The Puzzle of Resumption During Interrupted Narrative Reading: Towards More Adaptive Gaze-Based Attention-Aware Learning Technologies

Vorgelegt von

Francesca Zermiani

aus Padova, Italien

Erklärung (Statement of Authorship)

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbständig verfasst habe und dabei keine andere als die angegebene Literatur verwendet habe. Alle Zitate und sinngemäßen Entlehnungen sind als solche unter genauer Angabe der Quelle gekennzeichnet. Die eingereichte Arbeit ist weder vollständig noch in wesentlichen Teilen Gegenstand eines anderen Prüfungsverfahrens gewesen. Sie ist weder vollständig noch in Teilen bereits veröffentlicht.

I hereby declare that this text is the result of my own work and that I have not used sources without declaration in the text. Any thoughts from others or literal quotations are clearly marked. The thesis was not used in the same or in a similar version to achieve an academic grading or is being published elsewhere.

Place, Date: Signature:

Contents

Contents

List of Acronyms

AALT   Attention-Aware Learning Technologies
ACT-R   Adaptive Control of Thoughts-Rational
CI   Construction-Integration
CLT   Cognitive Load Theory
CTML   Cognitive Theory of Multimedia Learning
GRE   Graduate Record Examination
ICC   Interclass Correlation Coefficient
LT-WM   Long-Term Working Memory
NLP   Natural Language Processing
RHQ   Reading Habits Questionnaire
SSPAN   Symmetry Span
SAT   Suite of Assessments
SWAS   Story World Absorption Scale
vsWMC   visuo-spatial Working Memory Capacity
WMC   Working Memory Capacity

List of Figures

List of Tables

Abstract

Attention is crucial in managing cognitive load and facilitating information retention, yet in education, the observed shift from sustained attention to multitasking has introduced new challenges in adapting learning materials and teaching practice. These issues are also encountered in the context of reading, a fundamental task in the educational process that increasingly occurs in digital environments. Here, frequent interruptions and multimedia elements often disrupt the attentional focus, thereby challenging comprehension. While attention-aware learning technologies for reading show promise in mitigating interruption effects through gaze-based interventions, their designs often lack adaptability to individual differences in how learners re-focus on the text, known as task resumption. Despite substantial research on interruptions, the cognitive factors influencing learners' ability to resume reading effectively, as well as the influence of text characteristics on this process, are not sufficiently explored. The resulting data scarcity in the field of interrupted reading limits the development of personalized, adaptive tools to support diverse educational needs in interruption-prone settings. Besides, on a methodological perspective, research on interrupted reading and multimedia learning using narrative texts remain limited, despite these texts' rich stylistic and linguistic features that implicitly guide attention and provide engaging reading experiences. Using eye tracking experiments, resumption lag as behavioral measure of task resumption, and narratives as reading material, this thesis makes three major contributions. First, to address the need for studies investigating individual-related cognitive factors influencing resumption during interrupted reading, the thesis examines whether individual differences in visuo-spatial working memory capacity and prior knowledge can predict resumption lags. Findings indicate that the interaction of these factors lead to significantly shorter resumption lags after interruptions. Second, addressing the need for research on what text-specific features influence resumption processes, the thesis compares two types of fictional texts, one enactive and one descriptive, designed to evoke distinct mental imagery styles. Results demonstrate that enactive texts, contrary to expectations, prompt shorter fixation durations during resumption lags. Finally, to tackle the lack of data on interrupted reading, the thesis introduces *InteRead*, the first open-source, comprehensive eye tracking dataset dedicated to interrupted reading.

List of Tables

It mainly includes gaze data and extracted linguistic features from 50 participants reading a long fictional text with interruptions. The validation analyses conducted on the dataset reveal significant effects of interruptions, word characteristics, and individual differences on reading behavior. Taken together, the evidence shown in this thesis highlights the critical need to tailor educational strategies and adaptive reading technologies to both learners' cognitive profiles and text characteristics, fostering greater resilience to interruptions. It also provides a first step toward the development of a comprehensive corpus to support cross-disciplinary research on interrupted reading.

Deutsche Zusammenfassung

Aufmerksamkeit spielt eine entscheidende Rolle bei der Bewältigung kognitiver Beanspruchung und Förderung der Informationsspeicherung. In der Bildung hat der beobachtete Wandel von anhaltender Aufmerksamkeit bis hin zum Multitasking neue Herausforderungen bei der Anpassung von Lehrmaterialien und -methoden mit sich gebracht. Diese Probleme zeigen sich auch beim Lesen, einer grundlegenden Aufgabe im Bildungsprozess, die zunehmend in digitalen Umgebungen stattfindet. Häufige Unterbrechungen und multimediale Elemente stören dabei oft den Fokus, was das Textverständnis erschwert. Während aufmerksamkeitsbewusste Lerntechnologien für das Lesen vielversprechend darin sind, die Auswirkungen von Unterbrechungen durch blickbasierte Interventionen zu verringern, weist ihre Gestaltung jedoch häufig auf mangelnde Anpassungsfähigkeit an individuelle Unterschiede in der Wiederaufnahme der Aufgabe hin. Trotz umfangreicher Forschung zu Unterbrechungen sind die kognitiven Faktoren, die die Fähigkeit von Lernenden zur effektiven Wiederaufnahme der Arbeit beeinflussen, sowie die Rolle von Texteigenschaften in diesem Prozess nicht ausreichend erforscht. Der daraus resultierende Mangel an Daten im Bereich des Lesens mit Unterbrechungen schränkt die Entwicklung personalisierter und adaptiver Werkzeuge ein, die den unterschiedlichen Bildungsbedürfnissen in unterbrechungsanfälligen Umgebungen gerecht werden könnten. Auch aus methodischer Perspektive bleibt die Forschung zum Lesen mit Unterbrechungen und multimedialem Lernen mit narrativen Texten begrenzt, obwohl diese durch ihre stilistischen und sprachlichen Merkmale die Aufmerksamkeit lenken und ein interessantes Leseerlebnis bieten. Durch die Erfassung von Blickbewegungen, der Verzögerung von Wiederaufnahme des Lesens als Verhaltensmaß für die Wiederaufnahme der Arbeit und narrativen Texten als Lesematerial leistet diese Dissertation drei wesentliche Beiträge. Erstens untersucht sie, ob Unterschiede in der visuell-räumlichen Kapazität des Arbeitsgedächtnisses und im Vorwissen die Wiederaufnahme der Arbeit vorhersagen können um den Bedarf an Studien zu kognitiven, individuellen Einflussfaktoren auf die Wiederaufnahme der Arbeit beim Lesen mit Unterbrechungen zu decken. Die Ergebnisse zeigen, dass die Interaktion dieser Faktoren zu signifikant kürzeren Wiederaufnahmezeiten nach Unterbrechungen führt. Zweitens, um die Notwendigkeit von Forschung zu textbezogenen Merkmalen zu adressieren,

welche die Wiederaufnahme unterbrochener Aufgaben beeinflussen, vergleicht die Dissertation zwei Arten von fiktionalen Texten – einen enaktiven und einen deskriptiven –, die darauf ausgelegt sind, unterschiedliche Stile der mentalen Vorstellungsfähigkeit hervorzurufen. Die Ergebnisse zeigen, dass enaktive Texte, entgegen den Erwartungen, kürzere Fixationsdauern während der Phase zur Wiederaufnahme der Arbeit fördern. Drittens, um den Mangel an Daten zum unterbrochenen Lesen zu beheben, stellt die Dissertation mit *InteRead* den ersten öffentlich zugänglichen, umfassenden Eye-Tracking-Datensatz vor, der sich dem unterbrochenen Lesen widmet. Dieser Datensatz umfasst vor allem Blickdaten und extrahierte linguistische Merkmale von 50 Teilnehmenden, die einen langen fiktionalen Text mit eingestreuten Unterbrechungen lasen. Die Auswertung des Datensatzes zeigt signifikante Effekte von Unterbrechungen, Wortmerkmalen und individuellen Unterschieden auf das Leseverhalten. Zusammenfassend betonen die in dieser Dissertation gezeigten Ergebnisse die dringende Notwendigkeit, Bildungsstrategien und adaptive Lesetechnologien sowohl an die kognitiven Profile der Lernenden als auch an die Texteigenschaften anzupassen, um eine größere Widerstandsfähigkeit gegenüber Unterbrechungen zu fördern. Gleichzeitig stellen die Befunde erste Schritte in Richtung der Entwicklung einer umfassenden Datenbank dar, die interdisziplinäre Forschung zum unterbrochenen Lesen unterstützt.

Acknowledgments

First and foremost, I would like to express my deepest gratitude to my supervisor, Prof. Maria Wirzberger, for her support, insightful advice, positivity and constant motivation throughout this journey. Her enthusiasm for research, expertise, and ability to inspire have been crucial for my growth. She demonstrated what it means to be a good leader–through vision, courage, gratitude, transparency and, most importantly, resilience. Her passion and creativity in teaching, as well as her confidence in my teaching abilities long before I believed in them myself, have been both empowering and enlightening. Allow me to draw upon a metaphor we often used during our meetings: I am grateful for every time she was my rescue helicopter in moments when a deep jungle seemed to obscure my path. Her empathy and respect for my challenges, together with her capacity to refuel my spirits, helped me rise above the jungle, enabling me to see that my path was always there, I simply needed a new perspective. I feel truly fortunate to have had the privilege of learning and growing under her guidance.

I am also grateful to my co-supervisor, Prof. Andreas Bulling, for his constructive feedback, discussions about career development, and critical insights that significantly enhanced my research and prompted reflection on my future goals. He opened the doors to his team and its resources from the very beginning. His encouragement to actively participate in their research seminars not only expanded my knowledge and research interests but also paved the way for meaningful collaborations and friendships. These connections have been instrumental in shaping both my PhD journey and my personal growth.

Furthermore, I would like to express my gratitude to: Prof. Sibylle Baumbach for providing key resources that became fundamental to this thesis and for sharing her expertise in English literature; Prof. Titus von der Malsburg, for his insightful feedback, drawing from his extensive experience with eye tracking experiments and psycholinguistics; Prof. Benedikt Ehinger for his technical support on eye movement analyses; Prof. Moniek M. Kuijpers, for providing experimental resources, showing interest in my research, and sharing her expertise in quantitative narrative studies. Besides, I would like to thank Prof. Raquel G. Alhama, who has been a great mentor not only during my research internship but also during my PhD journey, offering a

care have been invaluable in keeping me grounded and helping me hang in there when the ride got rough.

Francesca Zermiani

1. Introduction

A student is deeply engaged in reading a textbook on a laptop when their phone vibrates with a notification. They glance at the screen to see a message from a friend, pulling them out of their focus. Later, a notification from a social media app pops up, tempting them to take a quick break. After a few minutes, a reminder from the calendar app appears on the laptop screen, notifying that the essay for the English course is due tomorrow. Following their friend's previous text message, the student receives a phone call, disrupting their reading flow as they decide whether to answer or ignore the call.

As this scenario highlights, attention plays a critical role in environments where we are required to frequently filter multiple types of information and reach a successful outcome. Education is one of such rapidly changing environments, where teachers have observed a shift from sustained attention to multitasking [Hayles, 2007]. This shift has posed new challenges in adapting learning materials and teaching models, often requiring them to be more concise and engaging [Hayles, 2007]. The study of attention in this context is therefore highly relevant due to its implications for enhancing learning outcomes and developing engaging learning material and supportive educational technologies. Research has shown that attentional processes are crucial for managing cognitive load and facilitating information retention [Sweller, 1988]. Moreover, individual differences in attentional capacity can affect how students handle interruptions and subsequently resume their primary tasks, which is a common challenge in learning environments [D'Mello and Graesser, 2012].

The role of attention has become particularly significant in the context of reading, as a demanding task to both learn and teach [Elleman and Oslund,

1. Introduction

2019]. Reading is a fundamental skill that underpins much of the educational process and learning materials, facilitating the acquisition of new knowledge and the development of critical thinking skills [Smith et al., 2021]. However, the shift towards digital devices for reading tasks has exponentially increased, introducing new challenges for maintaining attention [Delgado and Salmerón, 2021, Chevet et al., 2022a]. Digital reading often involves navigating hypertext, multimedia elements, and frequent distractions from notifications and other digital interruptions [Liu, 2022]. These interruptions can disrupt the reading flow, making it more difficult for learners to maintain focus and understand the material.

To mitigate the negative effect of disruptions, Attention-Aware Learning Technologies (AALT) have been developed. These technologies track and model students' attentional states, intervening when they are distracted [D'Mello, 2019]. Eye tracking has been proven to benefit the development of AALT for its capacity to detect eye movements, hence inferring more implicit attentional processes [Conati et al., 2013]. Fundamental research in eye tracking has indeed established a strong connection between overt attention and underlying cognitive processes, demonstrating that eye movements serve as a reliable indicator of attention allocation [Just and Carpenter, 1980, Rayner, 1998]. This has been particularly critical in advancing our understanding of reading-related cognitive mechanisms [Reichle et al., 1998]. As a consequence, several gaze-based AALT and applications have been developed for digital reading [D'Mello et al., 2016]. By identifying moments of inattention from gaze behavior and providing timely re-focusing interventions, these technologies aim to enhance the reading experience and support readers in maintaining their focus [D'Mello et al., 2012, 2017]. Furthermore, eye tracking data has been utilized in reading applications with diverse scopes, including identifying document types [Kunze et al., 2013], predicting comprehension levels [Mézière et al., 2023, Southwell et al., 2023], and minimizing the impact of interruptions [Jo et al., 2015, Mariakakis et al., 2015, Srivastava et al., 2021].

One key limitation in the development of these technologies becomes apparent in their one-size-fits-all designs, which still lack adaptability [Mari-

akakis et al., 2015, Hutt et al., 2021]. These technologies, while beneficial, often support only specific groups of students and fail to adjust their support mechanisms to different re-focusing strategies [Hutt et al., 2021]. While there is broad evidence on the process of re-focusing on a task after being interrupted, commonly defined as task resumption [Altmann and Trafton, 2002, Trafton et al., 2003], a systematic consideration of the factors influencing this process currently lacks Drews and Musters [2015]. While we have substantial knowledge about interruptions and their properties, the elements that make learners more or less efficient at handling interruptions remain largely unknown [Meys and Sanderson, 2013]. This gap hampers the development of adaptive AALT and applications that can respond to diverse learner needs and, consequently, enhance their reading experience.

1.1. Contributions Overview

To address these limitations, this dissertation makes three major contributions: (i) a study of reader-related factors influencing task resumption during interrupted reading; (ii) a study of text-related factors influencing task resumption during interrupted reading; (iii) the first publicly available dataset specifically designed to study interrupted reading across research fields.

Furthermore, focusing exclusively on narrative texts as reading material, this dissertation offers a novel contribution to the field of interrupted reading and multimedia learning, where such texts have been marginally explored. Narratives are not only rich in stylistic and linguistic features that inherently guide and manage readers' attention but are also strategically designed to engage readers through experiences such as absorption and imagery [Kuijpers et al., 2014, Jacobs and Willems, 2018, Kukkonen and Baumbach, 2022]. Narratives also show promise in educational settings, promoting creativity and literacy skills [Collins, 1999, Gallagher, 2011]. By examining these texts, this thesis highlights the potential of narratives as powerful tools for fostering attention, setting them apart from other materials typically used in multimedia learning studies.

For contribution (i), I investigate whether individual differences in vsWMC

and prior knowledge predict variations in resumption performance following interruptions. This performance is behaviorally quantified as the resumption lag, the interval between the conclusion of the interruption and the first relevant action taken to re-engage with the primary task [Altmann and Trafton, 2002, Trafton et al., 2003, Cane et al., 2012]. Previous research has shown that readers make use of their spatial working memory for the location of words within the text while reading [Cane et al., 2012] and draw on their prior knowledge for text comprehension [Rayner and McConkie, 1976, McCarthy and McNamara, 2021]. Participants' eye movements were tracked while they read a crime fiction extract and experienced interruptions, allowing for the analysis of how individual differences in vsWMC and prior knowledge affect resumption performance. Results indicate that the interaction of vsWMC and prior knowledge significantly predicts resumption time after interruptions, with higher vsWMC and prior knowledge leading to lower resumption lags.

For contribution (ii), I explore text-related factors by using two fictional texts designed with either a more enactive or descriptive mental imagery style. This distinction, based on Kuzmičová [2014]'s phenomenological typology theory, has been associated with variations in gaze behavior during reading, reflecting differences in the mental imagery processes evoked. Eye tracking data were collected as participants read these texts and experienced interruptions, providing insights into how different text features influence the resumption of reading tasks. The enactive style text is hypothesized to prompt longer resumption times following an interruption, accompanied by longer fixation durations. In contrast, the descriptive style text is expected to lead to shorter resumption times, with shorter fixation durations. Findings show that enactive style texts exhibit significantly shorter fixation durations during resumption lags.

For contribution (iii), I introduce *InteRead*, the first publicly available eye tracking dataset focused on interrupted narrative reading, providing gaze data from 50 participants, along with linguistic features extracted from the reading material [Zermiani et al., 2024a]. The dataset also provides demographic data, annotations of interruptions interspersed throughout the text

and resumption times following these interruptions. Consistent with prior research, the analyses performed to validate InteRead demonstrate that interruptions, word length, and word frequency significantly affect eye movements during reading. The analyses also display significant individual variability in reading time and resumption lags across participants. Overall, InteRead serves as a novel resource for advancing the understanding and research of factors affecting task resumption during reading, as well as the cognitive processes involved in narrative reading.

