerable to concurrent input. This is consistent with the idea that memorable items require fewer resources to encode and are more efficiently represented in WM (Bainbridge et al., 2026).

1.5.2 Different WM Strategies Carry Different Costs

Our findings suggest that visual and verbal thinkers use different representational strategies, and these strategies are most behaviorally consequential when WM is placed under strain. For instance, strong visual imagers paid larger costs for deprioritized WM content than strong verbal thinkers. Detailed visual codes may be harder to sustain across multiple items, whereas verbal codes may distribute more readily (though less precisely) across priority states (Souza & Oberauer, 2016; Souza & Skóra, 2017; van Moorselaar et al., 2014).

Strong visual imagers were also more vulnerable to visually similar distracting inputs, especially for weakly memorable WM samples. Strong verbal thinkers were more resistant to visual distraction and even benefited from the input. This suggests that similar visual input may compete with WM for representational territory in those who rely on visual maintenance strategies. But the same input may instead complement semantic representations, or ping muted visual representations, in those who use verbal labeling strategies. The diverging effects of visual input, depending on imagery profile and stimulus memorability, may help reconcile inconsistent evidence for distractor effects during visual WM (Lorenc et al., 2021; Xu, 2017).

These findings highlight that WM content need not be maintained in the format in which it was delivered (Adam et al., 2025). Individuals may transform the perceived content into whatever format suits them (Dong et al., 2026; Kiyonaga & Serences, 2025; Lee et al., 2013; Souza & Skóra, 2017; Yang et al., 2026). Yet not all thought formats are created equal, and the same external input may hurt or help memory depending on how it is represented.

1.5.3 Rethinking Imagery in WM

People across the visual imagery spectrum often perform similarly on visual WM tasks (Knight et al., 2026; Pounder et al., 2022; Weber et al., 2024), raising doubts about whether imagery tendencies have functional relevance. Our findings suggest that such null results may stem from methodological features rather than a true absence of imagery effects on WM.

Most imagery work focuses on imagery vividness (VVIQ) (Argueta et al., 2025), yet people also differ in how habitually they rely on visual or verbal representations (Roebuck & Lupyan, 2020). These facets need not align. Here imagery propensity predicted error patterns, whereas vividness predicted accuracy when forgettable items faced competition from similar visual distractors. Propensity may therefore index the representational strategy people default to, while vividness may index representational fidelity that matters most when WM is under strain. Considering unique imagery dimensions separately may reveal undiscovered relationships between imagery and WM.

WM research has also often used low-dimensional stimuli like colors and orientations (Pounder et al., 2022; Weber et al., 2024), which may fail to engage the higher-level visual processes where imagery differences are most pronounced (Cabbai et al., 2024; Chkhaidze, Reeder, et al., 2026; Liu et al., 2025). Indeed, imagery questionnaires tend to probe complex, scene-like imagery, whereas WM for basic features may be achievable with simple spatial or non-visual strategies (I. Phillips, 2025). The naturalistic images used here may therefore be critical to uncovering the imagery effects we observed.

1.5.4 Limitations and Future Directions

We focused on only low-level similarity between memoranda and visual input, derived from a particular image classification model. However, many other dimensions of relatedness between stimuli may influence WM performance. Likewise, here we focused on how intrinsic stimulus properties of the memoranda may influence performance, but features of the concurrent sensory input (other than visual similarity) may also govern its distracting power. Here we did not vary the length of the WM delay or the distractor onset, but such temporal dynamics may also impact memory. For instance, WM representations may become more semantic over time (Zhou et al., 2022), so distracting visual input may become less so at later time points. Finally, our sample underrepresents imagery extremes, which may be especially informative for understanding the scope of representational flexibility in WM.

The current data came from a study showing that competition during WM can also improve later memory, in that items associated with worse immediate WM performance were remembered better in LTM (Ballard et al., 2026). This hints that imagery-related differences in representational

strategy could also influence what is encoded into LTM.

1.5.5 Concluding Remarks

Information that is encoded visually may nonetheless be maintained in a variety of WM formats. The same stimulus may be maintained as a detailed visual code or a coarser semantic one, and those codes carry different costs under visual distraction and attentional prioritization demands. WM representations and performance are shaped by what is remembered, who is doing the remembering, and the context in which remembering unfolds.

Acknowledgement

Chapter 1, in full, is currently being prepared for submission for publication of the material as Visual and verbal thought tendencies shape working memory under distraction. Chkhaidze, A., Gaspariani, L., Piriyajitakonkij, M., Pappas, I., Ballard, I. C., & Kiyonaga, A. The dissertation author was the primary investigator and author of this material.

Chapter 2

Visual Imagery Modulates Neural, but Not Behavioral, Signatures of Mental Simulation During Language Comprehension

2.1 Introduction

I'm in the house.

It's nice out:

warm sun on cold snow.

First day of spring

or last of winter.

— *From Ron Padgett's "Poem"*

Reading these lines, you might find yourself picturing the white brightness of snow, feeling the low warmth of the sun, or sensing the quiet pull of the world outside. Embodied theories of cognition propose that meaning is partly constructed through this kind of reactivation of sensory and motor experiences (Barsalou, 1999; Glenberg & Kaschak, 2002; Pulvermüller, 2005; Zwaan, 2014). Yet not all minds picture alike. Some individuals report vivid and detailed mental imagery, whereas others report little to no voluntary visual imagery (Zeman, 2024). Here, we test whether visual imagery vividness shapes the extent to which people rely on perceptual simulation during language comprehension.

Beyond this introspective example, behavioral and neural studies provide evidence for sensorimotor grounding of meaning. Imageable words are recognized more quickly than less imageable words (Balota et al., 2004; Cortese & Fugett, 2004), and words referring to objects that people can easily physically interact with, such as *tool*, are judged more quickly and accurately than words referring to less interactive objects, such as *ship* (Siakaluk et al., 2008). Sentence meaning can also bias perception: motion language can alter subsequent motion perception, including sensitivity to opposite direction motion and motion-aftereffect-like biases (Meteyard et al., 2007; Dils & Boroditsky, 2010). People are also faster to verify pictures when their visual features, such as shape, orientation, color, or size, fit the features implied by a preceding sentence (de Koning et al., 2017; Stanfield & Zwaan, 2001; Zwaan et al., 2002; Zwaan & Pecher, 2012). Neuroimaging studies add parallel evidence: action words engage motor regions (Gerlach et al., 2002; Hauk et al., 2004), and color knowledge retrieved from words engages regions involved in color perception (Pulvermüller & Hauk, 2006; Simmons et al., 2007). Together, these findings suggest that language comprehension may rely on perceptual and motor simulation, not only on linguistic associations.

Recently, individual differences in mental imagery have been proposed as a test case for the embodied accounts of language (Muraki et al., 2023). If comprehension relies on perceptual simulation, simulation effects might scale with visual imagery capacity. Consistent with this possibility, individuals with aphantasia show reduced responses in several language-based tasks that plausibly depend on constructing perceptual or affective simulations. For example, they report reduced immersion in the events described by a story compared to strong visual imagers (Speed et al., 2024), weaker autonomic responses to frightening written scenarios (Wicken et al., 2021), and reduced motor-system engagement during action-language processing, indexed by the absence of TMS-evoked increases in corticospinal excitability (Dupont et al., 2024). Yet in behavioral paradigms commonly proposed to index language-induced simulation, visual imagery often does not predict how strongly people show these effects. This includes tasks in which participants verify object properties or judge whether a picture fits the situation implied by a sentence (Muraki & Pexman, 2021; Pecher, van Dantzig, & Schifferstein, 2009; Speed & Majid, 2018; Speed et al., 2026; Zwaan & Pecher, 2012). These findings have been taken to

suggest that visual imagery may not be required for automatic perceptual simulation during language comprehension. These tasks, however, may permit multiple routes to success.

In sentence-picture verification tasks, participants read a sentence and then judge whether a subsequent picture depicts an object mentioned in that sentence. On critical trials, the object has been mentioned, but its depicted shape either matches or mismatches the shape implied by the sentence (Zwaan & Pecher, 2012). The typical behavioral advantage for matching pictures is referred to as the match effect. This effect may reflect a visually specific simulation of the described situation, but it may also arise from linguistic associations or broader semantic expectations (Ostarek et al., 2019). For example, after reading about an owl in the sky, a reader may activate associations among owl, sky, flight, and wings, making a folded-wing picture less expected without constructing a detailed visual image of the owl. If different routes lead to the same response-time and accuracy pattern, behavior alone may miss mechanistic differences. Neural measures can therefore probe the processes underlying the shape-match effect and ask whether they differ across the imagery spectrum.

Here, we tested whether visual imagery vividness shapes behavioral and neural signatures of the match effect. Because prior behavioral studies have found little evidence that imagery vividness modulates sentence-picture verification effects, Experiment 1 first established whether our stimulus set produced the expected behavioral shape-match effect and whether this effect varied with imagery vividness. Experiment 2 then asked whether neural responses to pictures distinguish sentence-implied shape matches from mismatches, and whether imagery vividness modulates these neural responses in ways that may not be apparent in behavior. To address this question, we recorded EEG during the same task and modeled picture-locked responses across single trials, examining whether the shape-match effect differed across early perceptual and later semantic time windows.

2.2 Experiment 1: Behavioral Study

2.2.1 Methods

In Experiment 1, we conducted a preregistered online behavioral study testing whether pictures are verified more easily when their shape matches, rather than mismatches, the shape implied by a preceding sentence. Participants read sentences and then judged whether a subsequent picture depicted an object mentioned in the sentence. On critical trials, the object was always mentioned, but its shape either matched or mismatched the sentence-implied shape. We analyzed response time and accuracy on these trials to test whether our stimulus set produced the canonical shape-match effect, and whether its magnitude varied with visual imagery vividness as measured by the Vividness of Visual Imagery Questionnaire (VVIQ; Marks, 1973).

Participants

We conducted an a priori power analysis based on independent pilot data, which yielded a target final sample size of 120 participants. Full details of the analysis are available in our preregistration on OSF: <https://osf.io/4cp6k>. Briefly, we used a simulation-based power analysis for linear mixed-effects models using the `simr` package (Green & MacLeod, 2016) in R (Team, 2023). Our goal was to achieve $\geq 95\%$ power to detect the interaction between match effect and VVIQ scores. We recruited 135 participants to allow for anticipated exclusions (approximately 10%, based on pilot exclusion rates).

Participants were recruited through Prolific and were compensated at a rate of \$10 per hour. From the initial sample of 135, four participants were excluded for not completing the imagery questionnaires, and an additional 13 were excluded for failing one or more of the three attention checks embedded within the questionnaires. Two further participants were excluded for self-reporting as non-native English speakers, and one was excluded for having participated in a related prior study. To the remaining 115 participants, we then applied the preregistered task-based exclusion criteria. Trials with sentence-reading or picture-verification RTs below 300ms were excluded, as were trials where

picture-verification RTs exceeded three standard deviations above each participant's mean. Four additional participants were removed for retaining fewer than 30 valid trials. One participant with unusually long RTs and mean accuracy below 50% was also removed, leaving a final sample of 110 participants for analysis (65 female, 42 male, 3 non-binary; mean age = 39 ± 12.52).

All Experiment 1 procedures and analysis plans were preregistered before data collection. The study was approved by the Office of IRB Administration at UC San Diego, and all participants provided informed consent prior to the study.

Materials

Experimental stimuli consisted of 60 sentence-picture quadruplets adapted from Zeelenberg et al. (2025) and modeled after Zwaan et al. (2002). Each quadruplet consisted of two sentences and two grayscale photographs on a white background. One picture matched the shape implied by its paired sentence and the other mismatched it. For example, the sentence "There was an owl in the sky" could be paired with an image of an owl with wings outstretched (match) or wings folded (mismatch) (Figure 2.1a). Each participant saw only one of the four possible combinations for each item. An additional 60 filler pairs depicted objects not mentioned in the sentence; we refer to these filler trials as *unrelated* throughout. Four counterbalanced lists were created so that, across participants, each item appeared equally often in all conditions.

To assess individual differences in mental imagery, we used two questionnaires: the Vividness of Visual Imagery Questionnaire (VVIQ; Marks, 1973) and the Object-Spatial Imagery and Verbal Questionnaire (OSIVQ; Blazhenkova & Kozhevnikov, 2009). The VVIQ asks participants to visualize specific scenes and rate their vividness on a 5-point scale. The OSIVQ assesses quality of visual imagery and verbal thought. As preregistered, analyses in this study focus on the VVIQ.

Procedure

Each trial (Figure 2.1a) began with a central fixation cross (500ms), followed by a sentence. Participants pressed the spacebar once they had finished reading, triggering a 250ms blank screen, after

which a grayscale picture was presented until response. Participants judged whether the depicted object had been mentioned in the sentence ("y" for yes, "n" for no). All experimental trials required a "yes" response, and all unrelated trials required a "no" response. Each participant completed 120 trials (30 match, 30 mismatch, 60 unrelated) in randomized order. After completing the verification task, participants filled out the imagery questionnaires.

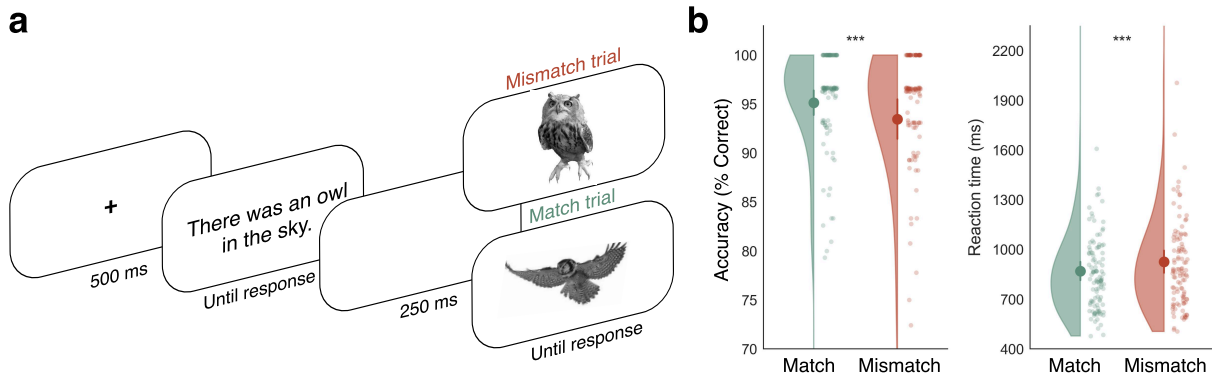


Figure 2.1. Sentence-picture verification task and behavioral results. (a) Sentence-picture verification task trial sequence. Participants viewed a fixation cross (500 ms), read a sentence implying a specific shape of an object, pressed the spacebar to advance, and after a 250 ms blank screen judged whether the depicted object had been mentioned in the sentence. The picture either matched the implied shape (e.g., owl with wings outstretched) or mismatched it (e.g., owl with wings folded). Each participant saw only one of the two possible pictures per item. (b) *Left*: Choice accuracy by trial type (match vs. mismatch). *Right*: Reaction time by trial type, computed on correct trials only. For both panels, half-violins show the distribution of subject-level means, individual dots represent participants, and central markers indicate group means with 95% confidence intervals (** $p < .001$).

Statistical Analysis

Only experimental trials, those requiring a "yes" response, were analyzed. Trials with sentence-reading times below 300ms were excluded, as were trials with picture-verification RTs below 300ms or exceeding three standard deviations above a participant's mean. RT analyses were further restricted to correct trials (5,803 of 6,146; 94.4%); accuracy analyses retained all trials. All continuous predictors were z-scored prior to model fitting.

Analyses were conducted in R (Team, 2023). We fit linear mixed-effects models for RT and generalized linear mixed-effects models for accuracy using `lmer` and `glmer` from the `lme4` package (Bates et al., 2015), with random intercepts for subjects and items. The primary predictors were sentence-picture pair type (match vs. mismatch), VVIQ scores, and their interaction. The structure of both RT and accuracy models was:

$$RT/accuracy \sim \text{Sentence-picture pair type} \times VVIQ + (1 \mid \text{Subject}) + (1 \mid \text{Item})$$

2.2.2 Results

We first asked whether the canonical match effect would replicate (Stanfield & Zwaan, 2001; Zwaan et al., 2002; Zwaan & Pecher, 2012; Zeelenberg et al., 2023). The match effect is operationalized as faster and more accurate responses to pictures matching the shape implied by the preceding sentence. Participants responded faster on match than on mismatch trials (Figure 2.1b, right; $\beta = -54.71$, $SE = 10.15$, $t = -5.39$, $p < .001$). They were also more accurate on match trials (Figure 2.1b, left; $\beta = 0.50$, $SE = 0.13$, $z = 3.69$, $p < .001$). Together, these results replicate the canonical match effect across both speed and accuracy, demonstrating that sentence-implied shape information influences subsequent picture verification.

We reasoned that, if representations driving the match effect rely on visual imagery, the magnitude of the effect might scale with imagery vividness as measured by VVIQ (Figure B.1; $M = 51.37 \pm 11.91$, range: 16–80). The match $\times$ VVIQ interaction was not significant for either RT ($\beta = -6.75$, $SE = 10.17$, $t = -0.66$, $p = .51$) or accuracy ($\beta = -0.14$, $SE = 0.14$, $z = -1.04$, $p = .30$), suggesting that the response-time and accuracy signatures of sentence-picture verification do not depend on vivid imagery.

One interpretation is that self-reported imagery vividness is simply not relevant to sentence-picture verification. Alternatively, comparable performance might also emerge from different underlying mechanisms. The match effect may partly reflect high-level semantic or conceptual compatibility rather than low-level perceptual simulation alone (Ostarek et al., 2019), suggesting the task can be

solved through multiple routes. Participants at different points along the imagery spectrum may rely on different representational routes, with some using visual simulation and others using semantic or spatial alternatives. In Experiment 2, we used EEG to test whether a neural signature of the match effect differs across the imagery spectrum, asking whether neural measures might reveal mechanistic differences that did not surface in participants' behavior.

2.3 Experiment 2: ERP Study

2.3.1 Methods

In Experiment 2, we recorded EEG while participants performed an EEG-adapted version of the sentence-picture verification task from Experiment 1. Sentences were presented word by word, and pictures were briefly presented after a fixed delay to support picture-locked ERP analyses. We analyzed neural responses to two contrasts: pictures of mentioned versus unmentioned objects, indexing sentence-picture relatedness, and pictures of mentioned objects whose shape matched versus mismatched the sentence-implied shape, indexing the match effect. We then asked whether these neural responses varied with visual imagery vividness.

Participants

We recruited 39 participants from the UC San Diego community and the online aphantasia community via Reddit, targeting a sample spanning a wide range of imagery abilities, including individuals with self-reported aphantasia and hyperphantasia. Participants were native or fluent English speakers, had normal or corrected-to-normal vision, and reported no history of neurological or psychiatric disorder. They were compensated at a rate of \$20 per hour.

Five participants were excluded before analysis: three due to having too few artifact-free trials, and two because delayed EEG recording onset prevented alignment of EEG epochs with trial identity. The final sample comprised 34 participants for EEG analysis (20 female, 10 male, 3 non-binary, 1 preferred not to say; mean age = 24.24 ± 6.50). Due to a coding error, RT data were unavailable for

one participant, yielding 33 participants for RT analyses. The study was approved by the Office of IRB Administration at UC San Diego, and all participants provided informed consent prior to the study.

Materials

Stimuli were identical to those used in Experiment 1: 60 sentence-picture quadruplets adapted from Zeelenberg et al. (2025). Each participant saw 30 match trials, 30 mismatch trials, and 60 unrelated trials, counterbalanced across four lists. Following the experimental task, participants completed the VVIQ (Marks, 1973) and OSIVQ (Blazhenkova & Kozhevnikov, 2009). As in Experiment 1, primary analyses focused on the VVIQ. We additionally included OSIVQ Verbal as a control measure of verbal thinking tendency, allowing us to test whether observed effects were specific to visual imagery or reflected broader individual differences in cognitive style. VVIQ scores ranged from 16 to 76 ($M = 44.06 \pm 20.14$; Figure 2.2, left) and OSIVQ Verbal scores ranged from 1.73 to 4.47 ($M = 3.02 \pm 0.66$; Figure 2.2, right).

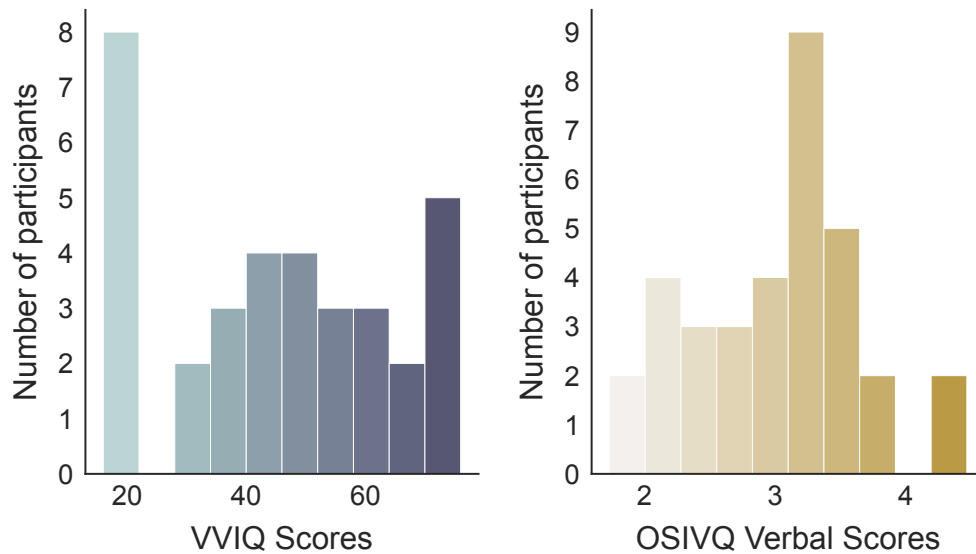


Figure 2.2. Distributions of VVIQ scores (*left*), indexing visual imagery vividness, and OSIVQ Verbal scores (*right*), indexing verbal thinking tendency, in the final EEG sample ($N = 34$).

Procedure

The sentence-picture verification task was adapted for EEG recording by replacing self-paced reading with rapid serial visual presentation (RSVP). Each trial began with a fixation cross (500ms), followed by the sentence presented one word at a time (200 ms per word, 300ms blank between words). After an 800ms delay following the final word, a grayscale picture appeared for 200ms, followed by a 500ms blank screen. A question mark then prompted participants to indicate whether the depicted object had been mentioned in the preceding sentence by pressing "y" for yes or "n" for no. Experimental trials (match and mismatch) required a "yes" response; unrelated trials required a "no" response. Trials were presented in randomized order across six blocks.

EEG Recording and Preprocessing

EEG was recorded at 512 Hz using a 64-channel electrode cap with electrodes placed at standard sites from the 10–20 system. Electrodes were referenced online to Fp1 and later re-referenced to the average of the left and right mastoid electrodes (TP9 and TP10). Blinks were monitored using an electrode below the right eye, and horizontal eye movements were monitored using a bipolar montage of electrodes placed at the right and left outer canthi. After re-referencing and dropping the mastoid channels, 62 scalp channels remained for analysis.

Continuous EEG was bandpass-filtered from 0.1 to 40 Hz and submitted to independent component analysis using the *runica* algorithm in EEGLAB (Delorme & Makeig, 2004). Components reflecting ocular artifacts were identified through visual inspection of component topographies and activation time courses; no more than three components per participant were removed, as excessive component removal risks introducing spurious corrections to task-related brain activity. Epochs were extracted from –200 to 800ms relative to picture onset and baseline-corrected to the 200ms pre-stimulus interval. Epochs exceeding a $\pm 150 \mu\text{V}$ peak-to-peak threshold at any channel were rejected. After artifact rejection, the picture-locked dataset comprised 3,835 epochs (unrelated: 1,896; match: 973; mismatch: 966; 6.0% rejection rate). Preprocessed epochs were imported into Python and analyzed

using MNE-Python (Gramfort et al., 2013). MNE-Python was used to compute condition-wise ERPs, extract single-trial mean amplitudes for the predefined time windows, and generate waveform and topographic visualizations.

Statistical Analysis

Although Experiment 2 was not preregistered, the model structure and analysis pipeline followed those of Experiment 1. The behavioral models used the same specifications, with random intercepts for subjects and items, and added a Helmert contrast comparing related to unrelated trials. The picture-locked ERP models extended this structure by using voltage as the dependent variable and adding the spatial coordinates (X , Y , Z) of each electrode as continuous predictors interacting with condition and imagery terms. Random effects matched those of the behavioral models.

Behavioral analysis. As in Experiment 1, we fit linear mixed-effects models for RT, using correct trials only, and generalized linear mixed-effects models for accuracy using the `lme4` package in R (Bates et al., 2015). Models included random intercepts for subjects and items. We excluded trials with RTs below 300ms or exceeding three standard deviations above each participant’s mean. RT analyses were conducted on 33 participants due to missing RT data for one participant. The primary predictors were sentence-picture pair type, VVIQ scores, and their interaction. Condition was coded using Helmert contrasts (Schad et al., 2020). We refer to match and mismatch trials collectively as *related* trials. The first contrast compared unrelated trials to the mean of related trials, indexing the relatedness effect; the second compared match to mismatch trials directly, indexing the match effect. Although the relatedness contrast was not a focus of Experiment 1, we included it here because it allowed us to examine the semantic relatedness effect between sentences and pictures in the ERPs.

Picture-locked ERP analysis. Picture-locked ERP analyses were conducted on single-trial mean voltage in two *a priori* time windows: an early window from 150–190ms, chosen to capture early posterior picture processing in the latency range of the N170 (Schendan & Lucia, 2010), and a later window from 300–500ms, chosen to capture N400-window semantic integration effects (Kutas & Federmeier, 2011). We refer to these as the N170 and N400 time windows, respectively, while

interpreting effects in terms of their timing and scalp distribution rather than assuming that each effect reflects a single canonical ERP component.

To model both spatially averaged voltage differences and scalp distribution, we included the spatial coordinates of each electrode as continuous predictors: X (left–right), Y (posterior–anterior), and Z (superior–inferior). This approach avoids treating electrodes as discrete levels of a fixed or random factor, which is complicated by spatial correlations between neighboring channels (Kappenman & Luck, 2011). By allowing X , Y , and Z to interact with condition and imagery predictors, we could test whether each effect varied systematically across the scalp without predefining specific regions of interest (ROIs) or conducting channel-by-channel tests (Vinaya et al., 2025; Winsler et al., 2018). Interactions with X indicate left–right variation, interactions with Y indicate posterior–anterior variation, and interactions with Z indicate superior–inferior variation.

As in the behavioral models, condition was coded using two planned contrasts. The first compared unrelated trials to the average of match and mismatch trials, indexing broad sentence-picture relatedness. The second compared mismatched to matched pictures, indexing shape-specific mismatch. VVIQ and OSIVQ Verbal scores were z-scored prior to model fitting. Each model was:

$$\begin{aligned} \text{Voltage} \sim & (\text{Relatedness effect} + \text{Match effect}) \times (\text{VVIQ} + \text{OSIVQ Verbal}) \times (X + Y + Z) \\ & + (1 \mid \text{Subject}) + (1 \mid \text{Item}) \end{aligned}$$

Main effects of the planned contrasts estimate voltage differences across the modeled scalp, whereas interactions with X , Y , or Z indicate systematic scalp distributions. For visualization, topographic maps were reconstructed from the model-estimated contrast and scalp-coordinate coefficients, following prior coordinate-based ERP modeling approaches (Vinaya et al., 2025; Winsler et al., 2018). Because these reconstructed maps summarize effects using three spatial dimensions rather than estimating each channel independently, they are more schematic than channel-level averages, but provide stable and interpretable summaries of the underlying scalp distributions (see Smith & Kutas, 2015 for discussion). Significance of all coefficients was assessed using a false discovery rate (FDR) correction (Benjamini &

Yekutieli, 2001) applied across the full set of model terms.

2.3.2 Results

Behavioral Results

We first asked whether the behavioral patterns from Experiment 1 would replicate. The relatedness effect (unrelated vs. related) was reliable in Experiment 2 as well: participants responded faster (Figure 2.3a, right; $\beta = -10.5$, $SE = 2.53$, $t = -4.14$, $p < .001$) and more accurately (Figure 2.3a, left; $\beta = 0.94$, $SE = 0.19$, $z = 4.88$, $p < .001$) on unrelated trials. The match effect, however, did not replicate. Match and mismatch trials yielded comparable RTs ($\beta = 2.3\text{ms}$, $SE = 4.89$, $t = 0.48$, $p = .63$) and accuracies ($\beta = 0.07$, $SE = 0.13$, $z = 0.56$, $p = .57$). This may reflect the smaller EEG sample ($N = 33$ vs. $N = 110$ in Experiment 1), as well as changes to the EEG task format, where self-paced reading was replaced with word-by-word RSVP followed by a fixed 800ms delay before picture onset.

As in Experiment 1, visual imagery did not reliably modulate behavior. No VVIQ interactions emerged on either RT (match $\times$ VVIQ: $\beta = -1.6$, $SE = 4.99$, $t = -0.33$, $p = .74$; relatedness $\times$ VVIQ: $\beta = -4.1$, $SE = 2.32$, $t = -1.78$, $p = .07$) or accuracy (match $\times$ VVIQ: $\beta = 0.21$, $SE = 0.14$, $z = 1.54$, $p = .12$; relatedness $\times$ VVIQ: $\beta = -0.08$, $SE = 0.14$, $z = -0.59$, $p = .56$). Full statistical output for the RT and accuracy models is provided in Tables B.1 and B.2.

Visual imagery did not interact with the match effect in either of the two experiments in the present study. However, even when participants reach comparable behavioral outcomes, they may do so through partly different mechanisms. We turn next to EEG to examine the neural signatures of relatedness and match effects, and to ask whether these signatures differ across imagery profiles during sentence–picture verification.

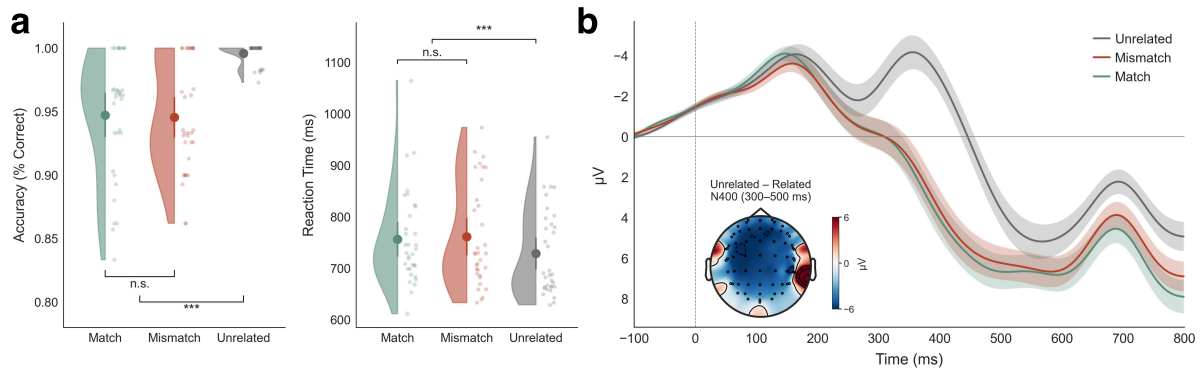


Figure 2.3. Behavioral and neural responses to sentence-picture verification in Experiment 2. **(a)** Accuracy (left) and reaction time on correct trials (right) by trial type (match, mismatch, unrelated) in the EEG sample ($N = 33$). Half-violins show the distribution of subject-level means, individual dots represent participants, and central markers indicate group means with 95% confidence intervals. The relatedness effect (unrelated vs. related) was significant for both measures ($***p < .001$), while the match effect (match vs. mismatch) did not reach significance (n.s.). **(b)** Grand-average ERPs ($N = 34$) at centro-parietal sites (Cz, C3, C4, CPz, CP1, CP2, Pz, P3, P4) for unrelated and related trials, where related trials include match and mismatch trials averaged together, time-locked to picture onset. Shaded regions show ± 1 SEM. Inset: scalp topography of the unrelated–related difference in the N400 window (300–500ms).

ERP Results

Relatedness effect. If sentence-picture verification engages semantic integration, unrelated pictures should elicit a larger N400 than related pictures, consistent with prior findings (Hamm et al., 2002; Hirschfeld et al., 2011). Indeed, unrelated pictures elicited more negative-going voltage than related pictures from 300–500ms ($\beta = -1.96$, $t = -6.23$, $p < .001$). This effect was also systematically distributed across the scalp, becoming more negative along the anterior (relatedness $\times$ Y: $\beta = -0.10$, $t = -17.53$, $p < .001$) and superior (relatedness $\times$ Z: $\beta = -0.28$, $t = -29.55$, $p < .001$) dimensions, producing the centro-frontal N400-window negativity consistent with prior ERP studies of picture-level semantic incongruity (Figure 2.3b; Federmeier & Kutas, 2001, 2002; Kutas & Hillyard, 1980; West & Holcomb, 2002). This relatedness effect did not interact with either VVIQ or OSIVQ Verbal scores ($ps > .05$; Figure B.2, top), suggesting that the neural index of semantic relatedness was robust across the imagery spectrum. Full statistical output is provided in Tables B.3

and B.4.

A weaker relatedness effect was observed already in the 150–190ms window. The unrelated–related contrast interacted with the anterior–posterior scalp axis, with unrelated pictures eliciting greater negativity at more anterior sites (relatedness $\times$ Y : $\beta = -0.019$, $SE = 0.006$, $t = -3.27$, $p = .001$, $p_{FDR} = .02$). This suggests that semantically related and unrelated pictures may be differentiated early in picture processing, although the effect was strongest in the later N400 window.

Match effect. We next asked whether the match effect, absent in behavior, was nonetheless reflected in the ERPs. Specifically, if the comparison between sentence-implied and depicted shape recruits visual processing, the match effect may emerge in early picture-locked ERPs. The match effect did not emerge as a spatially uniform voltage difference in either time window (Figure 2.3b; $ps > .05$ for N170 and N400). However, in the 150–190ms window, the contrast showed a reliable topographic distribution. Specifically, the Y -axis interaction (match $\times$ Y : $\beta = 0.03$, $t = 3.11$, $p = .028$) indicated that the contrast became increasingly negative at posterior sites, and the X -axis interaction (match $\times$ X : $\beta = -0.04$, $t = -4.15$, $p < .001$) indicated that this negativity was greater over the right hemisphere (Figure 2.4a, left). This pattern may reflect an enhanced N170 negativity to mismatched compared to matched pictures over right occipito-temporal regions, with a corresponding anterior positivity characteristic of the N170/vertex positive potential (VPP) geometry (Joyce & Rossion, 2005).

The match $\times$ Y effect persisted through the 300–500ms window (match $\times$ Y : $\beta = 0.04$, $t = 4.52$, $p < .001$). However, this later effect lacked the right-lateralized signature observed in the N170 window and instead showed a broader, bilaterally distributed posterior pattern (Figure B.2, middle; Tables B.3 and B.4).

Visual imagery modulates the early neural signature of shape mismatch. If participants with different imagery profiles detect mismatches through qualitatively different routes, with some relying more on visual simulation and others on alternative strategies, VVIQ may shape how the neural match effect is distributed across the scalp. Consistent with this possibility, VVIQ modulated the topographic distribution of the mismatch–match contrast specifically in the N170 window ($\beta = 0.04$, $t = 5.38$, $p < .001$; Figure 2.4a, middle). Specifically, higher VVIQ scores were associated with a

stronger posterior expression of the mismatch negativity, sharpening the bipolar pattern observed in the group average. No corresponding VVIQ modulation was observed in the N400 window: neither the overall VVIQ $\times$ match effect nor its interactions with scalp coordinates were significant (all $ps \geq .50$; Table B.4). Critically, this modulation was specific to visual imagery vividness: OSIVQ Verbal did not modulate the match effect in either time window (all $ps > .05$; Figure 2.4a, right; Tables B.3 and B.4). Neither imagery measure modulated the relatedness effect (all $ps > .40$; Tables B.3 and B.4), suggesting that imagery differences were not tied to broad sentence-picture relatedness, but to the early picture-processing response to shape mismatch.

To visualize the nature of the relationship between visual imagery and the match effect, we plotted mismatch–match topographies separately for VVIQ tertiles (Figure 2.4b). The strong imagery group showed a right occipito-temporal negativity for mismatched relative to matched pictures, accompanied by anterior positivity (Figure 2.4b, right). This pattern resembled the group-level match-effect topography (Figure 2.4a, left), but with a stronger overall expression. In moderate imagers, the effect was expressed more broadly over posterior sites, without the same clear right occipito-temporal focus. In weak imagers, posterior mismatch negativity was largely absent.

To further examine the lateralized posterior pattern in strong imagers, we plotted time-resolved waveforms over left and right occipito-temporal ROIs (Figure 2.4c). The waveforms confirmed that strong imagers showed the clearest mismatch–match divergence over right occipito-temporal sites in the 150–190ms window, whereas weak and moderate imagers showed little condition separation in this region. This posterior-negative/anterior-positive geometry in strong imagers resembles the scalp distribution of the N170/VPP complex, which has been linked to early visual object processing and modulated by visual imagery and imagery–perception interactions (Ganis & Schendan, 2008). We therefore describe this as an N170-window effect.

These findings suggest that similar behavioral performance masked differences in neural processing of shape mismatch across the imagery spectrum.