The results reveal key insights into how both reader- and text-related factors impact resumption performance. The findings discussed in this dissertation offer a fundamental understanding of resumption during interrupted narrative reading. Moreover, such insights can inform teaching practices and guide the further development of AALT to better align with individual learner needs and specific materials. Future research could explore additional factors and their impact on resumption performance during reading to improve our understanding of this process at both cognitive and linguistic levels. Moreover, the InteRead dataset opens new avenues for research. Future work could broaden the scope of this dataset by including participants challenged by maintaining their attentional focus and processing textual stimuli, and by investigating additional languages, thereby enhancing its applicability across diverse linguistic contexts and educational needs. Given the challenges digital reading presents for sustained attention [Delgado and Salmerón, 2021] and the increasing prevalence of multitasking in educational settings [Hayles, 2007], it is essential to study how different learners respond to interruptions and re-engage with text. This thesis represents a crucial step toward more personalized and effective educational tools, ultimately aiming to improve the learning experience.

1.2. Publications Overview

During the preparation of this thesis, I co-authored the following publications, emphasizing some of my research contributions. At the start of each relevant chapter, I will specify which paper underpin the results and insights

1. Introduction

discussed and clarify whether the content in the chapter summarizes the published work or introduces new unpublished contributions.

Parts of the research described in this thesis have been published in:

- Francesca Zermiani, Prajit Dhar, Ekta Sood, Fabian Kögel, Andreas Bulling, and Maria Wirzberger. InteRead: An Eye Tracking Dataset of Interrupted Reading. In Nicoletta Calzolari, Min-Yen Kan, Veronique Hoste, Alessandro Lenci, Sakriani Sakti, and Nianwen Xue, editors, *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 9154–9169, Torino, Italia, 2024a. ELRA and ICCL. URL https://aclanthology.org/2024.lrec-main.802

  – As the first author of this publication, I was primarily responsible for conceptualizing the research goals, conducting the experiment and collecting the data, as well as designing the research methods. Prajit Dhar and I collaboratively authored the original draft, with my contributions centered on introducing the research scope and dataset, outlining the methods and data collection, and framing the findings within the broader context of the field. My co-authors contributed through formal analysis, data validation and curation, as well as visualization, whereas supervision, resources, support on conceptualization and methods, and funding acquisition were provided by senior collaborators. All authors contributed to iterative writing reviewing and editing processes.

- Francesca Zermiani, Prajit Dhar, Florian Strohm, Sibylle Baumbach, Andreas Bulling, and Maria Wirzberger. Individual differences in visuo-spatial working memory capacity and prior knowledge during interrupted reading. *Frontiers in Cognition*, 3, 2024b. ISSN 2813-4532. doi: 10.3389/fcogn.2024.1434642

  – As the first author of this publication, I was primarily involved in conceptualizing the research objectives, developing the methodology, conducting the experiment and collecting the data, and performing formal analysis and validation of the results. I also led the writing of the

original draft. My co-authors contributed through formal analysis and software support for the experiment and annotation tool. Supervision, resources, support on overarching conceptualization and methods, and funding acquisition were provided by senior collaborators. All authors contributed to iterative writing reviewing and editing processes.

2. Theoretical Background

2.1. Information Processing and Instructional Design

To approach the topic discussed in this thesis from a theoretical perspective, we first need to identify crucial aspects of human cognition that shape how information is constructed and cognitive resources are managed in educational settings. A long tradition of research aimed at disentangling the cognitive structure of information processing along with its limitations [Kotseruba et al., 2016]. In particular, established models of memory [Anderson, 1983, Atkinson and Shiffrin, 1971, Baddeley, 1992] posited that working memory is extremely limited in processing duration, approximately 30 s [Cowan, 1988], and capacity of stored information, with a maximum of approximately four items available at the same time [Cowan, 2001]. Long-term memory storage and capacity are instead more stable and long-lasting, retaining structured and organized information [Atkinson and Shiffrin, 1971, Ericsson and Kintsch, 1995]. These two systems communicate with each other, as working memory is mostly considered the workspace where long-term memory representations are temporarily activated [Logie, 1996, Van der Linden, 1998]. During new knowledge processing, memory structures are formed and organized, often referred to as schemata, according to memory schema theories [Bartlett, 1932, Anderson, 1984]. Within the literature, schemata have been attributed various functions, broadly categorized as guiding behavior, facilitating memory encoding, enhancing information retrieval, and supporting the assimilation of new knowledge [Ghosh and Gilboa, 2014]. Research on the cognitive architecture and limitations of our information processing sys-

tem further provided the base for theories of instructional design, such as the prominent Cognitive Load Theory (CLT) [Sweller, 1988, 1994, 2010]. CLT introduced the construct of cognitive load as the load exercised on our working memory limited resources when processing information to be later stored in long-term memory. The theory further postulated that good instructional design avoids excessive working memory load, thereby promoting learning mechanisms [Sweller, 2010].

This research has been instrumental for navigating the increase use of multiple media and technologies in instructional design [Mayer, 2005, Mutlu-Bayraktar et al., 2019]. In recent years, education has indeed exponentially shifted towards digital platforms, driven by the growing application of technology-assisted learning tools such as digital classrooms, online assessments, and adaptive learning systems [Haleem et al., 2022]. These tools aim to promote sustainability and inclusivity in education as well as to provide essential support during critical periods like the COVID-19 pandemic, which dramatically accelerated their use [Kang, 2021, Haleem et al., 2022]. While these advancements have offered numerous benefits, they have also introduced new challenges and opportunities in the design of learning materials, as these educational technologies have raised the amount of accessible media and channels for presenting information [Clark and Mayer, 2016, Liu et al., 2018, Mutlu-Bayraktar et al., 2019].

Multimedia learning research has focused exactly on developing design principles for multimedia learning environments, considering how new information is processed and how to convey messages through both verbal and visual stimuli [Mayer, 1997, 2009, Mutlu-Bayraktar et al., 2019]. Connected to and rooted in foundational frameworks such as the CLT [Sweller, 1988], Baddeley [1992]'s model of working memory, and Paivio [1990]'s dual coding theory, the well-established Cognitive Theory of Multimedia Learning (CTML) [Mayer, 1997, 2005] comprises three primary cognitive assumptions: information is processed through different channels–one dedicated to visuo-spatial and the other to auditory-verbal processing; each channel has a limited capacity for information storage; and knowledge acquisition occurs by activating related knowledge, creating mental structures, and integrat-

ing them with prior knowledge [Mayer, 2005, Mutlu-Bayraktar et al., 2019]. According to the CTML, multimedia learning involves selecting relevant verbal and visual stimuli through our sensory memory, organizing them into verbally-based and visually-based models within our working memory, and finally connecting such models with existing structures in our long-term memory [Mayer, 1997, Mayer and Moreno, 2003, Mayer, 2005, Mutlu-Bayraktar et al., 2019].

Informed by the CLT, the CTML further developed several design principles to inform the creation of multimedia learning materials that effectively optimize learning outcomes [Mayer and Moreno, 2003, Mayer, 2005, Mutlu-Bayraktar et al., 2019]. These principles, defined and tested across extensive research (e.g., Moreno and Mayer [2000], Mautone and Mayer [2001], Mayer [2005], Butcher [2014], Mayer and Fiorella [2014], Mayer and Pilegard [2014], Leopold and Mayer [2015], Rodrigues and Silva [2022]), serve different purposes, such as reducing unnecessary cognitive load that does not support learning goals, effectively managing the selection and organization of the presented information, and fostering the integration of new information with prior knowledge [Mutlu-Bayraktar et al., 2019]. When properly applied, these principles can enhance attention and learning by making complex information more accessible and engaging, without overwhelming students' cognitive capacities [Mayer and Moreno, 2003, Mayer, 2005, Moreno and Mayer, 2007]. This highlights the importance of careful instructional design to ensure that multimedia elements complement rather than compete with each other for the learner's attention [Sweller, 1988, Mayer, 2005].

2.2. Reading

Reading is a cornerstone of education, essential for acquiring knowledge across diverse subjects and a critical cognitive skill that underpins the learning process [Elleman and Oslund, 2019, Smith et al., 2021]. The importance of reading extends beyond its educational value, as it is also a fundamental human cognitive activity deeply intertwined with several major cognitive processes, such as attention, word recognition, syntax parsing, eye movement

control [Rayner and Reichle, 2010]. For this reason, discourse processing and text comprehension have been the focus of extensive research for over 40 years [Van Dijk and Kintsch, 1983, Graesser et al., 1997, Schnotz, 2002].

Among the numerous developed frameworks, the Construction-Integration (CI) model [Kintsch and van Dijk, 1978, Van Dijk and Kintsch, 1983, Kintsch, 1991, 1998] is one of the most prominent. As readers engage with written material, they construct mental representations of its meaning through three major processing levels: the surface code level, storing linguistic details of the most recent words and syntactic information; the textbase level, where propositions are developed and integrated to infer the local meaning of the text; and ultimately, the situation model level, where different components of the texts are connected to prior mental representations for a holistic comprehension of the underlying themes and messages of the text [Van Dijk and Kintsch, 1983, Chevet et al., 2022a]. According to the CI model, readers thus construct mental representations of the events, actions, and characters described in the text, integrating this information with their prior knowledge and real-world experiences, according to their goals [Kintsch, 1991, Schnotz, 2002]. Despite being an automatic cognitive process, the construction and consolidation of the situation model of a text requires continuously selecting the most salient text components, making use of the limited resources of working memory [Kintsch and van Dijk, 1978, Kintsch, 1991, Chevet et al., 2022a]. Effective reading therefore requires sustained attention to relevant information, while also filtering out irrelevant details and resisting distraction [Reynolds and Anderson, 1982, Anderson, 1982, Yildiz and Çetinkaya, 2017].

2.2.1. Narrative Texts

Within the context of text processing, an important variable to address is the type of text being read. Text characteristics significantly impact reading [Smith et al., 2021]. The structure and organization of a text influence active processing and inference-making [Smith et al., 2021]. For instance, cohesive and coherent texts help identify relationships between sentences, linking in-

formation and reducing cognitive load [McNamara et al., 2011, Smith et al., 2021]. Text comprehension is also influenced by text genre, as the central process of linking prior knowledge and integrating it into the mental representations of a text, which forms meaning, is modulated by genre [Berkowitz and Taylor, 1981, Hartley, 1986, Wolfe and Woodwyk, 2010, McNamara et al., 2011].

Despite the several shapes a text can take, a macro-genre distinction is generally made between narrative and expository texts [Mar et al., 2021]. While expository texts aim at informing the readers about a certain subject, narrative texts entertain them through their events, interactions, characters and familiar plots [Mar et al., 2021]. Besides, narratives have shown to hold substantial potential for education, where they can foster creativity, motivation, critical thinking, and literacy skills, as reflected for instance in the power of storytelling in educational settings [Collins, 1999, Gallagher, 2011]. Narrative texts significantly differ from expository texts on both a structural and content level [Wolfe and Woodwyk, 2010, Mar et al., 2021]. Their structure and content closely mirror our everyday experiences, enhancing recall and comprehension [Mar et al., 2021, McNamara et al., 2011]. Stories typically follow a temporal and causal sequence, revolving around the characters and their actions [Graesser et al., 1991, Mar et al., 2021]. This familiar structure, which often includes emotional experiences associated with daily-life struggles and events, draws on extensive previous knowledge, supporting inferential thinking [Stein and Glenn, 1979, Graesser and Clark, 1985, Mar et al., 2021] and predictive processing [Kukkonen, 2020]. The emotional component particularly contributes to the potential of narratives for instructional design, in light of the observed benefits of positive emotions, induced through multimedia material, on learning as well as perceived task difficulty [Um et al., 2011, Rodrigues and Silva, 2022].

Additionally, the content of stories is centered on social relationships, psychology, and common themes like friendship, conflicts, and love: topics with which readers are deeply close to [Hogan, 2003, Mar et al., 2021]. This familiarity extends to the language used in narratives, as it closely resembles the vocabulary and expressions used in everyday conversations mirroring the

specific time period of the narration [Mar et al., 2021]. This strong connection between the structure and content of narratives and our prior experience of the narrated events contributes to the ease with which narratives are understood and remembered [Graesser et al., 1991, Wolfe and Woodwyk, 2010, Mar et al., 2021]. The language familiarity component of narratives further emphasizes their potential for instructional design. Materials presented in a conversational style have been associated with better learning performance, according to the so-called personalization principle in multimedia learning [Moreno and Mayer, 2000, Mayer, 2005].

Narratives are therefore particularly successful in binding readers' focus due to attention-grabbing linguistic devices [Mukařovský, 1964, Emmott et al., 2006, Sanford et al., 2006, Sanford and Emmott, 2012]. Such elements support readers in creating inferences and semantic representations of the text and include, for instance, fragmentation (i.e., graphically creating a sentence with phrases that cannot be classified as complete syntactic sentences) or mini-paragraphs (i.e., one-sentence paragraphs) [Emmott et al., 2006, Sanford and Emmott, 2012]. Another example can be found in detective fiction, where authors present murders and mysteries as problem-solving tasks for the readers, often referred to as the 'puzzle element' [Rzepka, 2005]. These strategies are reminiscent of the signaling effect in multimedia learning research, where instructional material containing signals, such as section headings and linking words of cause and effect, has been shown to improve comprehension and problem-solving performance [Mautone and Mayer, 2001, Richter et al., 2016]. Narrative texts therefore represent a rich material to uncover attention mechanisms in narrative reading through the analysis of linguistic attractors and distractors [Kukkonen and Baumbach, 2022, van de Ven, 2023].

2.2.2. Mental Imagery

Mental imagery is one of the key cognitive processes involved in reading, as it enables readers to create mental representations that simulate the sensory experience of the fictional worlds they encounter [Sadoski, 1983, Sadoski

et al., 1990, Jacobs and Willems, 2018]. This process refers to the reader's ability to create visual or sensory experiences in their mind based on the textual information provided [Jacobs and Willems, 2018]. It differs from explicit mental imagery, which involves deliberate and conscious allocation of cognitive resources to visualize a scene or object without direct sensory input [Jacobs and Willems, 2018, Mak and Willems, 2021]. In narrative reading instead, imagery is often more implicit, arising rapidly as the reader engages with the text [Jacobs and Willems, 2018, Mak and Willems, 2021]. It is less about creating vivid, picture-like images and more about simulating sensory and motor experiences that align with the content of the narrative [Jacobs and Willems, 2018, Mak and Willems, 2021]. When such representations are successfully prompted, they can lead to immersion [Kuijpers et al., 2014], reading pleasure and text recall [Sadoski, 1983, Sadoski et al., 1990, De Koning and van der Schoot, 2013]. A beneficial effect of mental imagery is also found in education, with the so-called imagination principle in instructional design [Sweller, 2003, Mayer, 2009, Leopold, 2021]: learning is more effective when students are asked to mentally imagine a learning content.

The process of eliciting non-conscious mental imagery is not uniform and can vary depending on the linguistic context, with stories being more likely to prompt mental imagery compared to isolated sentences [Jacobs and Willems, 2018]. This mental simulation is often reflected in the activation of related cortical areas in the brain while reading, which suggests that readers simulate the actions, emotions, and sensory experiences described in the text [Fischer and Zwaan, 2008, Jacobs and Willems, 2018, Mak et al., 2023]. For instance, reading action-related words can activate cortical motor areas of the brain [Fischer and Zwaan, 2008, Jacobs and Willems, 2018], while reading about emotional events can engage regions associated with emotion processing [Altmann et al., 2012, Jacobs and Willems, 2018]. Additionally, mental imagery during narrative reading can also be observed from gaze behavior [Mak and Willems, 2019, Magyari et al., 2020, Mak et al., 2023].

Several theories have attempted to model the intrinsic characteristics of narrative reading and what creates that peculiar experience of imageability while we read fictional stories [Scarry, 1999, Grünbaum, 2007, Jajdelska et al.,

2010, Gallese and Wojciehowski, 2011, Kuzmičová, 2012, 2013, 2014, Jacobs, 2015a,b]. What most of these models have in common is the focus on bodily movement language as key element to prompt mental imagery [Grünbaum, 2007, Jajdelska et al., 2010, Kuzmičová, 2012, 2013, 2014].

Grounded in embodied theories of language comprehension [Pecher and Zwann, 2005], Kuzmičová [2014] comprehensively defined the process of mental imagery as a consequence of describing objects in interaction with the character's actions or, in other words, through bodily movement language. This model identified two major varieties of mental imagery domain, namely verbal and referential [Kuzmičová, 2012, 2013, 2014]. Verbal imagery refers to the internal experience of perceiving or "hearing" the words in a narrative, either as if reading them aloud or as if listening to characters' voices during dialogues [Kuzmičová, 2014]. The referential variety instead emerges as the internal visualization or simulation of scenes, objects, or actions described in a narrative, either from an observer's perspective or as if experienced directly through the eyes of a character [Kuzmičová, 2014]. Referential imagery involves mental simulation that becomes conscious to the reader, thereby demanding cognitive resources [Kuzmičová, 2014]. Notably, these different varieties of mental imagery occur only at specific moments during reading, where various sensory inputs are concurrently activated and integrated across different modalities, both at the level of macro syntactic elements, such as sentences, paragraphs, and pages, and at the level of micro syntactic elements, e.g., words [Kuzmičová, 2014, Magyari et al., 2020]. Kuzmičová [2014]'s model further divided referential imagery between two subtypes: *enactment-imagery* and *description-imagery*[1].

Enactment-Imagery   Enactment-imagery occurs when the reader adopts the perspective of a character within the story, experiencing the events and actions as if they were happening to them [Kuzmičová, 2014]. This subtype of imagery employs multimodal channels, meaning it can involve visual, au-

---

[1]From this point onward throughout this thesis, the term "mental imagery" will specifically refer to Kuzmičová [2014]'s definition of the referential domain and its two subtypes, enactment- and description-imagery, introduced here.

ditory, kinesthetic, and tactile sensations, depending on the nature of the narrative [Kuzmičová, 2014]. The key trigger for enactment-imagery is textual cues involving object-directed bodily movements, where a character actively interacts with objects in the fictional world [Kuzmičová, 2014, Magyari et al., 2020]. For example, a sentence describing a character opening a door might prompt the reader to imagine the experience of the door handle, the movement of their hand, and the sound of the door creaking open [Magyari et al., 2020]. Enactment-imagery is typically associated with a higher level of immersion and transparency in the reading experience, as the reader becomes deeply engaged with the story, often feeling as though they are inside the narrative [Kuzmičová, 2014, Magyari et al., 2020].

Description-imagery In contrast, description-imagery involves the reader visualizing the scene from an external perspective, akin to observing the narrative from a distance [Kuzmičová, 2014]. This style of imagery is more static and focuses on the visual properties of objects and settings, without involving the reader in the characters' actions [Magyari et al., 2020]. For instance, a detailed description of a landscape might prompt the reader to visualize the scene but without feeling physically or emotionally involved in it [Magyari et al., 2020]. Description-imagery is usually associated with a lower level of immersion, as the reader remains a passive observer rather than an active participant in the story's world [Kuzmičová, 2014, Magyari et al., 2020].

Kuzmičová [2014]'s model suggests that these two styles of mental imagery are not mutually exclusive but can coexist within a single story, as different ends of a continuum, each offering a distinct way for readers to engage with the text [Magyari et al., 2020, Mak and Willems, 2021]. A story might constantly alternate between passages that evoke enactment-imagery and those that evoke description-imagery [Magyari et al., 2020, Mak and Willems, 2021]. The distinction between enactment- and description-imagery is also independent of the point of view adopted by the narrative, known as narrative perspective [Lanser, 1981], such as first-person or third-person per-

spective [Magyari et al., 2020]. One might indeed assume that first-person narratives would naturally lead to more enactment-imagery due to the closer identification with the protagonist [Magyari et al., 2020, Mak and Willems, 2021], as these narratives have been associated with higher immersion [Hartung et al., 2016]. However, Kuzmičová [2014]'s model emphasizes that it is the nature of the actions and descriptions within the text, rather than the narrative perspective itself, that determines the type of imagery evoked [Kuzmičová, 2012, 2013, Magyari et al., 2020]. Due to the coexistence of multiple variables like reader preferences and textual cues, the interplay of narrative perspective, immersion experience and mental imagery is therefore intricate and still requiring further empirical investigation [Magyari et al., 2020].

Taken together, these insights underscore the complexity of the cognitive processes and experiences that are involved in reading. While mostly perceived as an automatic task, different variables and mechanisms are at play, from text type and characteristics [Smith et al., 2021] to constructing mental representations [Kintsch and van Dijk, 1978] while maintaining attentional focus and disregarding irrelevant information [Reynolds and Anderson, 1982, Anderson, 1982]. This is especially critical for reading performed within digital environments, which include interruptions often prevailing over our required primary focus of attention [Chevet et al., 2022a].

2.3. Task Resumption

Let us go back to that student who received a message notification while being deeply engaged in reading a textbook. After exchanging a few messages with their friend, the student faces the challenge of returning to their textbook and regaining the focus that was previously disrupted. The mental shift from the social interaction back to reading requires effort, as their thoughts may still linger on the conversation. Besides, attentional disruptions are recurrent [Chevet et al., 2022b]. With multiple distractions pulling them away, the student must now refocus on the primary task at hand, navigating back through the material, recalling where they left off, and re-engaging with

the content. No matter how often one pauses their *primary task* to address an interrupting *secondary task*, the need to eventually resume the primary task is inevitable in order to complete it. *Task resumption* can only occur once the individual has recalled what they were doing before the interruption and identified the next relevant action to take [Trafton and Monk, 2007].

Research across several domains have investigated the process of resuming a task after an interruption. For instance, empirical research in the visual search domain has shown that participants experiencing short and frequent interruptions while viewing a search display detect targets more quickly in subsequent presentations of the same display compared to the initial one, a phenomenon known as the rapid resumption effect [Lleras et al., 2005]. In other words, interruptions during search tasks enhance performance in future search tasks on the same display [Lleras et al., 2005, Jungé et al., 2009]. The rapid resumption effect is attributed to the accumulation of visual information about the search display during interruptions, enabling participants to form hypotheses about the target's location, defined as the perceptual hypothesis [Lleras et al., 2005]. However, later studies raised questions regarding the actual extent of the benefit derived from rapid resumption in visual search [Shen and Jiang, 2006]. A central challenge in this field is indeed the lack of uniformity in how interruptions are defined and studied, leading to methodological diversity that complicates the development of a cohesive theoretical framework, as evidenced by a recent systematic review of the field [Cambronero-Delgadillo et al., 2024].

For these reason, and given that the studies reported in this thesis investigate resumption during a more complex task like reading, I will focus on the two most prominent and widely established frameworks in interruption literature. These include the the Memory for Goals model [Altmann and Trafton, 2002], addressing the memory mechanisms involved in task interruption and resumption, and the LT-WM theory [Ericsson and Kintsch, 1995], focusing on memory processes in expert performance, with a particular emphasis on reading. These frameworks offer insights into how interruptions impact task performance and how individuals manage interruptions.

2.3.1. Memory for Goals

In the Memory for Goals model [Altmann and Trafton, 2002, 2007], mental representations of an individual's intent to complete a specific task are referred to as goals. Every time a goal is required, the probability that our cognitive resources will be directed toward that goal increases, activating the related goal in our declarative memory, according to the construct of activation defined within the cognitive theory of Adaptive Control of Thoughts-Rational (ACT-R) [Anderson and Lebiere, 1998]. Goals are not static and their level of activation fluctuates over time influenced by different factors, such as the need to complete a subgoal, to attend another goal or simply due to an interruption [Altmann and Trafton, 2002, 2007]. The effect of such competing factors is further defined as the so-called interference level: as long as a goal is above such level, this goal guides an individual's behavior [Altmann and Trafton, 2002]. When a task is interrupted for as little as 30 s and the goal associated to the task drops below the interference level, that goal is temporarily suspended, hence its activation level quickly diminishes [Altmann and Trafton, 2002, 2007]. As a result, upon resumption of the main task, the time needed to pick up where one left off reflects the effort required to retrieve and reactivate the related suspended goal, namely the resumption cost [Altmann and Trafton, 2002]. Such effort increases as the activation level of the suspended goal decreases due to the interruption [Altmann and Trafton, 2007].

The framework also describes predictive constraints within this cognitive process, which can support information retrieval upon resumption, namely priming and strengthening. On the one hand, the priming constraint emphasizes cue availability, as it states that a suspended goal can only be retrieved when primed with a related cue [Altmann and Trafton, 2002, Hodgetts and Jones, 2006a, Altmann and Trafton, 2007, Trafton and Monk, 2007, Monk et al., 2008]. For priming to be effective, however, the cue linked to the goal must be available just before an interruption, establishing the connection, and again at resumption, allowing the cue to activate and prime the suspended goal [Altmann and Trafton, 2002, Hodgetts and Jones, 2006a, Alt-

mann and Trafton, 2007, Trafton and Monk, 2007, Monk et al., 2008]. The strengthening constraint, on the other hand, involves proactively increasing the activation level of the suspended goal by reinforcing the connection between environmental cues and that goal before the interruption [Altmann and Trafton, 2002, Hodgetts and Jones, 2006a, Altmann and Trafton, 2007, Trafton and Monk, 2007, Monk et al., 2008]. This reinforcement allows for the goal to be better maintained during the interruption and facilitates planning the actions to take toward the goal upon resumption [Altmann and Trafton, 2002, Trafton et al., 2003].

This process of preparing for resumption can be achieved during the *interruption lag*, defined as the time interval between a person first recognizing that an interruption is imminent, for example, hearing/seeing a message notification, and the moment they actually engage with the interruption, in this case the chat [Trafton et al., 2003]. Individuals can deploy two distinct strategies during the interruption lag: prospective goal encoding and retrospective rehearsal [Trafton et al., 2003]. Prospective goal encoding is closely related to the strengthening constraint and the action of increasing the basic activation level of the to-be-suspended goal, to enable planning of the upcoming actions upon resumption [Altmann and Trafton, 2002, Trafton et al., 2003]. Retrospective rehearsal instead involves mentally reviewing the state of the primary task at the moment of interruption, such as recalling the last action taken before the interruption occurred [Trafton et al., 2003]. Upon resumption, this memory would then ease the effort of determining the next appropriate step [Trafton et al., 2003].

Most importantly, Trafton et al. [2003] introduced a measure for task resumption, the *resumption lag*, described as the time between the end of the interruption and the first relevant action directed toward the suspended goal. The model's assumption of resumption cost has been primarily tested on relatively complex problem-solving tasks [Altmann and Trafton, 2002, Hodgetts and Jones, 2006a,b], procedural and decision-making tasks requiring to follow specific steps to complete an activity [Altmann and Trafton, 2004, Trafton et al., 2003, 2005, Altmann and Trafton, 2007, Monk et al., 2008, Altmann and Trafton, 2015, Wirzberger and Russwinkel, 2015, Hirsch et al., 2024],
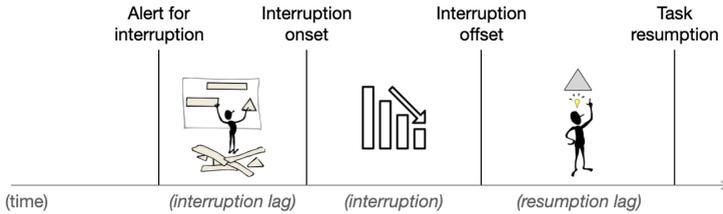
Figure 2.1.: Visualization of the main constructs and assumptions of the Memory for Goals activation-based model, as outlined in Altmann and Trafton [2002], Trafton et al. [2003]. After receiving the alert for an interruption, during the *interruption lag*, individuals can make use of prospective goal encoding or retrospective rehearsal to later accelerate resumption. While attending the interruption, the basic activation level of the suspended goal associated with the primary task rapidly decays, unless strengthened during the interruption lag. In the *resumption lag*, individuals re-activate the suspended goal's mental representations stored in memory through the available cues.

and learning tasks [Wirzberger et al., 2020]. Figure 2.1 displays a visual representation of the Memory for Goals framework as described in Altmann and Trafton [2002] and Trafton et al. [2003].

2.3.2. Long-Term Working Memory

Similarly to the Memory for Goals framework, the LT-WM theory [Ericsson and Kintsch, 1995, Delaney and Ericsson, 2016] also targets memory processes but from a distinct perspective and within a different context, leading to opposite assumptions. While the Memory for Goals model was initially developed to tackle memory processes in problem-solving tasks, the LT-WM theory aimed at explaining these processes in expert-level perfor-

mance. Among different expertise domains, [Ericsson and Kintsch, 1995] also focused on the reading domain, assuming that text comprehension is a highly developed skill mastered by most educated adults. After gaining expertise in a certain domain, experts can make use of long-term memory more efficiently and in integration with working memory, which is referred to as LT-WM Ericsson and Kintsch [1995], Gobet [2000], Delaney [2018]. This integration occurs through retrieval structures, which are retrieval cues formed during task encoding and further organized into structures through acquired expertise [Ericsson and Kintsch, 1995, Gobet, 2000, Delaney and Ericsson, 2016]. In order to create these structures, experts filter salient information during task encoding and develops meaningful associations, along with retrieval cues, that can be regenerated when required, enabling precise and reliable access to the original information [Delaney, 2018].

In the context of reading, information regarding previously read text stored in long-term memory remains available through retrieval cues maintained in short-term working memory [Ericsson and Kintsch, 1995, Cane et al., 2012, Delaney and Ericsson, 2016, Delaney, 2018]. These retrieval cues are updated as the reader processes new text, associating current text to prior representations in long-term memory [Ericsson and Kintsch, 1995, Delaney and Ericsson, 2016]. When an interruption occurs, these retrieval structures enable the reader to re-establish context upon resumption by using the new text as a retrieval cue to access previously read content [Ericsson and Kintsch, 1995]. This mechanism reflects how long-term memory serves as an extension of working memory, allowing information to be stored temporarily but accessed quickly to maintain comprehension [Glanzer et al., 1981, Ericsson and Kintsch, 1995].

Compared to the Memory for Goals, the LT-WM theory leads to a fundamentally different key assumption: interruptions do not cause any text comprehension impairment or information loss, as they can only impact short-term memory [Ericsson and Kintsch, 1995, Cane et al., 2012, Chevet et al., 2022a]. Figure 2.2 shows a visual representation of the primary assumptions of the LT-WM theory. Upon resumption, the process of retrieving information about previously read text from long-term memory is reflected in the
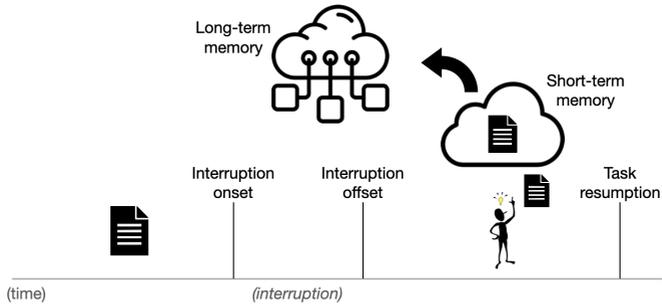
Figure 2.2.: Visualization of the main constructs and assumptions of the LT-WM theory in a reading context [Ericsson and Kintsch, 1995]. While reading, retrieval cues stored in short-term memory keep previously read information accessible in long-term memory, thereby functioning as an extended working memory. After an interruption, individuals can then easily retrieve this information. New text acts as a retrieval cue for the previously read material, enabling the reader to resume.

observed higher reading times [Glanzer et al., 1981, Ericsson and Kintsch, 1995, Cane et al., 2012, Cauchard et al., 2012, Chevet et al., 2022a]. Different studies have shown that, unless reading times are radically disrupted, interruptions during reading do not impair comprehension since information is not permanently lost [Glanzer et al., 1981, 1984, Fischer and Glanzer, 1986, Lorch, 1993, Oulasvirta and Saariluoma, 2006, Fox et al., 2008, Bowman et al., 2010, Cane et al., 2012, Pashler et al., 2013, Tran et al., 2013, Cho et al., 2015, Chevet et al., 2022a]. However, other studies show negative [Foroughi et al., 2015, 2016a] and even positive effects [Tran et al., 2013] of interruptions and multitasking on text recall and comprehension.

2.4. Attention-Aware Learning Technologies and Eye Tracking

In response to the challenges of maintaining attention on the primary task, AALT have been developed. Such technologies are innovative tools designed to monitor, respond to, and model different attentional states [D'Mello, 2019]. They have been applied across various areas, including driver monitoring systems [Liang et al., 2007] and information visualization systems [Steichen et al., 2014]. In educational settings, AALT are often designed to enhance the learning experience by adapting to real-time attentional fluctuations, ensuring that students optimally and dynamically assign their limited attentional resources to meet the demands of the material [D'Mello, 2019]. AALT and their system have been applied to different tools and technologies, such as animated conversational agents [D'Mello et al., 2012, Hutt et al., 2017] and educational computer games [Muir and Conati, 2012]. One of the primary goals of AALT is therefore to provide timely and adaptive interventions—such as prompts, hints, or adjustments to the content delivery—that help learners regain focus and improve their comprehension and retention of the material [Muir and Conati, 2012, Hutt et al., 2021].

Eye tracking technology has become a prominent tool in the development of different learning technologies and, particularly, in real-time modeling of students' attention through AALT, due to its ability to precisely track the location and infer the behavior of learners' attention from their eye movements [Gluck et al., 2000, D'Mello et al., 2012, Conati et al., 2013, Hutt et al., 2016]. Besides, eye tracking has been instrumental for capturing attention allocation during reading. Research has consistently demonstrated a strong connection between eye movements and the cognitive processes underlying reading [Rayner and McConkie, 1976, Just and Carpenter, 1980, Rayner et al., 1989, Rayner, 1998, Rayner et al., 2006]. These findings have led to the development of several practical applications aimed at optimizing the reading experience. For example, systems have been created to automatically detect document types based on reading patterns from eye movements [Kunze et al., 2013] and reduce the disruptive impact of interruptions in

mobile reading [Srivastava et al., 2021]. Additionally, research has explored optimal methods for displaying text on smart glasses in different contexts, such as while walking and sitting [Rzayev et al., 2018]. Gaze-based technologies have also been employed to address issues of inattention during learning [D'Mello et al., 2012], detecting and intervening in episodes of mind wandering during computerized reading tasks in both lab settings [D'Mello et al., 2016, 2017, Mills et al., 2021] and, more recently, in real-world environments [Hutt et al., 2017, 2021, 2024]. Furthermore, eye tracking has been applied to develop assistive tools that, for instance, support readers in regaining focus after interruptions [Jo et al., 2015, Mariakakis et al., 2015] or in recognizing and pronouncing difficult words [Sibert et al., 2000]. In addition to their application in AALT, gaze data have been also implemented within the field of Natural Language Processing (NLP) tasks, by capturing the reader's cognitive response to text, contributing with information that text features alone cannot fully provide [Barrett and Hollenstein, 2020, Mathias et al., 2020]. Among others, gaze data has improved reading comprehension prediction tasks [Malmaud et al., 2020, Mézière et al., 2023, Southwell et al., 2023], computational language models interpretation [Sood et al., 2020a,b, Eberle et al., 2022] and text understanding [Klerke et al., 2016, Barrett and Hollenstein, 2020].