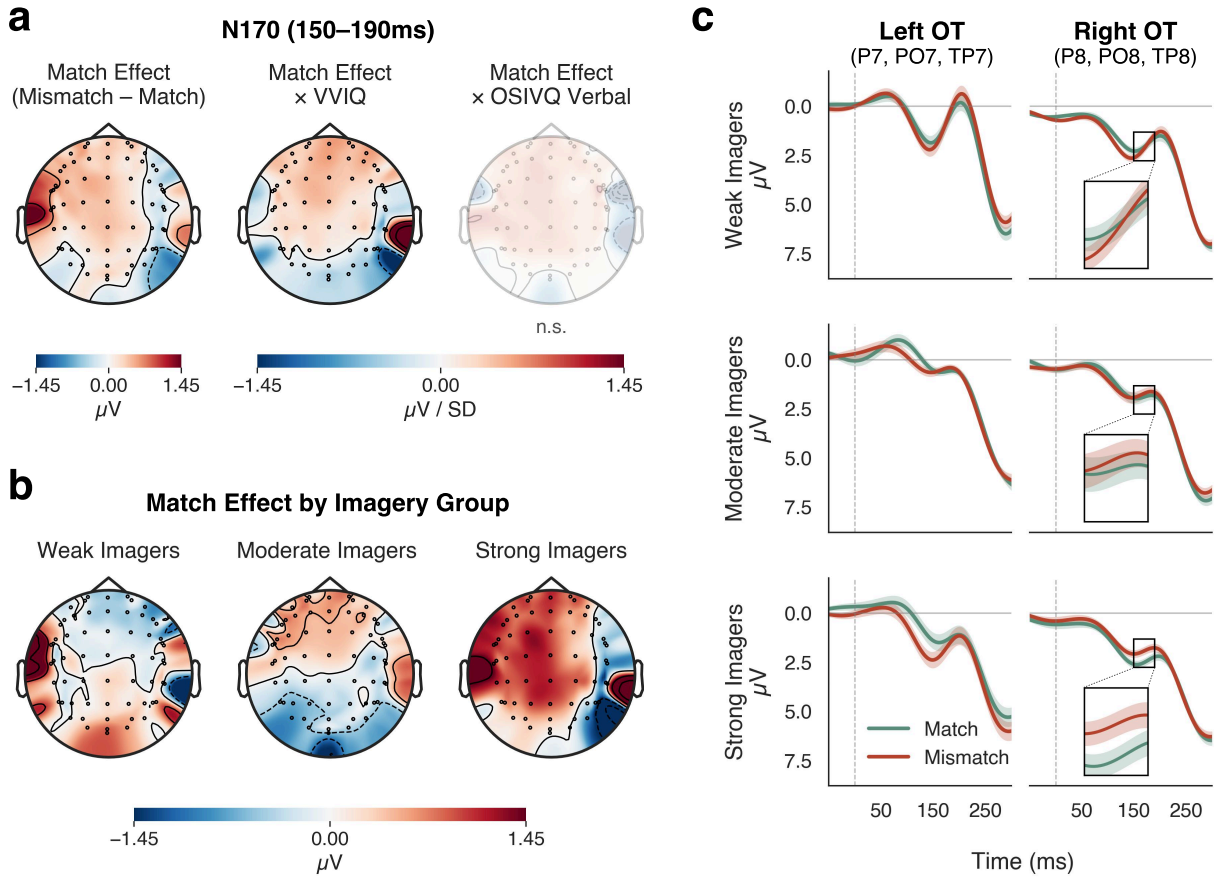


Figure 2.4. ERP signatures of the match effect and its modulation by visual imagery. (a) Group-level topographic distribution of the mismatch–match difference at the N170 (150–190ms post-picture onset; left) and its modulation by visual imagery ($\times$ VVIQ; middle) and verbal thinking tendency ($\times$ OSIVQ Verbal; right). The middle and right maps show per-channel regression coefficients of the mismatch–match difference on z-scored imagery scores, in units of μV per SD of the imagery score. The match effect was reliably modulated by visual imagery vividness, but not by OSIVQ Verbal (n.s., faded). (b) Mismatch–match topographies at the N170 shown separately for participants in the lower (weak imagers; $n = 12$), middle (moderate imagers; $n = 11$), and upper (strong imagers; $n = 11$) tertiles of VVIQ. Each topomap shows the per-channel difference between mismatch and match conditions, averaged across subjects within the tertile and over the 150–190ms window. (c) Match and mismatch waveforms time-locked to picture onset, averaged over left occipito-temporal (left OT: P7, PO7, TP7) and right occipito-temporal (right OT: P8, PO8, TP8) clusters, separately for each VVIQ tertile. Shaded regions show within-subject Cousineau–Morey standard errors (Cousineau, 2005; Morey, 2008). Black rectangles mark the N170 window (150–190ms). Tertile groupings in (b) and (c) are shown for visualization only; statistical analyses use continuous VVIQ scores.

2.4 Discussion

We examined how visual imagery vividness shapes sentence-picture verification. In Experiment 1, we found the canonical behavioral shape-match effect, but its magnitude did not vary with imagery vividness, as measured by VVIQ scores. In Experiment 2, we found distinct ERP signatures of object relatedness and shape match: unmentioned objects elicited a robust N400-window negativity, whereas shape mismatch elicited an early right-posterior negativity and a later posterior N400-window effect. Strong imagers showed a clear right-posterior mismatch negativity in the N170 time window, whereas weak and moderate imagers did not. However, the N400-window match effect was present at the group level and was not modulated by imagery, suggesting that shape mismatches were detected more broadly across the imagery spectrum. Thus, participants may have reached similar verification judgments through partly different processing routes. The pattern is consistent with the possibility that strong imagers detected shape mismatch through an earlier, more perceptually specific route, whereas weaker imagers may have relied more on later, less visually specific integration processes.

2.4.1 Neural Underpinnings of the Match Effect

We found that pictures whose shape conflicted with the sentence-implied shape elicited a posterior effect in the 150–190ms window. Mismatched pictures produced a right-posterior negativity relative to matched pictures, suggesting that sentence-implied shape influenced visual analysis of the picture. The latency and scalp distribution of this effect place it in the range of early visual-object responses, including the N170/VPP complex, which has been linked to structural and categorical processing of visual input (Joyce & Rossion, 2005; Schendan & Lucia, 2010). This timing suggests that sentence-derived visual expectations affected picture processing before later stages of conceptual evaluation and response selection. Imagery work further shows that imagined visual form can modulate early perceptual responses to subsequently viewed objects (Ganis & Schendan, 2008), consistent with the possibility that sentence-generated visual expectations shaped early picture processing here.

The shape-match effect was also evident in the 300–500ms window, where mismatched pictures

elicited a broader posterior negativity. This later response may reflect the conceptual consequences of recognizing the correct object in a state that conflicts with the sentence-implied shape. Picture-ERP studies have shown that object pictures can elicit earlier image-sensitive negativities as well as later N400-family effects, with later effects tracking semantic relatedness and integration more directly (McPherson & Holcomb, 1999; Kutas & Federmeier, 2011). In our data, the later match effect may therefore index semantic evaluation of the depicted object state, while the early posterior effect may index sensitivity to visual form.

The posterior distribution of the shape-match effect also distinguishes it from the relatedness effect, in which pictures depicting unmentioned objects elicited a larger, more anterior/superior N400-window negativity. This difference may partly reflect the composition of the contrasts. In the relatedness contrast, unrelated pictures differed from related pictures in object identity, semantic fit, and required response (“no” vs. “yes”), potentially introducing decision-related activity that overlapped with the N400-window negativity. Its anterior distribution is also consistent with ERP studies of image-based semantic processing, where incongruent or unrelated pictures often elicit frontally focused N300/N400-family negativities (McPherson & Holcomb, 1999; West & Holcomb, 2002; Wu & Coulson, 2011). By contrast, the shape-match comparison held object identity and response constant: both conditions involved sentence-mentioned objects and both required “yes” responses. The later posterior negativity in this contrast may therefore more specifically reflect the cost of reconciling the depicted object state with the state implied by the sentence.

Together, these findings suggest that sentence-implied visual meaning may be supported by both perceptual and conceptual information. The early posterior match effect points to visual-form sensitivity, whereas the later N400-window match effect may reflect semantic evaluation of whether the depicted object state fits the sentence context. This layered pattern aligns with hybrid accounts of meaning, in which comprehension draws on both grounded sensorimotor representations and distributional or linguistic knowledge, with their relative contribution varying across tasks, contexts, and kinds of information (Barsalou, 2008; Davis & Yee, 2021; Connell, 2019; Louwerse, 2011). Recent EEG work similarly suggests that distributional and sensorimotor information make distinct, time-

varying contributions to conceptual processing (Vinaya et al., 2025). In our task, sentence-implied shape may therefore be represented through interacting visual and conceptual routes: expectations about object form may guide early perceptual analysis, while later semantic processing evaluates whether the depicted state fits the sentence event.

2.4.2 Visual Imagery as a Stress Test for Perceptual Simulation

Visual imagery diversity has been proposed as a useful test case for embodied theories of cognition (Lupyan et al., 2023), including language comprehension: if comprehension depends on perceptual simulation, then differences in imagery may reveal how those simulations are recruited (Muraki et al., 2023). There are several possible outcomes (Figure 2.5). Imagery vividness may not matter for the relevant process at all (Figure 2.5, left). It may matter quantitatively, such that weak and strong imagers rely on the same perceptual route but differ in the strength of simulation (Figure 2.5, middle). Or it may matter qualitatively, such that different imagery profiles rely on different routes to meaning, with some comprehenders relying more on perceptually grounded representations and others relying more on semantic, linguistic, or distributional knowledge (Figure 2.5, right).

Behavioral evidence has often seemed to support the first interpretation. Imagery vividness typically does not predict the size of property-verification or sentence-picture verification effects (Pecher, van Dantzig, & Schifferstein, 2009; Speed et al., 2026; Zwaan & Pecher, 2012). Similar null effects appear in visual working memory, where weak or absent imagery often does not impair overall accuracy or capacity (Jacobs et al., 2018; Knight et al., 2026; Pounder et al., 2022; Weber et al., 2024). However, comparable performance need not imply comparable representational format. Recent visual working memory studies suggest that weak imagers may rely on different strategies or representational formats, including non-visual spatial, verbal, semantic, or sensorimotor strategies (Keogh et al., 2021; Reeder, Pounder, et al., 2024). Visually encoded information may therefore be maintained as a detailed visual trace or as a coarser semantic/verbal code, yielding comparable performance but different costs under distraction or attentional strain (Chkhaidze, Gaspariani, et al., 2026).

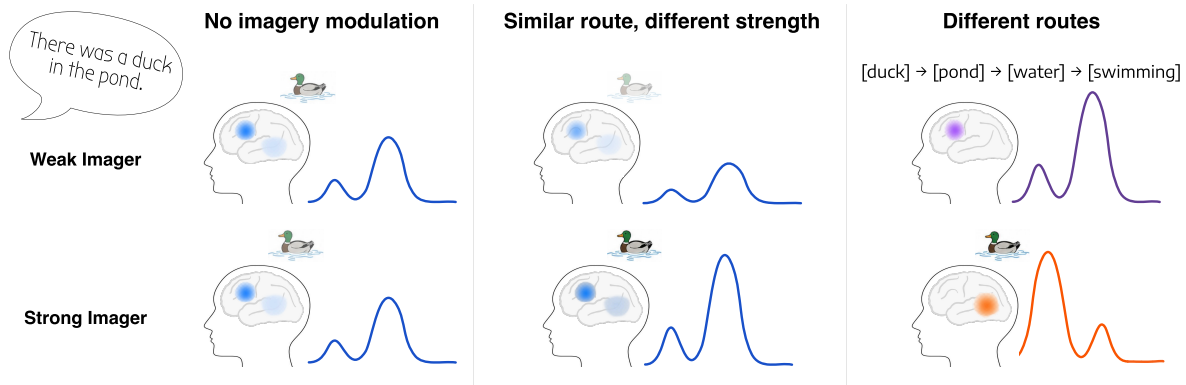


Figure 2.5. Alternative accounts of how visual imagery may relate to perceptual simulation during language comprehension. Imagery vividness may have no effect, modulate the strength of a shared perceptual-simulation route, or support different routes to the same behavioral judgment.

Our results are consistent with the idea that similar behavioral outcomes can arise from different processing routes. Visual imagery did not predict overt verification performance, but it did modulate the early posterior shape-match response, and this modulation was not observed for verbal thinking tendency. In strong visual imagers, the mismatch effect showed the same right-posterior distribution as the group-level N170-window match effect, but was more pronounced; in weak imagers, this early posterior pattern was largely absent. This suggests that the group-level early match effect was driven primarily by participants with stronger visual imagery. By contrast, the later N400-window match effect did not vary with imagery vividness, suggesting that weak imagers may still distinguish matched from mismatched pictures, but through a later, more semantic route rather than the early perceptual route evident in strong imagers. For instance, a reader may construct a visually specific representation of a duck swimming in a pond, but may also reach the same expectation through associations among duck, pond, water, and swimming (Figure 2.5, right). The relevant question may therefore be when perceptually specific simulation is recruited, for whom, and for which visual properties.

2.4.3 Limitations and Future Directions

Our ERP analyses were picture-locked, so they show how prior sentence context shaped processing of a subsequent picture, but not how different comprehenders encoded or represented

the sentence meaning as it unfolded. Future work could examine ERPs time-locked to sentence-final words to test whether imagery vividness modulates language processing before the picture appears. Second, we infer a representational route from ERP patterns rather than directly measuring strategy. Future studies could combine ERPs with post-task strategy reports, trial-level imagery ratings, or manipulations designed to bias participants toward visual versus verbal/semantic processing. Further, scalp topographies characterize the distribution of ERP effects but cannot unambiguously identify their neural generators. Converging evidence from MEG, fMRI, or source modeling would help determine which neural systems support the early posterior shape-match effect and its modulation by visual imagery. Finally, our imagery analyses were based on continuous VVIQ scores; the weak and strong imagery plots illustrate this continuous modulation rather than providing inferential evidence for discrete group differences. Larger samples with more participants at the aphantasic and hyperphantasic extremes would allow stronger tests of how representational flexibility varies across the imagery spectrum.

2.4.4 Concluding Remarks

Visual meaning implied by language may be represented in more than one format. The same sentence-picture judgment may be supported by a perceptually specific simulation of the implied object state, a coarser semantic expectation, or some combination of the two. Our findings suggest that the relative importance of each route may depend, in part, on the comprehender's visual imagery vividness.

Acknowledgement

Chapter 2, in full, is currently being prepared for submission for publication of the material as Visual imagery modulates neural, but not behavioral, signatures of mental simulation during language comprehension. Chkhaidze, A., Boroditsky, L., Trott, S., Coulson, S., & Kiyonaga, A. The dissertation author was the primary investigator and author of this material.

Chapter 3

From Dots to Faces: Individual Differences in Visual Imagery Capacity Predict the Content of Ganzflicker-induced Hallucinations

3.1 Introduction

In 1819, the physiologist Jan Purkinje made a striking observation: while waving his hand between his closed eyes and sunlight, he perceived “beautiful regular figures that are initially difficult to define but slowly become clearer” (Purkinje, 1819). His drawings of these experiences—geometric lattices, radiating patterns, and spiraling forms—represent the first scientific documentation of what we now call flicker-induced visual hallucinations: visual percepts that arise not from external stimuli but from the brain’s response to rhythmic light stimulation without pharmacological intervention (Hewitt et al., 2025). Flicker stimulation has become a widely used experimental method for non-invasively and non-pharmacologically inducing altered states of consciousness (ASCs): transient shifts in subjective experience that differ qualitatively from ordinary waking awareness (Fort et al., 2025). From Purkinje’s early introspections to modern laboratory studies (Figure 3.1), the experienced visuals in flicker paradigms have been so consistent that they have been codified as “form constants,” a universal vocabulary for simple visual hallucinations (Rule et al., 2011). This consistency suggests that, rather than transient perceptual noise, flicker-induced hallucinations reflect the intrinsic architecture of the human

visual system (Bressloff et al., 2001, 2002). As such, they offer a unique experimental window into other forms of internally generated visual experience, including visual mental imagery, a fundamental yet poorly understood cognitive trait.

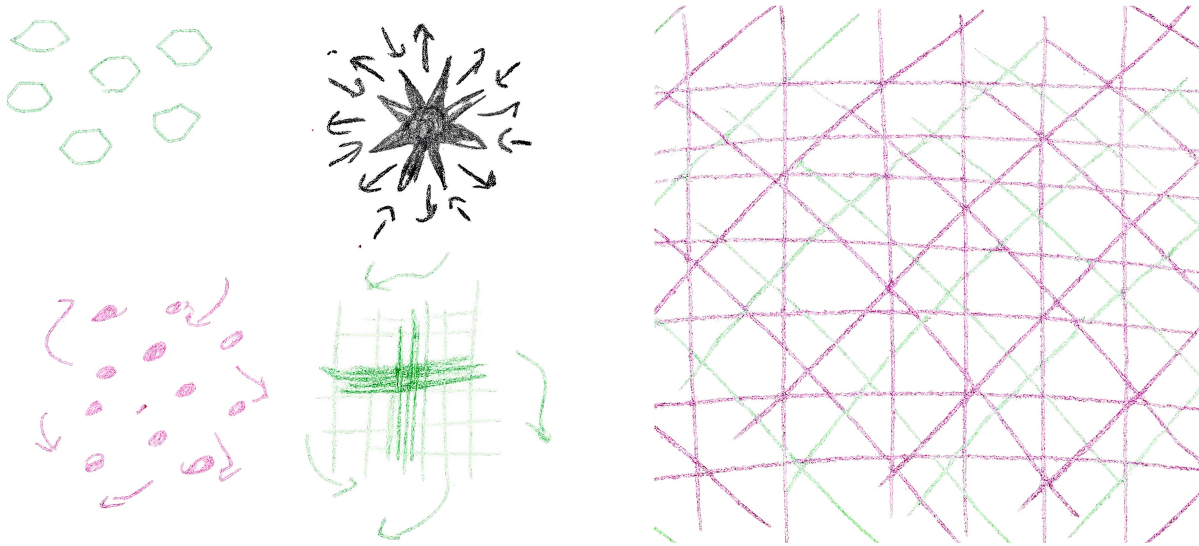


Figure 3.1. Ganzflicker-induced hallucination drawings from a pilot participant in our lab.

Verbal report: “Initially sort of a grid-like pattern, but with certain parts being more emphasized than others, and moving around across the screen. Then at some point sort of an array of pulsating dots, also moving around the screen. Then something like starbursts. Occasionally the grid-like pattern would return, and during some periods there was just a sort of non-descript dynamic patterns moving around quickly. At one point more toward the end, I was seeing something similar to chemical molecular structures. All of these were in constant motion and fluctuation.”

Mental imagery, the ability to internally simulate sensory experiences, is one of the most familiar phenomenological aspects of human experience. Argued to play a central role in core cognitive and affective functions (Pearson, 2019), visual mental imagery has been shown to relate to working memory (Keogh & Pearson, 2014), mind wandering (Christian et al., 2013), navigation (C. M. Bird et al., 2012), and language comprehension (Muraki et al., 2023). However, visual mental imagery capacity varies dramatically across individuals along what we term the *visual imagery spectrum*, referred to below simply as the imagery spectrum, ranging from *aphantasia*, or the complete absence of visual imagery, to *hyperphantasia*, or strong, lifelike mental pictures.

Such diversity raises fundamental questions about cognitive universality: if people differ in how they internally represent sensory information, then the underlying mechanisms of the cognitive phenomena that draw on these representations may also differ systematically. Indeed, recent work demonstrates that imagery differences can predict memory recall (Bainbridge et al., 2021), emotional reactivity (Wicken et al., 2021), and working memory strategies (Reeder, 2024), suggesting that understanding individual differences across the imagery spectrum may be crucial for developing more inclusive theories of human cognition (Lupyan et al., 2023). While research has extensively documented that imagery vividness varies across individuals (Zeman, 2024), we know remarkably little about what people with different imagery phenotypes *see*—that is, the content and structure of these internally generated visual experiences. Here, we used tools from natural language processing to analyze participants’ descriptions of their visual hallucinations, asking whether people across the imagery spectrum see different things in their mind’s eye.

3.1.1 The Layered Model of Imagery

Historically, imagery research has predominantly focused on scalar measures of vividness, such as the Vividness of Visual Imagery Questionnaire (VVIQ; Marks, 1973), rather than the compositional elements that constitute mental imagery, such as imagery manipulation and precision (Kosslyn et al., 1984). Initial neuroimaging studies of the imagery spectrum focused on early visual cortex, revealing that imagery vividness and precision correlate with the pattern and extent of activation in V1 during imagery tasks (Cui et al., 2007; Lee et al., 2012; Albers et al., 2013; Bergmann et al., 2016). However, more recent work has extended this investigation to the entire brain, confirming differences in early visual cortex when decoding imagined content (Cabbai et al., 2024), and revealing activation differences in the insula (Cabbai et al., 2024; Silvanto & Nagai, 2025), frontoparietal control networks (Spagna et al., 2021), and the fusiform gyrus (Liu et al., 2025), reflecting the distributed, multi-level nature of mental imagery processes.

Importantly, this distributed activity may translate into individual variation not just in imagery vividness, but in the content and structural complexity of what people can internally represent. While

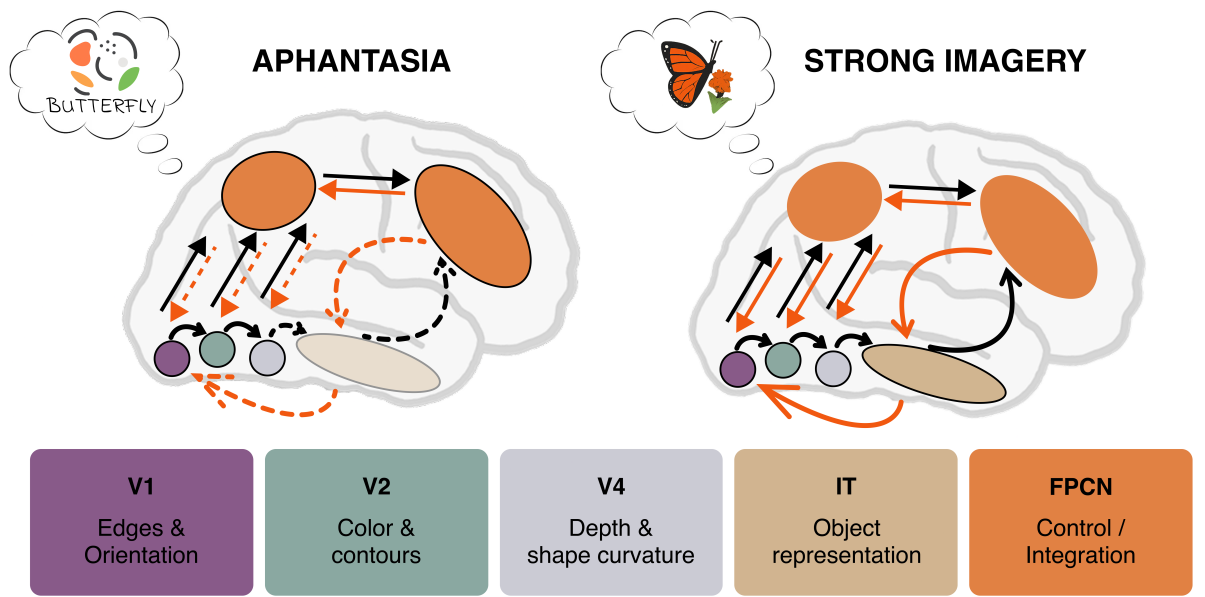


Figure 3.2. Layered model of visual imagery across the visual hierarchy. Proposed differences in connectivity across the visual hierarchy during internally generated visual experiences in individuals with aphantasia versus strong visual imagery. Black arrows indicate bottom–up feedforward connections; orange arrows indicate top–down modulatory connections. Solid lines represent strong connectivity; dashed lines represent reduced connectivity. Both individuals with aphantasia and individuals with moderate to strong imagery show intact bottom–up flow from early visual cortex (EVC: V1, V2, V4) through parietal to frontal regions, which may explain why people with no self-reported imagery can still experience simple visuals during Ganzflicker-induced hallucinations (Königsmark et al., 2021; Reeder, 2022). The key difference lies in the efficiency of coordinated activity across the hierarchy. In strong imagers, strong bidirectional connectivity between prefrontal areas and inferotemporal (IT) cortex, together with robust top–down modulation of visual cortical activity via the frontoparietal control network (FPCN), may enable integration of low-level features into coherent, semantically meaningful imagery. In aphantasia, these connections may be weaker, limiting conscious imagery to simpler visual content.

individuals across the imagery spectrum may similarly represent low-level visual features, such as lines, edges, and spatial patterns, those with weaker imagery or aphantasia may have reduced capacity for integrating these elements into complex, semantically meaningful images due to differences in top–down connectivity between control networks and visual areas during imagery (Milton et al., 2021; Liu et al., 2025). For example, when imagining a butterfly landing on a flower, early visual cortex might encode basic features like curved edges, color patches, and black lines across all individuals.

However, the successful construction of a vivid butterfly image may require higher-order semantic and object-processing areas to integrate these features into meaningful shapes, along with the recruitment of color-selective areas to contribute the rich hues in the scene, as well as coordination via frontoparietal control networks to bind these elements into a unified percept (Mechelli et al., 2004; Dijkstra et al., 2019; Spagna et al., 2021). This coordination may be compromised in weaker imagers, resulting in fragmented or impoverished butterfly imagery despite intact basic feature representations. We refer to this framework as the *layered model of visual imagery*, in which individuals across the imagery spectrum similarly represent low-level visual features in early visual cortex, but differ in their capacity to integrate these elements into complex, semantically meaningful representations (for details, see Figure 3.2).

3.1.2 Individual Differences in Imagery and the Induction of Visual Hallucinations

If individual differences in visual imagery reflect variation in the hierarchical coordination of visual representations, then these differences should manifest not only during voluntary imagery but also in other situations where internally generated content shapes perception, such as during flicker-induced hallucinations (Reeder, 2024). Since Purkinje’s pioneering forays, experimental methods to induce visual hallucinations have evolved, systematized, and diversified. Visual hallucinations can be successfully induced through both sensory deprivation (e.g. blindfolding; Merabet et al., 2004) and sensory stimulation (e.g. stroboscopic light stimulation; Hewitt et al., 2025). A combination of both, the “Ganzfeld” paradigm involves presenting participants with a uniform colored field, inducing hallucinations as the brain generates structure in the absence of patterned input (Metzger, 1929; Pütz et al., 2006; Wackermann et al., 2008; Schmidt & Prein, 2019; Schmidt et al., 2020). Stroboscopic light stimulation by contrast, delivers bright rhythmic stimulation, typically white light to closed eyes, producing hallucinations possibly through neural entrainment in the visual cortex (as reviewed in Hewitt et al., 2025). The “Ganzflicker”—a special case of the flicker-based paradigm—combines elements of both approaches: like stroboscopic light stimulation, it uses rhythmic flicker to drive visual experience; like Ganzfeld, it employs uniform chromatic stimulation, though unlike previous

suggestions, this does not have to fill the entire visual field to elicit hallucinations (Reeder, 2022). Specifically, the term refers to open-eye viewing of two uniform chromatic fields (typically red and black) alternating rhythmically (Königsmark et al., 2021). All these induction methods can successfully lead to ASCs (Fort et al., 2025), though the phenomenological character of the resulting experiences may vary (Shenyan et al., 2024; Fort et al., 2025).

Simple hallucinations—geometric shapes, color alterations, and elementary shapes—are the most commonly reported experiences in these light-induced hallucinations, across both Ganzfeld and flicker-based paradigms (Pütz et al., 2006; Wackermann et al., 2008; Allefeld et al., 2011; Rule et al., 2011; Schmidt & Prein, 2019; Schmidt et al., 2020; Bartossek et al., 2021; Hewitt et al., 2025; Pistolas & Wagemans, 2025). These experiences have been suggested to reflect the intrinsic dynamics of early visual cortex as computational models demonstrate that conditions of either rhythmic stimulation or reduced sensory input result in these patterned representations due to lateral connectivity and retinotopic organization in V1 (Ermentrout & Cowan, 1979; Bressloff et al., 2001, 2002).

Despite the predominance of simple hallucinations, complex hallucinatory content has also been reported in both Ganzfeld (Pütz et al., 2006; Shenyan et al., 2024) and flicker-induced paradigms (Allefeld et al., 2011; Schwartzman et al., 2019; Königsmark et al., 2021; Reeder, 2022; Shenyan et al., 2024) (Figure 3.3). Complex hallucinations involve recognizable semantic content, like faces, human figures, animals, objects, landscapes, and structured scenes. Notably, the likelihood of experiencing complex versus simple hallucinations appears to depend in part on the induction method. Shenyan et al. (2024) found that complex hallucinations were relatively more likely during Ganzfeld than Ganzflicker, suggesting that bottom-up rhythmic stimulation of the early visual cortex may not be sufficient to generate complex content. Instead, complex hallucinations may require additional involvement of higher-level visual areas—either through feedforward propagation up the visual hierarchy or through top-down interpretative processes (see e.g. Shenyan et al., 2024). Consistent with a role for top-down processing, (Pistolas & Wagemans, 2025) documented complex imagery (fish, sharks, ocean scenes) that appeared to be shaped by contextual cues such as illumination color and ambient sound. However, only a portion (20%) of participants were susceptible to such context-driven influences. Complex

hallucinations also occur in flicker paradigms: some participants undergoing stroboscopic light stimulation have reported semantically rich scenes such as wheat fields, faces of loved ones, and birds flying into space (Beauté et al., 2026). Thus, hallucination complexity appears to depend on more than just the induction method and may also vary as a function of stable individual traits such as visual imagery capacity (Beauté et al., 2026).



Figure 3.3. Ganzflicker-induced simple and complex hallucinations. Drawings by an artist based on her firsthand Ganzflicker experiences, illustrating both simple and complex visual hallucinations, including colored geometric patterns, a tunnel, and human figures. Reproduced with permission; drawings © Kirsten Baron.

Indeed, in recent years visual imagery capacity has emerged as a reliable predictor of flicker-induced hallucinations. Large-scale studies using the Ganzflicker paradigm show that individuals with stronger imagery report more frequent and more vivid hallucinations during Ganzflicker stimulation (Königsmark et al., 2021; Reeder, 2022). If complex hallucinations depend on layered mechanisms similar to those supporting voluntary visual imagery, then individuals with stronger imagery capacity should also be more likely to experience complex content. When asked to categorize their hallucinations using fixed response options, participants with strong imagery were more likely to report complex

hallucinations than those with weak imagery (Reeder, 2022).

While these studies established the first empirical link between visual imagery capacity and Ganzflicker-induced hallucinations, closed-ended response formats cannot fully reveal the experiential content of hallucinations. The layered model of imagery offers a testable prediction about hallucination content across the imagery spectrum: individuals with stronger imagery should report more complex, naturalistic hallucination content, while those with weaker imagery should predominantly report the simple geometric patterns linked to activity in the early visual cortex. In other words, the same processing differences that distinguish imagery phenotypes should be visible in the content of what people “see” during Ganzflicker—a prediction that can be tested by analyzing the semantic structure of open-ended phenomenological reports.

3.1.3 The Present Study

To characterize variability in hallucination content across the imagery spectrum, we conducted a secondary analysis of an open dataset containing > 6,000 participants’ free-text descriptions of their experiences following Ganzflicker stimulation (Reeder, 2022). After the Ganzflicker exposure, participants provided self-reported visual imagery vividness ratings to determine their position on the imagery spectrum, and answered structured questions about their experiences during Ganzflicker (Figure 3.4a). While Reeder (2022) focused primarily on these structured questionnaire data, participants also provided rich narrative reports describing their visual hallucinations (see Supplemental Material S1 for example descriptions), a corpus that has not previously been analyzed for content differences across the imagery spectrum. Analyzing the content and structure of these free-text descriptions requires a methodological approach that can accommodate the variability and richness of open-ended phenomenological reports while enabling systematic, replicable analysis. To address this challenge, we employed two complementary methods (Figure 3.4c).

First, we aimed to uncover the phenomenological content of Ganzflicker-induced hallucinations, asking whether it changes across the imagery spectrum. To locate latent themes in participants’ descriptions, we used Mapping of Subjective Accounts into Interpreted Clusters, a topic modeling

pipeline that relies on machine learning to identify coherent themes in text corpora, allowing data-driven discovery of hallucination content (Beauté et al., 2026). Following prior literature (Reeder, 2022; Hewitt et al., 2025), we expected a range of simple and complex themes to emerge. We hypothesized that participants with stronger imagery would describe more complex, naturalistic content (faces, scenes, structured hallucinations), while those with weaker imagery would report simpler, geometric patterns, arising from the individual differences in brain organization posited in the layered model of imagery.

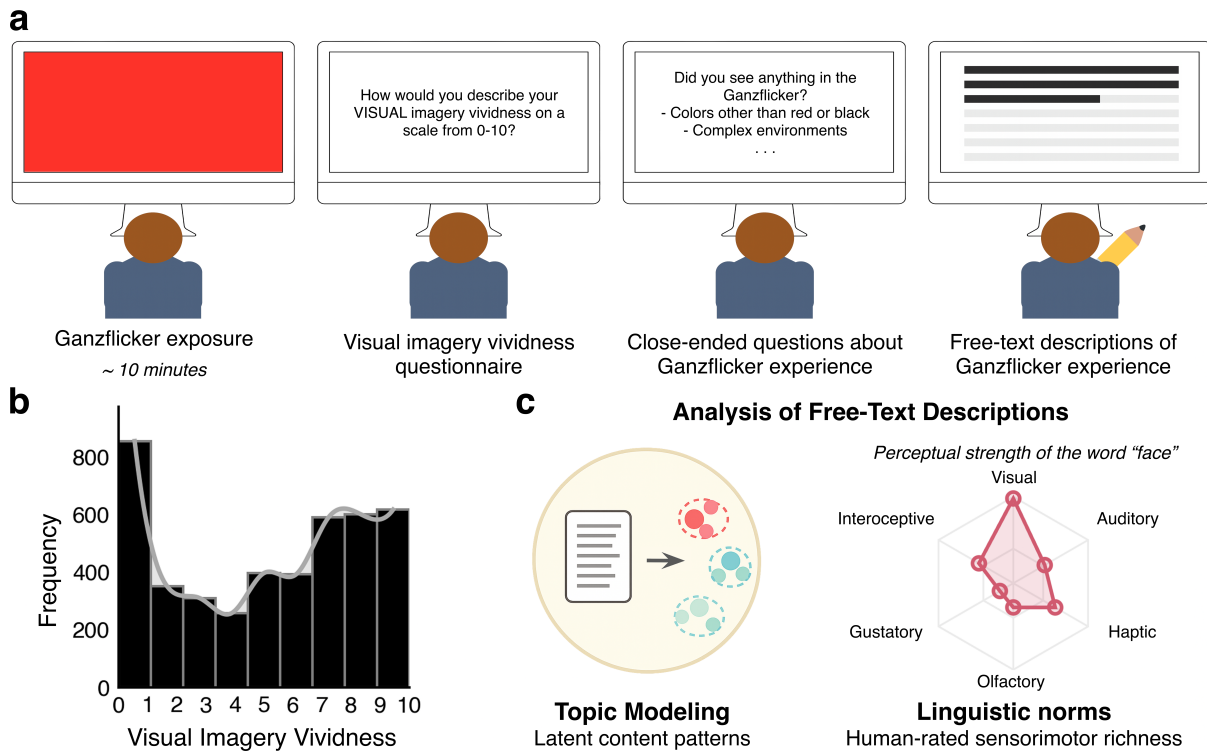


Figure 3.4. Ganzflicker task and analytical approach. (a) Experimental procedure from Reeder (2022). Participants viewed red–black flicker for approximately 10 minutes, completed a demographics questionnaire, and rated their visual imagery vividness on a 0–10 scale. Afterwards, they answered whether they saw anything during Ganzflicker, provided free-text descriptions of what they saw, and completed additional closed-ended questions about their experience. (b) Distribution of visual imagery vividness ratings across participants in this study. Bars show participant counts at each discrete vividness rating; the curve shows the overall shape of the distribution. (c) Analysis pipeline: topic modeling was used to identify semantic themes, and human-rated linguistic norms were used to quantify sensorimotor content in hallucination descriptions.

Second, we used the Lancaster Sensorimotor Norms (Lynott et al., 2020) to assess explicit sensorimotor language in participants’ descriptions. These norms provide crowd-sourced ratings of how strongly words evoke different sensory and motor experiences—for example, rating “thunder” as highly auditory, “velvet” as highly tactile, “kick” as foot-related, or “clap” as hand-related. While topic modeling may reveal thematic content at the level of whole descriptions, the Lancaster norms complement this analysis by indexing explicit sensorimotor vocabulary at the word level. Moreover, as they are derived from human ratings, they offer a psychologically grounded measure with established links to behavioral indices of word processing, including word naming and the lexical decision task (Connell & Lynott, 2012), as well as neural responses to words in the property verification task (Vinaya et al., 2025). Crucially, these norms allowed us to identify which sensory and motor dimensions emerge in Ganzflicker-induced hallucinations and test whether they vary systematically across the imagery spectrum. We expected participants with higher imagery vividness scores to use language with stronger sensorimotor associations, reflecting richer internal simulation.

3.2 Materials and Methods

3.2.1 Participants

We used a dataset from a Ganzflicker study (Reeder, 2022) with 6,664 individuals who underwent 10 minutes of continuous Ganzflicker stimulation and reported their experiences, alongside visual imagery vividness scores. Participants were recruited via a popular science article published in *The Conversation* describing the Ganzflicker phenomenon (Reeder, 2021), which was republished by at least 14 media outlets, including *Science Alert*, *The Daily Beast*, and *Big Think*, all of which contained a link to the Ganzflicker study hosted on Google Forms. This widespread media coverage led to a large, international volunteer sample, with the majority of readers based in the USA (73%), followed by the UK (5.2%), Canada (4.1%), and Australia (3.3%). A total of 68.1% of participants reported their gender as male, 27.6% as female, 2.4% as other, and 1.9% preferred not to say. Age was reported by 5,707 individuals ($M = 38.07$, $SD = 15.08$). The sample also included a higher-than-typical proportion

of individuals at the extremes of the imagery spectrum, including self-reported aphantasia (Figure 3.4b). We excluded participants whose descriptions were missing or who provided written descriptions in a language other than English, leaving 4,365 participants. As reported by Reeder (2024), the study was approved by the ethics committee of the medical faculty of Otto-von-Guericke University, and all participants gave informed consent.