The incorporation of gaze-based technologies into educational environments represents a significant advancement in developing adaptive, student-focused learning systems that can dynamically respond to attention, ultimately improving the efficacy of reading in the digital age. However, despite their potential, these systems still face significant challenges, particularly in their ability to account for individual differences among learners. Current designs often adopt a one-size-fits-all approach, limiting their effectiveness in adapting to diverse learning needs and attentional capacities, thereby personalizing their interventions [Mariakakis et al., 2015, Hutt et al., 2021]. Expanding these technologies to adjust dynamically to each learner's unique cognitive profile as well as to the internal characteristics of each learning material remains an open area of research.

3. Related Work

3.1. Resumption in Interrupted Reading

Reading is a complex cognitive process that demands a significant amount of information to be stored in our memory resources, especially when faced with interruptions and distractions [Cauchard et al., 2012]. Prior research has extensively investigated interrupted reading driven by the fundamental questions of how interruptions impact reading and how readers can effectively recover and integrate new information after an interruption [Cauchard et al., 2012]. Numerous studies have investigated different aspects of this process, such as the cognitive mechanisms involved in resumption, the interruption effect as well as the type and timing of interruptions.

The foundational work of Glanzer et al. [1981, 1984], Fischer and Glanzer [1986], Glanzer and Nolan [1986] provided significant insights into how interruptions affect reading. Their studies involved introducing interruptions during reading tasks through arithmetic problems or other independent reading activities, while also having a non-interrupted control condition [Glanzer et al., 1981, Cane et al., 2012, Cauchard et al., 2012]. These interruptions consistently resulted in a noticeable reading time increase when participants resumed reading the first post-interruption sentence, compared to the non-interrupted condition [Glanzer et al., 1981, Cane et al., 2012, Cauchard et al., 2012]. In particular, brief 10-s interruptions inflated reading time by about 350 ms, whereas more challenging interrupting tasks, like 30-s digit recall, led to an increase in reading time of approximately 1.5 s [Glanzer et al., 1981, Fischer and Glanzer, 1986]. This delay suggests that readers need time to retrieve the context or situation model of the text in order to continue pro-

cessing new information, as predicted by the LT-WM theory [Ericsson and Kintsch, 1995].

However, this foundational work [Glanzer et al., 1981, 1984, Fischer and Glanzer, 1986, Glanzer and Nolan, 1986] also identified conditions that could mitigate the negative interruption effect. For example, when participants were allowed to re-read the sentence preceding the interrupted one before reading the post-interruption text, the delay in resumption was significantly reduced [Glanzer et al., 1984]. Additionally, reminders of the high-level thematic content of the text through word cues [Lorch, 1993] or pictures representing the scene described in the text [Schneider and Dixon, 2009] also facilitated quicker resumption. These findings indicate that providing context or cues can help readers reconstruct the situation model more efficiently, allowing them for accessing information from working memory and integrating the new text [Cane et al., 2012, Cauchard et al., 2012, Chevet et al., 2022a].

Another fundamental aspect that has been largely investigated is whether interruptions during reading impact comprehension. Researchers have employed different methods to assess the construct of reading comprehension, with multiple-choice questions being the most commonly used, either developed by the authors or derived from established inventories, such the reading section of the Suite of Assessments (SAT) [The College Board, 2024] or the Graduate Record Examination (GRE) [Educational Testing Service, 2024] prompts [Fox et al., 2008, Bowman et al., 2010, Cane et al., 2012, Cauchard et al., 2012, Pashler et al., 2013, Tran et al., 2013, Cho et al., 2015, Foroughi et al., 2015, 2016a, Chevet et al., 2022a]. Other methods included tests like sentence paraphrasing [Glanzer et al., 1981, 1984, Schneider and Dixon, 2009] and sentence recognition [Oulasvirta and Saariluoma, 2006], verification tasks [Lorch, 1993], recall tasks [McNamara and Kintsch, 1996], and fill-in-the-blank questions [Fox et al., 2008]. Aligning with the LT-WM theory, a subset of this research showed that interruptions generally do not impair reading comprehension, as long as they do not significantly restrict the available reading time, regardless of the frequency or difficulty of the interrupting task [Oulasvirta and Saariluoma, 2006, Cane et al., 2012, Cauchard

et al., 2012, Chevet et al., 2022a]. This suggests that readers can still access information in LT-WM during interruptions of up to 30 s, thereby preserving a stable understanding of the text [Ericsson and Kintsch, 1995, Chevet et al., 2022a].

However, other studies have more recently challenged the LT-WM theory's assumptions by showing that interruptions can indeed affect comprehension, particularly when the comprehension task requires inference-making [Foroughi et al., 2015, 2016a]. In their study, Foroughi et al. [2015] found that while the capacity to recognize explicit text information remained stable, interruptions impaired the ability to link and integrate information from the text, hence inference making. These two different abilities were assessed through different sets of comprehension and recognition questions [Foroughi et al., 2015]. Their findings suggests that successful text comprehension requires activating information in working memory [Foroughi et al., 2015]. On the contrary, Cho et al. [2015] reported no significant effect of interruptions on comprehension, either for recognition or for information that required a higher degree of inference. This discrepancy may be due to differences in the reading comprehension assessment used in these studies [Chevet et al., 2022a]. Foroughi et al. [2015] employed more complex questions that required integrating multiple textual information, whereas Cho et al. [2015] used simpler inference questions.

The nature of the interruption–whether it involves a task-switch, a digital multimedia distraction, or an unrelated cognitive task–can also influence how fast a reader resumes reading [Chevet et al., 2022a]. For instance, interruptions involving background speech led to a higher increase in reading rate compared to reading with background music free of spoken content or in silence [Cauchard et al., 2012]. Besides, interruptions implemented through task-switching paradigms closely simulate the growing presence of digital distractions in our devices, enhancing ecological validity [Chevet et al., 2022a,b]. Studies have shown that such interruptions tend to increase the total time spent reading [Fox et al., 2008, Bowman et al., 2010, Pashler et al., 2013]. This increase is likely due to the readers' attempt to regain context from previously read text or to encode new information following an interruption

[Chevet et al., 2022a]. These findings support the LT-WM theory's premise that mental representations of text can be reconstructed following an interruption by leveraging contextual retrieval cues stored in short-term working memory, ensuring access to information in LT-WM. [Ericsson and Kintsch, 1995, Chevet et al., 2022a].

The timing of interruptions, whether they occur between paragraphs [Cho et al., 2015, Chevet et al., 2022a], sentences [Glanzer et al., 1981, 1984, Glanzer and Nolan, 1986, Lorch, 1993], or even in the middle of paragraphs [Cauchard et al., 2012, Chevet et al., 2022a] and sentences [McNamara and Kintsch, 1996, Cane et al., 2012, Pashler et al., 2013], did not appear to significantly affect comprehension. Nevertheless, interruptions occurring between two sentences have been shown to cause a smaller increase in reading times than mid-sentence interruptions [McNamara and Kintsch, 1996]. Additionally, the latter constitute more ecologically-valid interruptions as they are more likely to occur in our daily reading tasks [Cane et al., 2012].

Overall, while interruptions generally increase reading time, their impact on comprehension is influenced by factors such as the type of interruption [Cauchard et al., 2012] and the complexity of the comprehension task [Cho et al., 2015, Foroughi et al., 2015]. The mixed results in the literature suggest that more research is needed to fully understand the mechanisms underlying resumption in interrupted reading and how the interruptions experienced in digital educational environments can impact the processing of text.

3.1.1. Reader-Related Influencing Factors

Despite their significant role, individual differences in how individuals recover from interruptions are yet to be thoroughly investigated across different domains, thereby representing a significant research gap [Werner et al., 2011, Meys and Sanderson, 2013, Bai et al., 2014, Cambronero-Delgadillo et al., 2024]. Only a few studies have so far examined the influence of individual differences on resumption performance, mainly within the Memory for Goals framework, across procedural and demanding tasks, such as air traffic control [Seamster et al., 1993, Joslyn and Hunt, 1998], emergency or safety dispatch-

ing simulation [Joslyn and Hunt, 1998, Kanaan and Moacdieh, 2021], video cassette recorder programming [Werner et al., 2011], arithmetic tasks [Meys and Sanderson, 2013], abstract decision making [Bai et al., 2014], spatial navigation [Drews and Musters, 2015], and financial-management [Foroughi et al., 2016b].

These studies have primarily shown an impact of Working Memory Capacity (WMC) [Werner et al., 2011, Meys and Sanderson, 2013, Bai et al., 2014, Drews and Musters, 2015, Foroughi et al., 2016b], workload [Kanaan and Moacdieh, 2021], multitasking [Bai et al., 2014], and spatial abilities skills [Werner et al., 2011] on resumption performance. In the field of interrupted visual search, individuals with higher WMC were found to experience minimal disruption to search accuracy due to interruptions, irrespective of the interruption's duration [Labonté and Vachon, 2021].

In the context of interrupted reading, the number of studies investigating the impact of reader-related differences on resumption is also limited. The majority of studies that examined the impact of interruptions on reading comprehension and its mitigation did not consider the role of individual differences. In their study, however, Foroughi et al. [2016a] pre-screened participants' WMC by asking them to complete the automated operation span task [Unsworth et al., 2005]. Participants then had to read different prompts while either being interrupted or not, and finally had to answer a reading comprehension assessment including both comprehension and recognition questions. Their results indicated that interruptions had a detrimental effect on both comprehension and recognition questions only for subjects with lower WMC, who showed significantly worse related scores [Foroughi et al., 2016a]. Despite being the first to include WMC assessment to test LT-WM theory's assumptions during reading, this study did not provide any further insight on the impact of WMC on resumption performance. Similarly, Chevet et al. [2022a] investigated whether comprehension during interrupted reading is more strongly influenced by reading-specific memory capacities or general memory abilities. Participants completed both a reading span test [Daneman and Carpenter, 1980] and a digit span test from the Wechsler Adult Intelligence Scale [Wechsler, 2008], with the results analyzed in relation to error

rates on the reading comprehension questionnaire. The results revealed that comprehension was associated with general WMC, but not with the scores from the reading span test [Chevet et al., 2022a]. However, this study did not further analyze the relation between general WMC and resumption performance. An exception is represented by Altamura et al. [2022], who conducted a pilot study on the effect of instant messaging during reading and analyzed the impact of reader-related differences on resumption. Their preliminary findings pointed to a potential effect of variations in readers' comprehension level (high vs. low comprehenders), reading habits, and multitasking skills on resumption [Altamura et al., 2022].

Among different reader-related cognitive factors, vsWMC and prior knowledge are particularly relevant to investigate. The significance of spatial representations in task resumption has been well-documented, highlighting the role of perceptual processes that are not fully accounted for by the Memory for Goals and LT-WM theories [Ratwani and Trafton, 2008, Cane et al., 2012]. Perceptual processes have been nonetheless addressed in interrupted visual search research, suggesting that spatial configurations of the search display are maintained even during interruptions and subsequently confirmed or discarded upon resumption [Lleras et al., 2005]. However, in more complex non-reading tasks, the impact of spatial memory on resumption has shown generally positive, yet inconsistent results [Ratwani and Trafton, 2008, Werner et al., 2011, Meys and Sanderson, 2013]. For instance, while higher spatial ability scores were linked to faster resumptions in specific tasks, such as video recorder programming [Werner et al., 2011], no such effect was observed in arithmetic tasks [Meys and Sanderson, 2013]. This variability is likely attributed to the wide range of experimental designs and the reliance on reaction time as the main measure of resumption performance [Meys and Sanderson, 2013].

Building on earlier findings [Glenberg and Kruley, 1992], Schneider and Dixon [2009] explored the role of visuospatial information in reading comprehension, particularly focusing on how individual differences in processing such information affect resumption after interruptions. The authors investigated whether visuospatial cues, such as pictures related to the text, could

reduce the negative effects of interruptions on reading time. The presence of a related picture significantly minimized the interruption effect only for slow readers, suggesting that they took the time to integrate visuospatial cues into their mental models, thereby generating a more resilient and detailed representation in LT-WM [Schneider and Dixon, 2009]. While all readers construct mental models during reading, the extent to which they incorporate visuospatial information into these models seems to vary, impacting how they recover from interruptions [Schneider and Dixon, 2009].

Marking a crucial milestone in interrupted reading research, Cane et al. [2012] demonstrated that visually highlighting the last word read before an interruption significantly reduces resumption time. This finding suggests that resuming reading after an interruption involves not only retrieving information in LT-WM but also actively identifying the specific location in the text where the interruption took place [Cane et al., 2012]. In particular, when the last word was highlighted, readers spent less time re-reading the portion of text preceding the interruption point [Cane et al., 2012]. This result was further developed into a gaze-based digital bookmarking tool, which substantially decrease resumption times by providing a visual cue [Jo et al., 2015]. Further studies revealed that, upon resumption, readers frequently read again parts of text they had already read before the interruption, as observed from their eye movements, underscoring the process of locating the point of interruption in the text, hence the involvement of spatial memory [Cane et al., 2012, Chevet et al., 2022a]. Despite the clear association between visuospatial working memory and resumption in reading, previous studies have not yet relied on psychometric measurements of visuospatial memory, or have only focused on assessments involving mental spatial manipulation and visualization, which may not fully capture the capacity of retaining spatial configurations in working memory [Meys and Sanderson, 2013].

On the other hand, prior knowledge constitutes another relevant cognitive factor, with its critical role in forming meaning from text input [Kintsch and van Dijk, 1978, Van Dijk and Kintsch, 1983], thereby strongly predicting successful reading comprehension [Abdelaal and Sase, 2014]. Possessing a stronger prior knowledge regarding a text leads to an increased capacity in

discarding irrelevant information and, as a result, in processing and understanding the material [McCarthy et al., 2018, Schurer et al., 2020]. Besides, readers with greater prior knowledge exhibit a faster and easier information encoding in long-term memory while reading, hence a stronger resilience to interruptions [Ericsson and Kintsch, 1995]. Nevertheless, the influence of prior knowledge on resumption during reading has not been comprehensively studied. McNamara and Kintsch [1996] found that interruptions during difficult and unfamiliar paragraphs led to higher reading times per sentence, whereas familiar paragraphs were less affected. Moreover, while Chevet et al. [2022c] recently found no interaction between prior knowledge and interrupted reading, readers without prior knowledge about a text generally exhibited higher fixation counts per character. Existing work therefore primarily treated prior knowledge as familiarity with a text's content–often linked to participants' educational background–without yet considering the complex and multidimensional nature of prior knowledge in text comprehension [McCarthy and McNamara, 2021]. The influence of prior knowledge, assessed through a customized questionnaire, on resumption during reading has not yet been explored.

3.1.2. Text-Related Influencing Factors

Although research has consistently shown that the type of text influences reading, the specific impact of text type on interrupted reading has yet to be systematically explored. Previous studies have utilized a range of reading materials, predominantly focusing on short, expository texts. For instance, some work employed short texts such as high-school and college-level paragraphs from diverse sources, including folk tales, novels, and reading comprehension tests [Glanzer et al., 1981, 1984, Fischer and Glanzer, 1986, Glanzer and Nolan, 1986]. Similarly, Lorch [1993] used a 25-sentence practice text on photographic processes alongside experimental passages ranging from 25 to 79 sentences, covering topics like the great apes and comparisons between countries and children games. Other experiments focused on short paragraphs from encyclopedia entries, journal articles, and expository texts

[McNamara and Kintsch, 1996, Oulasvirta and Saariluoma, 2006], with some using texts from the SAT [The College Board, 2024] and/or GRE [Educational Testing Service, 2024] prompts to introduce varying levels of difficulty [Fox et al., 2008, Tran et al., 2013, Foroughi et al., 2015, 2016a]. Additionally, researchers incorporated longer, more detailed texts, such as a 3828-word passage on personality disorders [Bowman et al., 2010], or adapted paragraphs on multiple topics from textbooks, biographical texts, and fiction [Cane et al., 2012, Cauchard et al., 2012, Jo et al., 2015]. In more recent work, 1000-word procedural texts from encyclopedias [Chevet et al., 2022a], 700-word expository texts [Chevet et al., 2022c], hyperlinked online texts mimicking website navigation [Altamura et al., 2022] and passages from standard reading literature databases [Srivastava et al., 2021] were used to investigate the impact of interruptions and multitasking in digital reading. Notably, Chevet et al. [2022b] investigated attentional disruption and digital reading flow in an ecological context with extended narrative texts. They used two Sherlock Holmes stories to observe unconstrained reading, examining the effect of media-related and media-unrelated interruptions on reading flow and comprehension [Chevet et al., 2022b]. Despite this diverse range of reading materials used in previous studies, the potential influence of text type on how readers recover from interruptions remains unexplored.

Texts, in particular literary stories, can often make us see the narrated facts or events in our imagination, providing a sense of engagement that grips the readers [Sadoski et al., 1990, Kuzmičová, 2012, Kuijpers et al., 2014, Jacobs and Willems, 2018, Mak and Willems, 2019]. The benefit of stories' engagement has further been observed in education, with storytelling being a method to enhance reading, writing talking and drawing skills as well as the process of deriving meaning from our experiences [Collins, 1999]. Mental images during reading were also associated with overall text recall and comprehension, due to a tendency to produce mental images during crucial moments of a story [Sadoski, 1983, Sadoski et al., 1990]. Furthermore, in agreement with the imagination effect in educational settings [Sweller, 2003, Mayer, 2009, Leopold, 2021], when asked to form images related to a scientific text, students showed better performance in a subsequent learning

3. Related Work

test [Leopold and Mayer, 2015].

Prior empirical research have utilized narrative texts to explore the influence of mental simulation and imagery on reading behavior [Mak and Willems, 2019, Magyari et al., 2020]. Particularly, Mak and Willems [2019] investigated how different types of simulations–motor, perceptual, and mentalizing–affect reading behavior. Perceptual and motor simulations involve the mental re-enactment of sensory experiences and physical actions and have been shown to influence language processing [Mak and Willems, 2019]. In contrast, the authors referred to mentalizing as the simulation of others' thoughts, emotions, and intentions, hence attributing mental states to a story's characters [Mak and Willems, 2019]. Participants were asked to read three Dutch short stories, averaging 2600 words each, carefully selected for their ability to elicit perceptual and motor simulations as well as mentalizing, while their eye movements were tracked. Their findings revealed distinct relationships between the types of simulation and gaze duration. Specifically, motor simulation was linked to shorter gaze duration, indicating faster reading, while perceptual simulation and mentalizing were associated with longer gaze duration, suggesting a more demanding cognitive process [Mak and Willems, 2019].

In another study on eye movements and mental imagery during narratives, Magyari et al. [2020] explored how enactment- and description-imagery types [Kuzmičová, 2014] prompted by related text styles influence reading behavior. The authors hypothesized that these differences in narrative style would be reflected in eye movement patterns during reading, with enactment-imagery text features prompting shorter fixations, hence faster reading time, due to the more immersive experience they create [Magyari et al., 2020].

To test this hypothesis, Magyari et al. [2020] conducted two studies. In the first one, they analyzed word categories in four German-translated and adapted texts–two enactive and two descriptive–to determine if these texts contained different linguistic cues that might elicit distinct types of imagery [Magyari et al., 2020]. One text pair included two excerpts from the beginning of two stories from two different French authors. The second text pair instead included two excerpts from an American author. These texts

were on average 1260-word long. Their analysis revealed that enactive texts contained significantly more verbs, which are often associated with action and movement, while descriptive texts had a higher frequency of nouns and adjectives–albeit with inconsistencies across the two text pairs–correlating with the static nature of description. They therefore concluded that verbs might be a stronger textual predictor for enactment-imagery [Magyari et al., 2020].

In the second study, participants read these texts while their eye movements were recorded. Their comprehension, perceived mental imagery, reading experience and personality were also assessed. Participants reported the descriptive texts to be more difficult to visualize, likely due to the increased effort needed to imagine static scenes described from an external perspective, compared to the more intuitive adoption of characters' inner perspective facilitated by enactive texts, involving sensorimotor simulation [Magyari et al., 2020]. Interestingly, eye movement data revealed that, within the text pairs, only one of the enactive texts led to longer fixation durations on words, slowing down the reading time [Magyari et al., 2020]. The authors interpreted this finding as evidence that enactment-imagery may produce more intensive imagery generation, hence higher cognitive load during reading, as reflected in the increased reading time. On the other hand, descriptive texts, which posed greater challenges in visualizing the story, resulted in less mental imagery generation, hence reduced reader engagement [Magyari et al., 2020].

These findings are in contrast with Mak and Willems [2019]. While Mak and Willems [2019] found that motor simulations could facilitate faster reading by engaging readers more deeply in the narrative, Magyari et al. [2020] demonstrated that enactment-imagery might actually increase cognitive load, leading to slower reading time. These studies suggest that conjuring mental images while reading can vary significantly, depending on the type of imagery elicited by the text. This discrepancy highlights the complexity of mental imagery processes in reading, hence the need for further research to better understand the effects of different types of mental imagery. Such insights could shed light on how different text styles elicit distinct mental imagery, potentially influencing readers' engagement with the text. Further-

more, whether mental imagery could facilitate resumption during interrupted reading remains an open question.

3.2. Eye Tracking Data for Interrupted Reading

Eye tracking data has provided invaluable insights into the complex cognitive dynamics of reading [Just and Carpenter, 1980, Rayner and McConkie, 1976, Rayner, 1998]. Fostering related scientific research and technological advancements [D'Mello et al., 2012, D'Mello, 2019, Mathias et al., 2020], several publicly available eye tracking reading corpora offer data for different languages and materials. Different English-language datasets, for instance, consist of gaze data from adult participants, obtained during self-paced reading tasks on various short texts or sentences, among others: Frank et al. [2013], Mishra et al. [2016], Luke and Christianson [2018], Hollenstein et al. [2020], Sood et al. [2021]. Such eye movement datasets for reading exist in multiple languages, among others: Chinese [Pan et al., 2022], Danish [Hollenstein et al., 2022], German [Kliegl et al., 2004, Jäger et al., 2021], Hindi [Husain et al., 2015], Persian [Safavi et al., 2016], Russian [Sekerina et al., 2019b], English reading as second language [Berzak et al., 2022, Kuperman et al., 2023], as well as multilingual settings [Kennedy et al., 2003, Siegelman et al., 2022]. Notably, Cop et al. [2017] collected eye movement data from 33 monolingual and bilingual participants as they read an entire crime-fiction novel, revealing more intricate and comprehensive reading processes that are often missed in studies focused on sentence or short text reading. This resulted in the first bilingual eye tracking reading corpus [Cop et al., 2017].

However, the existing datasets often lack features representative of everyday learning environments, such as interruptions. Although prior work investigated eye movement and resumption from interruptions during the reading of naturalistic texts [Cane et al., 2012, Cauchard et al., 2012, Jo et al., 2015, Chevet et al., 2022a], the related collected data are not publicly

available, which creates a gap in the availability of eye tracking resources for educational research. This gap places a limit to the further training and development of gaze-based AALT and NLP-based educational technologies, especially those that aim to address adaptive individual learner support. Although these technologies can successfully monitor learners' performance and level of attention, they currently lack a comprehensive personalized support for task resumption [Hutt et al., 2021, Srivastava et al., 2021]. Leveraging eye tracking datasets specifically designed to study the effect of interruptions on reading could enhance such technologies with insights into individual responses to interruptions and text-specific interactions.

4. Research Gaps and Thesis Contributions

This thesis addresses critical research gaps in the field of interrupted reading, reporting findings on the resumption process in narrative text, with a focus on reader's individual cognitive factors, text-related influences, and the development of a publicly available eye tracking resource for further research. Despite the growing body of literature on resumption in reading, several areas remain insufficiently explored. These gaps include the impact of vsWMC and prior knowledge on resumption from interruptions during reading; the influence of different types of mental imagery prompted by text characteristics on resumption; ultimately, the lack of publicly available eye tracking reading datasets to foster cross-disciplinary research on interrupted reading.

4.1. Contribution 1: Reader-Related Cognitive Factors

One of the significant research gaps lies in understanding how individual differences in reader's vsWMC and prior knowledge affect resumption time during interrupted reading. While previous studies highlighted the importance of spatial memory processes for text comprehension and task resumption, they did not assess participants' vsWMC with psychometric tests [Meys and Sanderson, 2013]. Additionally, the role of prior knowledge in interrupted reading has been so far only addressed as familiarity with the text rather than multidimensional construct [McCarthy and McNamara, 2021]. This contribution explores whether individual differences in readers' vsWMC and

prior knowledge can predict variations in resumption time following interruptions. Before reading a crime fiction extract containing interruptions in the form of opinion questions, we assessed participants' vsWMC through a Symmetry Span (SSPAN) task [Shah and Miyake, 1996, Unsworth et al., 2009], and their prior knowledge about the text genre through a customized questionnaire, targeting different knowledge dimensions. By further annotating resumption lags from gaze data, Chapter 5 provides insights into how vsWMC and prior knowledge influence resumption performance.

The first contribution therefore addresses the following research question and related hypotheses:

RQ1 *Do individual variations in vsWMC and prior knowledge influence resumption lags following interruptions during reading?*

H1 Individuals with higher vsWMC scores exhibit lower resumption lags following an interruption during reading.

H2 Individuals with higher scores in prior knowledge on the content and genre of the text exhibit lower resumption lags following an interruption during reading.

4.2. Contribution 2: Text-Related Factors

Another research gap relates to the limited understanding of whether text features play a role in the resumption process. Particularly, while mental imagery constitutes a crucial aspect in narrative reading [Sadoski, 1983, Sadoski et al., 1990, Kuzmičová, 2014, Jacobs and Willems, 2018], prior research indicated that this process can vary significantly depending on the type of elicited imagery [Mak and Willems, 2019, Magyari et al., 2020, Mak et al., 2023]. However, it remains unclear whether different types of imagery triggered by textual cues also influence resumption lag, in the presence of interruptions during reading. This contribution thus tackles the influence of text-related factors by using two fictional texts designed to evoke either enactment- or descriptive-imagery [Magyari et al., 2020], based on Kuzmičová [2014]'s phenomenological typology theory. Participants read these texts while experi-

encing interruptions, in the form of opinion questions, strategically placed during passages expected to elicit mental imagery. Participants' eye movements were recorded during reading. Besides, we measured participants' reading comprehension and reading experience related to the texts through questionnaires previously implemented by Magyari et al. [2020]. By combining eye tracking data and reading-related variables, Chapter 6 thus presents findings on the effect of enactment- and description-imagery on resumption lag.

The second contribution thus addresses the following research question and emerging hypotheses:

RQ2 *Do narrative texts prompting enactment- vs. description-imagery have a different influence on resumption following interruptions while reading?*

H1 Compared to the descriptive style text, the enactive style text produces longer resumption lags following an interruption while reading.

H2 Compared to the descriptive style text, the enactive style text produces longer fixation durations during resumption following an interruption while reading.

4.3. Contribution 3: InteRead Dataset

An important challenge in the field of interrupted reading research is the scarcity of publicly available eye tracking reading datasets, specifically designed to simulate interruption-prone environments and to further examine the role of individual differences in resumption processes during reading. While previous studies have provided valuable insights into gaze behavior and resumption from interruptions, the related data are not publicly accessible to the broader research community. This gap hinders the development of adaptive, gaze-based technologies for education, particularly those aimed at addressing individual differences in reading to offer personalized support [Mariakakis et al., 2015, Hutt et al., 2021, Ganesh et al., 2023]. To address this gap, Chapter 7 introduces InteRead, the first publicly available eye track-

ing dataset focused on interrupted reading. The dataset includes gaze data from an interrupted narrative reading task, along with extracted linguistic features, demographic data, and annotations of resumption lags.

The chapter further presents three sets of validation analyses focusing on the effects of interruptions and linguistic features on gaze behavior, as well as individual variability in reading and resumption time within the dataset. These analyses tackle three research questions, laying the groundwork for future studies that can leverage the capabilities of InteRead: (1) How can we leverage InteRead to investigate the impact of interruptions on gaze behavior during reading? (2) How can we leverage InteRead to investigate the impact of linguistic features on gaze behavior during interrupted reading? (3) How can we leverage InteRead to investigate individual variability in interrupted reading? InteRead serves as a novel contribution and resource for advancing research on task resumption during reading and the cognitive processes involved in narrative reading. The availability of this dataset opens new avenues for cross-disciplinary collaborations, facilitating the development of more personalized and effective educational tools.

The third contribution thus tackles the overarching following research question:

RQ3 *How can we address data scarcity on individual differences in resumption during reading to advance the development of adaptive AALT?*

5. Reader-Related Factors in Resumption during Reading

This chapter[1] presents the first contribution of this thesis: a study investigating how individual differences in readers' vsWMC and prior knowledge influence resumption lag following interruptions during narrative reading. Previous studies have not thoroughly explored the relationship between vsWMC and resumption lag using psychometric tests, nor have they sufficiently examined the multiple dimensions of prior knowledge in this context.

The presented work builds upon these gaps and the importance of these cognitive factors in learning and text comprehension. The reported findings offer new insights into the reader-related factors at play in interrupted reading. In particular, they underscore how the interaction between vsWMC and prior knowledge more effectively predicts resumption lag than either factor alone.

The chapter addresses the following research question and hypotheses:

RQ1 *Do individual variations in vsWMC and prior knowledge influence resumption lags following interruptions during reading?*

H1 Individuals with higher vsWMC scores exhibit lower resumption lags following an interruption during reading.

H2 Individuals with higher scores in prior knowledge on the content and genre of the text exhibit lower resumption lags following an interruption during reading.

---

[1]The content of this chapter has already been published in Zermiani et al. [2024b] as well as partly in Zermiani et al. [2024a].

5.1. Methods

The entire study was approved by the Commission for Responsibility in Research of the University of Stuttgart (approval number Az. 22-018). All subjects provided informed consent and the conducted research followed the regulations outlined in Standard 8 of the Ethical Principles and Code of Conduct for Psychologists [American Psychological Association, 2020].

5.1.1. Participants

57 adult subjects (36 female, M age $= 27.51$ years, $SD = 5.55$ years and range $20 - 47$ years) were recruited through internal student mailing lists at the University of Stuttgart and social media. They were compensated by either a 10€ payment or a certificate of participation to obtain study credits for the "Medieninformatik" [Media Computer Science] study program of the University of Stuttgart. In order to participate, all subjects had to meet the following requirements: normal or corrected-to-normal eyesight; English proficiency–native speaker, C1, IELTS 6.5+, Toefl 95+, or other equivalent certificates; no diagnosed attention disorders; no diagnosed reading disorders.

5.1.2. Design

The study employed a within-subject design to investigate how individual differences in vsWMC and prior knowledge influence resumption lag during interrupted reading. vsWMC and prior knowledge served as measured independent variables, while resumption lag as the dependent variable. This design enabled us to test the hypothesized relationships between these cognitive factors and participants' speed in resuming reading after interruptions.

We operationalized vsWMC with the SSPAN task [Shah and Miyake, 1996, Kane et al., 2004], which has already been used in the context of reading [Olkoniemi et al., 2018]. To operationalize prior knowledge, we designed 12 questions as common practice in the research field, due to the limited time span of the studies conducted [Smith et al., 2021].

We measured resumption lag as the time span from the end of the interruption until the first stable reading pattern in the pre-interruption text, building on the definition by Jo et al. [2015]. To extract individual resumption lags from the pages containing an interruption, we manually annotated the gaze data. The aim was to identify the point in time when each participant resumed reading after an interruption, resulting in approximately six resumption points per subject. Cane et al. [2012] proposed a different definition of resumption lag, focused on re-reading times, while Jo et al. [2015] used a reading detection algorithm designed for auto-scrolling [Sharmin et al., 2013]. We therefore adopted a novel approach aimed at specifically capturing the initial stable reading pattern as the first relevant action for re-focusing, in line with Altmann and Trafton [2002], while excluding instances of extended re-reading, as these do not fit our definition of resumption lag.

5.1.3. Apparatus

The study took place in a laboratory room under constant lighting conditions. Participants seated at a desk in front of a screen, with mouse and keyboard. A room divider separated the participants' desk from the principal investigator's desk. We stabilized the participants' head movements with the use of an adjustable chin and headrest. We utilized these tools only during eye tracking phase.

Eye movements were recorded using a Tobii Pro Spectrum[2] screen-based eye tracker, with a sampling frequency rate of 1200 Hz. The whole experiment was displayed on the native Tobii Pro Spectrum screen (EIZO FlexScan EV2451) with a size of 52.8×29.7 cm, at a resolution of 1920×1080 px. Participants had an approximate viewing distance of 57 cm from the screen so that the position of their eyes was located roughly in the center of the Tobii Pro Spectrum headbox.

The eye tracker and its display were connected to a computer on the investigator's desk, which managed the experimental pipeline. We implemented

---

[2]Firmware version 2-6-1.

the experiment interface with $PsychoPy^3$ [Peirce et al., 2019], while $PyGaze^4$ [Dalmaijer et al., 2014] was used for managing the eye tracker. To calibrate the participants' eye movements, we used the default *PyGaze* five-point calibration procedure [Dalmaijer et al., 2014].

5.1.4. Materials

Reading Task

Participants were asked to read a selected excerpt from "The Adventure of the Speckled Band" [Doyle, 1892], a Sherlock Holmes story written in Victorian British English. This excerpt spanned 28 pages, averaging 154 words per page ($SD = 22.3$, range: 94–190 words per page), and maintained the original narrative structure, including direct dialogues within quotation marks. The text covered approximately half of the story and concluded just before Holmes's investigation of the crime scene, providing context on the primary characters, the central mystery, and the typical operating methods of the detective figure.

Each page was designed in a fixed format of 12 lines, except the final page, which contained seven lines. To ensure consistent visual presentation, the text was displayed in 20-point black Courier, a mono-spaced font, with a $2.5\times$ line spacing on a white background. The text layout was left aligned at 500 px and vertically centered. Participants pressed the space bar to advance to the next page; backward navigation was not permitted (see Figure 5.1).

Interrupting Task

Interruptions were presented as gray dialog boxes centered on the screen, with the background remaining white and the reading text temporarily removed. Participants were prompted to type a response to one of six unique opinion questions, with each question requiring completion within a 60-s time frame. These questions were selected from established items used in prior research [Pashler et al., 2013] and related to general topics from daily life.

---

[3]Package version 2022.2.3.
[4]Package version 0.7.4.
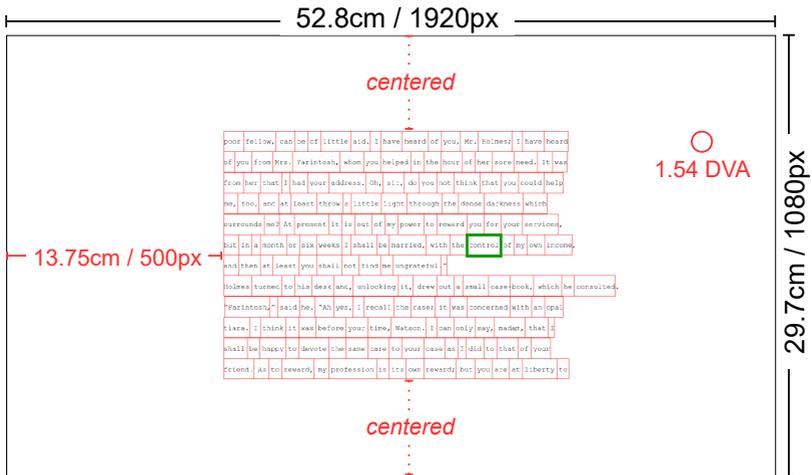
Figure 5.1.: Page setup, as depicted in Zermiani et al. [2024a], including detailed annotations and dimensions. Lines spanned 1.54 DVA. Red rectangular areas surrounding each word represent predefined regions of interest, commonly known as bounding boxes, defined by pixel coordinates. The bounding box for the target word is shown in green. These were not visible to participants.