3.2.2 Procedure

Participants viewed a Ganzflicker stimulus that alternated between red and black at 7.5 Hz for approximately 10 minutes. They were instructed to watch the Ganzflicker stimulus in full-screen mode and under dim lighting, accompanied by white noise to homogenize the acoustic environment, and to avoid distractions. Approximately 73% of participants reported compliance with the white-noise recommendation; however, the proportion of complex versus simple hallucinations did not differ between those who used white noise and those who did not (Königsmark et al., 2021; Reeder, 2022), suggesting that auditory masking did not substantially influence hallucination content. Further details on environment-control instructions are reported in Reeder (2022).

After the stimulus presentation, participants completed a demographic survey. To situate them on the imagery spectrum, they rated their visual imagery vividness using a single item: “How would you describe your VISUAL imagery vividness on a scale from 0 (no mental imagery) to 10 (as vivid as real perception)?” Participants also answered closed-ended questions assessing their Ganzflicker experiences and provided an open-ended free-text description of what they saw during exposure. In this chapter, we focus on these free-text descriptions.

3.2.3 Topic Modeling

To uncover latent themes in participants’ descriptions of Ganzflicker-induced hallucinations, we implemented a multi-step topic modeling pipeline based on BERTopic (Grootendorst, 2022), an unsupervised method that integrates transformer-based embeddings, dimensionality reduction, density-based clustering, and class-based term weighting (Figure 3.5). After clustering, each topic’s most

informative terms were extracted using class-based Term Frequency–Inverse Document Frequency (c-TF-IDF), which identifies words that best characterize a given cluster relative to the rest of the corpus. These keyword sets were used as the primary basis for human-interpretable topic labels. To further reduce researcher expectancy bias, we also employed GPT-4o-mini to generate independent topic labels based on the same representative keywords and exemplar sentences. Human- and model-generated labels showed close alignment (see Table 3.1). A more explicit description of the pipeline is detailed in Figure 3.5.

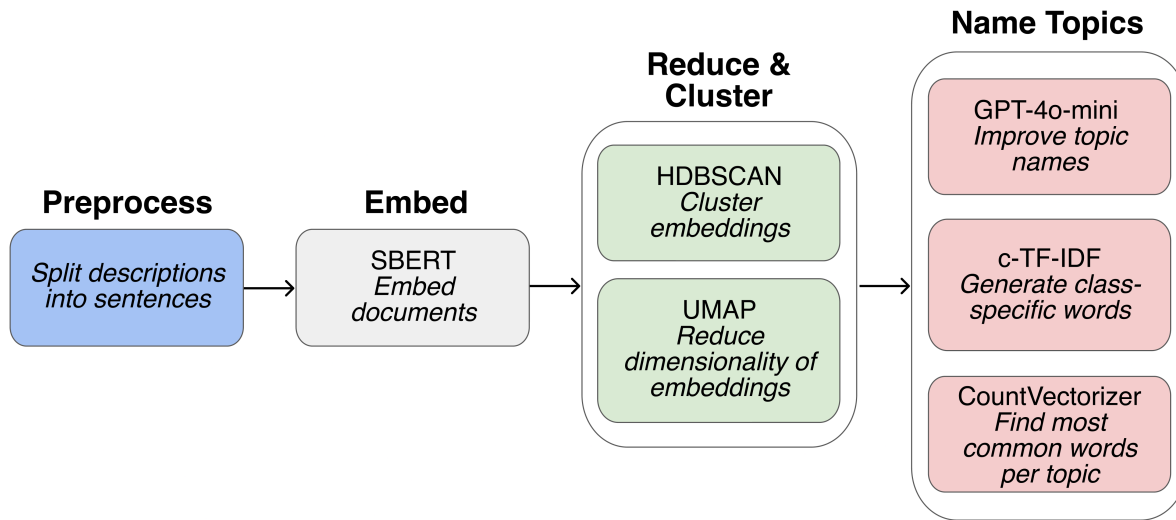


Figure 3.5. Topic modeling and analysis pipeline. Overview of the analytic pipeline used to characterize Ganzflicker-induced hallucination descriptions. Free-text descriptions were embedded using transformer-based language models, reduced in dimensionality, clustered into topics, and summarized using class-based Term Frequency–Inverse Document Frequency (c-TF-IDF). Topic labels were generated from representative keywords and exemplar descriptions, with GPT-4o-mini used to provide independent topic-label suggestions. Downstream analyses examined how topic expression and sensorimotor content varied across the visual imagery spectrum.

Table 3.1. Representative hallucination descriptions by topic.

Human labels	Model labels	Representative descriptions
Experiential descriptions	Trippy experience	“It was like watching a trippy video of colours and sounds”
Hallucination absence	Minimal content	“I didn’t see anything substantial though I thought I would”
Noise	White noise	“The white noise and screen together gave me an uneasy feeling but nothing noteworthy just as if this was a horror movie this would be the part where someone dies feeling”
Faces and figures	Human faces	“A potato that turned into an old man’s face and then some alien scenery that was pulsing”
Unpleasant images	Subjective experience	“It was not an enjoyable experience but i felt no overwhelming discomfort else i would have stopped the experiment”
Eyes	Eye patterns	“After several minutes i saw a left eye sketch-like which rotated and disappeared then reappeared several times”
Sensation duration	Duration events	“Images only lasted for a few seconds each”
Images	Images movement	“After a few seconds of this I usually would have to blink and the images would be gone”
Gaze	Eye variations	“As soon as I tried to look at them or focus on them they would disappear”
Webs	Spider webs	“patterns often in the form of spider webs”
Dots	Dots Pattern	“Tiny dots of other colours as part of the pattern like smaller than pixels and so many that they almost made up other colours”
Colors	Colors & Patterns	“The black ended up shifting between a greenish black and deep blue”
Movement	Visual distortion	“Mostly just different colors bouncing all around the screen in different shapes”
Flashes	Flashing lights	“Blue pink yellow flashes but mostly a continuous pink flash”

Table 3.1. Representative hallucination descriptions by topic (continued).

Human labels	Model labels	Representative descriptions
Lines	Lines & patterns	“Thick horizontal and vertical lines with fuzzy edges that intersected at the center of my field of view”
Patterns & shapes	Geometric Patterns	“I saw a bunch of shapes like squares circles lines”
Fractals	Fractals & patterns	“Like a kaleidoscope but much more faint and ever changing”
Spirals & rotation	Spiral rotation	“Everything constantly spinning”
Stars & galaxies	Stars & galaxies	“I saw purple dots that were circling around like stars in a galaxy also with spiral arms etc.”
Flowers	Flower imagery	“Perceived i was looking at a field of tulips and daisies and flowers as if through a crumpling red fold of cellophane”
Fire	Fire imagery	“I saw a starburst firework style shape occasionally and a couple of little sparks like a sparkler”
City skyline	Cityscape views	“Reminded me of looking at a bright light with my eyes closed”
Forest & trees	Forest scenery	“On the far edge and wrapping around the shore was a tree line that consisted of birch like trees tall and thinner with plenty of canopy but were not birches and some juniper trees it was accompanied with average pond foliage like cattails and reeds and such”
Water	Ocean reflections	“It wasn’t realistically colored or anything just felt like I was standing on the edge of a lake looking at something in the middle of it”
Space	Space travel	“Flying through a space station”
Tunnels	Tunnel experience	“A circle would appear then it turned into a long tunnel”
Hallways & doors	Hallway passage	“Also a very long dark corridor with light at the end it was moving around on the vertical axis”
Butterflies	Butterfly shape	“First was like a butterfly outline with neon lines”

Sentence-level tokenization and preprocessing. Each hallucination report was first split into sentences using `sent_tokenize()` from the Natural Language Toolkit (NLTK) (S. Bird et al., 2009), which applies the *Punkt* algorithm (Kiss & Strunk, 2006), an unsupervised model that uses punctuation and contextual cues to identify sentence boundaries. This step allowed us to capture distinct experiential elements that often co-occur within a single report. Treating entire descriptions as a single document would have conflated diverse perceptual content under one topic. The resulting sentence-level documents were lightly cleaned, converted to lowercase and stripped of excess whitespace, but no stemming, lemmatization, or stopword removal was performed. This minimal preprocessing ensured the preservation of contextual information important for semantic modeling, resulting in a corpus of $> 10,000$ sentence-level documents.

Sentence embeddings and dimensionality reduction. We generated semantic embeddings using Sentence-BERT (Reimers & Gurevych, 2019), which produces 384-dimensional vector representations optimized for capturing semantic similarity in short texts. This embedding process preserves subtle differences in meaning and context, allowing the model to distinguish between similar phrases with different phenomenological implications, such as “I felt time slow down” versus “I lost track of time.”

To prepare the embeddings for clustering, we applied Uniform Manifold Approximation and Projection (UMAP) to reduce their dimensionality (McInnes et al., 2018). UMAP preserves both local and global structure in the data, helping maintain the semantic relationships between sentences in the lower-dimensional space. We configured the reduced dimensionality (`n_components = 10`), the neighborhood size used to balance local versus global structure (`n_neighbors = 15`), and the minimum distance between embedded points (`min_dist = 0.1`) to yield an embedding structure that supported coherent and interpretable topic clustering.

Clustering with hierarchical density-based spatial clustering of applications with noise. The reduced sentence embeddings were clustered using hierarchical density-based spatial clustering of applications with noise (HDBSCAN) (Campello et al., 2013; McInnes et al., 2017), a non-parametric algorithm that identifies clusters of varying density without requiring a predefined number of clusters.

This feature is particularly advantageous when analyzing naturalistic text, where the number and size of coherent experiential categories are unknown. We set the minimum cluster size to 30, allowing the model to discover a diverse but stable set of themes while treating small or noisy clusters as outliers. Sentences not confidently assigned to any topic were labeled as outliers (topic = -1) and excluded from topic-based analyses to ensure the coherence and interpretability of topic clusters. A total of 32 sentences ($< 0.3\%$ of the corpus) were labeled as outliers; these were short Turkish-language descriptions that had not been flagged during automated language detection and were excluded from topic analyses. The final model identified 28 topics spanning a wide range of visual and conceptual content, ranging from simple geometric shapes to complex scenery (Figure 3.6; Table 3.1).

Topic extraction and coherence evaluation. To identify the most distinctive terms characterizing each topic, we applied c-TF-IDF (Grootendorst, 2022). This weighting scheme highlights words that are frequent within a topic and infrequent across other topics. Unlike standard TF-IDF, which considers individual documents, c-TF-IDF treats all documents within a topic as a single pooled entity, allowing for robust extraction of topic-defining terms. Specifically, our pipeline employed the BM25+ variant of c-TF-IDF, which provides additional smoothing and has been shown to improve performance on sparse or short-text data.

To evaluate the interpretability of topics, we computed the C_v coherence score (Röder et al., 2015), a widely used metric that assesses the degree of semantic similarity among the top-ranked words within each topic. C_v combines normalized pointwise mutual information, cosine similarity, and a sliding-window approach to assess word co-occurrence. Our model achieved a coherence score of 0.48, which falls within the commonly reported range of 0.4 to 0.7 in topic modeling studies involving short, noisy, or experiential texts (O’Callaghan et al., 2015; Röder et al., 2015). Topic structure was further examined using a two-dimensional topic embedding map (Figure 3.6) and a hierarchical dendrogram (Figure C.1), which visually confirmed the separability and cohesion of extracted themes.

Topic labeling. To reduce researcher bias—that is, the tendency for manual labeling to be influenced by theoretical priors—and to enhance interpretability of the keywords, we used GPT-4o-mini for LLM-assisted topic labeling. As outlined by Beauté et al. (2026), this method offers a scalable and

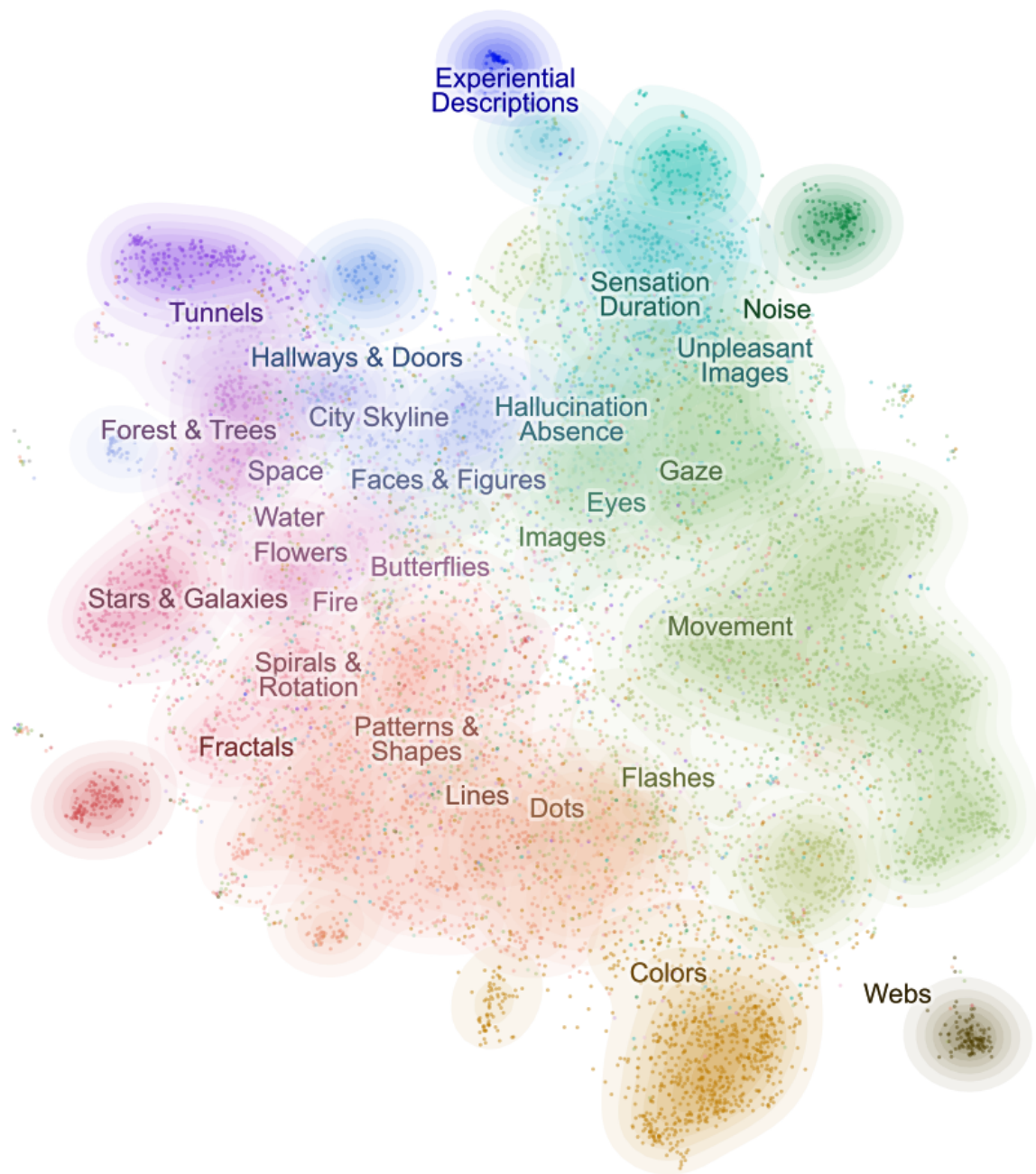


Figure 3.6. Topic representations and hierarchical relationships among hallucinations. UMAP projection of sentence embeddings colored by topic labels. Each point represents a sentence from a participant’s hallucination description embedded in a shared semantic space. Clusters correspond to semantically coherent hallucination types, ranging from simple geometric content to naturalistic scenes and more abstract descriptions.

relatively theory-neutral alternative to manual labeling for large phenomenological corpora. For each topic, we supplied the model with its top 10–15 c-TF-IDF keywords and a few representative sentences, prompting it to generate concise, human-readable topic labels. We also independently reviewed the keywords and representative example sentences for each topic and manually assigned labels. The two approaches produced highly consistent results, differing mainly in surface phrasing, such as “city skyline” versus “cityscape views.” Topic modeling figures (Figures 3.5, 3.6, C.1, C.2, and C.3) reflect the manually generated labels; Table 3.1 includes both human-generated and LLM-generated labels for comparison.

Participant-level topic features. Each sentence was assigned both a hard topic label and a soft probability distribution over all topics. We used BERTopic’s standard hard-label assignment, where the most probable topic for each sentence was determined as the $\arg \max$ of the topic-probability vector, $p(\text{topic} \mid \text{sentence})$. In addition, BERTopic provides the full topic-probability distribution for each sentence, indicating the likelihood that it belongs to each topic. To construct participant-level features, we aggregated these probabilities by identifying, for each participant $\times$ topic pair, the maximum probability that the model assigned to that topic across all of the participant’s sentences. This maximum-based aggregation preserves sensitivity to distinct, high-confidence topic expressions and prevents unrelated sentences from diluting the signal. The resulting participant-by-topic matrix contained continuous topic salience scores, where each cell reflected the strongest evidence that a given participant expressed that experiential theme, with scores near 1 indicating strong expression and scores near 0 indicating absence. These salience values indexed the presence and prominence of each topic at the participant level and served as inputs for subsequent regression and classification models testing whether topic structure predicted self-reported imagery vividness.

Predicting imagery vividness from hallucination content. To test whether the derived topic structure could distinguish individuals with different imagery phenotypes, we modeled the relationship between hallucination topics and participants’ vividness scores. We used Lasso regression with L_1 regularization (Tibshirani, 1996) to predict participants’ self-reported vividness scores based on their hallucination content. Lasso regression was selected here for its ability to identify sparse, stable

predictions in high-dimensional and potentially collinear feature spaces. Our analytic goal was not prediction per se, but interpretability: to identify robust topic–vividness associations that reflect the content structure of visual hallucinations across the imagery spectrum.

The predictor variables were the 28 z-scored topic probabilities obtained from the topic modeling analysis. The dataset was randomly split into 80% training and 20% testing subsets. The optimal regularization parameter (α) was selected via 10-fold cross-validation, testing 100 logarithmically spaced values between 0.001 and 10. All analyses were conducted in Python using `scikit-learn` (Pedregosa et al., 2011).

To further characterize hallucination content across distinct imagery phenotypes, we categorized participants into three imagery groups based on vividness scores: weak (0–3), moderate (4–7), and strong (8–10), following conventions in prior work (Zeman et al., 2020; Reeder, 2022). These bins approximate the distinction between aphantasia, typical imagery, and hyperphantasia, respectively. Final group sizes were: weak = 1,515, moderate = 1,634, and strong = 1,216. This approach allowed us to investigate which topics constituted the “fabric” of hallucination content in each imagery group and to visualize their distinctive content patterns.

We trained three separate Lasso-regularized binary classifiers to predict imagery group membership from hallucination content, represented as standardized topic probabilities. Each classifier was trained to distinguish one imagery group (weak, moderate, or strong) from all others using stratified train–test splits (80% training, 20% testing). We used the `LogisticRegressionCV` class from `scikit-learn` (Pedregosa et al., 2011) with the following parameters: `penalty = "l1"`, `solver = "saga"`, `scoring = "f1"`, `cv = 10`, `max_iter = 10000`, and `class_weight = "balanced"`. The regularization parameter C was optimized through 10-fold cross-validation across 30 logarithmically spaced values between 0.01 and 100. Class balancing was implemented during training by assigning higher weights to under-represented classes to ensure model sensitivity to signals in minority groups, even when overall accuracy remained limited.

Non-zero coefficients from Lasso-regularized classifiers represent the most robust content–vividness relationships that survived both regularization penalties and class-imbalance challenges. To

assess coefficient stability, we performed 1,000 bootstrapped model iterations for each imagery group using the `resample` function from `scikit-learn` (Pedregosa et al., 2011). In each iteration, we resampled the standardized feature matrix and target labels, fitted a logistic regression model with an L_1 penalty using the previously determined optimal regularization parameter, and recorded which features received non-zero coefficients. We retained only topics that survived regularization in at least 60% of iterations as our threshold for interpretability.

To assess model performance, we used the F1 score, the harmonic mean of precision and recall. F1 is a widely used evaluation metric in imbalanced classification settings where sensitivity to the minority class is important. This metric is particularly well suited to contexts where the goal is not maximizing raw accuracy, but rather understanding which features meaningfully distinguish categories while balancing false positives and false negatives (Saito & Rehmsmeier, 2015). To evaluate the statistical significance of model performance, we conducted permutation tests (Nichols & Holmes, 2002) for each classifier. In these tests, we randomly shuffled the imagery group labels, thereby breaking any systematic relationship between hallucination content and vividness category, and retrained the model 1,000 times on each shuffled dataset. This procedure generated a null distribution of F1 scores under the assumption that no true association exists. We then compared the F1 score from the original model to this null distribution to evaluate whether the model’s performance reflected real structure in the data.

3.2.4 Sensorimotor Content Analysis

To supplement our computational study with a human-judgment perspective, we conducted a complementary analysis examining the perceptual and embodied content of hallucination descriptions using the Lancaster Sensorimotor (LS) norms (Lynott et al., 2020), as described in Chkhaidze et al. (2023) (see Figure 3.8a for the pipeline).

Description preprocessing. We first preprocessed hallucination descriptions using standard natural language processing techniques in Python with NLTK (S. Bird et al., 2009). This involved: (i) tokenization, (ii) removal of punctuation and conversion to lowercase, (iii) removal of English

stopwords using the NLTK stopwords corpus, (iv) word lemmatization to reduce words to their base forms, and (v) spell correction using the `pyspellchecker` library (Barrus, 2025). Description length was calculated as the total number of tokens after preprocessing.

Quantifying descriptions using Lancaster Sensorimotor norms. The LS norms provide human-derived ratings for approximately 40,000 English words across 11 dimensions: six perceptual modalities (visual, auditory, gustatory, olfactory, haptic, and interoceptive) and five motor/action dimensions (foot, hand, head, mouth, and torso). For each word, human raters indicated how strongly they experience that concept through each sensory modality or through actions involving each body part. In addition to these individual modality ratings, the LS norms include two composite scores: *perceptual strength* and *action strength*. These reflect the maximum strength value assigned to the dominant modality for each word. For example, if a word is rated highest on the visual dimension, its perceptual strength is that visual rating. This approach is designed to capture the most salient sensory or motor modality through which a concept is typically experienced, and is theorized to reflect conceptual concreteness (Banks & Connell, 2022).

For each hallucination description, we matched preprocessed words to the LS database and calculated average scores across all six perceptual dimensions, all five motor dimensions, and the two composite measures. Descriptions with fewer than three valid LS norm-matched words, approximately 7% of the sample, were excluded, leaving 4,057 participants for this analysis.

Regression models. We asked three research questions. First, we tested whether the overall perceptual and action richness of hallucination descriptions predicted participants' self-reported imagery vividness. To do this, we fit a generalized linear model (GLM) predicting vividness from *perceptual strength* and *action strength*. The model was formulated as:

$$\text{Vividness}_i = \beta_0 + \beta_1 \cdot \text{PerceptualStrength}_i + \beta_2 \cdot \text{ActionStrength}_i + \epsilon_i$$

Second, to clarify which specific perceptual dimensions were driving the overall perceptual

strength effect, we fit a GLM including all six LS perceptual modality scores as simultaneous predictors:

$$\begin{aligned} \text{Vividness}_i = & \beta_0 + \beta_1 \cdot \text{Visual}_i + \beta_2 \cdot \text{Auditory}_i + \beta_3 \cdot \text{Gustatory}_i \\ & + \beta_4 \cdot \text{Olfactory}_i + \beta_5 \cdot \text{Haptic}_i + \beta_6 \cdot \text{Interoceptive}_i + \epsilon_i \end{aligned}$$

Lastly, we asked whether the overall action strength effect was similarly driven by any specific motor modality. We fit a separate GLM using the five motor dimension scores to isolate their relative contributions:

$$\begin{aligned} \text{Vividness}_i = & \beta_0 + \beta_1 \cdot \text{Head}_i + \beta_2 \cdot \text{Hand}_i + \beta_3 \cdot \text{Mouth}_i \\ & + \beta_4 \cdot \text{Foot}_i + \beta_5 \cdot \text{Torso}_i + \epsilon_i \end{aligned}$$

All predictors were z-scored prior to model fitting. GLM analyses were conducted in R using the `stats` package (Team, 2023).

Controlling for response length via mediation. Initial analysis revealed that description length was a significant positive predictor of vividness ratings ($\beta = .01$, $SE = .00$, $t = 6.80$, $p < .001$), indicating that participants with more vivid imagery provided longer descriptions. To account for this potential confounding influence, we conducted a series of causal mediation analyses using the `mediation` package in R (Tingley et al., 2014). Specifically, for each set of predictors—(i) composite perceptual and action strength, (ii) the six perceptual modality scores, and (iii) the five motor modality scores—we assessed whether the effect of sensorimotor richness on visual vividness was mediated by description length.

Each mediation analysis included two models: a mediator model predicting standardized description length from the sensorimotor predictor of interest, and an outcome model predicting visual vividness from both the sensorimotor predictor and standardized description length.

Mediator model, shown here for a single predictor:

$$\text{DescriptionLength}_i = \gamma_0 + \gamma_1 \cdot \text{SensorimotorPredictor}_i + \eta_i$$

Outcome model:

$$\text{Vividness}_i = \beta_0 + \beta_1 \cdot \text{SensorimotorPredictor}_i + \beta_2 \cdot \text{DescriptionLength}_i + \epsilon_i$$

We estimated the average direct effect (ADE), average causal mediation effect (ACME), total effect, and proportion mediated for each predictor using 5,000 nonparametric bootstrap simulations with percentile-based confidence intervals.

3.3 Results

3.3.1 Unsupervised Topic Modeling Identifies Systematic Themes Across the Imagery Spectrum

To map the latent content structure of participants' hallucination descriptions, we used topic modeling. This unsupervised approach identifies patterns of word co-occurrence and groups semantically similar descriptions together, allowing us to reveal the types of experiences participants reported without imposing predetermined categories (Blei et al., 2003). Applying this method to the corpus yielded 28 distinct experiential themes, each reflecting a coherent hallucination type, such as geometric patterns, faces, tunnels, and stars (Table 3.1).

The discovered topics encompassed a diverse range of hallucinations, from basic visual elements, such as lines, dots, and color flashes, to scenery, such as forests and skylines, and socially relevant objects, such as faces and figures (Figure 3.6; Table 3.1). Representative quotes from descriptions associated with each topic are shown in Table 3.1. Each row shows a free-text quote drawn from participant descriptions of their Ganzflicker-induced hallucinations, alongside the topic with which it was most strongly associated. For instance, the topic *spirals and rotation* included descriptions like “everything constantly spinning,” while the topic *space* was associated with reports like “flying through a space station.” These examples show that our unsupervised model successfully captured linguistically distinct yet conceptually similar experiences, ranging from motion and geometric percepts to naturalistic scenes and affective content.

Semantic structure in the two-dimensional UMAP embedding (McInnes et al., 2018) showed clearly delineated clusters, where spatial proximity reflects thematic similarity (Figure 3.6). Each point represents a sentence from a participant’s hallucination description, embedded into a shared semantic space and colored by its assigned topic. The visualization revealed several broad regions: geometric and motion-related content, including patterns and shapes, fractals, spirals and rotation, lines, dots, and movement, formed one area; naturalistic and environmental imagery, including forest and trees, water, flowers, space, city skylines, and tunnels, formed another; and metacognitive domains, including sensation duration and hallucination absence, appeared in a more separate region. Together, these analyses indicate that hallucination descriptions organize into coherent semantic domains with consistent relationships among topics. The resulting topic structure depicted in Figure 3.6 offers a data-driven foundation to examine whether hallucination content varies systematically across individuals with different imagery phenotypes.

We hypothesized that participants with stronger imagery would describe more complex, naturalistic hallucination content, whereas weaker imagers would report simpler visual distortions and geometric forms. We observed systematic content differences consistent with this prediction: strong imagers more often reported structured, semantically meaningful content, such as faces, hallways, and natural environments, while weak imagers tended to describe simpler features, such as geometric patterns, visual distortions, and flashes (Figure 3.7).

To test this quantitatively, we first modeled participants’ continuous visual imagery vividness scores using Lasso regression (Tibshirani, 1996). Given any particular hallucination description, the topic model output provides a probabilistic profile across all 28 topics, reflecting the estimated content composition of that report. We used these topic probabilities as features in our models to predict imagery scores. Topic probabilities explained approximately 5% of the variance in vividness scores ($R^2 = .05$). From the initial set of 28 topics, 25 were retained as non-zero predictors (see Figure C.2). Given the noisy nature of free-text reports and the multifaceted character of subjective experience, this effect reflects meaningful structure in the data despite its modest size. Because human psychology and behavior are inherently complex, small but reliable effects in psychological research are both expected

and interpretable (Funder & Ozer, 2019). Positive predictors included naturalistic or structured content—faces and figures ($\beta = 0.29$), hallways and doors ($\beta = 0.15$), stars ($\beta = 0.14$), city skylines ($\beta = 0.14$), and forest imagery ($\beta = 0.13$). Negative predictors consisted mainly of simple perceptual content, including movement ($\beta = -0.23$), flashes ($\beta = -0.17$), and lines ($\beta = -0.12$). These results indicate that richer, more naturalistic descriptions are associated with higher imagery vividness, while lower vividness is linked to simpler perceptual experiences.

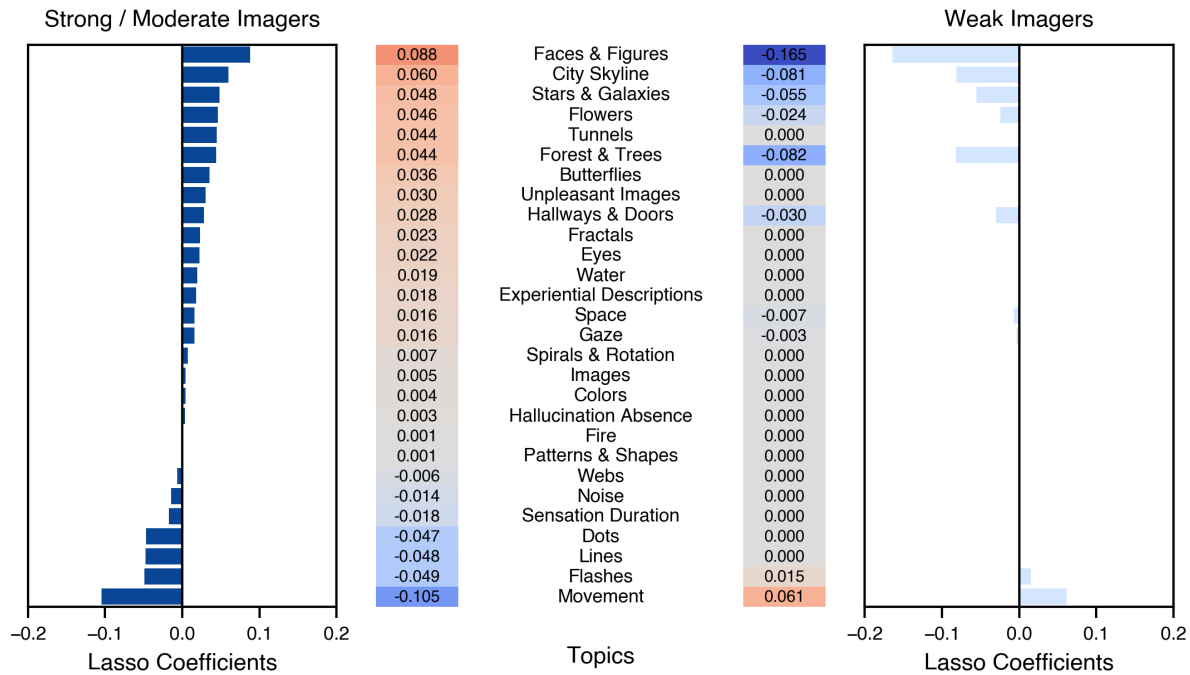


Figure 3.7. Hallucination content patterns distinguishing weak from moderate and strong imagery groups. Lasso regression coefficients averaged across the moderate and strong imagery groups and contrasted with the weak group. The plot shows only topics that survived regularization and appeared in at least 60% of 1,000 bootstrapped regressions. Colors indicate coefficient direction and stability, with positive values reflecting topics more prevalent in the moderate/strong imagery group and negative values reflecting topics more common in the weak group.

To characterize the hallucination content profiles of distinct imagery phenotypes, we categorized participants into three vividness groups: weak (0–3), moderate (4–7), and strong (8–10). This binning followed conventions in the imagery literature distinguishing aphantasia, typical imagery, and

hyperphantasia (Zeman, 2024). For each group, we computed the average topic probability across participants (Figure C.3). To assess whether these profiles were discriminable, we then trained Lasso-regularized classifiers to predict group membership from topic-probability features. We trained three separate one-vs.-rest classifiers, where each classifier learned to identify one imagery group versus all others, revealing which topics were most diagnostic of each imagery phenotype. The classifiers were evaluated using the F1 score, the harmonic mean of precision and recall, which is well suited to imbalanced classification tasks and helps assess how well models identify the minority class (Saito & Rehmsmeier, 2015). For statistical validation, we conducted permutation tests (Nichols & Holmes, 2002) in which group labels were randomly shuffled and each model was retrained 1,000 times. The weak-imagery classifier showed significant above-chance performance ($F1 = 0.54, p < .001$), indicating that those with the least vivid imagery can be distinguished from other participants based on their hallucination content. Classifiers for the moderate ($F1 = 0.45$) and strong groups ($F1 = 0.41$) did not exceed chance ($p > .50$), likely reflecting similarity between these two groups (Figure C.3).

To interpret classifier outputs, we examined bootstrapped Lasso coefficients for the topics and retained only those that appeared in at least 60% of 1,000 bootstrapped models (Figure 3.7). The weak-imagery profile was sparsely defined: only 8 of 28 topics were retained, mostly reflecting the absence of naturalistic content. Positive predictors for weak imagery included simple perceptual features, such as flashes and movement. In contrast, all 28 topics survived regularization and bootstrapping for the moderate and strong groups, and their strongest predictors overlapped with those from the continuous model. Notably, topics absent in weak-imagery descriptions, such as faces and natural scenes, were positive predictors of strong imagery, whereas topics characteristic of weak imagery were negative predictors of strong imagery (Figure 3.7).

Together, these findings suggest that free-text descriptions of hallucinations contain discriminative signals related to individual imagery phenotypes. The semantic content of these descriptions offers a high-dimensional, quantifiable window into how internal visual experiences differ across the imagery spectrum.

3.3.2 Human-Assessed Sensorimotor Content of Descriptions Varies With Imagery Vividness

We expected that participants with stronger imagery vividness would use language with stronger sensorimotor associations when describing their hallucinations, particularly in visual and motor dimensions, reflecting richer internal simulation.

To quantify this content, we obtained Lancaster scores for each word in participants' descriptions across six perceptual modalities, five motor dimensions, and two composite measures, then averaged these scores within each participant to create 13 participant-level sensorimotor features (see Figure 3.8a for the analysis pipeline). Across the dataset, visual strength was the most prominent perceptual modality, reflected in its relatively high average score ($M = 3.57$) and substantial individual variability (range = 3.87), followed by haptic ($M = 1.32$) and auditory ($M = 1.11$) dimensions. In the motor domain, head-related ($M = 2.66$) and hand-related ($M = 1.56$) concepts were most frequently represented. The composite measures of perceptual and action strength also showed high mean scores and variability, supporting their interpretability as global indices of sensorimotor richness. Full score distributions and descriptive statistics are reported in Table C.1 and Figure C.4.

We first modeled imagery vividness using composite perceptual and motor/action strength, controlling for description length. We entered the composite perceptual and motor/action strength scores from the Lancaster Sensorimotor norms into a GLM, along with standardized description length as a covariate. Both predictors were significant: higher perceptual strength was associated with greater imagery vividness ($\beta = 0.35$, $SE = 0.05$, $t = 6.54$, $p < .001$), as was higher motor strength ($\beta = 0.21$, $SE = 0.05$, $t = 4.17$, $p < .001$) (Figure 3.8b). Description length was also a significant positive predictor ($\beta = 0.42$, $SE = 0.05$, $t = 8.39$, $p < .001$), consistent with the idea that more vivid experiences elicit richer verbal responses.