5. Reader-Related Factors in Resumption during Reading

To ensure a robust level of interruption, we adopted opinion questions. This choice was informed by the seductive detail effect, which suggests that interesting but irrelevant information can divert attention and hinder learning [Harp and Mayer, 1998]. By presenting these questions as text items, we sought to capture participants' attention effectively, as seductive texts have been shown to be particularly engaging [Rey, 2012]. Additionally, requiring participants to generate written answers was designed to reduce the likelihood of subvocal rehearsal of the page content, thereby limiting engagement of the phonological loop [Baddeley, 1992]. This interactive interruption task, which required both thoughtful responses and written input, aimed to significantly occupy cognitive resources and disrupt verbal and visual processing, aligning with findings by Oulasvirta and Saariluoma [2006].

Following prior methodology [Cane et al., 2012, Jo et al., 2015], interruptions were triggered by participants' fixations on target words strategically placed across six predetermined pages (pages 3, 7, 11, 15, 19, and 24). To ensure interruptions occurred intermittently, these pages were spaced every three to four pages. Target word selection adhered to several criteria [Cane et al., 2012]: (i) target words were consistently chosen from the middle line (sixth line) on each page, (ii) first and last words of the middle line were excluded, (iii) words positioned at the beginning or end of sentences were avoided, ensuring that interruptions occurred in the middle of a sentence, and (iv) the specific target words and pages were fixed across all participants. Furthermore, function words, which are less frequently fixated during reading [Rayner and McConkie, 1976, Krejtz et al., 2016], were excluded from the target pool. An example of a target word position can be seen in Figure 5.1.

If a participant's eye movement did not adequately trigger the target word, the space-bar button–normally locked on pages with interruptions–was automatically unlocked. This precaution ensured participants could continue reading seamlessly without perceiving any inconsistency.

Prior Knowledge Questionnaire

We developed a 12-item questionnaire to evaluate participants' prior knowledge of Sherlock Holmes stories, as well as their general familiarity with crime fiction. To avoid neutral responses throughout the experiment, we used a 6-point Likert scale for all Likert-type questions. The questionnaire was structured to capture different dimensions of prior knowledge, divided into three main subsets:

1. Detailed Knowledge: Four multiple-choice questions focused on specific details from Sherlock Holmes stories. Each question included the option "I do not know" to discourage guessing.

2. Target-Domain Knowledge: Four yes/no statements assessed participants' prior exposure to crime fiction, including TV adaptations of Sherlock Holmes, the specific story used in the study, or other Sherlock Holmes stories.

3. General Knowledge: Four agreement statements measured on a 6-point Likert scale rated participants' general interest in the detective or crime genre.

Symmetry Span Task

The SSPAN task [Shah and Miyake, 1996], originally developed to assess vsWMC and spatial visualization skills, was implemented here following the version by Unsworth et al. [2009]. This approach ensured a thorough evaluation of vsWMC. In the primary storage task, participants were required to recall the sequence of colored squares presented in a 4×4 grid, with each square displayed for 650 ms.

In parallel, a distractor or processing task involved symmetry judgments, where participants evaluated the vertical symmetry of grids containing black squares in various configurations. The number of squares to recall and symmetry judgments to perform between two and five per trial [Unsworth et al., 2009]. Across three blocks of trials, we randomized the sequence of total squares to recall to prevent participants from anticipating this number.

To familiarize participants with the dual-task demands, they also completed three practice rounds prior to the main trials [Redick et al., 2012]: one for the primary storage task, one for the distractor task, and one for the combined task. This practice sequence ensured participants were adequately prepared to manage both tasks concurrently [Redick et al., 2012].

Reading Comprehension and Experience

The post-test questionnaire served as both a manipulation check and a retrospective self-assessment to evaluate participants' engagement with the task and their perception of the reading as well as experimental experience. It included four comprehension questions, designed by the authors, to verify that participants were actively engaged with the text throughout the reading task.

Moreover, six items on a 6-point Likert scale were used to measure participants' interest in the story and their irritation with the interruptions. Of these, four questions were adapted from the Story World Absorption Scale (SWAS) [Kuijpers et al., 2014], as validated tool for assessing reader enjoyment and immersion in a narrative. Besides, drawing on Wirzberger and Russwinkel [2015], the questionnaire included two questions about the strategies participants used to resume reading following each interruption: whether they searched for the exact interruption point in the text, tried to remember the content, or employed both strategies.

5.1.5. Procedure

After receiving all the information about the study and signing the informed consent, participants were asked to sit at the assigned desk in the laboratory room. They were then asked to start the experiment and answer standard demographic and general reading habits questions. They then completed the prior knowledge questionnaire. Subsequently, they followed the instructions for completing the SSPAN task, after three practice trials.

Before starting the interrupted reading task, participants were made aware that at some point during the reading process the text might disappear, and a window asking them to type their response to a question would appear for

60 s, after which they were to continue reading. They were informed that such questions pertained to everyday events or current topics. Furthermore, they were also advised to read the story carefully as they would have to complete a short comprehension test at the end. They then started reading the story while their eye movements were being recorded. At the end of the reading task, they completed the reading comprehension check and the reading experience questionnaire.

5.1.6. Scoring

Prior Knowledge

To extract the scores from the prior knowledge questionnaire, we assigned points according to the type of question: one point was assigned for each correct answer in the four multiple-choice questions pertaining to detailed knowledge. For the other two groups of questions, scores followed an ordinal assignment from 0 to 1 for the yes/no statements (target-domain knowledge), and from 0 to 5 for the Likert scale statements (general knowledge). The sum of these scores represents the prior knowledge measurement, resulting in a possible maximum score of 28 points per participant.

We assessed the internal consistency of our prior knowledge questionnaire by calculating the Cronbach's Alpha (α) score. We obtained an overall α of .83 for the entire prior knowledge questionnaire, including detailed knowledge ($\alpha = .68$), target-domain knowledge ($\alpha = .45$), and general knowledge ($\alpha = .90$).

vsWMC

Following established scoring procedures for cognitive span tasks [Redick et al., 2012, Foster et al., 2015], we extracted two main scores from the SSPAN task: the partial recall score and symmetry accuracy. The partial recall score, ranging from 0 to 42, relates to the main storage task and represents the sum of colored squares recalled in the correct position, regardless of whether the entire sequence of squares was accurately recalled in the correct order. This
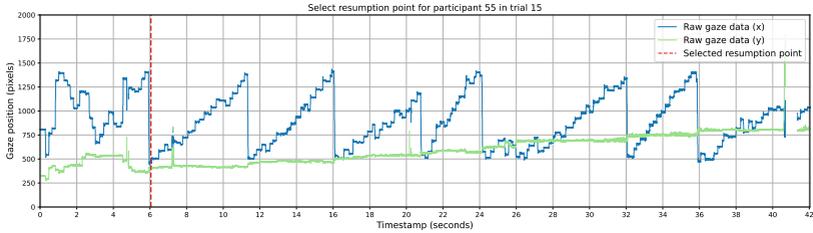
Figure 5.2.: Example of the annotation tool interface. The tool visualizes the x (blue) and y (green) gaze coordinates for each participant starting from the end of an interruption until the end of the page. The red line here highlights the resumption point selected by the rater, indicating the starting point of the first stable reading gaze pattern.

method provides a more robust and internally consistent measurement of vsWMC [Unsworth et al., 2009, Redick et al., 2012]. In addition, symmetry accuracy, measured as the proportion of correct symmetry judgments, reflects participants' proficiency in the processing task and has also been shown to correlate with performance on the storage task [Unsworth et al., 2005, 2009, Richmond et al., 2022].

Resumption Lag

To obtain resumption lags, we manually annotated the gaze data using a custom-built annotation tool (Figure 5.2) that visualized x and y gaze coordinates from the moment the interruption ended until the page conclusion. For each annotation, the tool allowed the annotators to identify the first stable reading pattern within the pre-interruption text (lines 1 to 6), ensuring it met our definition of resumption lag as outlined by Altmann and Trafton [2002]. Annotators selected the starting point of this reading pattern by clicking on it, and the tool automatically converted the x and y coordinates of this point into a timestamp. The resumption lag was then calculated by subtracting the timestamp of the interruption offset from the annotated timestamp and converting the obtained timestamp into seconds.

Table 5.1.: Estimates of ICC for manual annotations of resumption lags. Average-rating, absolute-agreement and two-way random-effects model was used for the calculation.

	Interclass Correlation	95% Confidence Interval		F Test With True Value 0			
		Lower Bound	Upper Bound	Value	*df*1	*df*2	Sig
Average measures	.916	.894	.933	12.200	324	268	.000

Two independent raters conducted the process without participating in data collection, ensuring unbiased assessment. The annotation process yielded on average six resumption lags per participant, one for each page featuring an interruption. The final resumption lag for each participant was derived as the average of the annotations provided by the two raters. An additional iteration verified that all selected starting points aligned with the pre-interruption text. We assessed the reliability of our annotations by calculating the ICC. Following the reporting guidelines by Koo and Li [2016], the ICC was computed in R [R Core Team, 2021], using the *irr* package[5] [Gamer et al., 2019], average-rating (k = 2), absolute-agreement and two-way mixed-effects model (see Table 5.1). The estimated ICC value of .916 is slightly above the 0.90 threshold, which generally indicates excellent reliability [Portney and Watkins, 2009, Koo and Li, 2016].

5.2. Results

A total of 10 participants were removed from the sample. Seven participants were excluded for the following reasons: (1) missing over 50% of gaze data points on pages containing interruptions, (2) failing to trigger more than two out of six planned interruptions, and (3) having excessively noisy gaze data on the interruption pages, which made it difficult to locate a resumption point in the annotation phase. Additionally, to ensure robust inclusion of the processing component in the SSPAN task, we applied a data-driven criterion, excluding participants who scored more than two standard deviations below the mean processing performance score [Richmond et al., 2022]. This

---

[5]Package version 0.84.1.

Table 5.2.: Descriptive statistics for main variables as reported in Zermiani et al. [2024b].

	M	*SD*	*Min*	*Max*
vsWMC	27.57	7.60	5	41
Prior knowledge	11.21	6.65	0	24
Resumption lag (s)	2.81	1.54	0.95	8.29

Note. Total sample size $n = 47$; M = mean, SD = standard deviation, Min = minimum, Max = maximum, s = seconds; the vsWMC score corresponds to the *partial recall score* from the SSPAN task, following Redick et al. [2012].

criterion led to the exclusion of three additional participants, resulting in a final sample size of $n = 47$ (see Table 5.2 for detailed descriptive statistics).

To statistically inspect the effects of vsWMC and prior knowledge on resumption lags, we used a linear mixed-effects model, implemented in R [R Core Team, 2021], with the *lme4* package[6] [Bates et al., 2015]. We computed *p*-values for fixed effects with the *lmerTest* package[7] [Kuznetsova et al., 2017]. The model specified resumption lag as the criterion variable, with partial recall scores from the SSPAN task, prior knowledge scores, and their interaction as predictors. Participant ID and page number were set as random effects to account for repeated measures across different pages by the same participants. Each participant contributed, on average, six resumption lags, leading to a total of 272 observations. All predictors were z-standardized.

The model yielded a conditional R^2 of 0.23. Analyzing predictors individually, vsWMC (β = -0.19, CI = -0.58 – 0.19, SE = 0.19, $t(43.71)$ = -0.99, p = .327) and prior knowledge (β = 0.02, CI = -0.34 – 0.38, SE = 0.18, $t(43.99)$ = 0.10, p = .916) did not yield significant contributions on resumption lags. However, the interaction between vsWMC and prior knowledge significantly predicted resumption lags (β = -0.34, CI = -0.65 – -0.02, SE = 0.16, $t(45.53)$ = -2.08, p = .043).

---

[6]Package version 1.1.28.
[7]Package version 3.1.3.

To further interpret the interaction, we conducted a median split on the prior knowledge scores ($Mdn = 10$), categorizing participants as high ($n = 23$, $M = 16.8$, $SD = 4.56$) or average-to-low prior knowledge readers ($n = 24$, $M = 5.88$, $SD = 2.82$), following prior work [Taub and Azevedo, 2019, Altamura et al., 2022, Chevet et al., 2022a]. Figure 5.3 illustrates how the interaction between vsWMC and prior knowledge on resumption time varies according to the reader's prior knowledge category. In separate Pearson correlation analyses, vsWMC and resumption lags showed a non-significant correlation for the average-to-low prior knowledge group ($r(22) = 0.06$, $p = .760$), while high prior knowledge readers demonstrated a non-significant trend towards moderate negative correlation ($r(21) = -0.40$, $p = .053$), suggesting that prior knowledge may strengthen the relationship between vsWMC and efficient resumption.

Following Kumle et al. [2021], we estimated the achieved statistical power for each predictor specified in the linear mixed-effects model, including the random-effects structure. The post-hoc power analysis was conducted in R [R Core Team, 2021] using the *simr* package[8] [Green and MacLeod, 2015]. We estimated power using likelihood ratio tests with 1000 simulations and an alpha level of .05. The analysis yielded an estimated power of 18.00% (CI = 15.67 - 20.52) to detect the main effect of vsWMC on resumption lags, 6.00% (CI = 4.61 - 7.66) for prior knowledge, and 58.00% (CI = 54.87 - 61.08) for the interaction between vsWMC and prior knowledge. These results reflect the achieved power based on the observed effect sizes in the sample [Kumle et al., 2021].

5.3. Discussion

This chapter examined the role of reader-related factors, specifically vsWMC and prior knowledge, in influencing reading resumption speed following inter-ruptions. Addressing current research gaps, the study employed the SSPAN task as a psychometric measure of vsWMC and assessed multi-dimensional

---

[8]Package version 1.0.7.
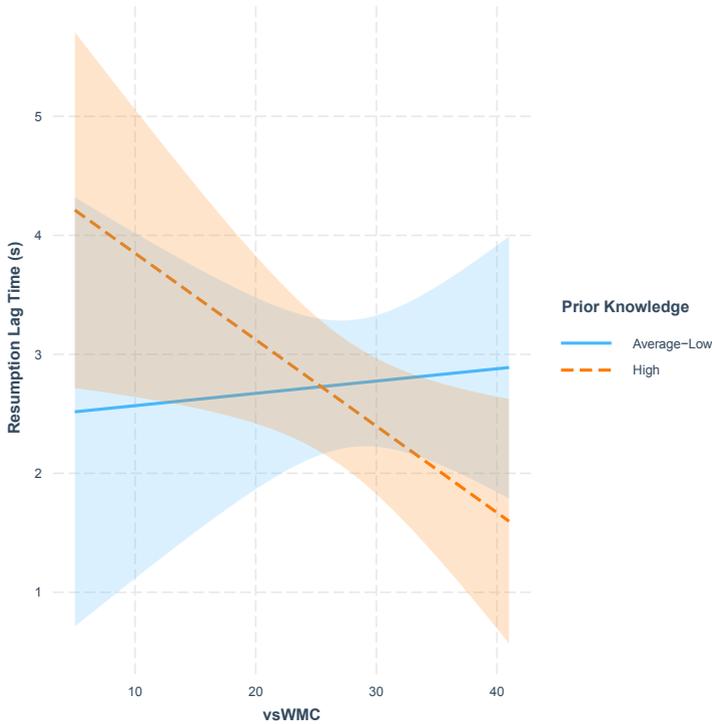
Figure 5.3.: Interaction effect of vsWMC and prior knowledge on resumption lags after performing a median split on the prior knowledge scores. The interaction plot reported in Zermiani et al. [2024b] displays the effect of the vsWMC score (x-axis) and resumption lag in seconds (y-axis), for the average-to-low (blue) and high (orange) prior knowledge groups. The bands correspond to the estimates' 95% confidence intervals.

aspects of prior knowledge, both relatively unexplored factors in interrupted reading research. Additionally, a novel tool for visualizing and annotating reading resumption from gaze data was introduced, providing reliable annotations and precise measures of resumption lags. Despite we could not confirm our hypotheses that higher vsWMC and greater prior knowledge would individually predict shorter resumption times, findings revealed a significant and more complex interaction between these factors.

5.3.1. Hypothesis 1

Initially, we hypothesized that higher vsWMC would result in shorter resumption lags (H1, see Section 5) . Our findings, however, indicated that vsWMC alone did not significantly predict resumption speed. This outcome aligns with prior research by Meys and Sanderson [2013], which found no significant impact of spatial memory on recovery time in an interrupted arithmetic tasks. Similarly, Werner et al. [2011] reported mixed results when assessing spatial ability and resumption performance, depending on the type of assessment method. This lack of significant findings might also partially stem from the genre of the reading material, which involved familiar characters and plot structures from the well-established Sherlock Holmes stories. Particularly, the predictable patterns and linguistic elements of crime fiction, which attract readers' attention [Rzepka, 2005, Kukkonen, 2020, Kukkonen and Baumbach, 2022] may have minimized reliance on vsWMC during resumption.