To determine whether the effects of sensorimotor content were confounded or mediated by description length, we conducted causal mediation analyses with 5,000 bootstrap simulations. For perceptual strength, the total effect ($\beta = 0.23$, $p < .001$) was composed of a positive direct

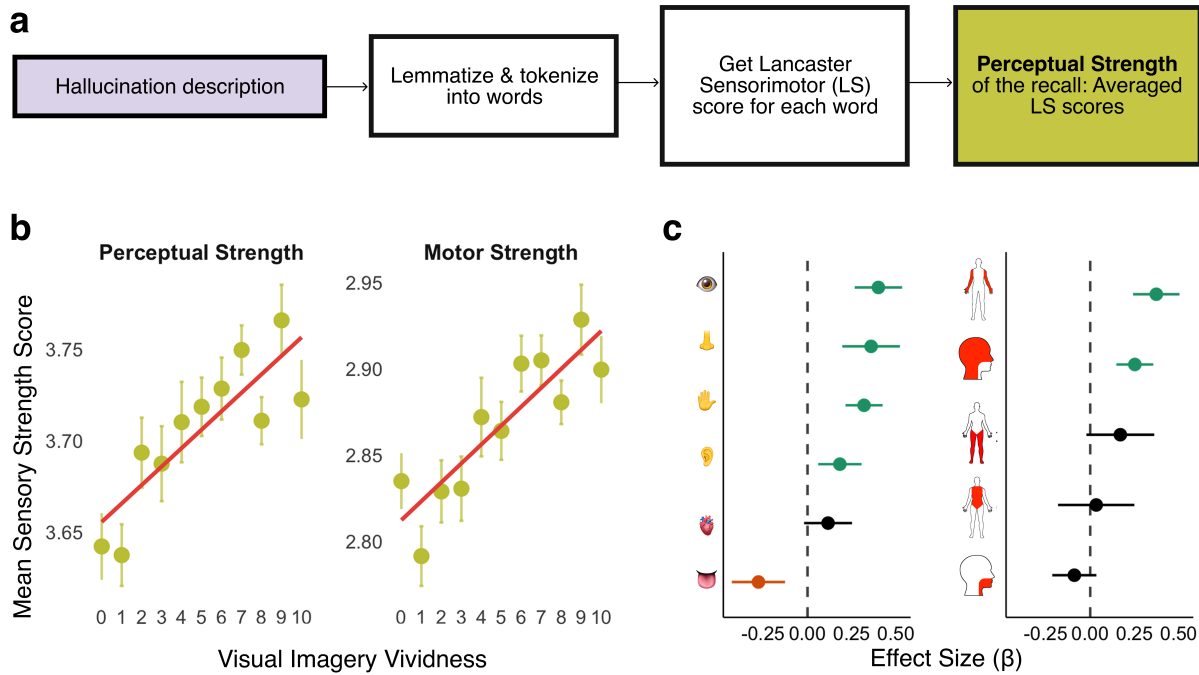


Figure 3.8. Lancaster sensorimotor analysis pipeline and results. **(a)** Analysis pipeline showing how hallucination descriptions were processed to extract sensorimotor content using the Lancaster Sensorimotor norms. Text was lemmatized and tokenized, each word received Lancaster sensorimotor scores, and perceptual strength was calculated by averaging LS scores of individual words in each description. **(b)** Relationship between visual imagery vividness and composite sensorimotor dimensions. Both perceptual strength (left) and motor strength (right) showed significant positive associations with self-reported imagery vividness. Points represent mean scores $\pm$ standard error for each vividness level. **(c)** Forest plots showing effect sizes (β coefficients) from generalized linear models predicting imagery vividness from individual sensorimotor dimensions, controlling for description length. The left panel shows perceptual modalities with icons representing visual (eye), olfactory (nose), haptic (hand), auditory (ear), interoceptive (heart), and gustatory (tongue) dimensions. The right panel shows motor dimensions with body-part icons for head, hand, mouth, torso, and foot. Points show β estimates, and error bars represent 95% confidence intervals (CIs). The dashed vertical line marks $\beta = 0$; effects whose CIs do not overlap this line are significant.

effect ($ADE = 0.35, p < .001$) and a significant negative indirect effect via description length ($ACME = -0.12, p < .001$), indicating suppressive mediation. That is, the full influence of perceptual content was underestimated when not controlling for description length, with over half the total effect (-54%) masked by its association with longer responses. In contrast, action strength showed a small but significant positive mediation effect ($ACME = 0.04, p < .001$), contributing to a total effect of 0.25 ($p < .001$), with 16% of this effect attributable to longer responses. These results suggest that both perceptual and action content are independently predictive of visual imagery vividness.

To identify which specific sensory modalities contributed to the observed effect of perceptual strength on visual imagery vividness, we fit a generalized linear model including all six Lancaster perceptual dimensions—visual, auditory, gustatory, olfactory, haptic, and interoceptive—as simultaneous predictors (see Figure C.5 for individual dimension trends). Description length was included as a covariate to control for potential verbosity differences.

The results revealed that five of the six perceptual modalities reliably predicted imagery vividness (Figure 3.8c, left). Stronger visual content in hallucination descriptions was associated with higher vividness ratings ($\beta = 0.40, p < .001$), as was greater haptic ($\beta = 0.32, p < .001$), olfactory ($\beta = 0.36, p < .001$), and auditory content ($\beta = 0.18, p = .003$). Interestingly, gustatory strength was negatively associated with vividness ($\beta = -0.28, p < .001$), while interoceptive content showed a positive but non-significant effect ($\beta = 0.12, p = .09$).

To account for description length, we ran mediation analyses for each modality. These analyses confirmed that the core effects remained robust even after controlling for verbosity. Specifically, the direct effect of visual strength remained strong and positive ($ADE = 0.40, p < .001$), even though the total effect was slightly suppressed by a significant negative mediation via description length ($ACME = -0.10$). This indicates that participants with more vivid imagery did not merely write longer descriptions; rather, the visual content of those descriptions was independently predictive of vividness. Haptic and olfactory strength also retained significant direct effects after mediation ($ADEs = 0.32$ and 0.36 , respectively; both $ps < .001$), suggesting that these less commonly studied

sensory modalities were meaningfully tied to self-reported individual differences in visual imagery. Although olfactory strength survived mediation analysis, its low mean and restricted range suggest that it should be interpreted with caution, as it may reflect isolated rather than widespread patterns of olfactory content. Auditory content also remained significant ($ADE = 0.18, p = .003$), though its mediated component was small but positive ($ACME = 0.022$). The effect of auditory content on imagery vividness was therefore mostly direct and robust, although there was also a small, significant indirect effect of description length. The effect of interoceptive content remained non-significant after mediation ($ADE = 0.12, p = .10$; CI includes 0), indicating no additional influence of verbosity. While gustatory strength showed a significant negative coefficient in the regression model, its mediation results and real-data trend (Figure C.5) suggest interpretational caution. The model returned a negative direct effect ($ADE = -0.28, p < .001$), but raw data binned by vividness showed a slight upward trend in gustatory content, likely driven by sparse outliers. This dimension also had the lowest mean ($M = 0.30$) and smallest variance (0.035) among all predictors, suggesting that gustatory language was rarely present and potentially contributed noise. Given this, we refrain from interpreting its effect as meaningful. Together, these findings suggest that visual imagery vividness is reflected not only in explicitly visual language but also in modality-rich descriptions involving haptic, olfactory, and auditory references, in addition to action-related language. The robustness of these effects after controlling for verbosity strengthens the claim that these are not mere artifacts of expressive ability or fluency.

To determine which body-related dimensions accounted for the effect of action strength on visual imagery vividness, we next modeled each of the five Lancaster motor dimensions—head, hand, mouth, torso, and foot strength—as simultaneous predictors of vividness (Figure C.5 shows individual trends for each motor dimension). Description length was included as a covariate. The model revealed significant positive effects for hand strength ($\beta = 0.37, p < .001$) and head strength ($\beta = 0.25, p < .001$), suggesting that vivid imagers were more likely to describe content referencing the head or hands, such as “looking around,” “reaching,” or “hands in front of me.” The effects of foot, mouth, and torso strength were not statistically significant ($ps > .05$) (Figure 3.8c, right).

We followed up this model with separate causal mediation analyses to account for the role of

description length. As in earlier models, mediation analyses were conducted for each motor dimension using 5,000 bootstrap simulations. Both hand strength and head strength remained robust predictors after controlling for length. For hand strength, the direct effect remained significant ($ADE = 0.37$, $p < .001$), while the indirect effect ($ACME = -0.056$, $p < .001$) indicated a suppressive relationship with length: longer responses slightly diluted the observed association, but did not eliminate it. A similar pattern held for head strength ($ADE = 0.25$, $p < .001$; $ACME = -0.031$, $p < .001$). In contrast, the effects of foot, mouth, and torso strength did not survive mediation analysis. Foot strength showed a small positive total effect (0.17), but neither the direct nor indirect paths reached significance ($p = .07$ and $p = .78$, respectively). Torso strength had a positive indirect effect ($ACME = 0.031$, $p < .001$), but its direct and total effects were non-significant, and its mediation estimate was highly unstable (proportion mediated = 0.48, 95% CI $[-3.77, 4.10]$). Mouth strength showed a suppressive pattern (proportion mediated = -1.65), but neither its direct nor total effect was significant ($p = .16$), and the instability of this estimate again likely reflects noise due to sparse representation.

These results highlight head- and hand-related content as the most consistent motor predictors of imagery vividness. Their predictive power held even when accounting for description length and shared variance with other motor dimensions. This pattern suggests that vivid internal experiences are more likely to include references to sensorimotor processes involving head movements and hand interactions.

Prior work shows that women report higher visual vividness scores than men on the VVIQ (Richardson, 1995; Aydın, 2020), and self-reported vividness declines across the lifespan (Gulyás et al., 2022). To evaluate whether these factors influenced imagery ratings in our sample, we regressed vividness scores on age and gender. Age was not a significant predictor ($\beta = -0.0002$, $p = .95$), but consistent with the prior literature, gender significantly predicted vividness scores, with women reporting reliably higher vividness than men ($\beta = -0.53$, $p < .001$). To ensure that the main effects reported above were not confounded by gender differences, we re-ran all Lancaster sensorimotor models including gender as a covariate. All sensorimotor predictors that were significant in the primary models remained significant with minimal change in effect size (Figure C.6).

3.4 Discussion

Through computational analysis of $> 4,000$ phenomenological reports, we characterized the content and structure of Ganzflicker-induced hallucinations and examined how this content varies across the visual imagery spectrum. Results from our complementary analytic approaches—topic modeling and sensorimotor norms—converged to reveal systematic differences in hallucination phenomenology consistent with the layered model of visual imagery. Below, we situate our findings within the broader literature on experimentally induced hallucinations and altered states of consciousness (ASCs), and discuss what these findings reveal about individual differences in visual imagery.

3.4.1 Phenomenological Profile of Ganzflicker-Induced Hallucinations

Our topic modeling identified 28 distinct experiential themes, providing a detailed characterization of Ganzflicker-induced hallucination phenomenology. This granular phenomenological mapping extends prior work on flicker-induced experiences, which has typically relied on structured questionnaires or broad categorical distinctions (Bartossek et al., 2021; Hewitt et al., 2025). Consistent with decades of research on light-induced hallucinations (Klüver, 1926; Bressloff et al., 2001, 2002; Hewitt et al., 2025), simple geometric content was among the most frequently reported experiences. Participants described classical “form constants” (Klüver, 1926): spirals, tunnels, and cobwebs that recur across diverse altered states and induction methods, from Ganzfeld (Wackermann et al., 2008; Pistolas & Wagemans, 2025) to stroboscopic light stimulation (Bartossek et al., 2021; Hewitt et al., 2025). However, our analysis also documented complex, semantically meaningful content: faces, human figures, city skylines, forest scenery, stars, and galaxies—naturalistic imagery that has previously been reported in flicker paradigms (Allefeld et al., 2011; Schwartzman et al., 2019; Königsmark et al., 2021; Shenyan et al., 2024; Beauté et al., 2026), but not characterized at this level of detail.

The hierarchical clustering of these 28 topics (Figure C.1) revealed three experiential domains that mirror the visual processing hierarchy: simple geometric content, such as lines, flashes, patterns, dots, and colors; naturalistic scenes, such as skylines, forests, water, and butterflies; and a cluster of social

topics, including faces, figures, and eyes, and metacognitive topics, including experiential descriptions and sensation duration. This data-driven organization aligns with theoretical frameworks proposing that hallucinations evolve from geometric forms to structured experiences as higher-order visual areas become engaged (Ffytche, 2008; Reeder, 2024).

We also examined whether Ganzflicker induced alterations in consciousness beyond visual hallucinations, following the taxonomy of altered-state features introduced by Fort et al. (2025). While our results primarily implicated the perception and imagery dimensions, participants also reported experiences that may relate to other altered-state features: feelings of being “transported to another place,” a “dreamlike state,” and “lost sense of space” share phenomenological overlap with the self-boundary feature, which includes depersonalization and derealization experiences, while “lost a sense of time” aligns with the time-sense feature. Participants also reported emotional changes ranging from enjoyable to unpleasant. These findings situate Ganzflicker within the broader altered-states literature, demonstrating that, like other non-pharmacological induction methods (Berkovich-Ohana et al., 2013; Kübel et al., 2021), it produces alterations extending beyond perception to encompass spatial disorientation, dreamlike absorption, and emotional engagement.

Importantly, experienced altered states varied systematically across the imagery spectrum. Vivid imagers were significantly more likely to report self-boundary alterations, including feelings of being transported, dreamlike states, and spatial disorientation (Figure C.7). By contrast, weak imagers were more likely to report no changes in consciousness. Emotional experiences also varied across individuals: vivid imagers rated their experience as more enjoyable (Figure C.8), even while describing more affectively intense visual content, such as “unpleasant images.” This finding suggests that vivid hallucinations may be experienced as engaging even when their content is disturbing, and fits with the claim that the emotion feature interacts with, but is conceptually distinct from, perceptual content (Fort et al., 2025). Imagery was not related to alterations in time perception, suggesting that time sense might be a relatively stable and broadly shared feature of Ganzflicker experience, invariant across the imagery spectrum. Beyond these broader experiential differences, hallucination content itself varied systematically with imagery capacity.

3.4.2 Hallucinated Content Maps Onto the Visual Processing Hierarchy Across the Imagery Spectrum

Imagery research has traditionally treated vividness as a single continuum, capturing how clearly or intensely mental images are experienced. Our findings suggest that individuals differ not just in the intensity of their imagery, but in the level of representational complexity they can achieve. Accordingly, these findings are consistent with the layered model of visual imagery outlined in the introduction (Figure 3.2). Inspired by neuroimaging evidence that individuals with aphantasia retain early visual representations during imagery (Cabbai et al., 2024; Chang et al., 2025), this model holds that all individuals represent basic visual features in early visual cortex. Because individuals with aphantasia show reduced functional connectivity between frontoparietal control networks and higher-order visual regions (Milton et al., 2021; Spagna et al., 2021; Liu et al., 2025), the layered model suggests that stronger imagers may more effectively engage higher-order regions to integrate basic visual features into coherent, semantically meaningful representations.

If imagery vividness is a scalar intensity parameter affecting all visual processing equally, we would expect proportional scaling of both simple and complex content across the imagery spectrum. Instead, we observed a qualitative shift: weak imagers described the building blocks of visual experience, while strong imagers reported content requiring top-down integration across the full cortical hierarchy. We formalize this mapping in Table 3.2, which links hallucination topics to their predicted processing levels based on their association with imagery vividness.

Even individuals reporting the complete absence of imagery described seeing something during Ganzflicker, but what they saw differed qualitatively from what stronger imagers reported. Weak imagers predominantly described simple visual experiences, such as geometric forms, lines, dots, flashes, and movement. This content is consistent with accounts of early visual cortex involvement during hallucinations, including form constants arising from V_1/V_2 dynamics (Ermentrout & Cowan, 1979; Bressloff et al., 2002; Burke, 2002; Rule et al., 2011), as well as motion-selective intermediate areas such as V_3/MT (Amaya et al., 2023, 2025). Strong imagers, by contrast, reported complex, naturalistic,

and socially relevant content: faces and figures, city skylines, forests, butterflies, and immersive spatial experiences. This type of content engages later extrastriate and ventral temporal regions that represent more complex and multidimensional stimuli (Kanwisher et al., 1997; Epstein & Kanwisher, 1998).

Table 3.2. Hallucination content mapped onto the visual processing hierarchy across the imagery spectrum.

Processing level	Area	Weak imagers/aphantasia	Moderate–strong imagers
Visual primitives	V1/V2	lines, dots, flashes	Lines, dots, flashes
Intermediate	V3/V4, MT/V5	movement, colors	Movement, colors, patterns and shapes, spirals and rotation, tunnels, webs
Semantically meaningful objects	IT	—	Butterflies, flowers, stars and galaxies, faces and figures, eyes, city skyline, forests and trees, hallways and doors, space
Control/integration	Frontoparietal, insula	—	Experiential descriptions, sensation duration

Neural accounts of flicker-induced hallucinations also support our proposed link between individual differences in visual imagery and the extent of hierarchical interaction in the visual system. For example, computational models propose that simple geometric hallucinations arise from dynamical interactions within early visual cortex, where local excitatory and long-range inhibitory connections generate repeating patterns that resemble Klüver’s form constants (Ermentrout & Cowan, 1979; Bressloff et al., 2002; Rule et al., 2011). Complex naturalistic content, by contrast, requires engagement of higher-order visual areas (Shenyan et al., 2024; Stoliker et al., 2025). In addition to

cortical hierarchical models, neural accounts have also highlighted a potential role for thalamocortical interactions in flicker-induced hallucinations (see Hewitt et al., 2025 for a review).

More effective thalamocortical engagement of higher-order visual areas could explain the individual differences in hallucination complexity observed in the present study. The thalamus acts as a central hub regulating cortical excitability and integrating information across networks (Hwang et al., 2017; Shine, 2021), with ventral projections targeting supragranular cortical layers implicated in conscious experience (J. W. Phillips et al., 2019; Suzuki & Larkum, 2020). Just as enhanced top-down connectivity has been shown to be relevant for voluntary imagery in hyperphantasia (Milton et al., 2021), fMRI studies reveal that periodic flicker stimulation produces thalamocortical hyperconnectivity between ventral thalamic nuclei and higher-order visual cortices, with connectivity strength correlating with subjective hallucination intensity (Amaya et al., 2023, 2025). Future studies could test whether imagery capacity, thalamocortical connectivity, and the complexity of flicker-induced hallucinations are related, as well as whether similar mechanisms shape other experiences that rely on the generative capacity of the visual system.

3.4.3 Sensorimotor Content of Hallucination Descriptions May Reflect Multisensory Simulation

To complement computational analyses with human-derived measures, we turned to the Lancaster Sensorimotor Norms to examine the perceptual and action-related content of participants' descriptions using crowd-sourced judgments. We found that descriptions from individuals with higher imagery vividness contained more words with strong visual, haptic, and auditory content, as well as head- and hand-related motor associations. Notably, stronger imagers used richer language not only in the visual domain but also for other perceptual content (see Figure 3.8c), consistent with evidence that imagery differences often extend beyond vision, with many individuals with aphantasia reporting weak imagery across sensory modalities (Zeman et al., 2020). This suggests that what we measure as “visual” imagery vividness may index broader variation in multisensory simulation capacity (Dance et al., 2021). In short, strong imagers used language with richer perceptual connotations, suggestive of

more grounded and content-dense internal simulations.

Potentially at odds with predictive-coding accounts proposing that interoception and agency play a central role in imagery generation (Silvanto & Nagai, 2025; Silvanto, 2025), we failed to find a relationship between interoceptive language and imagery strength. However, the Lancaster norms quantify explicit interoceptive language, whereas the relevant theoretical accounts emphasize implicit mechanisms such as agency or bodily prediction. These may not surface in verbal descriptions when participants are instructed to report what they saw. When directly asked about their conscious experience, however, stronger imagers were more likely to report spatial disorientation and dreamlike absorption, and to evaluate the experience as emotionally enjoyable compared to those with weak visual imagery (Figures C.7 and C.8). This confirms that embodied and affective experience do vary with imagery vividness, even when not captured by explicit language norms.

The most robust predictors of individuals' visual imagery vividness were visual, head-related, and hand-related language, consistent with our layered model, in which stronger imagers more effectively engage high-level perceptual systems that support complex, socially meaningful internal content. Faces and hands are among the most socially and evolutionarily significant concepts for humans (Bracci et al., 2010; Lacruz et al., 2019), each with its own dedicated neural circuitry (Kanwisher et al., 1997; Bracci et al., 2010). When higher-order areas are recruited during complex hallucinations, these strongest “defaults” may surface first. Because individuals with aphantasia show reduced activation of the fusiform gyrus during imagery tasks (Liu et al., 2025), stronger imagers might more effectively recruit specialized regions, such as the fusiform face area, during Ganzflicker-induced hallucinations, resulting in richer and more complex visual content, as well as possibly richer multisensory associations.

3.4.4 Limitations

Our study drew on a large online sample of approximately 4,000 participants, allowing us to capture rare individual differences, such as aphantasia and hyperphantasia, and to examine Ganzflicker experiences across diverse contexts. However, online studies come with inherent constraints. For example, we could not directly control participants' viewing environments. Although all participants viewed

the same 7.5 Hz red–black flicker stimulus, variations in the properties of their monitors, such as luminance, color saturation, and timing stability, as well as viewing conditions, such as lighting, distractions, and use of white noise, may have introduced heterogeneity into their perceptual responses. Without standardized conditions, the data could in principle represent a heterogeneous set of experiences rather than differences related to participants’ position on the imagery spectrum.

However, several factors support the interpretation that our findings reflect genuine Ganzflicker phenomenology. Prior analyses of the impact of environmental variables in this dataset, including white-noise use, lighting conditions, and display medium, failed to reveal significant effects on reported hallucination content (see Reeder, 2022 for details; see also Königsmark et al., 2021). Further, the hallucination content documented here aligns with that reported in prior Ganzflicker studies using different samples and setups (Königsmark et al., 2021), including one conducted in a controlled laboratory environment (Shenyan et al., 2024). Nevertheless, online data collection on its own cannot fully rule out the contribution of uncontrolled environmental factors, and future work with larger controlled laboratory samples would strengthen confidence in these findings.

Another limitation of our study relates to participant recruitment via a public article about Ganzflicker. This may have attracted individuals already interested in hallucination phenomena, as well as those at the extremes of the imagery spectrum, in keeping with the higher-than-typical proportion of self-reported aphantasia in our sample. This raises the possibility that observed results reflect demand or expectancy effects, with those interested in hallucinations eager to report altered states of consciousness and those who identify as aphantasic eager to produce descriptions reflecting their expectations rather than their actual experiences. However, Ganzflicker-induced hallucinations reliably emerge in laboratory settings without self-selection (Shenyan et al., 2024). Further, in the present study, even participants reporting no visual imagery described hallucinations, arguing against expectancy alone as an explanation (Reeder, 2022). Finally, hallucination content changed predictably with imagery vividness, a pattern inconsistent with simple demand effects.

Yet another limitation of the present study is our use of participants’ verbal descriptions as an index of their subjective visual experiences. Although language provides a powerful window

into the content of participants' hallucinations, descriptions necessarily conflate phenomenology with narrative expression. Individuals may differ not only in what they experience but also in how they verbalize those experiences. Notably, participants with more vivid imagery provided longer descriptions, raising the possibility that verbosity rather than experiential richness drives some effects. To address this, our analyses using the Lancaster Sensorimotor Norms controlled for description length, and vividness effects remained significant. Corroborating evidence also comes from closed-ended measures: participants with higher imagery vividness reported greater alterations in consciousness and more positive affect during Ganzflicker stimulation on questions that did not require generating linguistic descriptions (Figures C.7 and C.8). That said, verbal reports alone cannot adjudicate between differences in what people experience and differences in how they describe it. Future work incorporating objective measures, such as EEG, eye tracking, and physiological markers, will be critical for directly assessing perceptual dynamics during Ganzflicker stimulation.

3.4.5 Conclusion

Together, these findings offer a rich characterization of Ganzflicker-induced experience, from the diversity of visual hallucinations to broader alterations in consciousness, and reveal how both vary systematically across the visual imagery spectrum. Overall, results align with the theory of divergent predictive perception (Reeder, Sala, & van Leeuwen, 2024), which proposes that individuals across the imagery spectrum differ in how strongly top–down predictions shape perceptual experience. Under this account, strong imagers generate richer predictive models that more readily engage higher-order visual representations, while weak imagers rely more heavily on bottom–up input. The Ganzflicker paradigm, which provides rhythmic stimulation with minimal structured input, may be particularly sensitive to these differences: when bottom–up information is impoverished, individual variation in top–down generative capacity becomes visible in the content of conscious experience.

Beyond Ganzflicker, these findings may have broader implications for understanding human cognition. Individual differences in imagery are not merely subjective curiosities but have been shown to influence a wide range of cognitive processes, from shaping emotions (Wicken et al., 2021) and moral

judgments (Amit & Greene, 2012) to affecting how people adopt scientific theories (Sulik et al., 2025). By incorporating phenomenological variability into our models, we move toward a more complete and inclusive science of the mind. Further, this work demonstrates that computational analysis of free-text reports can reveal the hidden structure of complex mental phenomena, offering a scalable approach for studying individual differences in cognitive processes that are otherwise difficult to access (Feuerriegel et al., 2025). By treating language as a window into the mind's eye, we can begin to capture how people with different imagery phenotypes internally construct their visual world.

Acknowledgement

Chapter 3, in full, is a reprint of the material as it appears in Chkhaidze, A., Reeder, R. R., Gag, C., Kiyonaga, A., & Coulson, S. (2026). From dots to faces: individual differences in visual imagery capacity predict the content of Ganzflicker-induced hallucinations. *Neuroscience of Consciousness*, 2026(1), niago16. The dissertation author was the primary investigator and author of this material.

Chapter 4

Some Bottlenecks in Imagery Research

4.1 The Promise and Puzzle of Visual Imagery Diversity

Close your eyes and imagine a bowl of ripe strawberries. For many people, some kind of visual image comes to mind, though its vividness, detail, and stability may vary. For others, thought proceeds without any visual experience. These differences, spanning the range from aphantasia, or lack of imagery, to hyperphantasia, or lifelike imagery (Zeman, 2024), were long treated as introspective curiosities rather than meaningful sources of cognitive variation (Pylyshyn, 1973, 2002). Recent work has begun to change that view.

Individuals with weaker visual imagery show reduced imagery-based perceptual priming (Keogh & Pearson, 2018, 2024), diminished or absent imagery-related pupillary light responses when imagining or maintaining dark versus bright stimuli (Dong & Kiyonaga, 2024; Kay et al., 2022), and weaker prefrontal–visual resting-state connectivity compared to strong imagers (Milton et al., 2021). Visual imagery has thus become an empirically tractable dimension of cognitive variation, with measurable behavioral, physiological, and neural correlates.

This matters because internal visual representations have been proposed to support a wide range of cognitive functions, including episodic memory, spatial navigation, language comprehension, and visual working memory (for a review, see Pearson, 2019). Embodied accounts of language comprehension, for instance, propose that understanding involves perceptual simulation of described events (Barsalou, 1999; Pulvermüller, 2013; Zwaan, 2014), sometimes with enough specificity to interact

with later visual processing (Dils & Boroditsky, 2010; Lupyan, 2012; Lupyan et al., 2020). Individual differences in imagery therefore provide a natural experiment: they allow us to ask whether cognitive outcomes often assumed to rely on a common representational currency can instead emerge from different representational mechanisms (Lupyan et al., 2023; Muraki et al., 2023).

Here, the literature becomes puzzling. On the one hand, imagery differences appear in behaviors as varied as emotion processing (Wicken et al., 2021), autobiographical memory (Greenberg & Knowlton, 2014), and moral decision-making (Amit & Greene, 2012). Yet in processes where visual imagery seems most obviously relevant, including visual working memory and visual attention, the effect of individual differences in imagery is often weak, inconsistent, or absent (Cabbai et al., 2024; Knight et al., 2026; Pounder et al., 2022; Vanbuckhave et al., 2026; Weber et al., 2024). If imagery differences manifest in domains as complex and abstract as moral decision-making, their inconsistent presence in visual working memory demands an explanation.

We suggest that this puzzle is not a reason to doubt that imagery differences are cognitively meaningful, but a reason to examine how imagery is defined, measured, and studied. In our view, inconsistencies in the literature partly reflect four bottlenecks: (1) imagery is predominantly treated as a single vividness continuum; (2) measures of rich scene and object imagery are used to predict tasks requiring low-level feature precision; (3) imagery differences may be sought at the wrong level of the visual hierarchy; and (4) equivalent task performance is taken to imply equivalent cognitive mechanisms. We outline each bottleneck and suggest concrete ways forward.

4.2 Bottleneck 1: Visual Imagery Is Ill-Defined

What do we talk about when we talk about visual imagery? In practice, imagery diversity has been operationalized predominantly in terms of vividness. Of 71 aphantasia papers reviewed by Argueta et al. (2025), 53 used the Vividness of Visual Imagery Questionnaire (VVIQ; Marks, 1973) as their primary measure of imagery capacity. This focus is not without justification—vividness has real behavioral and neural correlates (Kay et al., 2022; Keogh & Pearson, 2018, 2024; Milton et al., 2021).

But vividness of what, exactly?

Visual imagery, like visual perception, may not be unitary. Rather than reflecting a single global capacity, it may be composed of dissociable components. In perception, object identity and spatial relations rely on partially dissociable systems (Mishkin et al., 1983), and a similar distinction may extend to imagery (Blazhenkova & Pechenkova, 2019; Farah et al., 1988; Kozhevnikov et al., 2002; Palermo et al., 2022). This distinction is cognitively meaningful: individuals with aphantasia show poorer recall of object or sensory details alongside relatively preserved spatial memory (Bainbridge et al., 2021). I. Phillips (2025) suggests that this spared spatial imagery “helps resolve the puzzle of aphantasia”: how individuals reporting no visual imagery can perform more-or-less normally on tasks presumed to depend on imagery. On this account, many such tasks may be solvable through spatial, sensorimotor, or semantic strategies, even when rich visual-object imagery is weak or absent (see also Reeder, Pounder, et al., 2024).

Yet even the object–spatial distinction does not fully capture imagery variation. Current instruments often conflate conceptually different aspects of imagery within the same scale. The Object-Spatial Imagery Questionnaire (OSIQ; Blazhenkova, Kozhevnikov, & Motes, 2006), for instance, mixes phenomenology (“My images are very colourful and bright”), self-rated skill (“I was very good in 3D geometry as a student”), and preference (“if asked to choose between engineering and visual arts, I would choose visual arts”). More broadly, imagery vividness, capacity, and propensity need not align. A person may have vivid visual imagery yet rarely recruit it, relying instead on verbal or schematic representations; conversely, someone may frequently attempt to use imagery while experiencing it as weak or unstable.

This distinction was central to the working memory study described in Chapter 1. In a naturalistic visual working-memory task, imagery propensity predicted the kinds of errors participants made, whereas vividness predicted performance most clearly when fragile memory traces were challenged by similar visual distractors (Chkhaidze, Gaspariani, et al., 2026). Participants with stronger verbal thinking tendencies showed error patterns consistent with more semantic or label-based maintenance, while stronger visual imagery was associated with greater sensitivity to visual similarity under distrac-

tion. Thus, propensity and vividness appeared to capture different aspects of cognition: propensity indexed the representational strategy people habitually relied on, whereas vividness mattered more for the fidelity or robustness of visual representations under strain.

The field has largely treated imagery variation as a continuum from aphantasia to hyperphantasia. The evidence reviewed here suggests it may be more productive to think of it as a multidimensional space, in which individuals differ not only in how vivid their imagery is, but also in what kind of imagery they have and how habitually they recruit it. The multidimensional view of imagery is supported by work showing substantial heterogeneity within aphantasia. Delem et al. (2025) found that individuals classified as having aphantasia varied not only in visual imagery, but also in the extent to which other imagery abilities and cognitive strategies were affected. Thus, low visual vividness can reflect different representational profiles rather than a single underlying cognitive trait.

Progress will require measures that make these distinctions explicit. Rather than relying on a single global vividness score, researchers should combine instruments that capture different facets of imagery: vividness (VVIQ; Marks, 1973), object versus spatial imagery (OSIQ; Blazhenkova et al., 2006), and habitual reliance on visual versus verbal formats (IRQ; Roebuck & Lupyan, 2020). This would allow stronger tests of which components of imagery matter for which tasks, and would make null findings more interpretable. A null relationship between VVIQ and task performance, for example, may mean that visual vividness is irrelevant, but it may also mean that the task depends on spatial imagery, imagery propensity, or a level of representation not captured by the VVIQ.

4.3 Bottleneck 2: Mismatch Between Imagery Measures and Cognitive Tasks

The VVIQ asks participants to imagine a relative's face, a rising sun, a familiar shop front, and a mountain landscape: rich, complex visual-object scenarios. Yet it is routinely used to predict behavior on tasks that require much narrower feature-level representations, such as maintaining orientation or color in working memory, or using color-based attentional templates (Cabbai et al., 2024; Keogh

et al., 2021; Knight et al., 2026; Reeder, Pounder, et al., 2024; Weber et al., 2024). A null relationship between scene-vividness scores and orientation memory is therefore not necessarily evidence that imagery is irrelevant; it may reflect a mismatch between the representational scale of the measure and the task. Indeed, imagery-related differences appear more consistently in domains involving naturalistic, semantically rich, or affectively meaningful material, including scene recall (Bainbridge et al., 2021), real-world object memory (Chkhaidze, Gaspariani, et al., 2026), emotional language (Wicken et al., 2021), and moral decision-making (Amit & Greene, 2012). Answering whether and when imagery matters for cognition therefore requires aligning imagery measures with the representational demands of the task. Addressing this mismatch requires movement from both sides: developing measures that better capture low-level feature imagery, and choosing tasks that match the representational scale of existing measures.

First, when a task depends on low-level feature maintenance, imagery measures should be matched to that level of representation. Studies of orientation, color, spatial frequency, or contrast working memory should not rely exclusively on global scene-vividness questionnaires. The field currently lacks well-validated, short measures of low-level feature imagery, and this absence is itself part of the problem. A useful next step would be to adapt existing imagery probes to simple feature domains: how precisely can someone voluntarily imagine a tilted grating, a shade of red, or a high-frequency texture? How stable are these representations across a delay? How much do they interfere with perception?

Second, when researchers use broad vividness measures such as the VVIQ, tasks should be chosen to match the representational scale those measures actually sample. Naturalistic objects, scenes, autobiographical episodes, and emotionally meaningful language may be more sensitive to imagery differences because they require the construction of rich, integrated representations. In this sense, null effects in low-level tasks and positive effects in naturalistic tasks need not conflict. They may indicate that imagery diversity is not expressed uniformly across tasks, but depends on the match between the imagery measure and the representational demands of the task.

4.4 Bottleneck 3: Beyond Representational Absence in Aphantasia

One influential assumption has been that weaker imagery reflects weaker recruitment of early visual cortex. Early neuroimaging work found that imagery vividness correlates with activation in early visual areas, encouraging the view that weak imagery reflects a failure of top-down signals to generate low-level visual representations (Albers et al., 2013; Bergmann et al., 2016; Cui et al., 2007; Lee et al., 2012). On this view, imagery differences should be most visible near the earliest stages of the visual hierarchy.

Recent evidence challenges this picture. The content of imagined stimuli can be decoded from early visual cortex in both individuals with aphantasia and typical imagers, with decoding accuracy that does not differ between groups (Cabbai et al., 2024; Chang et al., 2025). Low-level visual features appear to be represented across the imagery spectrum. Group differences emerge not in early visual cortex but at higher levels of the visual hierarchy, including the fusiform gyrus, insula, and frontoparietal control networks (Silvanto & Nagai, 2025; Spagna et al., 2021; Liu et al., 2025). Behavioral evidence points in the same direction: during flicker-induced visual hallucinations, even individuals who report weak or absent voluntary imagery report experiencing some simple hallucinations (Königsmark et al., 2021; Reeder, 2022). In Chapter 3, I examined differences in hallucination content across the imagery spectrum by analyzing free-text descriptions of Ganzflicker-induced visual experiences from over 4,000 participants. I found that weaker imagers more often described relatively simple perceptual content, including dots, flashes, colors, geometric forms, and moving patterns, whereas stronger imagers more often described complex, naturalistic content, including faces, bodies, animals, objects, landscapes, and scenes.

This pattern motivated the layered model proposed in Chapter 3. On this view, internally generated visual experience can vary across layers of representational complexity, from low-level features and simple perceptual patterns to integrated objects, agents, and scenes. Imagery vividness may therefore reflect not only the strength of early visual activation, but also the extent to which feature-level visual

representations are coordinated with higher-order visual, semantic, and frontoparietal systems to form coherent, meaningful content (Chkhaidze, Reeder, et al., 2026; Dijkstra et al., 2019; Mechelli et al., 2004; Spagna et al., 2021). Thus, weaker imagery may be less integrated or less structurally complex, even when basic visual features are still internally represented.

This reframing has direct implications for how imagery differences should be studied. If imagery differences emerge primarily in the integration of features into objects and scenes, paradigms that index only early visual processing may be poorly suited to detect them. Tasks and neural measures that target object and scene construction, high-level visual content, and interactions between visual and frontoparietal systems may be better positioned to reveal where imagery differences actually arise.

4.5 Bottleneck 4: Similar Outcomes, Different Routes

The functional role of imagery diversity may be masked when studies focus only on task endpoints. Research often compares accuracy or reaction time across individuals with different imagery profiles; when group differences fail to emerge, the conclusion is that imagery does not matter. But equivalent performance need not imply equivalent mechanisms. Imagery differences may shape how people arrive at an answer, even when they ultimately arrive at the same one.

Visual working memory provides a clear example. Individuals with weak or absent imagery can often perform comparably to typical imagers, but report relying on different representational strategies during maintenance (Keogh et al., 2021; Reeder, Pounder, et al., 2024). Chapter 1 provides empirical evidence that such strategy differences can leave measurable signatures in behavior. In a visual working-memory task using naturalistic images, overall performance differences across imagery profiles were limited, but imagery tendencies predicted how participants erred and how they were affected by visual distraction. Stronger verbal thinking tendencies were associated with error patterns consistent with more semantic or label-based maintenance: participants appeared more likely to preserve broad category information while losing the fine-grained visual detail needed to identify the exact exemplar. By contrast, stronger visual imagery was associated with greater sensitivity to visually similar distraction,

suggesting that visually maintained representations may be especially vulnerable when competing sensory input overlaps with the remembered item.

These findings suggest that equivalent working-memory performance should not be taken as evidence for equivalent representational mechanisms. Different routes may support comparable accuracy under many conditions, but they may diverge when the task requires especially precise visual information or when the representation is placed under strain. Consistent with this idea, *aphantasia* has been associated with selective impairments in visual working memory when the task requires a high degree of precision (Jacobs et al., 2018). Chapter 1 points to a similar principle: imagery vividness mattered most when less memorable items had to be protected from visually similar distraction. Thus, imagery may not always determine whether a task can be solved, but it may shape what kind of representation supports performance and under what conditions that representation becomes fragile.

Chapter 2 provides a complementary example from language comprehension. In a sentence-picture verification task, visual imagery vividness did not reliably predict the behavioral size of the match effect. On the surface, this could be taken to suggest that imagery does not meaningfully shape language-induced perceptual simulation. However, the EEG results revealed a different pattern. Visual imagery vividness modulated early posterior neural responses to shape mismatch, consistent with greater recruitment of perceptually specific visual representations in stronger imagers (cf. Ganis & Schendan, 2008). Yet this imagery-related difference was not simply a matter of weak imagers failing to distinguish match from mismatch altogether. In the later N400-range window, which is commonly associated with semantic integration (Kutas & Federmeier, 2011), match–mismatch differences were present even among participants with weak or absent self-reported imagery. Thus, a behavioral null does not necessarily imply that perceptual simulation is absent or equivalent across individuals. It may instead indicate that participants can reach the same judgment through partially different routes.

Uncovering these differences requires paradigms that reveal the mechanistic route, not just the behavioral outcome. Error analysis is one approach: in mental arithmetic, for example, phonological strategies produce errors driven by sound similarity, whereas spatial or abacus-based strategies produce distinct error signatures (Frank & Barner, 2012). Similar logic can be applied to imagery research.

If two individuals reach the same level of accuracy but make different kinds of errors, slow down under different conditions, or recruit different neural systems, then equivalent performance may conceal different representational mechanisms. Combining behavioral endpoints with process-sensitive measures—including error patterns, response-time distributions, strategy reports, eye movements, and neural recruitment—can reveal where, when, and how different cognitive routes diverge.

4.6 Concluding Remarks

The tools and frameworks that have guided imagery research were built for a simpler picture of imagery than the one now emerging. Vividness remains a valuable starting point, but it cannot carry the explanatory burden alone. The path forward is not to abandon existing measures or findings, but to refine the links among constructs, tasks, and mechanisms: to distinguish object from spatial imagery, vividness from propensity, low-level features from integrated scenes, and task success from the route by which success is achieved. By studying imagery diversity with sharper tools, the field can move from asking whether imagery matters in general to identifying which forms of imagery matter, when they matter, and how different minds use different representational routes to think and act.

Acknowledgement

Chapter 4, in full, is currently being prepared for submission for publication of the material as *Some bottlenecks in imagery research*. Chkhaidze, A., & Kiyonaga, A. The dissertation author was the primary investigator and author of this material.

Conclusion

Human thought is often treated as if it has a common format: a shared representational currency that supports memory, language, imagination, and reasoning. Yet people's own descriptions of their minds suggest a more varied picture. Some think vividly in images, others rely more on words, and others report little or no visual imagery at all. This dissertation used that diversity as a tool for studying cognition.

The central puzzle motivating this work is that imagery differences often seem like they should matter, but do not always appear where we expect them. People with weak or absent visual imagery can often perform normally on tasks that seem likely to require visual representation. This could suggest that imagery is not functionally important. The studies in this dissertation point to a different conclusion. Imagery may not always determine whether a person can complete a task successfully, but it can shape how that task is solved, what kind of information is represented, and when those representations become fragile.

4.7 Summary of Contributions

Prior work has shown that people with weak or absent imagery often report using nonvisual strategies during visual working memory tasks. Chapter 1 builds on this work by showing that these strategy differences can leave measurable signatures in behavior. Depending on their imagery and verbal thought tendencies, individuals appeared to rely on different representational routes, which shaped both the kind of information preserved in working memory and its vulnerability to perceptual competition and attentional demands.

Chapter 2 extends this logic to language comprehension. Prior work on embodied language has often treated perceptual simulation as a general mechanism of comprehension, but behavioral evidence for imagery-related differences has been mixed. This chapter shows that imagery differences may emerge at the level of neural processing even when overt behavior is similar. Visual imagery vividness shaped early posterior responses in a time window linked to visual-perceptual processing, consistent with greater recruitment of visually specific representations in stronger imagers. At the same time, later meaning-related effects were present more broadly across the imagery spectrum, suggesting that language processing can also proceed through less visually specific routes. This chapter therefore supports a flexible, dual-route view of language comprehension, in which perceptual simulation may contribute to understanding, but is not the only path to successful meaning construction.

Chapter 3 shifts from task performance to the content of internally generated experience. Much of imagery research focuses on how vivid imagery feels across individuals, but this chapter shows that imagery vividness may also relate to what internally generated visual experience contains. Weaker and stronger imagers differed in the complexity and organization of the Ganzflicker-induced visual hallucinations they described. Weaker imagers described simple visual patterns, whereas stronger imagers described complex objects, agents, and scenes. These findings motivate a layered view of imagery. Weak imagery may not reflect weak visual representations in the early visual cortex. Instead, the difference may lie partly in how low-level visual features are integrated into more complex, object-like and scene-like representations.

Chapter 4 synthesizes these findings within the broader imagery literature. Its central contribution is to explain why imagery effects may often be inconsistent across studies, by proposing several bottlenecks in how imagery has been defined, measured, and interpreted. Imagery is often reduced to a single vividness score, paired with tasks that require different representational demands, studied at the wrong level of processing, and evaluated through endpoint behavioral measures that may miss differences in mechanism. Chapter 4 therefore reframes imagery diversity as variation in representational routes, and offers a framework for asking when, how, and for whom different forms of imagery matter.

4.8 Limitations and Future Directions

Several limitations and open questions remain. First, the studies in this dissertation rely partly on indirect measures of internal representations. Error patterns, distraction effects, ERPs, and verbal reports are informative, but they do not fully identify the underlying representational format. Future work should use more direct representational measures, including neuroimaging approaches such as fMRI, to better localize the neural systems and generators giving rise to imagery-related representational differences.

Second, more work is needed to understand the trade-offs between visual and verbal representational routes. The findings from Chapter 1 suggest that neither format is uniformly better. Visual imagery may be useful for preserving fine-grained perceptual detail, but detailed visual representations may also be more vulnerable when they must be maintained alongside other information or protected from visually similar distraction. Verbal or label-based representations may preserve less visual detail, but they may be easier to maintain under competition because they reduce the burden on visual representational resources. An important question for future work is whether people can flexibly shift between these routes depending on task demands. When fine-grained visual detail is required, visual imagery may be advantageous; when multiple items must be managed or distraction is high, verbal coding may be more efficient. Future experiments should therefore manipulate when visual versus verbal representations are useful or costly, rather than treating imagery and verbal thinking only as stable individual differences.

Third, future studies should include broader samples across the imagery spectrum. Studies in this dissertation examined continuous imagery differences rather than recruiting large groups of individuals with aphantasia or hyperphantasia. This dimensional approach is valuable, but some effects may be clearest at the extremes. Future work should combine both approaches, examining continuous variation while also including sufficiently powered groups with weak, absent, and exceptionally vivid imagery.

Finally, future work should continue developing tasks that are more diagnostic of representa-

tional routes. If a task can be solved through visual detail, verbal labels, semantic knowledge, spatial structure, or some shifting combination of these formats, then accuracy and reaction time alone may not tell us much about the format of cognition. More informative paradigms will place pressure on the representation itself: by examining errors, interference, neural timing, precision, content reports, drawings, or computational measures of what people describe and produce. The goal should not only be to ask whether imagery matters, but to design tasks that reveal when it matters and how.

4.9 Closing Remarks

The central lesson of this dissertation is that imagery diversity does not simply divide people into those who can and cannot perform tasks thought to depend on visual imagery. Rather, it shows that cognition can be successfully implemented through multiple representational routes. These routes may be hidden when we look only at overall performance, but become visible when we ask what kind of information is represented, when the representation becomes fragile, and how people solve the task. By making these hidden representational routes visible, individual differences in imagery offer a window into the flexibility of cognition.

Appendix A

Chapter 1

A.1 Supplementary Tables

These supplementary materials provide full statistical summaries that support the interpretation and transparency of Chapter 1. Tables A.1–A.9 report the full mixed-effects model outputs for each dependent variable (RT, accuracy, error type), fitted separately for each imagery questionnaire (IRQ, OSIVQ, VVIQ).

Table A.1. Response time model results for IRQ predictors.**IRQ: RT Model**

Predictor	<i>b</i>	<i>SE</i>	95% CI	<i>t</i>	<i>p</i>
(Intercept)	0.3950066	0.0082182	[0.379, 0.411]	48.0650196	0.0000000 ***
IRQ Visual	0.0000972	0.0096108	[-0.019, 0.019]	0.0101133	0.9919309
IRQ Verbal	0.0117632	0.0094700	[-0.007, 0.030]	1.2421568	0.2141840
Visual Similarity	-0.0029665	0.0013142	[-0.006, -0.000]	-2.2571553	0.0240023 *
Attentional Priority	-0.0287008	0.0013103	[-0.031, -0.026]	-21.9033681	0.0000000 ***
ResMem Scores	-0.0236142	0.0012707	[-0.026, -0.021]	-18.5834321	0.0000000 ***
IRQ Visual × Visual Similarity	0.0037912	0.0015003	[0.001, 0.007]	2.5270037	0.0115068 *
IRQ Verbal × Visual Similarity	-0.0037753	0.0014895	[-0.007, -0.001]	-2.5346028	0.0112602 *
IRQ Visual × Attentional Priority	-0.0034700	0.0015249	[-0.006, -0.000]	-2.2755973	0.0228739 *
IRQ Verbal × Attentional Priority	-0.0003427	0.0014839	[-0.003, 0.003]	-0.2309130	0.8173833
IRQ Visual × ResMem Scores	0.0012907	0.0014758	[-0.002, 0.004]	0.8745690	0.3818123
IRQ Verbal × ResMem Scores	-0.0004659	0.0014539	[-0.003, 0.002]	-0.3204691	0.7486140
Visual Similarity × ResMem Scores	-0.0043996	0.0012240	[-0.007, -0.002]	-3.5944562	0.0003254 ***
Attentional Priority × ResMem Scores	-0.0024342	0.0013021	[-0.005, 0.000]	-1.8693920	0.0615736 .
Visual Similarity × Attentional Priority	0.0008800	0.0013148	[-0.002, 0.003]	0.6693098	0.5033006
IRQ Visual × Visual Similarity × ResMem Scores	-0.0020549	0.0014122	[-0.005, 0.001]	-1.4551120	0.1456440
IRQ Verbal × Visual Similarity × ResMem Scores	-0.0002984	0.0014086	[-0.003, 0.002]	-0.2118057	0.8322594

Table A.2. Accuracy model results for IRQ predictors.

IRQ: Accuracy Model

Predictor	OR	SE	95% CI	z	p	
(Intercept)	3.4461095	0.1536033	[3.158, 3.761]	27.7577645	0.0000000	***
IRQ Visual	0.9009549	0.0470632	[0.813, 0.998]	-1.9966695	0.0458611	*
IRQ Verbal	1.0187410	0.0524387	[0.921, 1.127]	0.3607165	0.7183114	
Visual Similarity	1.0060772	0.0094626	[0.988, 1.025]	0.6441790	0.5194594	
Attentional Priority	1.2310305	0.0106673	[1.210, 1.252]	23.9864924	0.0000000	***
ResMem Scores	1.1657886	0.0103999	[1.146, 1.186]	17.1953395	0.0000000	***
IRQ Visual × Visual Similarity	0.9908224	0.0112831	[0.969, 1.013]	-0.8096547	0.4181386	
IRQ Verbal × Visual Similarity	1.0126571	0.0113515	[0.991, 1.035]	1.1220372	0.2618466	
IRQ Visual × Attentional Priority	1.0185460	0.0106661	[0.998, 1.040]	1.7548057	0.0792926	.
IRQ Verbal × Attentional Priority	0.9859107	0.0101232	[0.966, 1.006]	-1.3819337	0.1669921	
IRQ Visual × ResMem Scores	0.9881703	0.0105940	[0.968, 1.009]	-1.1100153	0.2669924	
IRQ Verbal × ResMem Scores	1.0092445	0.0106893	[0.989, 1.030]	0.8688192	0.3849460	
Visual Similarity × ResMem Scores	1.0480054	0.0089133	[1.031, 1.066]	5.5130448	0.0000000	***
Attentional Priority × ResMem Scores	1.0124031	0.0085076	[0.996, 1.029]	1.4668911	0.1424057	
Visual Similarity × Attentional Priority	0.9967911	0.0086283	[0.980, 1.014]	-0.3713062	0.7104095	
IRQ Visual × Visual Similarity × ResMem Scores	1.0130190	0.0105203	[0.993, 1.034]	1.2455292	0.2129373	
IRQ Verbal × Visual Similarity × ResMem Scores	0.9898471	0.0101112	[0.970, 1.010]	-0.9990128	0.3177885	

Table A.3. Error-type model results for IRQ predictors.**IRQ: Error Type Model**

Predictor	OR	SE	95% CI	z	p	
(Intercept)	2.1362877	0.1021472	[1.945, 2.346]	15.8750385	0.0000000	***
IRQ Visual	0.8864089	0.0497226	[0.794, 0.989]	-2.1495325	0.0315922	*
IRQ Verbal	1.1370704	0.0629115	[1.020, 1.267]	2.3217141	0.0202483	*
Visual Similarity	1.0286454	0.0191542	[0.992, 1.067]	1.5167287	0.1293352	
Attentional Priority	1.2057491	0.0199724	[1.167, 1.246]	11.2954169	0.0000000	***
ResMem Scores	0.9238838	0.0169748	[0.891, 0.958]	-4.3089110	0.0000164	***
IRQ Visual × Visual Similarity	1.0166516	0.0234787	[0.972, 1.064]	0.7150944	0.4745507	
IRQ Verbal × Visual Similarity	0.9688052	0.0216530	[0.927, 1.012]	-1.4179601	0.1562024	
IRQ Visual × Attentional Priority	1.0291946	0.0206297	[0.990, 1.070]	1.4356372	0.1511056	
IRQ Verbal × Attentional Priority	0.9751091	0.0194475	[0.938, 1.014]	-1.2638344	0.2062895	
IRQ Visual × ResMem Scores	0.9976453	0.0220426	[0.955, 1.042]	-0.1067011	0.9150261	
IRQ Verbal × ResMem Scores	0.9993140	0.0215507	[0.958, 1.042]	-0.0318188	0.9746166	
Visual Similarity × ResMem Scores	0.9445831	0.0166432	[0.913, 0.978]	-3.2356821	0.0012135	**
Attentional Priority × ResMem Scores	1.0317061	0.0167212	[0.999, 1.065]	1.9259074	0.0541159	.
Visual Similarity × Attentional Priority	1.0029371	0.0163993	[0.971, 1.036]	0.1793604	0.8576548	
IRQ Visual × Visual Similarity × ResMem Scores	1.0080272	0.0228778	[0.964, 1.054]	0.3522782	0.7246296	
IRQ Verbal × Visual Similarity × ResMem Scores	0.9896631	0.0211479	[0.949, 1.032]	-0.4862550	0.6267864	

Table A.4. Response time model results for OSIVQ predictors.

OSIVQ: RT Model

Predictor	<i>b</i>	<i>SE</i>	95% CI	<i>t</i>	<i>p</i>	
(Intercept)	0.3952248	0.0081895	[0.379, 0.411]	48.2599384	0.0000000	***
OSIVQ Visual	0.0111321	0.0084276	[-0.005, 0.028]	1.3209010	0.1865399	
OSIVQ Verbal	0.0064454	0.0084364	[-0.010, 0.023]	0.7639910	0.4448759	
Visual Similarity	-0.0032020	0.0013108	[-0.006, -0.001]	-2.4426753	0.0145820	*
Attentional Priority	-0.0289452	0.0013075	[-0.032, -0.026]	-22.1382877	0.0000000	***
ResMem Scores	-0.0236405	0.0012672	[-0.026, -0.021]	-18.6553628	0.0000000	***
OSIVQ Visual × Visual Similarity	0.0022001	0.0013090	[-0.000, 0.005]	1.6808088	0.0928057	.
OSIVQ Verbal × Visual Similarity	0.0015026	0.0013408	[-0.001, 0.004]	1.1206792	0.2624293	
OSIVQ Visual × Attentional Priority	-0.0030963	0.0013183	[-0.006, -0.001]	-2.3487062	0.0188423	*
OSIVQ Verbal × Attentional Priority	0.0037959	0.0013429	[0.001, 0.006]	2.8266228	0.0047059	**
OSIVQ Visual × ResMem Scores	0.0016059	0.0012743	[-0.001, 0.004]	1.2602483	0.2075851	
OSIVQ Verbal × ResMem Scores	0.0010735	0.0013114	[-0.001, 0.004]	0.8186106	0.4130121	
Visual Similarity × ResMem Scores	-0.0044780	0.0012212	[-0.007, -0.002]	-3.6667422	0.0002459	***
Attentional Priority × ResMem Scores	-0.0024360	0.0013023	[-0.005, 0.000]	-1.8705747	0.0614093	.
Visual Similarity × Attentional Priority	0.0010886	0.0013154	[-0.001, 0.004]	0.8275738	0.4079155	
OSIVQ Visual × Visual Similarity × ResMem Scores	-0.0026119	0.0012287	[-0.005, -0.000]	-2.1257747	0.0335264	*
OSIVQ Verbal × Visual Similarity × ResMem Scores	0.0008713	0.0012619	[-0.002, 0.003]	0.6904432	0.4899184	

Table A.5. Accuracy model results for OSIVQ predictors.

OSIVQ: Accuracy Model

Predictor	OR	SE	95% CI	z	p
(Intercept)	3.4446927	0.1534458	[3.157, 3.759]	27.7656105	0.0000000 ***
OSIVQ Visual	0.9154145	0.0421142	[0.836, 1.002]	-1.9210345	0.0547274 .
OSIVQ Verbal	0.9896582	0.0454243	[0.905, 1.083]	-0.2264899	0.8208204
Visual Similarity	1.0076933	0.0094570	[0.989, 1.026]	0.8166232	0.4141438
Attentional Priority	1.2324158	0.0106414	[1.212, 1.253]	24.2021810	0.0000000 ***
ResMem Scores	1.1655545	0.0103605	[1.145, 1.186]	17.2346604	0.0000000 ***
OSIVQ Visual × Visual Similarity	0.9923353	0.0100783	[0.973, 1.012]	-0.7575923	0.4486951
OSIVQ Verbal × Visual Similarity	0.9868215	0.0095342	[0.968, 1.006]	-1.3730799	0.1697275
OSIVQ Visual × Attentional Priority	1.0141258	0.0094559	[0.996, 1.033]	1.5043652	0.1324873
OSIVQ Verbal × Attentional Priority	0.9546062	0.0085140	[0.938, 0.971]	-5.2087437	0.0000002 ***
OSIVQ Visual × ResMem Scores	0.9976283	0.0095455	[0.979, 1.017]	-0.2481703	0.8040026
OSIVQ Verbal × ResMem Scores	0.9918919	0.0091189	[0.974, 1.010]	-0.8855398	0.3758656
Visual Similarity × ResMem Scores	1.0475766	0.0088797	[1.030, 1.065]	5.4834166	0.0000000 ***
Attentional Priority × ResMem Scores	1.0128103	0.0085145	[0.996, 1.030]	1.5141298	0.1299929
Visual Similarity × Attentional Priority	0.9953380	0.0086223	[0.979, 1.012]	-0.5394275	0.5895919
OSIVQ Visual × Visual Similarity × ResMem Scores	1.0186030	0.0094058	[1.000, 1.037]	1.9961096	0.0459220 *
OSIVQ Verbal × Visual Similarity × ResMem Scores	0.9896739	0.0087162	[0.973, 1.007]	-1.1785633	0.2385721

Table A.6. Error-type model results for OSIVQ predictors.

OSIVQ: Error Type Model

Predictor	OR	SE	95% CI	z	p
(Intercept)	2.1551394	0.1035611	[1.961, 2.368]	15.9793122	0.0000000 ***
OSIVQ Visual	0.9297757	0.0465821	[0.843, 1.026]	-1.4533216	0.1461345
OSIVQ Verbal	1.0248577	0.0503765	[0.931, 1.129]	0.4995215	0.6174121
Visual Similarity	1.0300457	0.0191782	[0.993, 1.068]	1.5899578	0.1118443
Attentional Priority	1.2050712	0.0199159	[1.167, 1.245]	11.2870690	0.0000000 ***
ResMem Scores	0.9247190	0.0169299	[0.892, 0.959]	-4.2748893	0.0000191 ***
OSIVQ Visual × Visual Similarity	0.9980141	0.0209033	[0.958, 1.040]	-0.0949101	0.9243863
OSIVQ Verbal × Visual Similarity	0.9939572	0.0187411	[0.958, 1.031]	-0.3214614	0.7478607
OSIVQ Visual × Attentional Priority	1.0399802	0.0192285	[1.003, 1.078]	2.1202309	0.0339866 *
OSIVQ Verbal × Attentional Priority	0.9311962	0.0158561	[0.901, 0.963]	-4.1864247	0.0000283 ***
OSIVQ Visual × ResMem Scores	0.9875028	0.0198762	[0.949, 1.027]	-0.6248078	0.5320972
OSIVQ Verbal × ResMem Scores	1.0067204	0.0185633	[0.971, 1.044]	0.3632417	0.7164244
Visual Similarity × ResMem Scores	0.9467878	0.0165609	[0.915, 0.980]	-3.1260820	0.0017715 **
Attentional Priority × ResMem Scores	1.0319191	0.0167803	[1.000, 1.065]	1.9322228	0.0533320 .
Visual Similarity × Attentional Priority	1.0007631	0.0164225	[0.969, 1.033]	0.0464844	0.9629242
OSIVQ Visual × Visual Similarity × ResMem Scores	0.9964863	0.0200676	[0.958, 1.037]	-0.1747852	0.8612484
OSIVQ Verbal × Visual Similarity × ResMem Scores	1.0119828	0.0183488	[0.977, 1.049]	0.6569505	0.5112127

Table A.7. Response time model results for VVIQ predictors.

VVIQ: RT Model

Predictor	<i>b</i>	<i>SE</i>	95% CI	<i>t</i>	<i>p</i>
(Intercept)	0.3951030	0.0082462	[0.379, 0.411]	47.9136242	0.0000000 ***
VVIQ	0.0079359	0.0089171	[-0.010, 0.025]	0.8899620	0.3734902
Visual Similarity	-0.0031514	0.0013159	[-0.006, -0.001]	-2.3948266	0.0166315 *
Attentional Priority	-0.0286152	0.0013114	[-0.031, -0.026]	-21.8204969	0.0000000 ***
ResMem Scores	-0.0236466	0.0012725	[-0.026, -0.021]	-18.5833000	0.0000000 ***
VVIQ × Visual Similarity	0.0008905	0.0013896	[-0.002, 0.004]	0.6408442	0.5216266
VVIQ × Attentional Priority	-0.0029350	0.0013849	[-0.006, -0.000]	-2.1193569	0.0340648 *
VVIQ × ResMem Scores	0.0012519	0.0013565	[-0.001, 0.004]	0.9229401	0.3560425
Visual Similarity × ResMem Scores	-0.0044233	0.0012267	[-0.007, -0.002]	-3.6058997	0.0003113 ***
Attentional Priority × ResMem Scores	-0.0024414	0.0013021	[-0.005, 0.000]	-1.8749829	0.0608003 .
Visual Similarity × Attentional Priority	0.0009206	0.0013148	[-0.002, 0.003]	0.7001647	0.4838274
VVIQ × Visual Similarity × ResMem Scores	-0.0012875	0.0013033	[-0.004, 0.001]	-0.9878539	0.3232285

Table A.8. Accuracy model results for VVIQ predictors.

VVIQ: Accuracy Model

Predictor	OR	SE	95% CI	<i>z</i>	<i>p</i>	
(Intercept)	3.4566336	0.1550253	[3.166, 3.774]	27.6551388	0.0000000	***
VVIQ	0.9223561	0.0448587	[0.838, 1.015]	-1.6618491	0.0965430	.
Visual Similarity	1.0085236	0.0095191	[0.990, 1.027]	0.8992290	0.3685307	
Attentional Priority	1.2308531	0.0106959	[1.210, 1.252]	23.9023033	0.0000000	***
ResMem Scores	1.1654152	0.0104218	[1.145, 1.186]	17.1178064	0.0000000	***
VVIQ × Visual Similarity	0.9850478	0.0104557	[0.965, 1.006]	-1.4193010	0.1558113	
VVIQ × Attentional Priority	1.0060048	0.0098604	[0.987, 1.026]	0.6108071	0.5413273	
VVIQ × ResMem Scores	0.9981935	0.0100282	[0.979, 1.018]	-0.1799829	0.8571660	
Visual Similarity × ResMem Scores	1.0447985	0.0089150	[1.027, 1.062]	5.1359993	0.0000003	***
Attentional Priority × ResMem Scores	1.0123802	0.0085065	[0.996, 1.029]	1.4643607	0.1430954	
Visual Similarity × Attentional Priority	0.9967629	0.0086251	[0.980, 1.014]	-0.3747008	0.7078830	
VVIQ × Visual Similarity × ResMem Scores	1.0258090	0.0099206	[1.007, 1.045]	2.6348518	0.0084174	**

Table A.9. Error-type model results for VVIQ predictors.

VVIQ: Error Type Model

Predictor	OR	SE	95% CI	<i>z</i>	<i>p</i>
(Intercept)	2.1521927	0.1050078	[1.956, 2.368]	15.7095724	0.0000000 ***
VVIQ	0.9669878	0.0513940	[0.871, 1.073]	-0.6316142	0.5276390
Visual Similarity	1.0269994	0.0191701	[0.990, 1.065]	1.4272567	0.1535059
Attentional Priority	1.2067633	0.0200544	[1.168, 1.247]	11.3093236	0.0000000 ***
ResMem Scores	0.9217874	0.0169572	[0.889, 0.956]	-4.4270870	0.0000096 ***
VVIQ × Visual Similarity	1.0190514	0.0214804	[0.978, 1.062]	0.8953123	0.3706202
VVIQ × Attentional Priority	1.0112575	0.0193568	[0.974, 1.050]	0.5848392	0.5586559
VVIQ × ResMem Scores	1.0165390	0.0210315	[0.976, 1.059]	0.7928617	0.4278584
Visual Similarity × ResMem Scores	0.9446804	0.0165896	[0.913, 0.978]	-3.2406045	0.0011928 **
Attentional Priority × ResMem Scores	1.0316486	0.0167119	[0.999, 1.065]	1.9234295	0.0544261 .
Visual Similarity × Attentional Priority	1.0034605	0.0164096	[0.972, 1.036]	0.2112500	0.8326922
VVIQ × Visual Similarity × ResMem Scores	1.0029757	0.0203466	[0.964, 1.044]	0.1464662	0.8835534

A.2 VVIQ and Attentional Priority

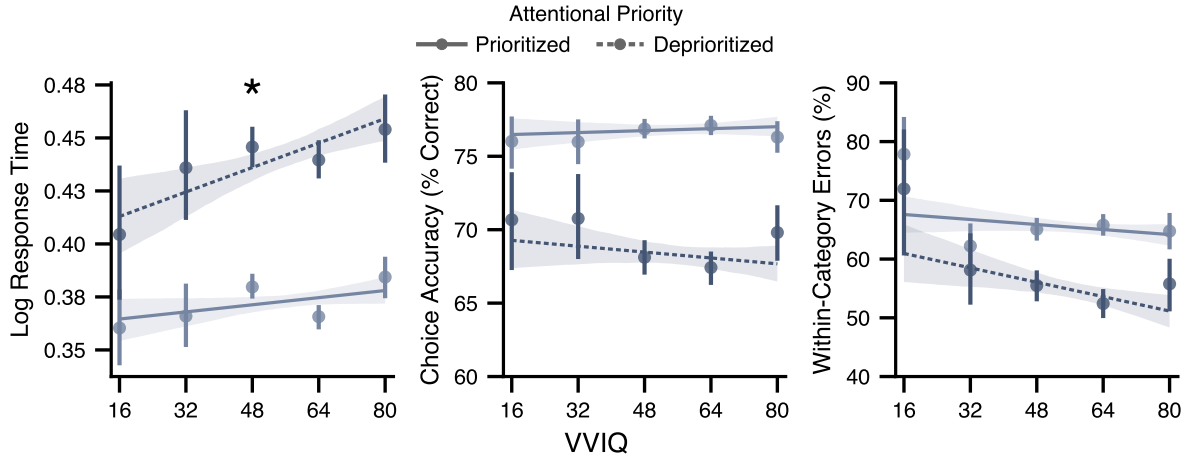


Figure A.1. WM performance as a function of imagery vividness (VVIQ) and attentional priority. Response times (left), accuracy (middle), and within-category error rates (right) plotted against VVIQ scores, separated by attentional priority of the tested WM item (solid = prioritized trials, dashed = deprioritized trials). VVIQ scores were binned into five equal-width bins for visualization purposes, with points representing mean values within each bin. Lines represent regression fits, and error bars represent 95% confidence intervals. The asterisk indicates a significant interaction between VVIQ scores and attentional priority ($p < .05$). This figure extends Figure 1.3c by showing vividness effects not depicted in the main text.

A.3 Three-way Interaction Between Imagery, Memorability, and Target-Distractor Similarity

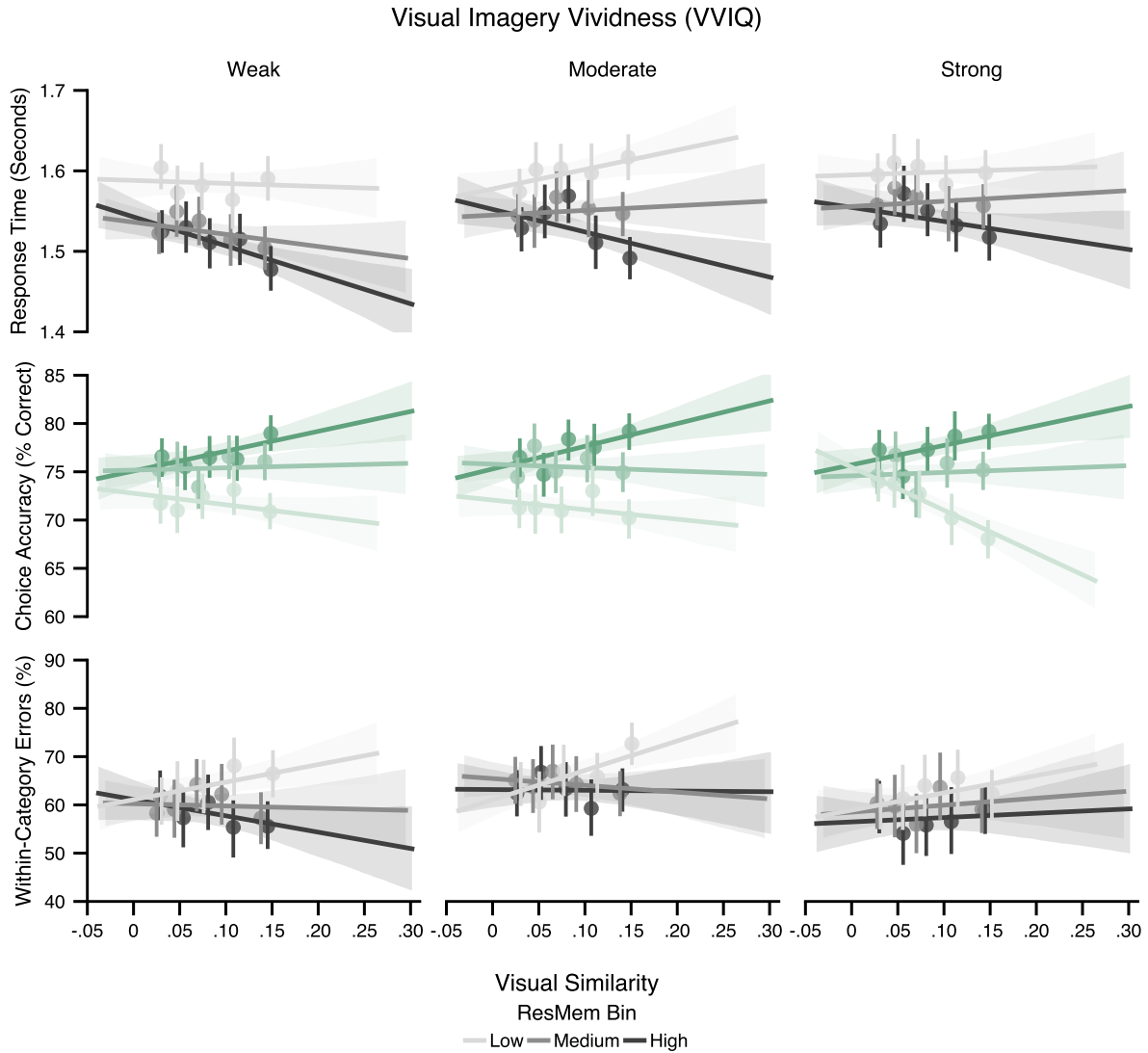


Figure A.2. Visual imagery vividness (VVIQ) and target-distractor similarity effects across memorability levels. Response time, choice accuracy, and within-category errors are plotted as a function of target-distractor similarity, shown separately for low-, medium-, and high-memorability items and for weak, moderate, and strong VVIQ groups. Plotting conventions follow Figure 1.5. Gray panels indicate effects that did not reach significance in the corresponding model.

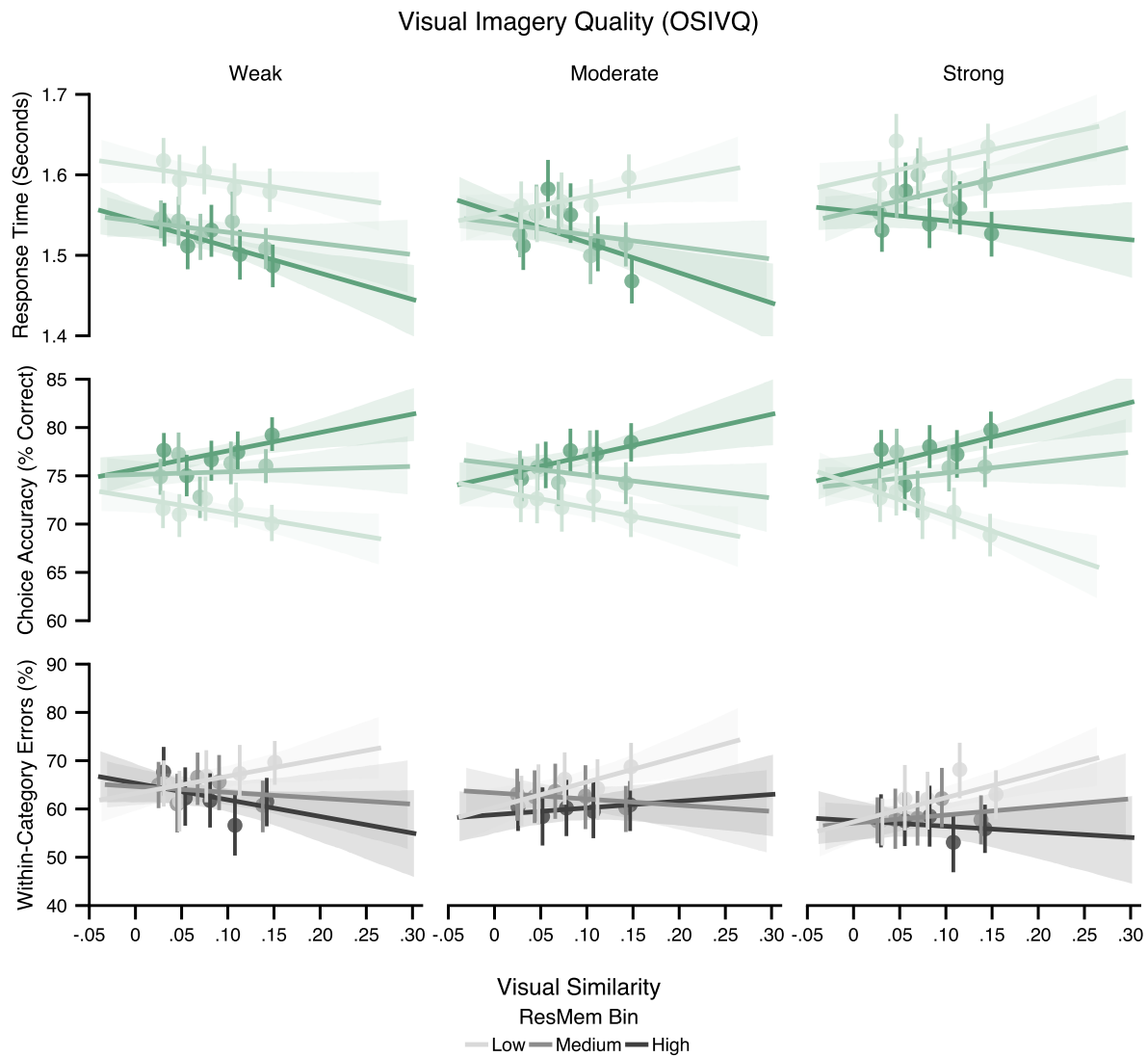


Figure A.3. Visual imagery quality (OSIVQ) moderates the interaction between stimulus memorability and target-distractor similarity. Response time, choice accuracy, and within-category errors are plotted as a function of target-distractor similarity, separated by memorability level and OSIVQ group. Plotting conventions follow Figure 1.5. Gray panels indicate effects that did not reach significance in the corresponding model.