Although previous studies have highlighted the role of spatial memory in visual search and reading tasks [Lleras et al., 2005, Ratwani and Trafton, 2008, Cane et al., 2012], our findings based on a psychometric test support the limited role of vsWMC as individual predictor of recovery speed in narrative reading contexts. In support of our observed results, studies like Glenberg and Kruley [1992] and Schneider and Dixon [2009] showed that visuospatial information would aid comprehension and model reactivation following interruptions only for slow readers, who may rely on additional cues to integrate and construct detailed mental models. Individual variations in vsWMC

would therefore become meaningful for resumption performance during reading when in combination with other reader-related variables, such as reading speed [Schneider and Dixon, 2009].

5.3.2. Hypothesis 2

Our second hypothesis stated that higher prior knowledge would lead to shorter resumption lags (H2, see Section 5). However, the study revealed that prior knowledge alone did not significantly predict resumption speed. This result is consistent with earlier findings by Chevet et al. [2022c], which showed that although a lack of prior knowledge influenced gaze behavior, it did so independent of interruptions. Schurer et al. [2020] similarly reported that prior knowledge did not affect changes in attentional focus during reading.

These studies, however, mostly involved expository texts or short paragraphs, where readers' familiarity or non-familiarity with the topic covered in the text was used as measure of prior knowledge. Since expository texts typically require a stronger integration of prior knowledge than narratives [Wolfe and Woodwyk, 2010], our findings suggest that prior knowledge may play a less critical role in resuming narrative reading, especially in the presence of interruptions and when assessed with a tailored questionnaire.

5.3.3. Interaction of vsWMC and Prior Knowledge

Despite the lack of substantial main effects, our results showed a significant interaction between vsWMC and prior knowledge on resumption lag, where readers with both high vsWMC and high prior knowledge regarding Sherlock Holmes and detective fiction resumed more quickly. This interaction aligns with findings by Foroughi et al. [2016b], who suggested that readers with lower WMC have more difficulties in filtering relevant information during task resumption, hence reducing their performance. They further concluded that readers likely require substantial expertise regarding a text's specific type and content in order to efficiently recover information from long-term working memory upon resumption [Foroughi et al., 2016b]. Our results extend this insight to suggest that the integration of vsWMC and prior knowledge

may support more efficient discrimination of relevant text, enabling faster reactivation of mental representations, hence shorter resumption lags.

This interaction could therefore indicate that readers with higher vsWMC and more extensive prior knowledge would make use of cognitive strategies that support faster retrieval and integration of task-related information. These results resonate with prior theoretical frameworks on multimedia learning that emphasize the integration of visuospatial and knowledge-based processing for constructing mental models [Schnotz, 2002]. The extent to which readers integrate visuospatial elements into their mental models, however, might be influenced by their cognitive capacities and reading skills [Glenberg and Kruley, 1992, Schneider and Dixon, 2009], which could explain the observed interaction between vsWMC and prior knowledge. High-vsWMC readers with relevant prior knowledge may create more comprehensive mental models that allow for more swift and effective recovery after interruptions.

5.3.4. Implications

The findings support the need for theories like the Memory for Goals and the LT-WM to also account for and incorporate perceptual and visuospatial processing as integral components of the resumption process [Ratwani and Trafton, 2008, Cane et al., 2012]. Although spatial mechanisms may be involved in task resumption mostly in interaction with pre-existence knowledge and other cognitive skills, theoretical frameworks of resumption should outline a more comprehensive and multi-modal representation of this phenomenon.

In the context of the Memory for Goals framework, vsWMC could be introduced within the priming constraint and cue availability mechanism. This constraint posits that the suspended goal can be reactivated if primed with environmental relevant cues [Altmann and Trafton, 2002, Trafton et al., 2005]. As suggested by [Ratwani and Trafton, 2008], vsWMC may support an effective filtering and retrieval of the available environment cues, thereby priming the suspended task goal. Spatial memory information would then be integrated with information stored in long-term memory. Within the

LT-WM theory instead, new textual information functions as retrieval cue to recover previously read information available in short-term memory. vsWMC and spatial information may therefore be integrated in this retrieval process, adding a more perceptual modality and dimension to the accessible cues.

On a more applied perspective, the results underscore the potential of adaptive support systems and AALT that cater to individual differences in vsWMC and prior knowledge. While highlighting the word at the interruption point with visual cues significantly aids resumption [Cane et al., 2012, Jo et al., 2015], future research should examine how the effectiveness of this support may vary depending on reader-related factors such as prior knowledge and vsWMC. For instance, gaze-based AALT and reading applications could leverage these insights to offer personalized support in resuming interrupted reading, particularly through tailored visuospatial cues or interactive tools that align with the reader's cognitive strengths and background.

Furthermore, the observed evidence emphasizes the importance of developing visuospatial skills alongside traditional reading comprehension strategies in educational settings, a critical area often neglected in classroom instruction [Mathewson, 1999]. Personalized support that accounts for the combined influence of vsWMC and prior knowledge may be particularly beneficial in multimedia and computer-based learning environments, where students frequently encounter interruptions and digital distractions and must resume learning sessions.

5.3.5. Limitations

One limitation of this study is that it involved only one type of narrative text, which may limit the generalizability of findings to other genres or text structures. The use of a Sherlock Holmes story, with its famous characters and plot, may have reduced the influence of vsWMC and prior knowledge on resumption. A broader variety of reading materials, including unfamiliar or descriptively rich texts, might yield different results. In addition, the use of fixed interruption location on the middle line of each page could have led to predictability in interruption positioning, potentially impacting resumption

behaviors, although prior research found that interruption location did not significantly impact reading comprehension [Chevet et al., 2022a].

Another limitation emerges from the median split approach to distinguish between the average-low and high prior knowledge groups. Previous studies have highlighted the significance of incorporating a medium group, in addition to low and high groups, to better capture and explain variations in the data [Irwin and McClelland, 2003, Royston et al., 2006]. While following this approach could provide deeper insights, our decision to use a simpler split was guided by prior related work [Taub and Azevedo, 2019, Altamura et al., 2022] as well as the need for balanced group sizes. This approach was intended primarily for additional inspection and visualization rather than forming the sole basis of analysis.

Finally, the manual annotation of gaze data for detecting resumption times might have introduced potential human bias and errors among the annotators. While the annotation process was standardized, variability across raters might have influenced consistency. Future studies could integrate reading fixation detection algorithms, like those developed by Sharmin et al. [2013] for gaze-based auto-scrolling reading and applied in the context of interrupted reading [Jo et al., 2015], to enhance the accuracy of resumption lag detection, thereby improving the reliability and robustness of gaze-based resumption measurements.

5.3.6. Conclusion and Outlook

Addressing the lack of research on how individual cognitive factors affect resumption lag during interrupted reading, the presented study offered novel insights with implications for teaching practices and educational technologies. The results demonstrated a significant interaction of vsWMC and prior knowledge in predicting resumption lags, inviting for further research into reader-related cognitive factors and their effect for interrupted reading.

For instance, since narrative and expository texts may demand different cognitive processes, future research should assess whether these differences also extend to interrupted reading, depending on the reader's vsWMC and

prior knowledge. Besides, additional work should also examine how vsWMC and prior knowledge interact with other text features. Similarly, less familiar genres, texts with rich descriptive details vs. more suspenseful texts, or texts from different historical periods could be examined. The need for vsWMC and prior knowledge might increase or decrease depending on the text type, potentially diversifying their role in resumption during reading. Moreover, while previous studies demonstrated the positive impact of highlighting the interrupted word as a visual cue [Cane et al., 2012, Jo et al., 2015], future studies should explore how the effectiveness of this support varies according to individual differences in vsWMC and prior knowledge. Understanding these variations could further inform the design of adaptive AALT and reading applications that offer more targeted support to users.

In conclusion, this chapter has highlighted the interplay between vsWMC and prior knowledge in supporting reading resumption, contributing to a deeper understanding of the cognitive processes underlying interrupted reading. These findings underscore the potential for future research to leverage individual cognitive capacities to improve reading technologies and educational strategies focused on engagement and efficient task resumption. Nevertheless, the role of text characteristics in interrupted reading remains insufficiently investigated, building the foundation for subsequent research, reported in Chapter 6.

6. Text-Related Factors in Resumption during Reading

This chapter[1] presents the second contribution of this thesis by investigating whether different narrative text styles, specifically texts prompting enactment-imagery vs. description-imagery, can affect resumption time following interruptions. Using an eye tracking study, we compared two narrative excerpts distinguished by their stylistic characteristics: an enactive text, expected to evoke enactment-imagery through descriptions of characters actively engaging with their environment, and a descriptive text, expected to elicit description-imagery by emphasizing the visual properties of environments, with minimal character interaction [Kuzmičová, 2012, 2014, Magyari et al., 2020].

The findings reported in this chapter reveal that imagery type significantly affects fixation duration during resumption lags, suggesting that enactive texts may facilitate re-engagement after interruptions compared to descriptive texts. These results contribute to the field of interrupted reading by introducing mental imagery, induced by text features, as an influential factor in resumption, an aspect previously unexplored in the related literature. Furthermore, the observed findings underscore the importance of considering text features in interrupted reading experimental designs and instructional design, with potential implications for adaptive educational technologies and reading interfaces.

In particular, this chapter tackles the following research question and hypotheses:

---

[1]The content of this chapter has not been published before.

RQ2 *Do narrative texts prompting enactment- vs. description-imagery have a different influence on resumption following interruptions while reading?*

H1 Compared to the descriptive style text, the enactive style text produces longer resumption lags following an interruption while reading.

H2 Compared to the descriptive style text, the enactive style text produces longer fixation durations during resumption following an interruption while reading.

6.1. Methods

The entire study was approved by the Commission for Responsibility in Research of the University of Stuttgart (approval number Az. 22-018). All subjects provided informed consent and the conducted research followed the regulations outlined in Standard 8 of the Ethical Principles and Code of Conduct for Psychologists [American Psychological Association, 2020].

6.1.1. Participants

We recruited 64 participants (42 female, M age $= 28.66$ years, $SD = 9.72$ years and range $18 - 66$ years) through internal mailing lists at the University of Stuttgart and social media. Each volunteer received either a €15 payment or a certificate of participation applicable for study credits for the "Medieninformatik" [Media Computer Science] study program of the University of Stuttgart. To ensure eligibility, all participants completed a pre-screening survey and met the following criteria: normal or corrected-to-normal vision, native-level German proficiency, and no diagnosed attention or reading disorders.

6.1.2. Design

The study employed a within-subject design, with each participant exposed to two narrative texts representing different styles, previously used in Mag-

yari et al. [2020]: a descriptive style text and an enactive style text, presented in random order. These two text conditions were chosen to operationalize the independent variable, i.e., type of elicited mental imagery, as prior research demonstrated that texts with distinctive stylistic characteristics prompt different types of mental imagery [Kuzmičová, 2012, 2014, Jacobs and Willems, 2018, Mak and Willems, 2019, Magyari et al., 2020]. Descriptive passages induce description-imagery, with a focus on the visual properties of the narrated settings, with scarce involvement of the characters, who appear as observers [Kuzmičová, 2014]. Enactive passages instead depict bodily-movements and actively involved characters, interacting with their surroundings, thereby evoking enactment-imagery [Kuzmičová, 2014].

The primary dependent variable was resumption lag, measured as the time from the end of an interruption to the first stable reading pattern in the pre-interruption text, following prior work [Jo et al., 2015, Zermiani et al., 2024a]. We manually annotated the gaze data to obtain individual resumption lags from the pages containing an interruption. The aim was to identify the point in time when each participant resumed reading after an interruption, resulting in approximately 16 resumption points per subject, eight for each text.

In addition to resumption lag, we further operationalized the resumption process with fixation duration, specifically focusing on fixations occurred within the resumption lag interval. Prior work showed significant variations in fixation duration following interruptions in reading, making it a relevant proxy for the resumption process [Cane et al., 2012, Jo et al., 2015, Chevet et al., 2022a]. Besides, fixation duration is frequently used as indicator of higher-level reading processes, such as semantic integration and processing difficulty [Rayner and McConkie, 1976, Just and Carpenter, 1980, Rayner, 1998, Reichle et al., 1998]. By including fixation duration, we also aimed to maintain consistency with Magyari et al. [2020], who used it as proxy for general reading behavior associated with mental imagery generation.

6.1.3. Apparatus

The study was conducted in a laboratory room under constant lightning conditions. The participant's desk was equipped with a screen, a mouse and a keyboard. A room divider separated this desk from the investigator's workstation. Chin and headrest were provided only during gaze data collection. We captured eye movements with the Tobii Pro Spectrum[2] screen-based eye tracker, operating at a sampling frequency of 1200 Hz. We displayed the experiment on the Tobii Pro Spectrum's native screen (EIZO FlexScan EV2451), measuring 52.8×29.7 cm, with a resolution of 1920×1080 px. Participants maintained an approximate viewing distance of 60 cm, positioning their eyes near the center of the Tobii Pro Spectrum's headbox.

We connected the eye tracker and its display a computer, controlling the experimental workflow. As in the previous study [Zermiani et al., 2024a], the experiment interface was developed using *PsychoPy*[3] [Peirce et al., 2019], while eye tracker management was handled via *PyGaze*[4] [Dalmaijer et al., 2014]. Participants' eye movements were calibrated using the default five-point calibration procedure provided by *PyGaze* [Dalmaijer et al., 2014].

6.1.4. Materials

Reading Tasks

All participants read two text excerpts, one expected to prompt description-imagery and the other expected to prompt enactment-imagery. This text pair was obtained from and previously used by Magyari et al. [2020], who conducted their study on two text pairs. Among the two text pairs, we selected the two extracts from the German translation of the novel "The End of the Story" [Davis, 1994] – "*Das Ende der Geschichte*", translated by Klaus Hoffer, published in 2009 [Magyari et al., 2020]. We opted for this text pair for two main reasons. First, these texts were written by the same writer and extracted from the same novel, ensuring consistency in style

---

[2]Firmware version 2-6-1.
[3]Package version 2022.2.3.
[4]Package version 0.7.4.

and context. Second, they have demonstrated stronger reliability as stimuli for distinguishing gaze behavior differences between enactive and descriptive narrative styles, compared to the other text pair used in Magyari et al. [2020].

This novel encloses the memories and reflections of an unnamed narrator who tries to process the end of an intense love affair with a younger man by writing about it. The two excerpts referred to the beginning of the story, particularly pages 2–7 for the enactive style text, and pages 9–21 for the descriptive style one [Magyari et al., 2020]. The plot is not suspenseful and encompasses simple daily-life events like walking through a city, as such mental imagery experience is predicted to become more prominent [Magyari et al., 2020]. The extracts were manipulated by Magyari et al. [2020] to incorporate a higher rate of respectively descriptive and enactive style text parts. Magyari et al. [2020] also ensured that the manipulations did not compromise coherence and understandability and that the German translation was of good quality. The authors provided us with the original annotations of the parts of the texts expected to evoke enactment-imagery and description-imagery [Magyari et al., 2020], based on Kuzmičová [2014]'s imagery typology model.

The descriptive style text included a total of 1181 words and 18 pages, while the enactive one had 1227 words and 17 pages [Magyari et al., 2020]. To ensure consistency, we maintained the same page structure of Magyari et al. [2020], who adjusted line breaks to avoid splitting words as well as line numbers to place sentence endings at page breaks, whenever possible. Paragraph division instead followed the original version of the novel [Magyari et al., 2020]. Pages contained 9 to 11 lines, with a maximum text height and line width of approximately 22×21 cm. However, to also ensure consistency with our previous study and relevant prior research on interrupted reading [Cane et al., 2012, Chevet et al., 2022a], in contrast to Magyari et al. [2020] we did not display page numbers and used a mono-spaced 20-point black Courier font, with a 2.5-point line spacing (see Figure 6.1). The text was left aligned at 500 px and participants pressed the spare bar to move to the next page, while moving backward was not allowed.

For the practice reading task instead, we used an excerpt from the novel
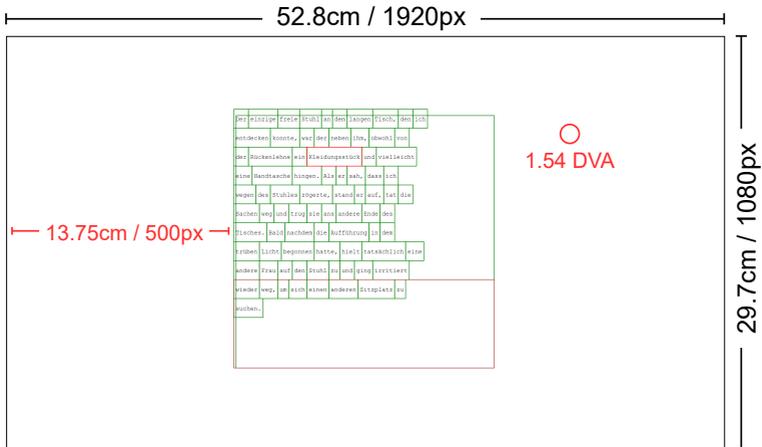
Figure 6.1.: Page setup, including detailed annotations and dimensions. Lines covered 1.54 DVA. Green rectangular areas surrounding each word and the entire text represent predefined regions of interest, commonly known as bounding boxes, defined by pixel coordinates. The bounding box for the target word is shown in red. An additional bounding box in red at the bottom of the text was added to unlock the space-bar as soon as hit with eye movements. This function was activated only in those cases when the participant did not trigger, hence skipped, the interruption. Subjects did not see any bounding box during the experiment.

"Buddenbrooks" [Mann, 2006], following Magyari et al. [2020]. The text covered five pages and was displayed with the same above-described format and layout.