Appendix B

Chapter 2

B.1 Distribution of Visual Imagery Vividness Scores in Experiment 1

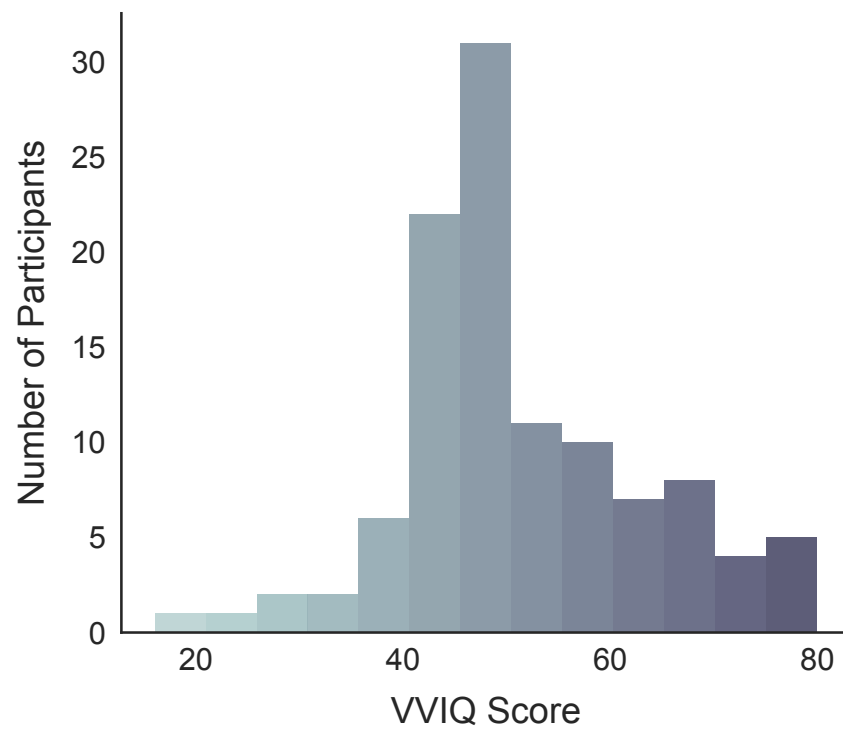


Figure B.1. Distribution of visual imagery vividness scores in Experiment 1. Distribution of VVIQ scores, indexing visual imagery vividness ($N = 110$).

B.2 Picture-Locked LMM Effects Across the Scalp

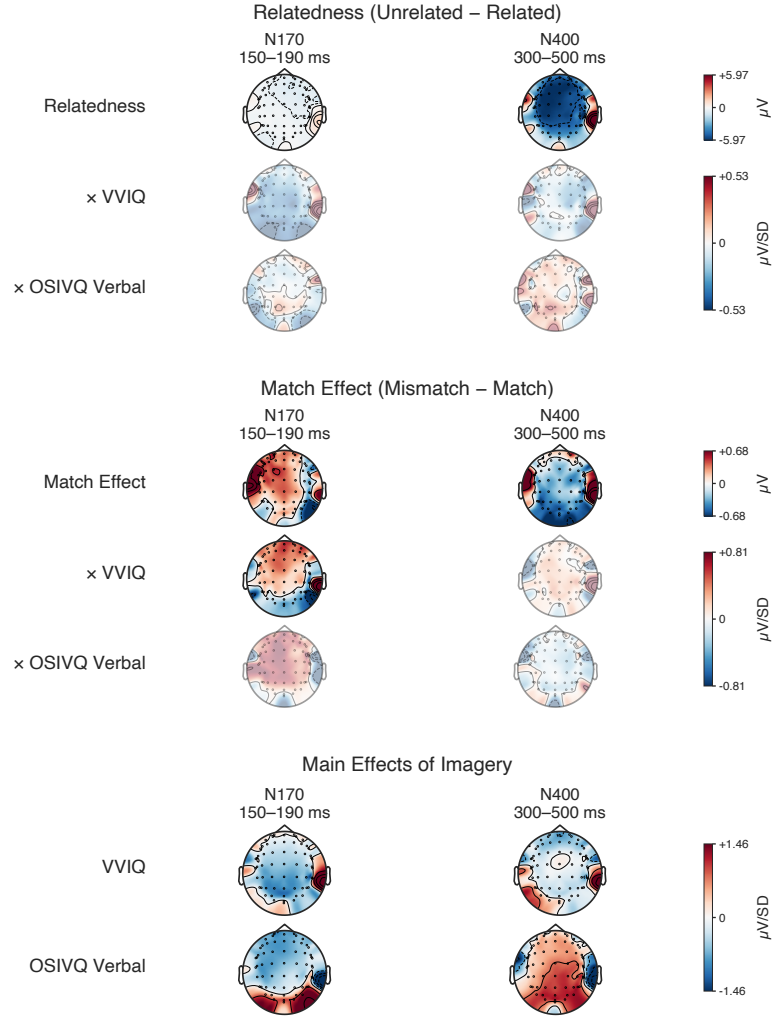


Figure B.2. Picture-locked LMM effects across the scalp. Topographic maps of beta coefficients from the picture-locked mixed-effects models for the N170 (150–190 ms; left) and N400 (300–500 ms; right) time windows. Top: Relatedness effect (unrelated–related) and its interactions with VVIQ and OSIVQ Verbal. Middle: Match effect (mismatch–match) and its interactions with VVIQ and OSIVQ Verbal. Bottom: Main effects of VVIQ and OSIVQ Verbal on the grand-mean ERP. Color scales are in μV for the main condition effects and μV per SD of the imagery score for interaction and main imagery effects. Topographies are reconstructed from the model-estimated beta coefficients for the spatial X, Y, and Z interaction terms. Non-significant effects are shown muted. All reported significance values are FDR-corrected. Full numerical output for all coefficients is provided in Tables B.3 and B.4.

B.3 Behavioral Model Tables

Tables B.1 and B.2 report the mixed-effects model outputs for the behavioral analyses in Experiment 2. The models tested whether visual imagery vividness modulated behavioral signatures of mental simulation in reaction time and accuracy.

Table B.1. Reaction time model results for Experiment 2.

Reaction Time						
Fixed effect	β (ms)	SE	95% CI	t	p	
<i>Intercept</i>	<i>748.83</i>	<i>14.87</i>	<i>[719.68, 777.98]</i>	<i>50.35</i>	<i>< .001</i>	<i>***</i>
Mismatch vs Match	2.37	4.89	[-7.22, 11.95]	0.48	.629	
Unrelated vs Related	-10.45	2.53	[-15.41, -5.50]	-4.14	< .001	***
VVIQ	-1.98	15.07	[-31.51, 27.56]	-0.13	.896	
Mismatch vs Match $\times$ VVIQ	-1.65	4.99	[-11.43, 8.14]	-0.33	.741	
Unrelated vs Related $\times$ VVIQ	-4.14	2.32	[-8.68, 0.41]	-1.78	.075	

Table B.2. Accuracy model results for Experiment 2.

Accuracy							
Fixed effect	β (log-odds)	SE	OR	95% CI (OR)	z	p	
<i>Intercept</i>	<i>5.02</i>	<i>0.34</i>	<i>151.93</i>	<i>[77.37, 298.34]</i>	<i>14.59</i>	<i>< .001</i>	<i>***</i>
Mismatch vs Match	0.07	0.13	1.07	[0.83, 1.38]	0.55	.580	
Unrelated vs Related	0.94	0.19	2.55	[1.75, 3.71]	4.87	< .001	***
VVIQ	-0.16	0.19	0.85	[0.59, 1.24]	-0.83	.408	
Mismatch vs Match $\times$ VVIQ	0.21	0.14	1.23	[0.95, 1.61]	1.55	.121	
Unrelated vs Related $\times$ VVIQ	-0.08	0.14	0.92	[0.69, 1.22]	-0.59	.557	

B.4 ERP Model Tables

Tables B.3 and B.4 report the full mixed-effects model outputs for the picture-locked ERP analyses. These models estimate condition effects, imagery effects, and spatial interactions across the scalp for the N₁₇₀ and N₄₀₀ time windows.

Table B.3. Mixed-effects model output for picture-locked N170 voltage (150–190 ms).

N170 (150–190 ms)							
Fixed effect	β	SE	95% CI	t	p (raw)	p (FDR)	
<i>Intercept</i>	<i>0.385</i>	<i>0.374</i>	<i>[-0.347, 1.117]</i>	<i>1.03</i>	<i>.31</i>	<i>1.00</i>	
VVIQ	0.056	0.363	[-0.654, 0.767]	0.16	.88	1.00	
OSIVQ Verbal	0.281	0.363	[-0.430, 0.992]	0.77	.44	1.00	
Relatedness (U vs R)	-0.119	0.222	[-0.554, 0.317]	-0.53	.59	1.00	
Match (M vs MM)	0.078	0.112	[-0.142, 0.298]	0.69	.49	1.00	
X	0.078	0.003	[0.071, 0.084]	23.42	< .001	< .001	***
Y	-0.321	0.003	[-0.327, -0.315]	-104.35	< .001	< .001	***
Z	-0.338	0.005	[-0.347, -0.328]	-68.89	< .001	< .001	***
VVIQ $\times$ Relatedness (U vs R)	-0.142	0.080	[-0.300, 0.015]	-1.77	.08	.66	
VVIQ $\times$ Match (M vs MM)	-0.260	0.114	[-0.483, -0.036]	-2.28	.02	.25	
OSIVQ Verbal $\times$ Relatedness (U vs R)	-0.087	0.082	[-0.247, 0.073]	-1.07	.29	1.00	
OSIVQ Verbal $\times$ Match (M vs MM)	-0.009	0.116	[-0.237, 0.219]	-0.08	.94	1.00	
VVIQ $\times$ X	0.002	0.003	[-0.005, 0.008]	0.47	.64	1.00	
VVIQ $\times$ Y	0.020	0.003	[0.014, 0.026]	6.29	< .001	< .001	***
VVIQ $\times$ Z	-0.059	0.005	[-0.068, -0.049]	-11.78	< .001	< .001	***
OSIVQ Verbal $\times$ X	0.030	0.003	[0.023, 0.037]	8.80	< .001	< .001	***
OSIVQ Verbal $\times$ Y	-0.088	0.003	[-0.094, -0.082]	-27.83	< .001	< .001	***
OSIVQ Verbal $\times$ Z	-0.053	0.005	[-0.063, -0.043]	-10.43	< .001	< .001	***
Relatedness (U vs R) $\times$ X	-0.014	0.006	[-0.026, -0.002]	-2.22	.03	.28	
Relatedness (U vs R) $\times$ Y	-0.019	0.006	[-0.031, -0.008]	-3.27	.001	.02	*
Relatedness (U vs R) $\times$ Z	-0.022	0.009	[-0.040, -0.004]	-2.34	.02	.23	
Match (M vs MM) $\times$ X	-0.037	0.009	[-0.054, -0.019]	-4.15	< .001	< .001	***
Match (M vs MM) $\times$ Y	0.026	0.008	[0.009, 0.042]	3.11	.002	.03	*
Match (M vs MM) $\times$ Z	0.019	0.013	[-0.007, 0.045]	1.44	.15	1.00	
VVIQ $\times$ Relatedness (U vs R) $\times$ X	0.015	0.006	[0.002, 0.027]	2.28	.02	.25	
VVIQ $\times$ Relatedness (U vs R) $\times$ Y	0.014	0.006	[0.003, 0.026]	2.39	.02	.21	
VVIQ $\times$ Relatedness (U vs R) $\times$ Z	-0.017	0.009	[-0.035, 0.002]	-1.77	.08	.66	
VVIQ $\times$ Match (M vs MM) $\times$ X	-0.015	0.009	[-0.033, 0.003]	-1.66	.10	.80	
VVIQ $\times$ Match (M vs MM) $\times$ Y	0.045	0.008	[0.029, 0.061]	5.38	< .001	< .001	***
VVIQ $\times$ Match (M vs MM) $\times$ Z	0.026	0.013	[0.000, 0.053]	1.99	.05	.44	
OSIVQ Verbal $\times$ Relatedness (U vs R) $\times$ X	0.007	0.006	[-0.006, 0.020]	1.05	.29	1.00	
OSIVQ Verbal $\times$ Relatedness (U vs R) $\times$ Y	0.010	0.006	[-0.002, 0.022]	1.65	.10	.80	
OSIVQ Verbal $\times$ Relatedness (U vs R) $\times$ Z	0.007	0.010	[-0.012, 0.026]	0.73	.47	1.00	
OSIVQ Verbal $\times$ Match (M vs MM) $\times$ X	-0.013	0.009	[-0.031, 0.004]	-1.47	.14	1.00	
OSIVQ Verbal $\times$ Match (M vs MM) $\times$ Y	0.021	0.008	[0.005, 0.038]	2.52	.01	.15	
OSIVQ Verbal $\times$ Match (M vs MM) $\times$ Z	0.037	0.014	[0.011, 0.064]	2.75	.006	.08	

Table B.4. Mixed-effects model output for picture-locked N₄₀₀ voltage (300–500 ms).

N400 (300–500 ms)							
Fixed effect	β	SE	95% CI	t	p (raw)	p (FDR)	
<i>Intercept</i>	1.481	0.554	[0.396, 2.567]	2.67	.01	.14	
VVIQ	-0.119	0.538	[-1.174, 0.936]	-0.22	.83	1.00	
OSIVQ Verbal	0.248	0.538	[-0.807, 1.303]	0.46	.65	1.00	
Relatedness (U vs R)	-1.964	0.315	[-2.581, -1.346]	-6.23	< .001	< .001	***
Match (M vs MM)	-0.081	0.114	[-0.305, 0.143]	-0.71	.48	1.00	
X	0.043	0.003	[0.037, 0.050]	12.85	< .001	< .001	***
Y	-0.250	0.003	[-0.256, -0.243]	-79.69	< .001	< .001	***
Z	0.027	0.005	[0.017, 0.037]	5.46	< .001	< .001	***
VVIQ $\times$ Relatedness (U vs R)	-0.169	0.082	[-0.329, -0.008]	-2.05	.04	.40	
VVIQ $\times$ Match (M vs MM)	-0.223	0.116	[-0.451, 0.004]	-1.92	.05	.50	
OSIVQ Verbal $\times$ Relatedness (U vs R)	0.028	0.083	[-0.135, 0.191]	0.33	.74	1.00	
OSIVQ Verbal $\times$ Match (M vs MM)	-0.241	0.119	[-0.474, -0.009]	-2.04	.04	.41	
VVIQ $\times$ X	-0.011	0.003	[-0.017, -0.004]	-3.11	.002	.03	*
VVIQ $\times$ Y	-0.035	0.003	[-0.041, -0.029]	-10.94	< .001	< .001	***
VVIQ $\times$ Z	-0.011	0.005	[-0.021, -0.001]	-2.17	.03	.31	
OSIVQ Verbal $\times$ X	0.018	0.003	[0.011, 0.024]	5.09	< .001	< .001	***
OSIVQ Verbal $\times$ Y	-0.043	0.003	[-0.050, -0.037]	-13.44	< .001	< .001	***
OSIVQ Verbal $\times$ Z	0.054	0.005	[0.044, 0.064]	10.56	< .001	< .001	***
Relatedness (U vs R) $\times$ X	0.018	0.006	[0.006, 0.031]	2.84	.005	.06	
Relatedness (U vs R) $\times$ Y	-0.104	0.006	[-0.116, -0.093]	-17.53	< .001	< .001	***
Relatedness (U vs R) $\times$ Z	-0.281	0.010	[-0.300, -0.262]	-29.55	< .001	< .001	***
Match (M vs MM) $\times$ X	0.000	0.009	[-0.018, 0.017]	-0.05	.96	1.00	
Match (M vs MM) $\times$ Y	0.038	0.008	[0.021, 0.054]	4.52	< .001	< .001	***
Match (M vs MM) $\times$ Z	-0.015	0.013	[-0.042, 0.011]	-1.15	.25	1.00	
VVIQ $\times$ Relatedness (U vs R) $\times$ X	-0.003	0.007	[-0.016, 0.010]	-0.46	.65	1.00	
VVIQ $\times$ Relatedness (U vs R) $\times$ Y	0.006	0.006	[-0.006, 0.017]	0.91	.36	1.00	
VVIQ $\times$ Relatedness (U vs R) $\times$ Z	0.001	0.010	[-0.018, 0.020]	0.11	.92	1.00	
VVIQ $\times$ Match (M vs MM) $\times$ X	0.007	0.009	[-0.011, 0.025]	0.77	.44	1.00	
VVIQ $\times$ Match (M vs MM) $\times$ Y	0.000	0.009	[-0.016, 0.017]	0.02	.98	1.00	
VVIQ $\times$ Match (M vs MM) $\times$ Z	0.024	0.014	[-0.002, 0.051]	1.80	.07	.65	
OSIVQ Verbal $\times$ Relatedness (U vs R) $\times$ X	-0.008	0.007	[-0.021, 0.005]	-1.25	.21	1.00	
OSIVQ Verbal $\times$ Relatedness (U vs R) $\times$ Y	-0.003	0.006	[-0.015, 0.009]	-0.42	.67	1.00	
OSIVQ Verbal $\times$ Relatedness (U vs R) $\times$ Z	0.011	0.010	[-0.008, 0.030]	1.14	.25	1.00	
OSIVQ Verbal $\times$ Match (M vs MM) $\times$ X	0.003	0.009	[-0.016, 0.021]	0.27	.78	1.00	
OSIVQ Verbal $\times$ Match (M vs MM) $\times$ Y	-0.001	0.009	[-0.018, 0.015]	-0.17	.87	1.00	
OSIVQ Verbal $\times$ Match (M vs MM) $\times$ Z	-0.005	0.014	[-0.032, 0.022]	-0.33	.74	1.00	

Appendix C

Chapter 3

C.1 Example Hallucination Descriptions

“Swirls, a faded, ghost-like tulip that moved in circular ways, balls of light. Later, there were straight, line-like fragments, black ones around corners. The red combined with sunset-like colors to yellow on edges ... I was briefly reminded of a total solar eclipse I saw a few years back, not sure why ... Screen expanding and contracting, lines changing into small squares moving into center of screen ... I also then tried to create something to see, that did not work, all was random.”

“Saw old stone buildings, first a small square room spinning and then I was outside of the room and the building was spinning, later I started seeing like the outside of the building (it was like it was a castle?) there were some stone columns and all was blue-ish stone, someone was standing looking at me but it was just a shadow, I was flying above it so the person was small and I couldn’t see anything.”

“I saw purple dots that were circling around like stars in a galaxy, also with spiral arms etc. They were going quite fast, I wanted them to slow down (to imitate a galaxy) but they didn’t. Then in the next few seconds the purple dots would form other shapes like waves or pillars that were turning as well.”

“There was this bird, almost like a Native American symbol, rising up from the SE corner that was blue and gold, looking down on a man in armor with his back to me holding a sword. Another was a space station—futuristic and circular with ships flying in and out like a port, then eventually meteors crashing into it. I also kept seeing various images of knights and soldiers—war and fighting.”

“It started with a sort of a fractal, coming out of a white circle in the middle, with eight grey branches coming outwards, which then split into eight more, etc.”

“I felt like my eyes were closed, but I could still see all this, the triangles violently shifting, breaking into a circle of white, surrounded by a cyan halo. It was around then the white noise stopped, and the circle turned into a heart, a black outline around a red and white circle.”

C.2 Hierarchical Structure of Topics

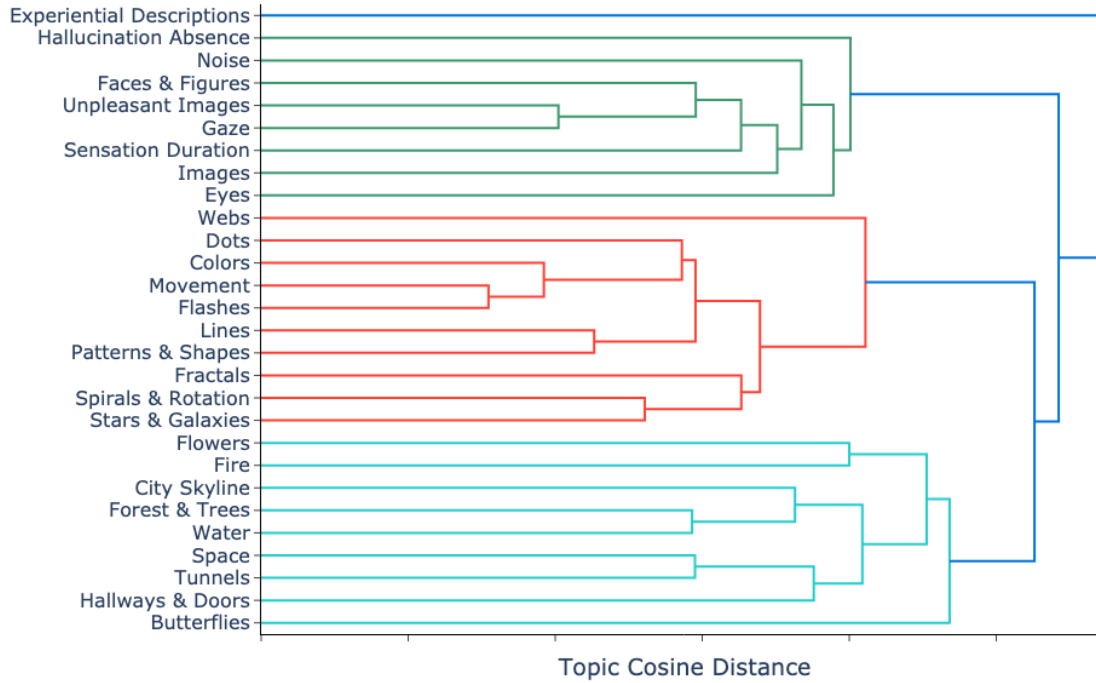


Figure C.1. Hierarchical relationships among topics. Dendrogram illustrating hierarchical relationships among topics. The dendrogram was constructed using cosine distances between topic embeddings, revealing the hierarchical semantic structure that emerges from the topic modeling process. The clustering revealed three emergent experiential clusters: social and metacognitive, simple visuals, and naturalistic scenes.

C.3 Lasso Regression Coefficients for Continuous Vividness Prediction

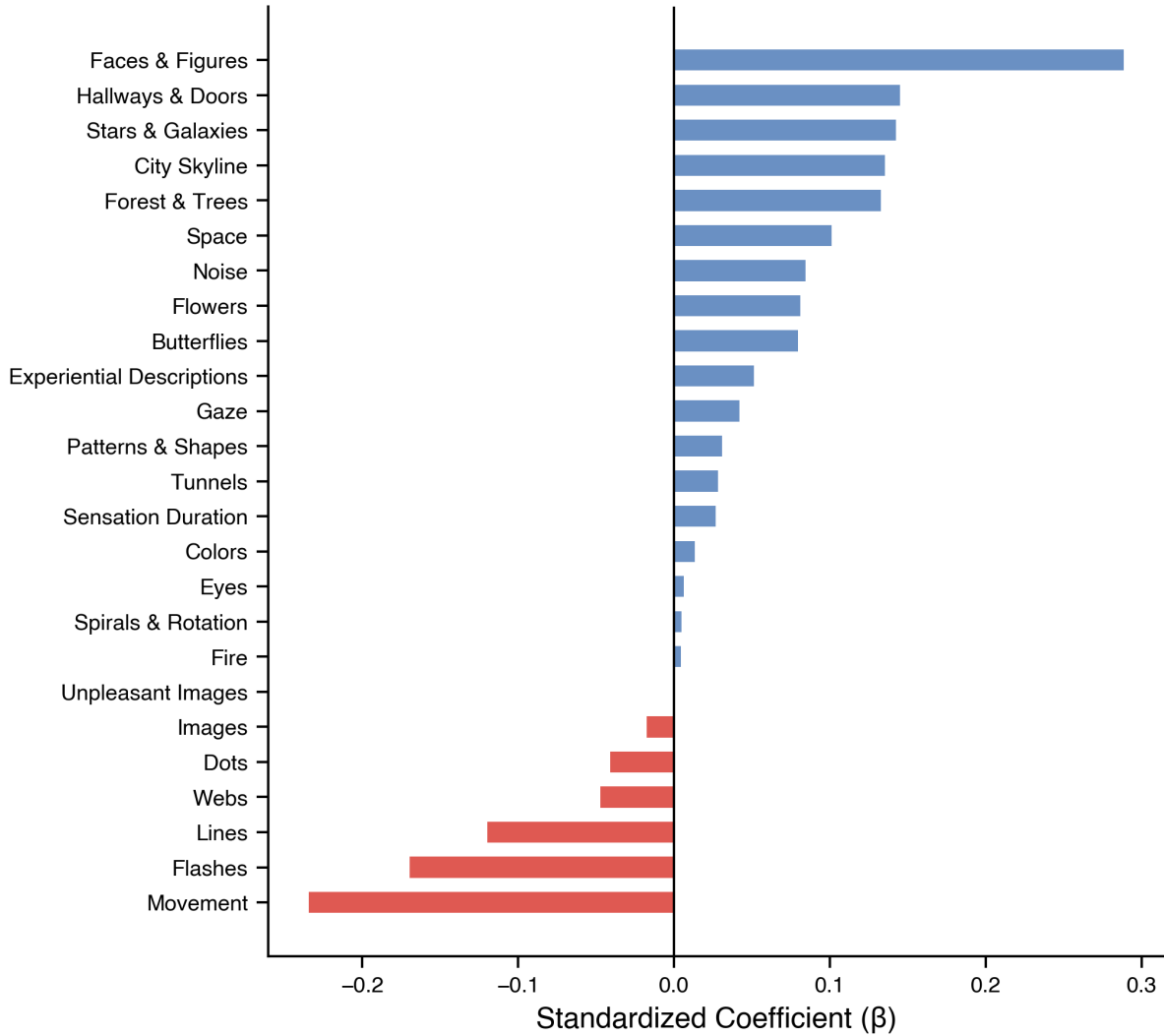


Figure C.2. Lasso regression coefficients for continuous visual vividness prediction. Forest plot showing standardized Lasso regression coefficients (β) for continuous vividness prediction. Only non-zero coefficients after cross-validated L1 regularization are displayed ($n = 25$ of 28 topics). Positive coefficients indicate topics associated with higher self-reported vividness, whereas negative coefficients indicate topics associated with lower vividness. Complex visual scenes and dynamic imagery were most predictive of higher vividness scores, while basic visual elements and perceptual distortions were associated with lower scores. The model achieved $R^2 = .05$ (MSE = 8.63) with optimal $\alpha = 0.03$.

C.4 Imagery Group-Level Topic Expression

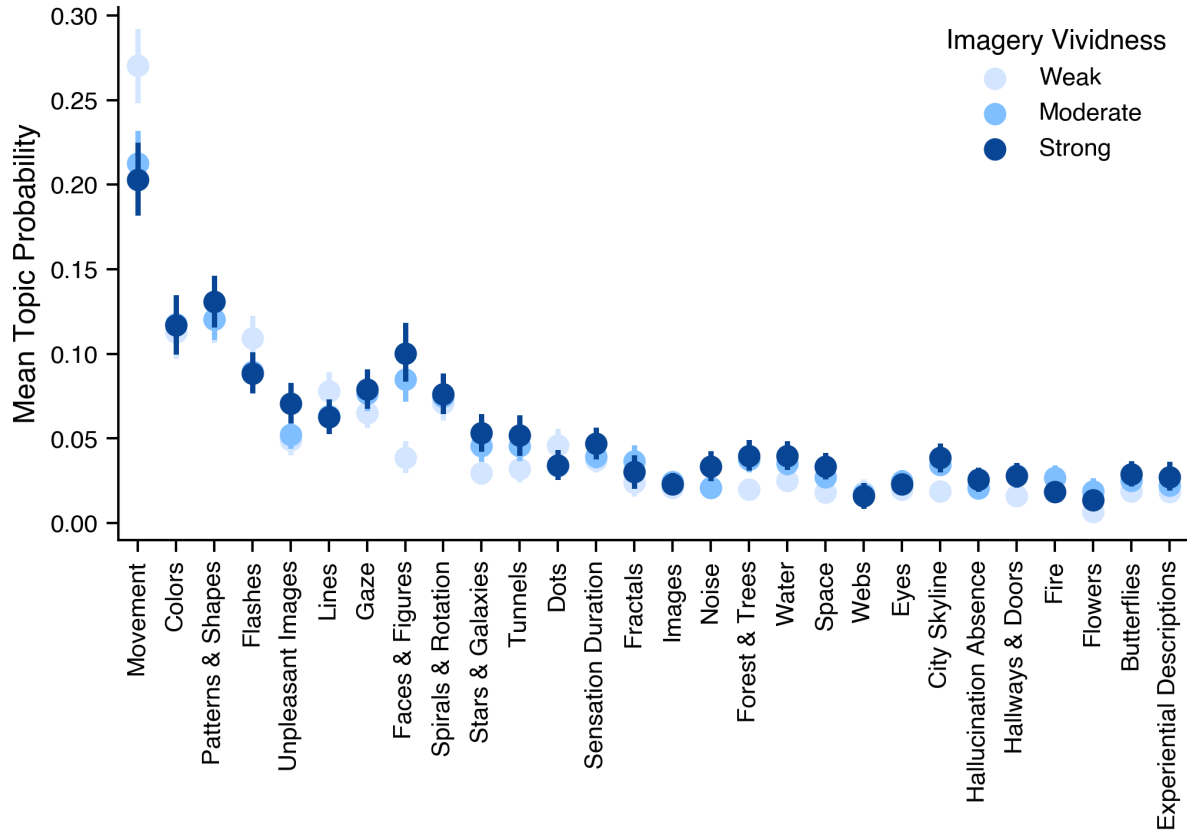


Figure C.3. Imagery group-level topic expression. Mean topic probabilities ($\pm 95\%$ CI) by imagery vividness group. Strong imagers were more likely to describe complex content, such as faces and natural or urban scenes, compared to weak imagers, while weak imagers described simpler visual phenomena, such as visual flashes and lines. Moderate imagers' descriptions showed less coherence.

C.5 Lancaster Sensorimotor Norm Dimensions

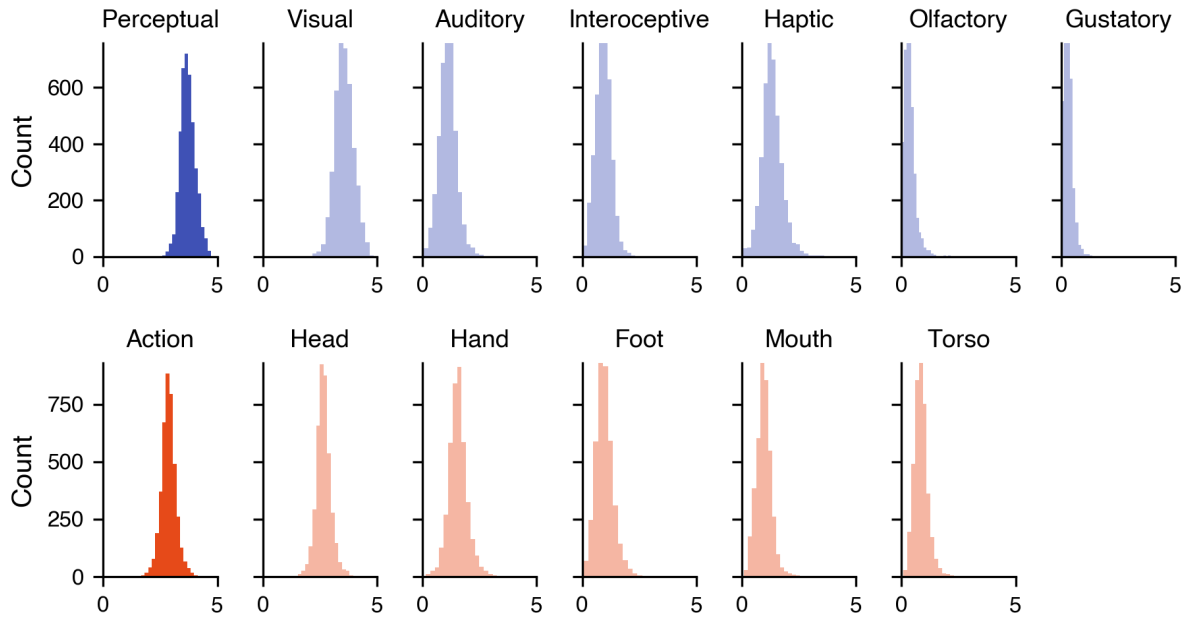


Figure C.4. Distributions of Lancaster Sensorimotor Norm dimensions across hallucination descriptions. Histograms show the distribution of values for 13 sensorimotor dimensions drawn from the Lancaster Sensorimotor Norms, averaged across words within each hallucination description. Dimensions are grouped by type: sensory dimensions are shown in the top row and motor dimensions in the bottom row. Solid-colored histograms represent composite measures of perceptual strength and action strength; lighter shades indicate specific subdimensions. These plots reflect the range and density of embodied and perceptual content present in the dataset. Perceptual and motor variables vary in overall magnitude and dispersion, with visual terms dominating perceptual content and head- and hand-related content dominating motor content in the hallucination descriptions.

Table C.1. Descriptive statistics for Lancaster Sensorimotor Norm dimensions. Mean, range, and variance for each sensorimotor dimension computed across all hallucination descriptions ($n = 4,057$). Higher scores indicate stronger association with the corresponding modality or body part. Composite scores, including perceptual and motor strength, reflect the scores of the dominant perceptual and motor modalities, respectively.

Dimension	Mean	Min	Max	Range	Variance
Perceptual					
Visual Strength	3.57	1.00	4.87	3.87	0.16
Haptic Strength	1.32	0.05	3.62	3.57	0.19
Auditory Strength	1.11	0.06	3.79	3.73	0.14
Interoceptive Strength	0.93	0.05	3.53	3.48	0.12
Olfactory Strength	0.36	0.00	2.16	2.16	0.06
Gustatory Strength	0.30	0.00	2.48	2.48	0.04
Motor					
Head Strength	2.66	1.53	4.76	3.23	0.10
Hand Strength	1.56	0.14	4.05	3.91	0.17
Foot Strength	0.97	0.05	4.36	4.31	0.15
Mouth Strength	0.96	0.10	3.60	3.5	0.11
Torso Strength	0.85	0.08	3.48	3.4	0.09
Composite Measures					
Perceptual Strength	3.71	2.00	4.87	2.87	0.11
Motor Strength	2.87	1.66	4.76	3.1	0.11

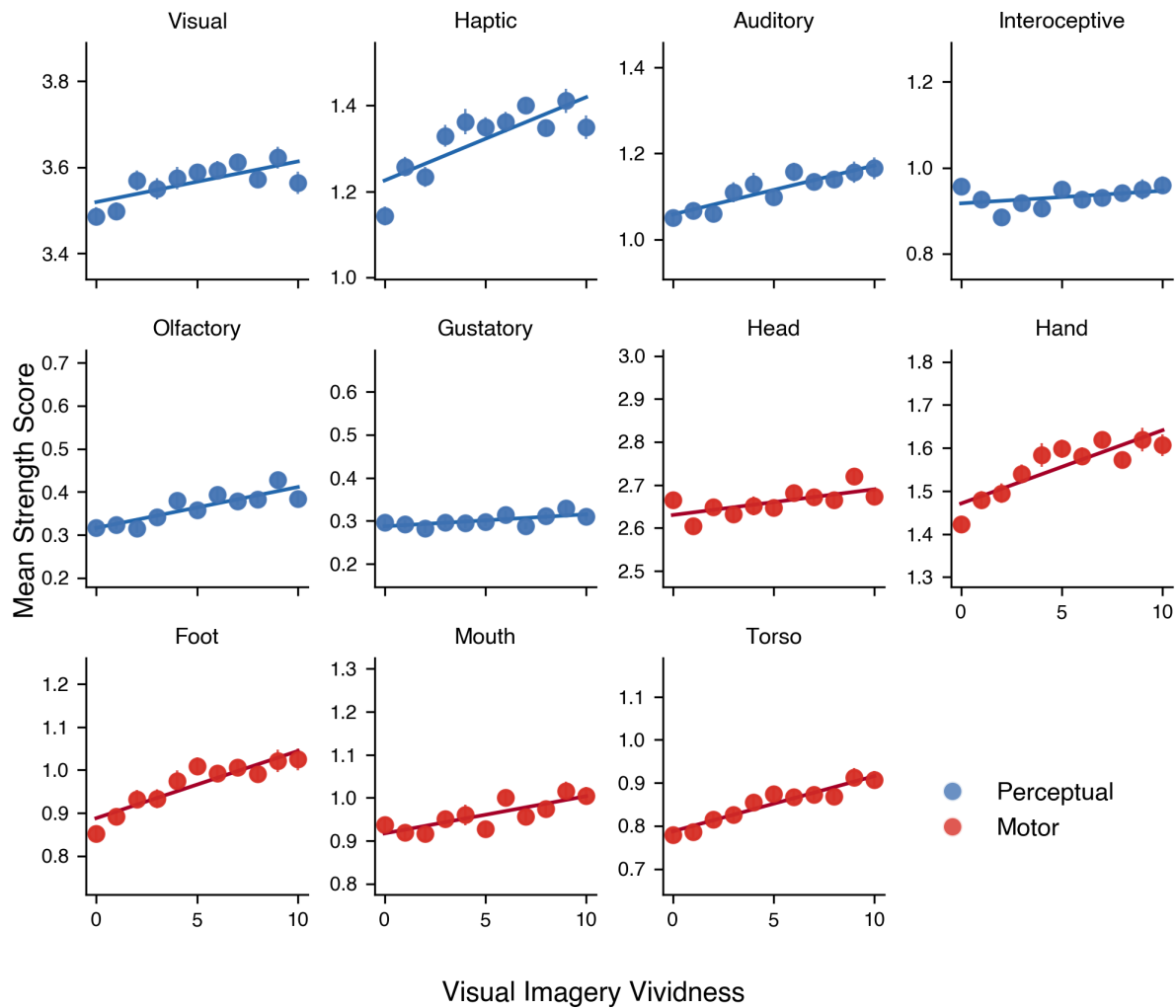


Figure C.5. Individual sensorimotor dimensions by visual imagery vividness. Individual relationships between visual imagery vividness and each of the 11 Lancaster Sensorimotor Norm dimensions. Top panels show the six perceptual modalities (visual, haptic, auditory, interoceptive, olfactory, gustatory), and bottom panels show the five motor dimensions (head, hand, foot, mouth, torso). Points represent mean scores $\pm$ standard error for each vividness level. Regression lines show linear trends. Perceptual dimensions are shown in blue, and motor dimensions are shown in orange/red. Note the varying y-axis scales across dimensions, reflecting different baseline frequencies of sensorimotor content in hallucination descriptions.

C.6 Robustness of Sensorimotor Effects

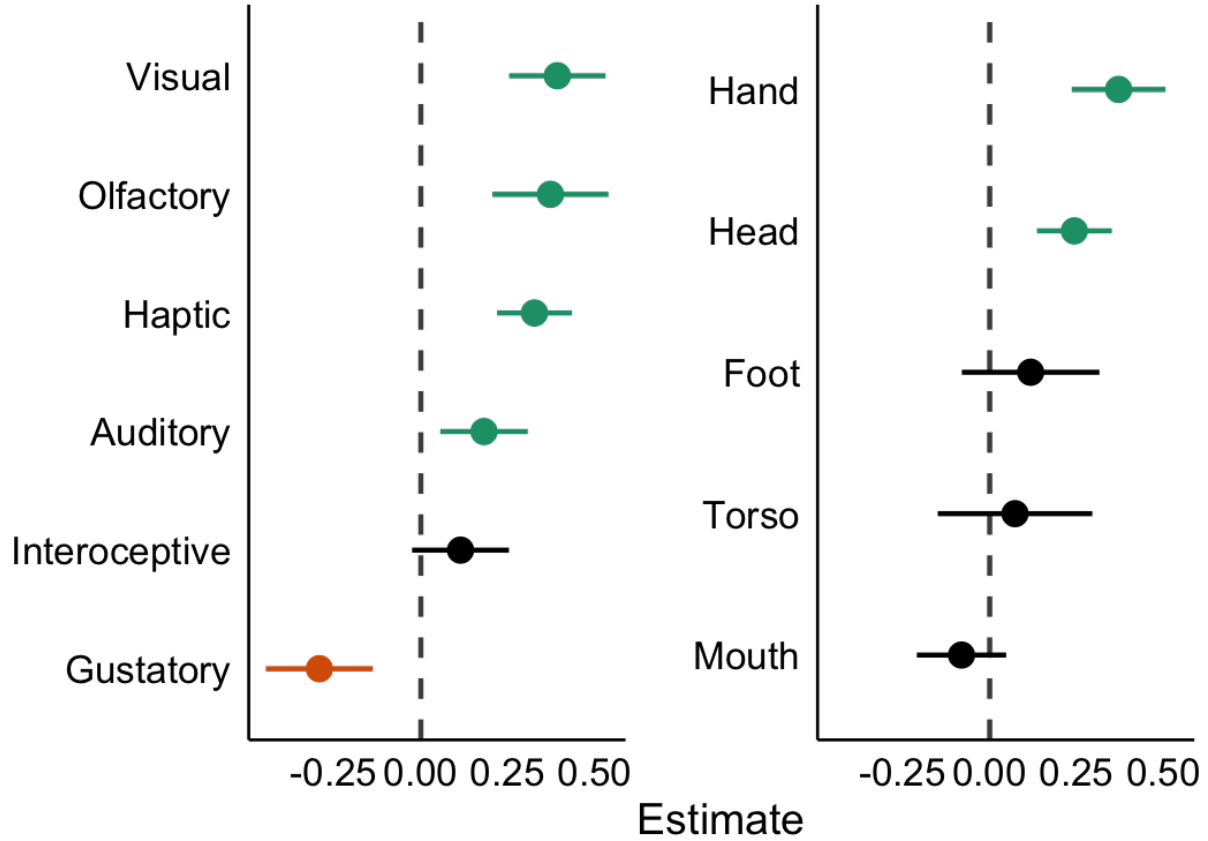


Figure C.6. Sensorimotor effects remain unchanged after controlling for gender. Effect sizes (β coefficients $\pm 95\%$ CIs) from perceptual (left) and motor (right) Lancaster models including gender as a covariate. Teal points indicate significant positive predictors of imagery vividness, orange indicates significant negative predictors, and black indicates non-significant predictors. All core sensorimotor effects observed in the primary models remained robust after controlling for gender.

C.7 Ganzflicker Experience and Imagery Vividness

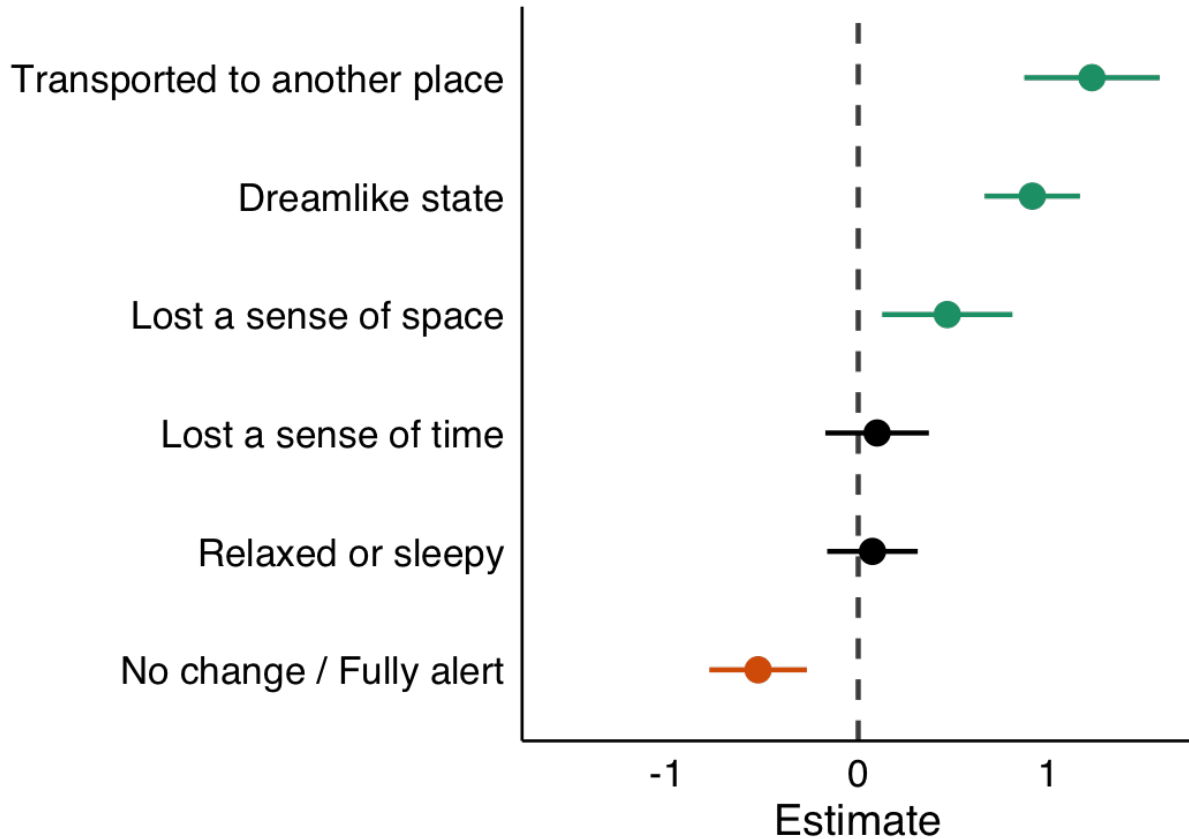


Figure C.7. Altered consciousness states during Ganzflicker are associated with stronger visual imagery. Coefficient plot showing regression estimates and 95% confidence intervals for the association between consciousness-change categories and visual imagery vividness. After completing the Ganzflicker paradigm and providing descriptions, participants answered “How did your normal state of consciousness change during the experience?” and could select multiple options from six categories. Green points indicate statistically significant positive associations, and the red point indicates a significant negative association. Participants who reported feeling “Transported to another place” ($p < .001$), experiencing a “Dreamlike state” ($p < .001$), or having “Lost a sense of space” ($p < .01$) had significantly higher imagery vividness scores. Conversely, those reporting “No change / Fully alert” had significantly lower vividness scores ($p < .001$). “Lost a sense of time” and “Relaxed or sleepy” were not significantly associated with vividness scores. These results suggest that individuals with more vivid internal visual experiences are more likely to report altered states of consciousness during Ganzflicker, possibly reflecting a more immersive and engaging experience with internally generated visual content.

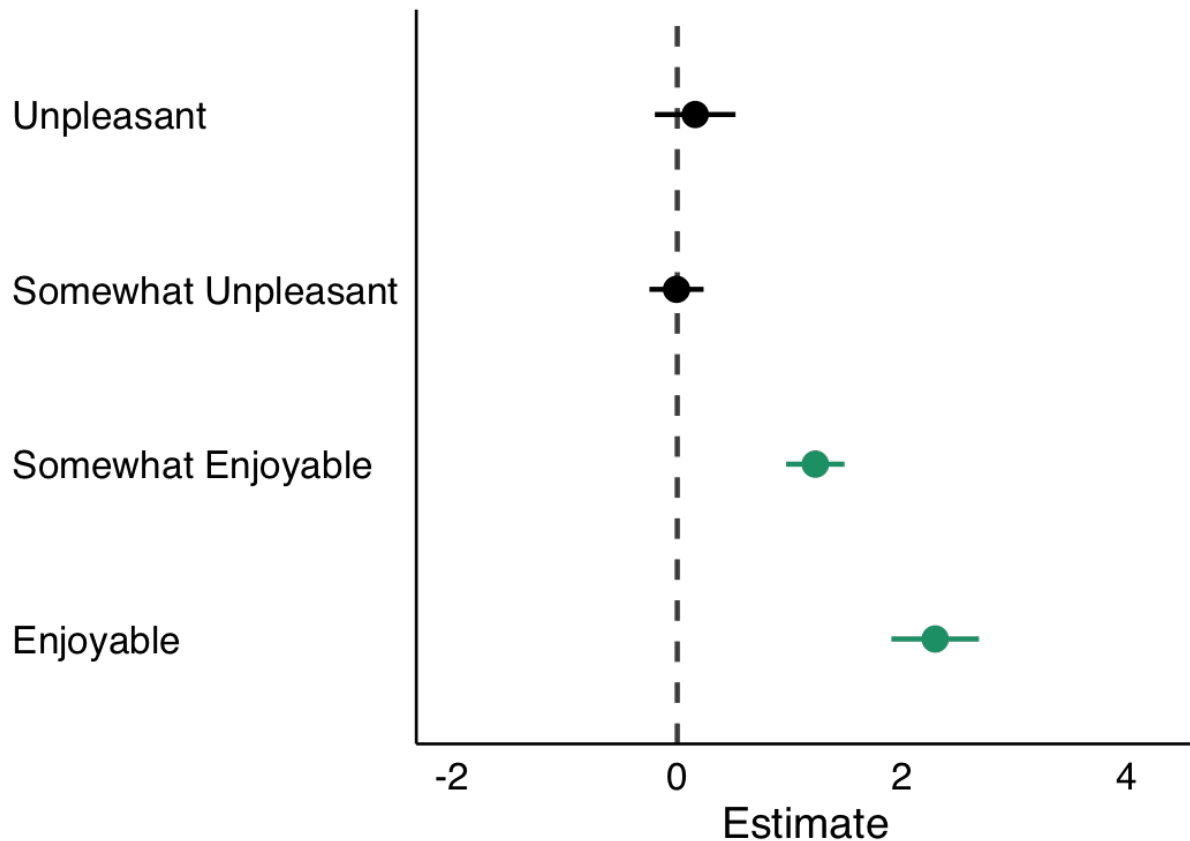


Figure C.8. Participants with stronger visual imagery report more positive emotional evaluations of the Ganzflicker experience. Coefficient plot showing regression estimates and 95% confidence intervals for the association between emotional-evaluation categories and visual imagery vividness, relative to neutral responses (“No feeling about it one way or the other”; reference category). After completing the Ganzflicker paradigm, participants answered “How did you find the experience emotionally?” with response options ranging from “Unpleasant, wanted it to stop” to “Enjoyable, didn’t want it to stop.” Green points indicate statistically significant positive associations. Participants who rated the experience as “Somewhat enjoyable, but easy to stop” had vividness scores 1.23 points higher ($p < .001$), while those rating it “Enjoyable, didn’t want it to stop” had vividness scores 2.29 points higher ($p < .001$) compared to neutral responses. Unpleasant categories showed no significant difference from neutral. These results suggest that individuals with more vivid internal visual experiences tend to find visually stimulating paradigms more emotionally engaging and positive.

References

- Adam, K. C. S., Klatt, L.-I., Miller, J. A., Rösner, M., Fukuda, K., & Kiyonaga, A. (2025). Beyond Routine Maintenance: Current Trends in Working Memory Research. *Journal of Cognitive Neuroscience*, 37(6), 1035–1052. doi: 10.1162/jocn_a_02298
- Albers, A. M., Kok, P., Toni, I., Dijkerman, H. C., & de Lange, F. P. (2013). Shared representations for working memory and mental imagery in early visual cortex. *Current biology: CB*, 23(15), 1427–1431. doi: 10.1016/j.cub.2013.05.065
- Allefeld, C., Pütz, P., Kastner, K., & Wackermann, J. (2011). Flicker-light induced visual phenomena: Frequency dependence and specificity of whole percepts and percept features. *Consciousness and Cognition*, 20(4), 1344–1362. doi: 10.1016/j.concog.2010.10.026
- Amaya, I. A., Behrens, N., Schwartzman, D. J., Hewitt, T., & Schmidt, T. T. (2023). Effect of frequency and rhythmicity on flicker light-induced hallucinatory phenomena. *PLOS ONE*, 18(4), e0284271. doi: 10.1371/journal.pone.0284271
- Amaya, I. A., Nierhaus, T., & Schmidt, T. T. (2025). Thalamocortical interactions reflecting the intensity of flicker light–induced visual hallucinatory phenomena. *Network Neuroscience*, 9(1), 1–17. doi: 10.1162/netn_a_00417
- Amit, E., & Greene, J. D. (2012). You see, the ends don’t justify the means: Visual imagery and moral judgment. *Psychological Science*, 23(8), 861–868. doi: 10.1177/0956797611434965
- Argueta, P., Dominguez, J., Zachman, J., Worthington, P., & Kana, R. K. (2025). “The Giant Black Elephant with white Tusks stood in a field of Green Grass”: Cognitive and brain mechanisms underlying aphantasia. *Consciousness and Cognition*, 127, 103790. doi: 10.1016/j.concog.2024.103790
- Aydın, Ç. (2020). GENDER DIFFERENCES IN VISUAL IMAGERY: OBJECT AND SPATIAL

- IMAGERY. *Dokuz Eylül Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 22(3), 1045–1064. doi: 10.16953/deusosbil.531463
- Baddeley, A. D. (1983). Working memory. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, 302(1110), 311–324. doi: 10.1098/rstb.1983.0057
- Bainbridge, W. A. (2019). Memorability: How what we see influences what we remember. In *The psychology of learning and motivation: Knowledge and vision* (pp. 1–27). Elsevier Academic Press. doi: 10.1016/bs.plm.2019.02.001
- Bainbridge, W. A., Dilks, D. D., & Oliva, A. (2017). Memorability: A stimulus-driven perceptual neural signature distinctive from memory. *NeuroImage*, 149, 141–152. doi: 10.1016/j.neuroimage.2017.01.063
- Bainbridge, W. A., Pounder, Z., Eardley, A. F., & Baker, C. I. (2021). Quantifying aphantasia through drawing: Those without visual imagery show deficits in object but not spatial memory. *Cortex*, 135, 159–172. doi: 10.1016/j.cortex.2020.11.014
- Bainbridge, W. A., Walther, D. B., Fukuda, K., & Goetschalckx, L. (2026). Memorability of visual stimuli and the role of processing efficiency. *Nature Reviews Psychology*, 5(1), 47–58. doi: 10.1038/s44159-025-00512-3
- Baker, B., Lansdell, B., & Kording, K. P. (2022). Three aspects of representation in neuroscience. *Trends in Cognitive Sciences*, 26(11), 942–958. doi: 10.1016/j.tics.2022.08.014
- Ballard, I., Gaspariani, L., Piriyaikitakonkij, M., Pappas, I., & Kiyonaga, A. (2026). *Sensory competition during working memory strengthens long-term memory* (No. 68q2n_v2). PsyArXiv.
- Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., & Yap, M. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology. General*, 133(2), 283–316. doi: 10.1037/0096-3445.133.2.283
- Banks, B., & Connell, L. (2022). Multi-dimensional sensorimotor grounding of concrete and abstract categories. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 378(1870), 20210366. doi: 10.1098/rstb.2021.0366
- Barber, L., Reniers, R., & Uptegrove, R. (2021). A review of functional and structural neuroimaging

- studies to investigate the inner speech model of auditory verbal hallucinations in schizophrenia. *Translational Psychiatry*, 11(1), 582. doi: 10.1038/s41398-021-01670-7
- Barrus, T. (2025). *Barrust/pyspellchecker*.
- Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, 22(4), 577–660. doi: 10.1017/S0140525X99002149
- Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, 59, 617–645. doi: 10.1146/annurev.psych.59.103006.093639
- Bartossek, M. T., Kemmerer, J., & Schmidt, T. T. (2021). Altered states phenomena induced by visual flicker light stimulation. *PLOS ONE*, 16(7), e0253779. doi: 10.1371/journal.pone.0253779
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67, 1–48. doi: 10.18637/jss.v067.i01
- Bayet, L., Zinszer, B. D., Reilly, E., Cataldo, J. K., Pruitt, Z., Cichy, R. M., ... Aslin, R. N. (2020). Temporal dynamics of visual representations in the infant brain. *Developmental Cognitive Neuroscience*, 45, 100860. doi: 10.1016/j.dcn.2020.100860
- Beauté, R., Schwartzman, D. J., Dumas, G., Crook, J., Macpherson, F., Barrett, A. B., & Seth, A. K. (2026). Mapping of Subjective Accounts into Interpreted Clusters (MOSAIC): Topic Modelling and LLM applied to Stroboscopic Phenomenology. *Neuroscience of Consciousness*, 2026(1), niag008. doi: 10.1093/nc/niag008
- Benjamini, Y., & Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *The Annals of Statistics*, 29(4), 1165–1188. doi: 10.1214/aos/1013699998
- Bergmann, J., Genç, E., Kohler, A., Singer, W., & Pearson, J. (2016). Smaller Primary Visual Cortex Is Associated with Stronger, but Less Precise Mental Imagery. *Cerebral Cortex*, 26(9), 3838–3850. doi: 10.1093/cercor/bhvi86
- Berkovich-Ohana, A., Dor-Ziderman, Y., Glicksohn, J., & Goldstein, A. (2013). Alterations in the sense of time, space, and body in the mindfulness-trained brain: A neurophenomenologically-guided MEG study. *Frontiers in Psychology*, 4. doi: 10.3389/fpsyg.2013.00912

- Bettencourt, K. C., & Xu, Y. (2016). Decoding the content of visual short-term memory under distraction in occipital and parietal areas. *Nature Neuroscience*, 19(1), 150–157. doi: 10.1038/nn.4174
- Bird, C. M., Bisby, J. A., & Burgess, N. (2012). The hippocampus and spatial constraints on mental imagery. *Frontiers in Human Neuroscience*, 6. doi: 10.3389/fnhum.2012.00142
- Bird, S., Klein, E., & Loper, E. (2009). *Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit*. O'Reilly Media.
- Blazhenkova, O., & Kozhevnikov, M. (2009). The new object-spatial-verbal cognitive style model: Theory and measurement. *Applied Cognitive Psychology*, 23(5), 638–663. doi: 10.1002/acp.1473
- Blazhenkova, O., Kozhevnikov, M., & Motes, M. A. (2006). Object-spatial imagery: A new self-report imagery questionnaire. *Applied Cognitive Psychology*, 20(2), 239–263. doi: 10.1002/acp.1182
- Blazhenkova, O., & Pechenkova, E. (2019). The two eyes of the blind mind: Object vs. spatial aphantasia? *Russian Journal of Cognitive Science*, 6(4), 51–65.
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *J. Mach. Learn. Res.*, 3(0), 993–1022.
- Bracci, S., Ietswaart, M., Peelen, M. V., & Cavina-Pratesi, C. (2010). Dissociable neural responses to hands and non-hand body parts in human left extrastriate visual cortex. *Journal of Neurophysiology*, 103(6), 3389–3397. doi: 10.1152/jn.00215.2010
- Brady, T. F., & Störmer, V. S. (2024). Comparing memory capacity across stimuli requires maximally dissimilar foils: Using deep convolutional neural networks to understand visual working memory capacity for real-world objects. *Memory & Cognition*, 52(3), 595–609. doi: 10.3758/s13421-023-01485-5
- Brady, T. F., Störmer, V. S., & Alvarez, G. A. (2016). Working memory is not fixed-capacity: More active storage capacity for real-world objects than for simple stimuli. *Proceedings of the National Academy of Sciences*, 113(27), 7459–7464. doi: 10.1073/pnas.1520027113
- Bressloff, P. C., Cowan, J. D., Golubitsky, M., Thomas, P. J., & Wiener, M. C. (2001). Geometric visual hallucinations, Euclidean symmetry and the functional architecture of striate cortex.

- Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 356(1407), 299–330. doi: 10.1098/rstb.2000.0769
- Bressloff, P. C., Cowan, J. D., Golubitsky, M., Thomas, P. J., & Wiener, M. C. (2002). What Geometric Visual Hallucinations Tell Us about the Visual Cortex. *Neural Computation*, 14(3), 473–491. doi: 10.1162/089976602317250861
- Buchsbaum, B. R., & D’Esposito, M. (2019). A sensorimotor view of verbal working memory. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, 112, 134–148. doi: 10.1016/j.cortex.2018.11.010
- Burke, W. (2002). The neural basis of Charles Bonnet hallucinations: A hypothesis. *Journal of Neurology, Neurosurgery, and Psychiatry*, 73(5), 535–541. doi: 10.1136/jnnp.73.5.535
- Cabbai, G., Racey, C., Simner, J., Dance, C. J., Ward, J., & Forster, S. (2024). Sensory representations in primary visual cortex are not sufficient for subjective imagery. *Current Biology*, 34(21), 5073–5082.e5. doi: 10.1016/j.cub.2024.09.062
- Campello, R. J. G. B., Moulavi, D., & Sander, J. (2013). Density-Based Clustering Based on Hierarchical Density Estimates. In J. Pei, V. S. Tseng, L. Cao, H. Motoda, & G. Xu (Eds.), *Advances in Knowledge Discovery and Data Mining* (pp. 160–172). Springer. doi: 10.1007/978-3-642-37456-2_14
- Cao, R., Brunner, P., Chakravarthula, P. N., Wahlstrom, K. L., Inman, C., Smith, E. H., ... Wang, S. (2025). A neuronal code for object representation and memory in the human amygdala and hippocampus. *Nature Communications*, 16(1), 1510. doi: 10.1038/s41467-025-56793-y
- Chalmers, D. J. (1993). Connectionism and compositionality: Why Fodor and Pylyshyn were wrong. *Philosophical Psychology*, 6(3), 305–319. doi: 10.1080/09515089308573094
- Chang, S., Zhang, X., Cao, Y., Pearson, J., & Meng, M. (2025). Imageless imagery in aphantasia revealed by early visual cortex decoding. *Current Biology*, 35(3), 591–599.e4. doi: 10.1016/j.cub.2024.12.012
- Chetverikov, A. (2026). *Online behavioral studies are safe for now: Unusual RTs do not imply bots (A reply to Van der Stigchel et al., 2026)* (No. gjw5u_v1). PsyArXiv.
- Chkhaidze, A., Coulson, S., & Kiyonaga, A. (2023). Individual Differences in Preferred Thought

Formats Predict Features of Narrative Recall. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 45(45).

Chkhaidze, A., Gaspariani, L., Ballard, I., & Kiyonaga, A. (2026). *Visual and verbal thought tendencies shape working memory under distraction* (No. k2che_vi). PsyArXiv.

Chkhaidze, A., Reeder, R. R., Gag, C., Kiyonaga, A., & Coulson, S. (2026). From dots to faces: Individual differences in visual imagery capacity predict the content of Ganzflicker-induced hallucinations. *Neuroscience of Consciousness*, 2026(1), niago16. doi: 10.1093/nc/niago16

Christian, B. M., Miles, L. K., Parkinson, C., & Macrae, C. N. (2013). Visual perspective and the characteristics of mind wandering. *Frontiers in Psychology*, 4, 699. doi: 10.3389/fpsyg.2013.00699

Christophel, T. B., Klink, P. C., Spitzer, B., Roelfsema, P. R., & Haynes, J.-D. (2017). The Distributed Nature of Working Memory. *Trends in Cognitive Sciences*, 21(2), 111–124. doi: 10.1016/j.tics.2016.12.007

Clark, A., & Toribio, J. (1994). Doing without representing? *Synthese*, 101(3), 401–431. doi: 10.1007/BF01063896

Connell, L. (2019). What have labels ever done for us? The linguistic shortcut in conceptual processing. *Language, Cognition and Neuroscience*, 34(10), 1308–1318. doi: 10.1080/23273798.2018.1471512

Connell, L., & Lynott, D. (2012). Strength of perceptual experience predicts word processing performance better than concreteness or imageability. *Cognition*, 125(3), 452–465. doi: 10.1016/j.cognition.2012.07.010

Cortese, M. J., & Fugett, A. (2004). Imageability ratings for 3,000 monosyllabic words. *Behavior Research Methods, Instruments & Computers*, 36(3), 384–387. doi: 10.3758/BF03195585

Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. In *Tutorials in Quantitative Methods for Psychology* (Vol. 1, pp. 42–45). doi: 10.20982/tqmp.01.1.p042

Cui, X., Jeter, C. B., Yang, D., Montague, P. R., & Eagleman, D. M. (2007). Vividness of mental imagery: Individual variability can be measured objectively. *Vision Research*, 47(4), 474–478. doi: 10.1016/j.visres.2006.11.013

- Dance, C. J., Ward, J., & Simner, J. (2021). What is the Link Between Mental Imagery and Sensory Sensitivity? Insights from Aphantasia. *Perception*, 50(9), 757–782. doi: 10.1177/03010066211042186
- Davis, C. P., & Yee, E. (2021). Building semantic memory from embodied and distributional language experience. *WIREs Cognitive Science*, 12(5), e1555. doi: 10.1002/wcs.1555
- Dawes, A. J., Keogh, R., Robuck, S., & Pearson, J. (2022). Memories with a blind mind: Remembering the past and imagining the future with aphantasia. *Cognition*, 227, 105192. doi: 10.1016/j.cognition.2022.105192
- de Koning, B. B., Wassenburg, S. I., Bos, L. T., & Van der Schoot, M. (2017). Size Does Matter: Implied Object Size is Mentally Simulated During Language Comprehension. *Discourse Processes*, 54(7), 493–503. doi: 10.1080/0163853X.2015.1119604
- De Cesare, A., Cavicchi, S., Cristadoro, G., & Lippi, M. (2021). Do Humans and Deep Convolutional Neural Networks Use Visual Information Similarly for the Categorization of Natural Scenes? *Cognitive Science*, 45(6), e13009. doi: 10.1111/cogs.13009
- Delem, M., Turkben, S., Cavalli, E., Cousineau, D., & Plancher, G. (2025). Unsupervised clustering reveals spatial and verbal cognitive profiles in aphantasia and typical imagery. *Neuropsychologia*, 219, 109279. doi: 10.1016/j.neuropsychologia.2025.109279
- Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, 134(1), 9–21. doi: 10.1016/j.jneumeth.2003.10.009
- Dennett, D. C. (1969). *Content and Consciousness*. Humanities P.
- Dijkstra, N., Bosch, S. E., & van Gerven, M. A. J. (2019). Shared Neural Mechanisms of Visual Perception and Imagery. *Trends in Cognitive Sciences*, 23(5), 423–434. doi: 10.1016/j.tics.2019.02.004
- Dils, A. T., & Boroditsky, L. (2010). Visual motion aftereffect from understanding motion language. *Proceedings of the National Academy of Sciences*, 107(37), 16396–16400. doi: 10.1073/pnas.1009438107
- Dong, Y., Hung, Y.-c., Xie, C., & Kiyonaga, A. (2026). *Windows to the goal: Pupillary working memory*

signatures prospectively adapt to task demands. bioRxiv. doi: 10.64898/2025.12.06.692784

- Dong, Y., & Kiyonaga, A. (2024). Ocular working memory signals are flexible to behavioral priority and subjective imagery strength. *Journal of Neurophysiology*, 132(1), 162–176. doi: 10.1152/jn.00446.2023
- Dupont, W., Papaxanthis, C., Madden-Lombardi, C., & Lebon, F. (2024). Explicit and implicit motor simulations are impaired in individuals with aphantasia. *Brain Communications*, 6(2), fcae072. doi: 10.1093/braincomms/fcae072
- Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature*, 392(6676), 598–601. doi: 10.1038/33402
- Ermentrout, G. B., & Cowan, J. D. (1979). A mathematical theory of visual hallucination patterns. *Biological Cybernetics*, 34(3), 137–150. doi: 10.1007/BF00336965
- Ester, E. F., Serences, J. T., & Awh, E. (2009). Spatially global representations in human primary visual cortex during working memory maintenance. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 29(48), 15258–15265. doi: 10.1523/JNEUROSCI.4388-09.2009
- Farah, M. J., Hammond, K. M., Levine, D. N., & Calvanio, R. (1988). Visual and spatial mental imagery: Dissociable systems of representation. *Cognitive Psychology*, 20(4), 439–462. doi: 10.1016/0010-0285(88)90012-6
- Federmeier, K. D., & Kutas, M. (2001). Meaning and modality: Influences of context, semantic memory organization, and perceptual predictability on picture processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27(1), 202–224. doi: 10.1037/0278-7393.27.1.202
- Federmeier, K. D., & Kutas, M. (2002). Picture the difference: Electrophysiological investigations of picture processing in the two cerebral hemispheres. *Neuropsychologia*, 40(7), 730–747. doi: 10.1016/S0028-3932(01)00193-2
- Feuerriegel, S., Maarouf, A., Bär, D., Geissler, D., Schweisthal, J., Pröllochs, N., ... Van Bavel, J. J. (2025). Using natural language processing to analyse text data in behavioural science. *Nature Reviews Psychology*, 4(2), 96–111. doi: 10.1038/s44159-024-00392-z
- Ffytche, D. H. (2008). The hodology of hallucinations. *Cortex; a Journal Devoted to the Study of the*

Nervous System and Behavior, 44(8), 1067–1083. doi: 10.1016/j.cortex.2008.04.005

Fort, L. D., Costines, C., Wittmann, M., Demertzi, A., & Schmidt, T. T. (2025). Classification schemes of altered states of consciousness. *Neuroscience & Biobehavioral Reviews*, 175, 106178. doi: 10.1016/j.neubiorev.2025.106178

Frank, M. C., & Barner, D. (2012). Representing exact number visually using mental abacus. *Journal of Experimental Psychology: General*, 141(1), 134–149. doi: 10.1037/a0024427

Funder, D. C., & Ozer, D. J. (2019). Evaluating Effect Size in Psychological Research: Sense and Nonsense. *Advances in Methods and Practices in Psychological Science*, 2(2), 156–168. doi: 10.1177/2515245919847202

Ganis, G., & Schendan, H. E. (2008). Visual mental imagery and perception produce opposite adaptation effects on early brain potentials. *NeuroImage*, 42(4), 1714–1727. doi: 10.1016/j.neuroimage.2008.07.004

Ganis, G., Thompson, W. L., & Kosslyn, S. M. (2004). Brain areas underlying visual mental imagery and visual perception: An fMRI study. *Cognitive Brain Research*, 20(2), 226–241. doi: 10.1016/j.cogbrainres.2004.02.012

Gayet, S., Paffen, C. L. E., & Van der Stigchel, S. (2013). Information Matching the Content of Visual Working Memory Is Prioritized for Conscious Access. *Psychological Science*, 24(12), 2472–2480. doi: 10.1177/0956797613495882

Gerlach, C., Law, I., & Paulson, O. B. (2002). When Action Turns into Words. Activation of Motor-Based Knowledge during Categorization of Manipulable Objects. *Journal of Cognitive Neuroscience*, 14(8), 1230–1239. doi: 10.1162/089892902760807221

German, J. S., & Jacobs, R. A. (2020). Can machine learning account for human visual object shape similarity judgments? *Vision Research*, 167, 87–99. doi: 10.1016/j.visres.2019.12.001

Gillies, G., Park, H., Woo, J., Walther, D. B., Cant, J. S., & Fukuda, K. (2023). Tracing the emergence of the memorability benefit. *Cognition*, 238, 105489. doi: 10.1016/j.cognition.2023.105489

Glenberg, A. M. (1997). What memory is for. *The Behavioral and Brain Sciences*, 20(1), 1-19; discussion 19-55. doi: 10.1017/S0140525X97000010

- Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin & Review*, 9(3), 558–565. doi: 10.3758/BF03196313
- Goetschalckx, L., & Wagemans, J. (2019). MemCat: A new category-based image set quantified on memorability. *PeerJ*, 7, e8169. doi: 10.7717/peerj.8169
- Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., ... Hämäläinen, M. (2013). MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience*, 7, 267. doi: 10.3389/fnins.2013.00267
- Green, P., & MacLeod, C. J. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, 7(4), 493–498. doi: 10.1111/2041-210X.12504
- Greenberg, D. L., & Knowlton, B. J. (2014). The role of visual imagery in autobiographical memory. *Memory & Cognition*, 42(6), 922–934. doi: 10.3758/s13421-014-0402-5
- Grootendorst, M. (2022). *BERTopic: Neural topic modeling with a class-based TF-IDF procedure*.
- Guariglia, C., & Pizzamiglio, L. (2007). The Role of Imagery in Navigation: Neuropsychological Evidence. In F. Mast & L. Jäncke (Eds.), *Spatial Processing in Navigation, Imagery and Perception* (pp. 17–28). Springer US. doi: 10.1007/978-0-387-71978-8_2
- Gulyás, E., Gombos, F., Sütöri, S., Lovas, A., Ziman, G., & Kovács, I. (2022). Visual imagery vividness declines across the lifespan. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, 154, 365–374. doi: 10.1016/j.cortex.2022.06.011
- Hamm, J. P., Johnson, B. W., & Kirk, I. J. (2002). Comparison of the N300 and N400 ERPs to picture stimuli in congruent and incongruent contexts. *Clinical Neurophysiology*, 113(8), 1339–1350. doi: 10.1016/S1388-2457(02)00161-X
- Hardwick, R. M., Caspers, S., Eickhoff, S. B., & Swinnen, S. P. (2018). Neural correlates of action: Comparing meta-analyses of imagery, observation, and execution. *Neuroscience and Biobehavioral Reviews*, 94, 31–44. doi: 10.1016/j.neubiorev.2018.08.003
- Harrison, S. A., & Tong, F. (2009). Decoding reveals the contents of visual working memory in early visual areas. *Nature*, 458(7238), 632–635. doi: 10.1038/nature07832

Hauk, O., Johnsrude, I., & Pulvermüller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron*, 41(2), 301–307. doi: 10.1016/s0896-6273(03)00838-9

Hebart, M. N., Contier, O., Teichmann, L., Rockter, A. H., Zheng, C. Y., Kidder, A., ... Baker, C. I. (2023). THINGS-data, a multimodal collection of large-scale datasets for investigating object representations in human brain and behavior. *eLife*, 12, e82580. doi: 10.7554/eLife.82580

Hewitt, T., Amaya, I. A., Beauté, R., Seth, A. K., Schmidt, T. T., & Schwartzman, D. J. (2025). Stroboscopically induced visual hallucinations: Historical, phenomenological, and neurobiological perspectives. *Neuroscience of Consciousness*, 2025(1), niafo20. doi: 10.1093/nc/niafo20

Hirschfeld, G., Zwitserlood, P., & Dobel, C. (2011). Effects of language comprehension on visual processing - MEG dissociates early perceptual and late N400 effects. *Brain and Language*, 116(2), 91–96. doi: 10.1016/j.bandl.2010.07.002

Hume, D. (2000). *A Treatise of Human Nature*. Oxford University Press.

Hwang, K., Bertolero, M. A., Liu, W. B., & D’Esposito, M. (2017). The Human Thalamus Is an Integrative Hub for Functional Brain Networks. *Journal of Neuroscience*, 37(23), 5594–5607. doi: 10.1523/JNEUROSCI.0067-17.2017

Jacob, G., Pramod, R. T., Katti, H., & Arun, S. P. (2021). Qualitative similarities and differences in visual object representations between brains and deep networks. *Nature Communications*, 12(1), 1872. doi: 10.1038/s41467-021-22078-3

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

Joyce, C., & Rossion, B. (2005). The face-sensitive N170 and VPP components manifest the same brain processes: The effect of reference electrode site. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology*, 116(11), 2613–2631. doi: 10.1016/j.clinph.2005.07.005

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The Fusiform Face Area: A Module in Human Extrastriate Cortex Specialized for Face Perception. *Journal of Neuroscience*, 17(11), 4302–4311. doi: 10.1523/JNEUROSCI.17-11-04302.1997

- Kappenman, E. S., & Luck, S. J. (2011). ERP Components: The Ups and Downs of Brainwave Recordings. In E. S. Kappenman & S. J. Luck (Eds.), *The Oxford Handbook of Event-Related Potential Components* (p. o). Oxford University Press. doi: 10.1093/oxfordhb/9780195374148.013.0014
- Kay, L., Keogh, R., Andrillon, T., & Pearson, J. (2022). The pupillary light response as a physiological index of aphantasia, sensory and phenomenological imagery strength. *eLife*, 11, e72484. doi: 10.7554/eLife.72484
- Keogh, R., & Pearson, J. (2014). The sensory strength of voluntary visual imagery predicts visual working memory capacity. *Journal of Vision*, 14(12), 7. doi: 10.1167/14.12.7
- Keogh, R., & Pearson, J. (2018). The blind mind: No sensory visual imagery in aphantasia. *Cortex*, 105, 53–60. doi: 10.1016/j.cortex.2017.10.012
- Keogh, R., & Pearson, J. (2024). Revisiting the blind mind: Still no evidence for sensory visual imagery in individuals with aphantasia. *Neuroscience Research*, 201, 27–30. doi: 10.1016/j.neures.2024.01.008
- Keogh, R., Wicken, M., & Pearson, J. (2021). Visual working memory in aphantasia: Retained accuracy and capacity with a different strategy. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, 143, 237–253. doi: 10.1016/j.cortex.2021.07.012
- Khosla, A., Raju, A. S., Torralba, A., & Oliva, A. (2015). Understanding and Predicting Image Memorability at a Large Scale. In *2015 IEEE International Conference on Computer Vision (ICCV)* (pp. 2390–2398). doi: 10.1109/ICCV.2015.275
- Kiss, T., & Strunk, J. (2006). Unsupervised Multilingual Sentence Boundary Detection. *Computational Linguistics*, 32(4), 485–525. doi: 10.1162/coli.2006.32.4.485
- Kiyonaga, A., & Egner, T. (2014). The Working Memory Stroop Effect: When Internal Representations Clash With External Stimuli. *Psychological Science*, 25(8), 1619–1629. doi: 10.1177/0956797614536739
- Kiyonaga, A., & Serences, J. T. (2025). Sensory reformatting for a working visual memory. *Trends in Cognitive Sciences*, 29(12), 1120–1135. doi: 10.1016/j.tics.2025.09.006

- Klüver, H. (1926). Mescal Visions and Eidetic Vision. *The American Journal of Psychology*, 37(4), 502–515. doi: 10.2307/1414910
- Knight, K. F., Milton, F., & Zeman, A. (2026). Aphantasia and visual working memory: No direct evidence of impaired visual working memory in aphantasics, either in behavioral performance or the accuracy of a multivoxel pattern classifier. *Neuropsychologia*, 226, 109430. doi: 10.1016/j.neuropsychologia.2026.109430
- Königsmark, V. T., Bergmann, J., & Reeder, R. R. (2021). The Ganzflicker experience: High probability of seeing vivid and complex pseudo-hallucinations with imagery but not aphantasia. *Cortex*, 141, 522–534. doi: 10.1016/j.cortex.2021.05.007
- Kosslyn, S. M. (1980). *Image and Mind*. Harvard University Press.
- Kosslyn, S. M. (1994). *Image And Brain: The Resolution of the Imagery Debate*. The MIT Press. doi: 10.7551/mitpress/3653.001.0001
- Kosslyn, S. M., Brunn, J., Cave, K. R., & Wallach, R. W. (1984). Individual differences in mental imagery ability: A computational analysis. *Cognition*, 18(1), 195–243. doi: 10.1016/0010-0277(84)90025-8
- Kosslyn, S. M., Pascual-Leone, A., Felician, O., Camposano, S., Keenan, J. P., Thompson, W. L., ... Alpert, N. M. (1999). The role of area 17 in visual imagery: Convergent evidence from PET and rTMS. *Science*, 284(5411), 167–170. doi: 10.1126/science.284.5411.167
- Kozhevnikov, M., Hegarty, M., & Mayer, R. E. (2002). Revising the Visualizer-Verbalizer Dimension: Evidence for Two Types of Visualizers. *Cognition and Instruction*, 20(1), 47–77. doi: 10.1207/S1532690XCI2001_3
- Kragel, P. A., Koban, L., Barrett, L. F., & Wager, T. D. (2018). Representation, Pattern Information, and Brain Signatures: From Neurons to Neuroimaging. *Neuron*, 99(2), 257–273. doi: 10.1016/j.neuron.2018.06.009
- Kübel, S. L., Fiedler, H., & Wittmann, M. (2021). Red visual stimulation in the Ganzfeld leads to a relative overestimation of duration compared to green. *PsyCh Journal*, 10(1), 5–19. doi: 10.1002/pchj.395
- Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400

- component of the event-related brain potential (ERP). *Annual Review of Psychology*, 62, 621–647. doi: 10.1146/annurev.psych.093008.131123
- Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, 207(4427), 203–205. doi: 10.1126/science.7350657
- Lacruz, R. S., Stringer, C. B., Kimbel, W. H., Wood, B., Harvati, K., O’Higgins, P., ... Arsuaga, J.-L. (2019). The evolutionary history of the human face. *Nature Ecology & Evolution*, 3(5), 726–736. doi: 10.1038/s41559-019-0865-7
- Lee, S.-H., Kravitz, D. J., & Baker, C. I. (2012). Disentangling visual imagery and perception of real-world objects. *NeuroImage*, 59(4), 4064–4073. doi: 10.1016/j.neuroimage.2011.10.055
- Lee, S.-H., Kravitz, D. J., & Baker, C. I. (2013). Goal-dependent dissociation of visual and prefrontal cortices during working memory. *Nature Neuroscience*, 16(8), 997–999. doi: 10.1038/nn.3452
- Lewis-Peacock, J. A., Drysdale, A. T., Oberauer, K., & Postle, B. R. (2012). Neural evidence for a distinction between short-term memory and the focus of attention. *Journal of Cognitive Neuroscience*, 24(1), 61–79. doi: 10.1162/jocn_a_00140
- Liu, J., Zhan, M., Hajhajate, D., Spagna, A., Dehaene, S., Cohen, L., & Bartolomeo, P. (2025). Visual mental imagery in typical imagers and in aphantasia: A millimeter-scale 7-T fMRI study. *Cortex*, 185, 113–132. doi: 10.1016/j.cortex.2025.01.013
- Locke, J. (1948). An essay concerning human understanding, 1690. In *Readings in the history of psychology* (pp. 55–68). Appleton-Century-Crofts. doi: 10.1037/11304-008
- Lorenc, E. S., Mallett, R., & Lewis-Peacock, J. A. (2021). Distraction in Visual Working Memory: Resistance is Not Futile. *Trends in Cognitive Sciences*, 25(3), 228–239. doi: 10.1016/j.tics.2020.12.004
- Lorenc, E. S., Sreenivasan, K. K., Nee, D. E., Vandenbroucke, A. R. E., & D’Esposito, M. (2018). Flexible Coding of Visual Working Memory Representations during Distraction. *Journal of Neuroscience*, 38(23), 5267–5276. doi: 10.1523/JNEUROSCI.3061-17.2018
- Louwerse, M. M. (2011). Symbol Interdependency in Symbolic and Embodied Cognition. *Topics in Cognitive Science*, 3(2), 273–302. doi: 10.1111/j.1756-8765.2010.01106.x

- Lupyan, G. (2012). Linguistically Modulated Perception and Cognition: The Label-Feedback Hypothesis. *Frontiers in Psychology*, 3. doi: 10.3389/fpsyg.2012.00054
- Lupyan, G., Abdel Rahman, R., Boroditsky, L., & Clark, A. (2020). Effects of Language on Visual Perception. *Trends in Cognitive Sciences*, 24(11), 930–944. doi: 10.1016/j.tics.2020.08.005
- Lupyan, G., Uchiyama, R., Thompson, B., & Casasanto, D. (2023). Hidden Differences in Phenomenal Experience. *Cognitive Science*, 47(1), e13239. doi: 10.1111/cogs.13239
- Lynott, D., Connell, L., Brysbaert, M., Brand, J., & Carney, J. (2020). The Lancaster Sensorimotor Norms: Multidimensional measures of perceptual and action strength for 40,000 English words. *Behavior Research Methods*, 52(3), 1271–1291. doi: 10.3758/s13428-019-01316-z
- Marks, D. F. (1973). Visual imagery differences in the recall of pictures. *British Journal of Psychology*, 64(1), 17–24. doi: 10.1111/j.2044-8295.1973.tb01322.x
- McInnes, L., Healy, J., & Astels, S. (2017). Hdbscan: Hierarchical density based clustering. *The Journal of Open Source Software*, 2, 205. doi: 10.21105/joss.00205
- McInnes, L., Healy, J., & Melville, J. (2018). *UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction* (No. arXiv:1802.03426). arXiv. doi: 10.48550/arXiv.1802.03426
- McPherson, W. B., & Holcomb, P. J. (1999). An electrophysiological investigation of semantic priming with pictures of real objects. *Psychophysiology*, 36(1), 53–65. doi: 10.1017/s0048577299971196
- Mechelli, A., Price, C. J., Friston, K. J., & Ishai, A. (2004). Where Bottom-up Meets Top-down: Neuronal Interactions during Perception and Imagery. *Cerebral Cortex*, 14(11), 1256–1265. doi: 10.1093/cercor/bhh087
- Merabet, L. B., Maguire, D., Warde, A., Alterescu, K., Stickgold, R., & Pascual-Leone, A. (2004). Visual Hallucinations During Prolonged Blindfolding in Sighted Subjects. *Journal of Neuro-Ophthalmology*, 24(2), 109.
- Meteyard, L., Bahrami, B., & Vigliocco, G. (2007). Motion detection and motion verbs: Language affects low-level visual perception. *Psychological Science*, 18(11), 1007–1013. doi: 10.1111/j.1467-9280.2007.02016.x

- Metzger, W. (1929). Optische Untersuchungen am Ganzfeld. II. Zur Phänomenologie des homogenen Ganzfelds. [Optical researches on the total field [of vision]. II. On the phenomenology of homogeneous total fields.]. *Psychologische Forschung*, 13, 6–29. doi: 10.1007/BF00406757
- Milton, F., Fulford, J., Dance, C. J., Gaddum, J., Heuerman-Williamson, B., Jones, K., ... Zeman, A. (2021). Behavioral and Neural Signatures of Visual Imagery Vividness Extremes: Aphantasia versus Hyperphantasia. *Cerebral Cortex Communications*, 2(2), tgab035. doi: 10.1093/texcom/tgab035
- Mirski, R., & Bickhard, M. H. (2019). Encodingism is Not Just a Bad Metaphor. *Behavioral and Brain Sciences*, 42. doi: 10.1017/S0140525X19001286
- Mishkin, M., Ungerleider, L. G., & Macko, K. A. (1983). Object vision and spatial vision: Two cortical pathways. *Trends in Neurosciences*, 6, 414–417. doi: 10.1016/0166-2236(83)90190-X
- Morey, R. D. (2008). Confidence Intervals from Normalized Data: A correction to Cousineau (2005). In *Tutorials in Quantitative Methods for Psychology* (Vol. 4, pp. 61–64). doi: 10.20982/tqmp.04.2.p061
- Muraki, E. J., & Pexman, P. M. (2021). Simulating semantics: Are individual differences in motor imagery related to sensorimotor effects in language processing? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 47(12), 1939–1957. doi: 10.1037/xlm0001039
- Muraki, E. J., Speed, L. J., & Pexman, P. M. (2023). Insights into embodied cognition and mental imagery from aphantasia. *Nature Reviews Psychology*, 2(10), 591–605. doi: 10.1038/s44159-023-00221-9
- Naselaris, T., Olman, C. A., Stansbury, D. E., Ugurbil, K., & Gallant, J. L. (2015). A voxel-wise encoding model for early visual areas decodes mental images of remembered scenes. *NeuroImage*, 105, 215–228. doi: 10.1016/j.neuroimage.2014.10.018
- Nedergaard, J. S. K., & Lupyan, G. (2024). Not Everybody Has an Inner Voice: Behavioral Consequences of Anendophasia. *Psychological Science*, 35(7), 780–797. doi: 10.1177/09567976241243004
- Needell, C. D., & Bainbridge, W. A. (2022). Embracing New Techniques in Deep Learning for Estimating Image Memorability. *Computational Brain & Behavior*, 5(2), 168–184. doi: 10.1007/s42113-022-00126-5

- Nichols, T. E., & Holmes, A. P. (2002). Nonparametric permutation tests for functional neuroimaging: A primer with examples. *Human Brain Mapping, 15*(1), 1–25. doi: 10.1002/hbm.1058
- O’Callaghan, D., Greene, D., Carthy, J., & Cunningham, P. (2015). An analysis of the coherence of descriptors in topic modeling. *Expert Systems with Applications, 42*(13), 5645–5657. doi: 10.1016/j.eswa.2015.02.055
- O’Craven, K. M., & Kanwisher, N. (2000). Mental Imagery of Faces and Places Activates Corresponding Stimulus-Specific Brain Regions. *Journal of Cognitive Neuroscience, 12*(6), 1013–1023. doi: 10.1162/08989290051137549
- O’Shea, H. (2022). Mapping relational links between motor imagery, action observation, action-related language, and action execution. *Frontiers in Human Neuroscience, 16*. doi: 10.3389/fnhum.2022.984053
- Ostarek, M., Joosen, D., Ishag, A., de Nijs, M., & Huettig, F. (2019). Are visual processes causally involved in “perceptual simulation” effects in the sentence-picture verification task? *Cognition, 182*, 84–94. doi: 10.1016/j.cognition.2018.08.017
- Overkott, C., & Souza, A. S. (2022). Verbal descriptions improve visual working memory but have limited impact on visual long-term memory. *Journal of Experimental Psychology: General, 151*(2), 321–347. doi: 10.1037/xge0001084
- Palermo, L., Boccia, M., Piccardi, L., & Nori, R. (2022). Congenital lack and extraordinary ability in object and spatial imagery: An investigation on sub-types of aphantasia and hyperphantasia. *Consciousness and Cognition, 103*, 103360. doi: 10.1016/j.concog.2022.103360
- Pan, Y., Lin, B., Zhao, Y., & Soto, D. (2014). Working memory biasing of visual perception without awareness. *Attention, Perception, & Psychophysics, 76*(7), 2051–2062. doi: 10.3758/s13414-013-0566-2
- Pearson, J. (2019). The human imagination: The cognitive neuroscience of visual mental imagery. *Nature Reviews Neuroscience, 20*(10), 624–634. doi: 10.1038/s41583-019-0202-9
- Pearson, J., & Keogh, R. (2019). Redefining visual working memory: A cognitive-strategy, brain-region approach. *Current Directions in Psychological Science, 28*(3), 266–273. doi: 10.1177/0963721419835210

- Pearson, J., & Kosslyn, S. M. (2015). The heterogeneity of mental representation: Ending the imagery debate. *Proceedings of the National Academy of Sciences of the United States of America*, 112(33), 10089–10092. doi: 10.1073/pnas.1504933112
- Pecher, D., van Dantzig, S., & Schifferstein, H. N. J. (2009). Concepts are not represented by conscious imagery. *Psychonomic Bulletin & Review*, 16(5), 914–919. doi: 10.3758/PBR.16.5.914
- Pecher, D., van Dantzig, S., Zwaan, R. A., & Zeelenberg, R. (2009). Language comprehenders retain implied shape and orientation of objects. *Quarterly Journal of Experimental Psychology*, 62(6), 1108–1114. doi: 10.1080/17470210802633255
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*.
- Pereira Seabra, J., Chopurian, V., Souza, A. S., & Christophel, T. B. (2025). Verbal Encoding Strategies in Visuo-Spatial Working Memory. *Journal of Cognition*, 8(1), 2. doi: 10.5334/joc.406
- Phillips, I. (2025). Spared spatial imagery solves the puzzle of aphantasia. *Trends in Cognitive Sciences*, 29(11), 961–962. doi: 10.1016/j.tics.2025.05.004
- Phillips, J. W., Schulmann, A., Hara, E., Winnubst, J., Liu, C., Valakh, V., ... Hantman, A. W. (2019). A repeated molecular architecture across thalamic pathways. *Nature Neuroscience*, 22(11), 1925–1935. doi: 10.1038/s41593-019-0483-3
- Piriyajitakonkij, M., Itthipuripat, S., Ballard, I., & Pappas, I. (2025). What Makes a Face Look Like a Hat: Decoupling Low-Level and High-Level Visual Properties with Image Triplets. In A. Del Bue, C. Canton, J. Pont-Tuset, & T. Tommasi (Eds.), *Computer Vision – ECCV 2024 Workshops* (pp. 224–233). Springer Nature Switzerland. doi: 10.1007/978-3-031-91578-9_16
- Pistolas, E., & Wagemans, J. (2025). And then there was light in the ganzfeld: Clarifying the methods, experiences, and modulating factors of hallucinations and decays. *Neuroscience of Consciousness*, 2025(1). doi: 10.1093/nc/niafo21
- Pounder, Z., Jacob, J., Evans, S., Loveday, C., Eardley, A. F., & Silvanto, J. (2022). Only minimal differences between individuals with congenital aphantasia and those with typical imagery on neuropsychological tasks that involve imagery. *Cortex*, 148, 180–192. doi: 10.1016/j.cortex.2021.12.010

- Powell, M. J. D. (2009). *The bobyqa algorithm for bound constrained optimization without derivatives* (Tech. Rep. No. Technical Report DAMTP 2009/NA06). Cambridge, UK: Department of Applied Mathematics and Theoretical Physics, University of Cambridge.
- Pratts, J., Pobric, G., & Yao, B. (2023). Bridging phenomenology and neural mechanisms of inner speech: ALE meta-analysis on egocentricity and spontaneity in a dual-mechanistic framework. *NeuroImage*, 282, 120399. doi: 10.1016/j.neuroimage.2023.120399
- Pulvermüller, F. (2005). Brain mechanisms linking language and action. *Nature Reviews Neuroscience*, 6(7), 576–582. doi: 10.1038/nrn1706
- Pulvermüller, F. (2013). How neurons make meaning: Brain mechanisms for embodied and abstract-symbolic semantics. *Trends in Cognitive Sciences*, 17(9), 458–470. doi: 10.1016/j.tics.2013.06.004
- Pulvermüller, F., & Fadiga, L. (2010). Active perception: Sensorimotor circuits as a cortical basis for language. *Nature Reviews Neuroscience*, 11(5), 351–360. doi: 10.1038/nrn2811
- Pulvermüller, F., & Hauk, O. (2006). Category-specific Conceptual Processing of Color and Form in Left Fronto-temporal Cortex. *Cerebral Cortex*, 16(8), 1193–1201. doi: 10.1093/cercor/bhj060
- Purkinje, J. E. (1819). *Beiträge Zur Kenntniss Des Sehens in Subjectiver Hinsicht*. Johann Gottfried Calve.
- Pütz, P., Braeunig, M., & Wackermann, J. (2006). EEG correlates of multimodal ganzfeld induced hallucinatory imagery. *International Journal of Psychophysiology*, 61(2), 167–178. doi: 10.1016/j.ijpsycho.2005.09.002
- Pylyshyn, Z. W. (1973). What the mind's eye tells the mind's brain: A critique of mental imagery. *Psychological Bulletin*, 80(1), 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), 19–28. doi: 10.3758/BF03196930
- Pylyshyn, Z. W. (1981). The imagery debate: Analogue media versus tacit knowledge. *Psychological Review*, 88(1), 16–45. doi: 10.1037/0033-295X.88.1.16

- Pylyshyn, Z. W. (2002). Mental imagery: In search of a theory. *Behavioral and Brain Sciences*, 25(2), 157–238. doi: 10.1017/S0140525X02000043
- Rademaker, R. L., Bloem, I. M., De Weerd, P., & Sack, A. T. (2015). The impact of interference on short-term memory for visual orientation. *Journal of Experimental Psychology. Human Perception and Performance*, 41(6), 1650–1665. doi: 10.1037/xhp0000110
- Rademaker, R. L., Chunharas, C., & Serences, J. T. (2019). Coexisting representations of sensory and mnemonic information in human visual cortex. *Nature Neuroscience*, 22(8), 1336–1344. doi: 10.1038/s41593-019-0428-x
- Reeder, R. R. (2021). *Pseudo-hallucinations: Why some people see more vivid mental images than others – test yourself here*.
- Reeder, R. R. (2022). Ganzflicker Reveals the Complex Relationship Between Visual Mental Imagery and Pseudo-Hallucinatory Experiences: A Replication and Expansion. *Collabra: Psychology*, 8(1), 36318. doi: 10.1525/collabra.36318
- Reeder, R. R. (2024). Current and future directions for research on hallucinations and delusions. *Scientific Reports*, 14(1), 8328. doi: 10.1038/s41598-024-57472-6
- Reeder, R. R., Pounder, Z., Figueroa, A., Jüllig, A., & Azañón, E. (2024). Non-visual spatial strategies are effective for maintaining precise information in visual working memory. *Cognition*, 251, 105907. doi: 10.1016/j.cognition.2024.105907
- Reeder, R. R., Sala, G., & van Leeuwen, T. M. (2024). A novel model of divergent predictive perception. *Neuroscience of Consciousness*, 2024(1), niae006. doi: 10.1093/nc/niae006
- Reimers, N., & Gurevych, I. (2019). *Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks* (No. arXiv:1908.10084). arXiv. doi: 10.48550/arXiv.1908.10084
- Richardson, J. T. E. (1995). Gender differences in the Vividness of Visual Imagery Questionnaire: A meta-analysis. *Journal of Mental Imagery*, 19(3-4), 177–187.
- Roberts, B. R. T., MacLeod, C. M., & Fernandes, M. A. (2023). Symbol superiority: Why \$ is better remembered than ‘dollar’. *Cognition*, 238, 105435. doi: 10.1016/j.cognition.2023.105435

- Röder, M., Both, A., & Hinneburg, A. (2015). Exploring the Space of Topic Coherence Measures. In *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining* (pp. 399–408). Association for Computing Machinery. doi: 10.1145/2684822.2685324
- Roebuck, H., & Lupyan, G. (2020). The Internal Representations Questionnaire: Measuring modes of thinking. *Behavior Research Methods*, 52(5), 2053–2070. doi: 10.3758/s13428-020-01354-y
- Rollo, B., Malatesta, G., D'Anselmo, A., Lucafò, C., & Tommasi, L. (2026). Disturbing the sound of silence: Bilateral temporal cortex stimulation and auditory mental imagery. *Brain and Cognition*, 192, 106378. doi: 10.1016/j.bandc.2025.106378
- Rule, M., Stoffregen, M., & Ermentrout, B. (2011). A Model for the Origin and Properties of Flicker-Induced Geometric Phosphenes. *PLOS Computational Biology*, 7(9), e1002158. doi: 10.1371/journal.pcbi.1002158
- Saito, T., & Rehmsmeier, M. (2015). The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets. *PLOS ONE*, 10(3), e0118432. doi: 10.1371/journal.pone.0118432
- Schacter, D. L., Addis, D. R., Hassabis, D., Martin, V. C., Spreng, R. N., & Szpunar, K. K. (2012). The future of memory: Remembering, imagining, and the brain. *Neuron*, 76(4), 677–694. doi: 10.1016/j.neuron.2012.11.001
- Schacter, D. L., & Madore, K. P. (2016). Remembering the past and imagining the future: Identifying and enhancing the contribution of episodic memory. *Memory Studies*, 9(3), 245–255. doi: 10.1177/1750698016645230
- Schad, D. J., Vasishth, S., Hohenstein, S., & Kliegl, R. (2020). How to capitalize on a priori contrasts in linear (mixed) models: A tutorial. *Journal of Memory and Language*, 110, 104038. doi: 10.1016/j.jml.2019.104038
- Schendan, H. E., & Lucia, L. C. (2010). Object-sensitive activity reflects earlier perceptual and later cognitive processing of visual objects between 95 and 500 ms. *Brain Research*, 1329, 124–141. doi: 10.1016/j.brainres.2010.01.062
- Schmidt, T. T., Jagannathan, N., Ljubljanc, M., Xavier, A., & Nierhaus, T. (2020). The multimodal Ganzfeld-induced altered state of consciousness induces decreased thalamo-cortical coupling. *Scientific Reports*, 10(1), 18686. doi: 10.1038/s41598-020-75019-3

- Schmidt, T. T., & Prein, J. C. (2019). The Ganzfeld experience—A stably inducible altered state of consciousness: Effects of different auditory homogenizations. *PsyCh Journal*, 8(1), 66–81. doi: 10.1002/pchj.262
- Schwartzman, D. J., Schartner, M., Ador, B. B., Simonelli, F., Chang, A. Y.-C., & Seth, A. K. (2019). *Increased spontaneous EEG signal diversity during stroboscopically-induced altered states of consciousness*. bioRxiv. doi: 10.1101/511766
- Scimeca, J. M., Kiyonaga, A., & D’Esposito, M. (2018). Reaffirming the Sensory Recruitment Account of Working Memory. *Trends in Cognitive Sciences*, 22(3), 190–192. doi: 10.1016/j.tics.2017.12.007
- Serences, J. T. (2016). Neural mechanisms of information storage in visual short-term memory. *Vision Research*, 128, 53–67. doi: 10.1016/j.visres.2016.09.010
- Serences, J. T., Ester, E. F., Vogel, E. K., & Awh, E. (2009). Stimulus-specific delay activity in human primary visual cortex. *Psychological Science*, 20(2), 207–214. doi: 10.1111/j.1467-9280.2009.02276.x
- Shenyan, O., Lisi, M., Greenwood, J. A., Skipper, J. I., & Dekker, T. M. (2024). Visual hallucinations induced by Ganzflicker and Ganzfeld differ in frequency, complexity, and content. *Scientific Reports*, 14(1), 2353. doi: 10.1038/s41598-024-52372-1
- Shepard, R. N., & Cooper, L. A. (1986). *Mental images and their transformations*. The MIT Press.
- Shine, J. M. (2021). The thalamus integrates the macrosystems of the brain to facilitate complex, adaptive brain network dynamics. *Progress in Neurobiology*, 199, 101951. doi: 10.1016/j.pneurobio.2020.101951
- Shoham, A., Grosbard, I. D., Patashnik, O., Cohen-Or, D., & Yovel, G. (2024). Using deep neural networks to disentangle visual and semantic information in human perception and memory. *Nature Human Behaviour*, 8(4), 702–717. doi: 10.1038/s41562-024-01816-9
- Siakaluk, P. D., Pexman, P. M., Sears, C. R., Wilson, K., Lockheed, K., & Owen, W. J. (2008). The Benefits of Sensorimotor Knowledge: Body–Object Interaction Facilitates Semantic Processing. *Cognitive Science*, 32(3), 591–605. doi: 10.1080/03640210802035399
- Silvanto, J. (2025). Interoception, insula, and agency: A predictive coding account of aphantasia. *Frontiers in Psychology*, 16. doi: 10.3389/fpsyg.2025.1564251

- Silvanto, J., & Nagai, Y. (2025). How Interoception and the Insula Shape Mental Imagery and Aphantasia. *Brain Topography*, 38(2), 27. doi: 10.1007/s10548-025-01101-6
- Simmons, W. K., Ramjee, V., Beauchamp, M. S., McRae, K., Martin, A., & Barsalou, L. W. (2007). A common neural substrate for perceiving and knowing about color. *Neuropsychologia*, 45(12), 2802–2810. doi: 10.1016/j.neuropsychologia.2007.05.002
- Simonyan, K., & Zisserman, A. (2015). *Very Deep Convolutional Networks for Large-Scale Image Recognition* (No. arXiv:1409.1556). arXiv. doi: 10.48550/arXiv.1409.1556
- Slotnick, S. D., Thompson, W. L., & Kosslyn, S. M. (2005). Visual Mental Imagery Induces Retinotopically Organized Activation of Early Visual Areas. *Cerebral Cortex*, 15(10), 1570–1583. doi: 10.1093/cercor/bh1035
- Smith, N. J., & Kutas, M. (2015). Regression-based estimation of ERP waveforms: I. The rERP framework. *Psychophysiology*, 52(2), 157–168. doi: 10.1111/psyp.12317
- Souza, A. S., & Oberauer, K. (2016). In search of the focus of attention in working memory: 13 years of the retro-cue effect. *Attention, Perception, & Psychophysics*, 78(7), 1839–1860. doi: 10.3758/s13414-016-1108-5
- Souza, A. S., Overkott, C., & Matyja, M. (2021). Categorical distinctiveness constrains the labeling benefit in visual working memory. *Journal of Memory and Language*, 119, 104242. doi: 10.1016/j.jml.2021.104242
- Souza, A. S., & Skóra, Z. (2017). The interplay of language and visual perception in working memory. *Cognition*, 166, 277–297. doi: 10.1016/j.cognition.2017.05.038
- Spagna, A., Hajhajate, D., Liu, J., & Bartolomeo, P. (2021). Visual mental imagery engages the left fusiform gyrus, but not the early visual cortex: A meta-analysis of neuroimaging evidence. *Neuroscience & Biobehavioral Reviews*, 122, 201–217. doi: 10.1016/j.neubiorev.2020.12.029
- Speed, L. J., Eekhof, L. S., & Mak, M. (2024). The role of visual imagery in story reading: Evidence from aphantasia. *Consciousness and Cognition*, 118, 103645. doi: 10.1016/j.concog.2024.103645
- Speed, L. J., & Majid, A. (2018). An Exception to Mental Simulation: No Evidence for Embodied Odor Language. *Cognitive Science*, 42(4), 1146–1178. doi: 10.1111/cogs.12593

- Speed, L. J., Mak, M., & McRae, K. (2026). Investigating mental simulation during sentence comprehension in aphantasia. *Neuropsychologia*, 227, 109443. doi: 10.1016/j.neuropsychologia.2026.109443
- Stanfield, R. A., & Zwaan, R. A. (2001). The Effect of Implied Orientation Derived from Verbal Context on Picture Recognition. *Psychological Science*, 12(2), 153–156. doi: 10.1111/1467-9280.00326
- Stokes, M. G. (2015). ‘Activity-silent’ working memory in prefrontal cortex: A dynamic coding framework. *Trends in Cognitive Sciences*, 19(7), 394–405. doi: 10.1016/j.tics.2015.05.004
- Stokes, M. G., Thompson, R., Cusack, R., & Duncan, J. (2009). Top-Down Activation of Shape-Specific Population Codes in Visual Cortex during Mental Imagery. *Journal of Neuroscience*, 29(5), 1565–1572. doi: 10.1523/JNEUROSCI.4657-08.2009
- Stoliker, D., Preller, K. H., Novelli, L., Anticevic, A., Egan, G. F., Vollenweider, F. X., & Razi, A. (2025). Neural mechanisms of psychedelic visual imagery. *Molecular Psychiatry*, 30(4), 1259–1266. doi: 10.1038/s41380-024-02632-3
- Sulik, J., Rim, N., Pontikes, E., Evans, J., & Lupyan, G. (2025). Differences in psychologists’ cognitive traits are associated with scientific divides. *Nature Human Behaviour*, 1–15. doi: 10.1038/s41562-025-02153-1
- Suzuki, M., & Larkum, M. E. (2020). General Anesthesia Decouples Cortical Pyramidal Neurons. *Cell*, 180(4), 666–676.e13. doi: 10.1016/j.cell.2020.01.024
- Team, R. C. (2023). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.
- Tibshirani, R. (1996). Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1), 267–288. doi: 10.1111/j.2517-6161.1996.tb02080.x
- Tingley, D., Yamamoto, T., Hirose, K., Keele, L., & Imai, K. (2014). Mediation: R Package for Causal Mediation Analysis. *Journal of Statistical Software*, 59, 1–38. doi: 10.18637/jss.v059.i05
- Van der Stigchel, S., Strauch, C., de Zwart, B., & Van Maanen, L. (2026). Will online behavioral research follow the fate of online survey research? *Proceedings of the National Academy of Sciences*, 123(8), e2535585123. doi: 10.1073/pnas.2535585123

- van Moorselaar, D., Theeuwes, J., & Olivers, C. N. L. (2014). In competition for the attentional template: Can multiple items within visual working memory guide attention? *Journal of Experimental Psychology. Human Perception and Performance*, 40(4), 1450–1464. doi: 10.1037/a0036229
- Vanbuckhave, C., Eikner, J. S., Laeng, B., Onnis, L., & Mathôt, S. (2026). Pupil Size Reflects Trial-Level Variability in Imagery Vividness During Immersive Storytelling but Not (or Hardly) Individual Differences in Trait Imagery. *Psychophysiology*, 63(4), e70298. doi: 10.1111/psyp.70298
- Vinaya, H., Trott, S., Pecher, D., Zeelenberg, R., & Coulson, S. (2025). Words and Worlds Both: Dynamic Effects of Distributional and Sensorimotor Information in Semantic Processing. *Open Mind*, 9, 2149–2174. doi: 10.1162/OPMI.a.316
- Wackermann, J., Pütz, P., & Allefeld, C. (2008). Ganzfeld-induced hallucinatory experience, its phenomenology and cerebral electrophysiology. *Cortex*, 44(10), 1364–1378. doi: 10.1016/j.cortex.2007.05.003
- Weber, S., Christophel, T. B., Görden, K., Soch, J., & Haynes, J.-D. (2024). Working memory signals in early visual cortex are present in weak and strong imagers. *Human Brain Mapping*, 45(3), e26590. doi: 10.1002/hbm.26590
- West, W. C., & Holcomb, P. J. (2002). Event-related potentials during discourse-level semantic integration of complex pictures. *Brain Research. Cognitive Brain Research*, 13(3), 363–375. doi: 10.1016/S0926-6410(01)00129-X
- Wicken, M., Keogh, R., & Pearson, J. (2021). The critical role of mental imagery in human emotion: Insights from fear-based imagery and aphantasia. *Proceedings. Biological Sciences*, 288(1946), 20210267. doi: 10.1098/rspb.2021.0267
- Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review*, 9(4), 625–636. doi: 10.3758/BF03196322
- Winsler, K., Midgley, K. J., Grainger, J., & Holcomb, P. J. (2018). An electrophysiological megastudy of spoken word recognition. *Language, Cognition and Neuroscience*, 33(8), 1063–1082. doi: 10.1080/23273798.2018.1455985
- Wittgenstein, L. (1953). *Philosophical investigations. Philosophische Untersuchungen*. Macmillan.

- Wu, Y. C., & Coulson, S. (2011). Are depictive gestures like pictures? commonalities and differences in semantic processing. *Brain and Language*, 119(3), 184–195. doi: 10.1016/j.bandl.2011.07.002
- Xie, W., & Zhang, W. (2017). Familiarity increases the number of remembered Pokémon in visual short-term memory. *Memory & Cognition*, 45(4), 677–689. doi: 10.3758/s13421-016-0679-7
- Xie, W., & Zhang, W. (2018). Familiarity Speeds Up Visual Short-term Memory Consolidation: Electrophysiological Evidence from Contralateral Delay Activities. *Journal of Cognitive Neuroscience*, 30(1), 1–13. doi: 10.1162/jocn_a_01188
- Xu, Y. (2017). Reevaluating the Sensory Account of Visual Working Memory Storage. *Trends in Cognitive Sciences*, 21(10), 794–815. doi: 10.1016/j.tics.2017.06.013
- Yang, S., Dong, Y., & Kiyonaga, A. (2026). *Flexible Working Memory in the Peripheral Nervous System*. bioRxiv. doi: 10.1101/2025.09.26.678884
- Ye, C., Guo, L., Wang, N., Liu, Q., & Xie, W. (2024). Perceptual encoding benefit of visual memorability on visual memory formation. *Cognition*, 248, 105810. doi: 10.1016/j.cognition.2024.105810
- Zeelenberg, R., Pecher, D., van der Meijden, M. E. M., Trott, S., & Bergen, B. (2025). Non-native language comprehenders encode implied shapes of objects in memory. *Cortex*, 182, 100–111. doi: 10.1016/j.cortex.2024.09.008
- Zeman, A. (2024). Aphantasia and hyperphantasia: Exploring imagery vividness extremes. *Trends in Cognitive Sciences*, 28(5), 467–480. doi: 10.1016/j.tics.2024.02.007
- Zeman, A., Milton, F., Della Sala, S., Dewar, M., Frayling, T., Gaddum, J., ... Winlove, C. (2020). Phantasia-The psychological significance of lifelong visual imagery vividness extremes. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, 130, 426–440. doi: 10.1016/j.cortex.2020.04.003
- Zhang, Z., & Lewis-Peacock, J. A. (2023a). Bend but don't break: Prioritization protects working memory from displacement but leaves it vulnerable to distortion from distraction. *Cognition*, 239, 105574. doi: 10.1016/j.cognition.2023.105574
- Zhang, Z., & Lewis-Peacock, J. A. (2023b). Prioritization sharpens working memories but does not protect them from distraction. *Journal of Experimental Psychology: General*, 152(4), 1158–1174.

doi: 10.1037/xge0001309

- Zhou, C., Lorist, M. M., & Mathôt, S. (2022). Categorical bias as a crucial parameter in visual working memory: The effect of memory load and retention interval. *Cortex*, 154, 311–321. doi: 10.1016/j.cortex.2022.05.007
- Zohar, E., Kozak, S., Abeles, D., Shahar, M., & Censor, N. (2024). Convolutional neural networks uncover the dynamics of human visual memory representations over time. *Cerebral Cortex*, 34(11), bhae447. doi: 10.1093/cercor/bhae447
- Zokaei, N., Board, A. G., Manohar, S. G., & Nobre, A. C. (2019). Modulation of the pupillary response by the content of visual working memory. *Proceedings of the National Academy of Sciences*, 116(45), 22802–22810. doi: 10.1073/pnas.1909959116
- Zwaan, R. A. (2014). Embodiment and language comprehension: Reframing the discussion. *Trends in Cognitive Sciences*, 18(5), 229–234. doi: 10.1016/j.tics.2014.02.008
- Zwaan, R. A., & Pecher, D. (2012). Revisiting Mental Simulation in Language Comprehension: Six Replication Attempts. *PLOS ONE*, 7(12), e51382. doi: 10.1371/journal.pone.0051382
- Zwaan, R. A., Stanfield, R. A., & Yaxley, R. H. (2002). Language comprehenders mentally represent the shapes of objects. *Psychological Science*, 13(2), 168–171. doi: 10.1111/1467-9280.00430