Interrupting Task

We again adopted opinion questions as interrupting task, in the form of dialogue boxes asking participants to type their answer within 60 s. For this study, we implemented new questions generated by ChatGPT-3.5 [OpenAI, 2023] (see Supplementary Material A). A total of 16 unique questions were generated, closely resembling the general opinion topics used in prior research [Pashler et al., 2013]. The questions addressed societal issues, such as justice, privacy, the role of technology, environmental responsibility, and inclusivity in various spheres like the workplace, media, and public safety. These questions were subsequently translated into German and reviewed by a native speaker to ensure clarity and appropriateness. Unlike the previous study, participants were interrupted eight times per text, increasing the frequency of task resumption events.

The selection of target words was refined to ensure interruptions occurred within text segments associated with sensory experience and mental imagery. Since this study aimed to investigate the effect of different imagery types on resumption, interruptions were strategically placed at the peak of those passages anticipated to evoke specific types of mental imagery. By placing interruptions at these imagery-peak points, we aimed to maximize the potential for interrupting vivid mental imagery. This design choice was guided by prior research on mental imagery generation, which demonstrated that the vividness of mental imagery improved as readers devoted more time generating mental images from descriptions of common objects [D'Angiulli and Reeves, 2003].

Particularly, we selected each target word to be approximately in the middle of passages originally annotated as descriptive or enactive by Kuzmičová [2014], Magyari et al. [2020] (see Figure 6.2). Descriptive passages were those in which the narrator primarily focused on visual properties of environments, with human characters minimally involved or typically presented only as ob-

```
DE
Der einzige freie Stuhl an dem langen Tisch, den ich
entdecken konnte, war der neben ihm, obwohl von
der Rückenlehne ein Kleidungsstück und vielleicht
eine Handtasche hingen. Als er sah, dass ich
wegen des Stuhles zögerte, (…)

EN
The only free chair at the long table that I
could spot was the one next to him, although
an item of clothing and perhaps
a handbag were hanging over its backrest. When
he saw me hesitate because of the chair, (...)
```

Figure 6.2.: Example of a target word selection for the descriptive style text. The German version above represents the beginning of page 4 of the descriptive style text used in the experiment. The English version provided below. The text in bold marks Kuzmičová's annotation of a passage with description-imagery potential [Magyari et al., 2020]. The underlined word marks the selected target word within the third line. This word met our selection criteria and assumed that the reader would be in the middle of a mental imagery experience.

servers [Magyari et al., 2020]. By contrast, enactive passages featured human characters actively interacting with their surroundings, manipulating or perceiving objects through senses, reflecting a more inner perspective [Magyari et al., 2020]. Unlike previous studies, where target words were consistently selected within the middle line of the page [Cane et al., 2012, Jo et al., 2015, Zermiani et al., 2024a], target words in this study varied in line position, aligning with the annotated imagery-prompting segments. We maintained the remaining criteria for target word selection described in Section 5.1.4: words at the beginning or end of sentences and lines, as well as function words, were excluded; the selected target words and interruption pages remained constant across participants.

Reading-Related Behavioral Measures

To ensure methodological consistency with Magyari et al. [2020], we adopted their reading comprehension and experience assessments. The reading comprehension questionnaire consisted of three multiple-choice questions per text, specifically designed to evaluate the general understanding of the narrative. All the questions were developed by Magyari et al. [2020].

Additionally, a reading experience questionnaire was used after each text, containing all 18 items from the original study to capture various subjective experiences. Participants rated each item on a 5-point Likert scale, expressing their agreement on aspects of mental imagery, immersion, emotional engagement, and cognitive processing [Magyari et al., 2020]. The questionnaire combined items developed by Magyari et al. [2020] along with selected items from established scales [Appel et al., 2002, Vorderer et al., 2004, Lüdtke et al., 2014], incorporating modifications where necessary by Magyari et al. [2020].

Items in the reading experience questionnaire measured included the following dimensions as reported in Magyari et al. [2020]:

- Mental imagery: items 1-4 evaluated different aspects of imagery evoked by the text;

- Attention and immersion: items 5-7 assessed respectively eventual attentional issues during reading, immersion, and transportation;

- Emotional and cognitive engagement: items probed the feeling of suspense (item 8), emotional involvement (item 9-10), and cognitive involvement (items 12-13);

- Cognitive access and coherence: items evaluated coherence awareness (item 14), cognitive access (items 15-16), and atmosphere provided by the text (item 17);

- Empathy: item 18 evaluated the feeling of empathy toward situations found in the text.

Finally, to examine participants' reading habits, we used the German version of the adapted Reading Habits Questionnaire (RHQ) [Kuijpers et al., 2020, Loi et al., 2023]. In this questionnaire, participants rated how often

they read 20 specific text genres on a 7-point scale, including: $0 =$ never in the past year; $1 =$ once in the past year; $2 =$ roughly every three months in the past year; $3 =$ roughly once a month in the past year; $4 =$ roughly once a week in the past year; $5 =$ more than once a week in the past year; $6 =$ almost every day in the past year. The text genres covered in RHQ comprised: poetry, classics, short stories, contemporary novels, historical novels, love stories, psychological thrillers, crime fiction, horror stories, adventure/action stories, science fiction, supernatural fiction, fantasy; novels for adolescents and young adults, children's literature, comics, humorous books, coming-of-age novels, fan fiction, reference books and non-fiction books [Loi et al., 2023].

6.1.5. Procedure

Once the participants had been briefed on the study, they signed an informed consent and were directed to their assigned desk in the laboratory. They then began the experiment with a set of standard demographic questions, followed by the practice reading task. In this task, participants were only instructed to read the text. They also familiarized themselves with the eye tracker and the calibration process.

Afterwards, the interrupted reading task began. Participants were informed that, at certain points during reading, a window might appear prompting them to type a response to a question within 60 seconds before resuming reading the text. They were told that the questions would cover everyday topics or current events. Moreover, participants were advised to read attentively, as they would be asked to complete a comprehension test afterward.

Participants were also informed that they would read two texts. However, they were unaware of each texts' scope, characteristics, or the specific novel from which they were taken. After reading each text, presented in a randomized order for each subject, participants immediately completed the reading comprehension and reading experience questionnaires. Following the completion of both texts and their respective questionnaires, participants were required to respond to the RHQ, which included a question on whether they

had previously read the novel "The End of the Story" [Davis, 1994].

6.1.6. Scoring

The number of correct answers on the two sets of reading comprehension multiple-choice questions, one for each text, served as measure for reading comprehension, following previous approaches in interrupted reading [Foroughi et al., 2015]. To calculate this score for each participant and questionnaire, we first classified each recorded response as correct or incorrect. By counting the instances of correct responses, we then obtained two total reading comprehension scores per participant, one for each text.

For the reading experience questionnaire, we reproduced the scoring method from Magyari et al. [2020]. Items originally belonging to the same scale were summed to obtain a unique score. In particular, we summed the scores for emotional involvement (items 9-10), cognitive involvement (items 12-13), and ease of cognitive access (items 15-16) [Magyari et al., 2020]. Since items 12 and 16 were negative statements, their scores were inverted prior to obtaining sum scores. Following Magyari et al. [2020], we treated all imagery items (items 1-4) as individual scores, as each item captured a distinct aspect of mental imagery.

For obtaining the resumption lags, we manually annotated the gaze data using the annotation tool to select the resumption point, which was enhanced to align with the study's design. We programmed the tool to group the resumption lags automatically based on the specific text, facilitating efficient case inspection when needed. In addition, a line check was incorporated at the initial stage, ensuring that the annotation could only proceed if the selected point fell within the pre-interruption text. This adjustment improved the process by eliminating the need for an extra verification step to confirm the line location of the selected point, thereby accelerating the overall annotation workflow.

Two independent raters separately annotated resumption lags. Neither of them took part in data collection. The annotation process produced on average 16 resumption lags per participant, eight for each text. The average

Table 6.1.: Estimates of ICC for manual annotations of resumption lags. Average-rating, absolute-agreement and two-way random-effects model was used for the calculation.

	Interclass Correlation	95% Confidence Interval		F Test With True Value 0			
		Lower Bound	Upper Bound	Value	$df1$	$df2$	Sig
Average measures	.711	.670	.747	3.490	895	838	.000

of the annotations provided by the two raters was considered as final measure for resumption lag. We again assessed the reliability of the annotations through the ICC. Following the reporting guidelines by Koo and Li [2016], we computed the ICC score in R [R Core Team, 2021], using the *irr* package[5] [Gamer et al., 2019], with average-rating (k = 2), absolute-agreement and two-way mixed-effects model (see Table 6.1). The estimated ICC value of .711 is within the 0.5–0.75 interval, indicating moderate reliability [Portney and Watkins, 2009, Koo and Li, 2016].

To obtain fixation durations related to resumption, we first extracted fixations from raw gaze data. We used the established REMoDNaV toolkit and eye movements detection algorithm [Dar et al., 2020]. We then calculated the related gaze metric, fixation duration (ms), from the detected fixations. Finally, we selected only fixation durations that occurred within the resumption lag intervals, previously determined through manual annotations.

6.2. Results

In total, we excluded four participants and four specific interruption pages from data analysis, resulting in a final sample size of $n = 60$. The participant exclusions were as follows: one participant encountered a technical issue and viewed the same text twice; another experienced calibration problems, resulting in insufficient gaze data; and two participants failed to trigger more than three interruptions. Additionally, we excluded four individual pages from four different participants due to annotation challenges that made it difficult to identify resumption points visually.

---

[5]Package version 0.84.1.

6.2.1. Imagery Type Effect on Resumption

To examine the effect of imagery type on resumption lags, we used a linear mixed-effects model, implemented in R [R Core Team, 2021], with the *lme4* package[6] [Bates et al., 2015]. We computed *p*-values for fixed effects with the *lmerTest* package[7] [Kuznetsova et al., 2017]. The model specified resumption lag as the criterion variable, with the categorical variable of imagery type, i.e., descriptive or enactive style text, as predictor. Participant ID and page number were included as random effects to account for repeated measurements across different pages by the same subject. With each participant contributing, on average, eight resumption lags per text, we obtained a total of 896 observations. The model obtained a conditional R^2 of 0.18. The imagery type did not significantly predict resumption lags (β = -0.12, CI = -0.25 – 0.01, SE = 0.07, $t(357.79)$ = -1.81, p = .072). On average, participants took 1.09 s (SD = 0.86) to resume from interruptions while reading the enactive text, and 1.20 s (SD = 1.21) for the descriptive text (Figure 6.3).

Following Kumle et al. [2021], we estimated the achieved statistical power for the predictor specified in the linear mixed-effects model, including the random-effects structure. The post-hoc power analysis was conducted in R [R Core Team, 2021] using the *simr* package[8] [Green and MacLeod, 2015]. We estimated power using likelihood ratio tests with 1000 simulations and an alpha level of .05. The analysis yielded an estimated power of 43.60% (CI = 40.50 - 46.74) to detect the main effect of imagery type on resumption lags. These results reflect the achieved power based on the observed effect sizes in the sample [Kumle et al., 2021].

In addition to resumption time, we further inspected the resumption process by analyzing fixation duration. We therefore applied the same linear mixed-effects model to study whether the imagery type had an effect on the duration of the fixations belonging to the resumption lag phase. The model had fixation duration as criterion variable and the categorical variable of imagery type as predictor. Participant ID and page number were again included

---

[6]Package version 1.1.32.
[7]Package version 3.1.3.
[8]Package version 1.0.7.
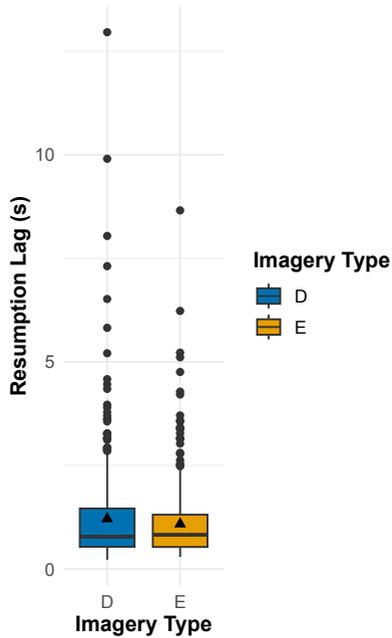
Figure 6.3.: Resumption lags (y-axis) based on imagery type (x-axis). D and E respectively stand for descriptive and enactive style text. The triangles indicate the mean. Black dots indicate outliers in each distribution.
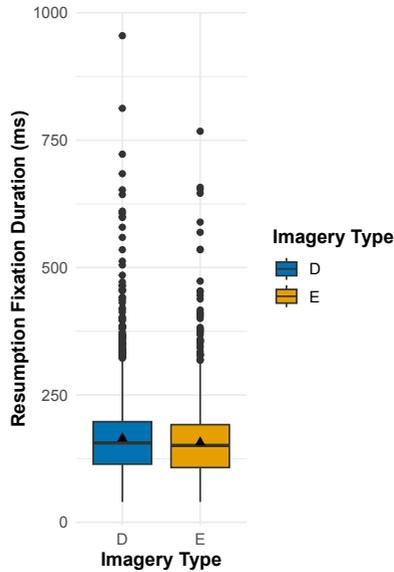
Figure 6.4.: Fixation duration from the resumption lag interval (y-axis) based on imagery type (x-axis). D and E respectively stand for descriptive and enactive style text. The triangles indicate the mean. Black dots indicate outliers in each distribution.

as random effects. The model, with $n_{obs} = 4165$, obtained a conditional R^2 of 0.08. The imagery type had a significant effect on fixation duration during resumption, with fixation durations being significantly lower for the enactive style text (β = -0.07, CI = -0.14 – -0.01, SE = 0.03, $t(445.59)$ = -2.34, p = .018). Upon resumption, the average fixation duration was 156.77 ms (SD = 72.55) for the enactive style text, compared to 164.50 ms (SD = 81.18) for the descriptive style text (see Figure 6.4).

We applied the same simulations to estimate the achieved statistical power for the predictor specified in the linear mixed-effects model. We estimated power using likelihood ratio tests with 1000 simulations and an alpha level of .05. The analysis yielded an estimated power of 68.70% (CI = 65.72 - 71.57) to detect the main effect of imagery type on fixation duration during

resumption. These results reflect the achieved power based on the observed effect sizes in the sample [Kumle et al., 2021].

6.2.2. Reading Comprehension

We additionally reproduced Magyari et al. [2020]'s analyses on the post-test measure for reading comprehension. Following their method and due to the non-normal distribution of the reading comprehension score, we applied the non-parametric Wilcoxon Signed-Rank Test [Wilcoxon, 1945] to check for differences in text comprehension. The test indicated a significant difference in the total correct answers to the reading comprehension questionnaire between the two texts ($Mdn_{enactive} = 3$, $Mdn_{descriptive} = 3$, V $= 307$, $z =$ -3.56, $p < .001$), with a large effect size ($r = .80$).

To understand the direction of this difference, we further inspected the distribution of the score. On average, subjects correctly answered 2.43 ($SD = 0.70$, range $0 - 3$) multiple-choice questions in the reading comprehension test for the enactive style text, while they correctly answered 2.83 ($SD = 0.42$, range $1 - 3$) comprehension questions for the descriptive style text.

6.2.3. Reading Experience

We also replicated Magyari et al. [2020]'s analyses for reading experience. Due to the non-normal distribution of all the reading experience scores, we again ran the Wilcoxon Signed-Rank Test to compare the scores between the different texts. We adjusted the *p*-values for multiple comparisons with the Benjamini-Hochberg adjustment [Benjamini and Hochberg, 1995]. The results for all the reading experience items are displayed in Table 6.2.

To better interpret the observed differences in reading experience ratings, we further obtained the descriptive statistics for all questionnaire items, displayed in Table 6.3. Generally, for items where the Wilcoxon Signed-Rank Test [Wilcoxon, 1945] indicated significant differences, ratings were higher for the enactive style text compared to the descriptive one. The only exception was item 3, assessing the difficulty of generating mental images, where ratings were higher for the descriptive style text compared to enactive one.

Table 6.2.: Results of the Wilcoxon Signed-Rank Test for the ratings of the reading experience questionnaire between the descriptive and enactive texts. Items 9-10, 12-13 and 15-16 were summed in line with Magyari et al. [2020].

	V	z	r	p
1 Imagery	126.5	-1.28	0.29	.245
2 Imagery	108.5	-2.25	0.49	**.045**
3 Imagery	580	-2.34	0.51	**.039**
4 Imagery	186	-0.38	0.09	.745
5 Attentional Focus	318.5	-1.43	0.32	.210
6 Immersion	280	-1.11	0.25	.308
7 Spatial Presence	248	-0.29	0.07	.763
8 Suspense	237	-2.72	0.60	**.019**
9-10 Emotional Involvement	67	-5.01	1.11	**<.001**
11 Overall Reading Pleasure	177	-3.87	0.85	**<.001**
12-13 Cognitive Involvement	342.5	-2.19	0.49	**.048**
14 Coherence	342.5	-1.89	0.43	.086
15-16 Ease of Cognitive Access	316.5	-2.79	0.63	**.019**
17 Atmosphere	125	-2.65	0.60	**.018**
18 Empathy	92	-4.01	0.90	**<.001**

Note. V = test statistic, z = Z score, r = effect size, p = p-value. All p-values were adjusted with the Benjamini-Hochberg procedure [Benjamini and Hochberg, 1995]. P-values $<.05$ are written in bold.

Table 6.3.: Descriptive statistics for the ratings of the reading experience questionnaire between the descriptive and enactive texts. Items 9-10, 12-13 and 15-16 were summed in line with Magyari et al. [2020]. Items in bold indicate those that showed significant results in the Wilcoxon Signed-Rank Test.

	Mdn		*M*		*SD*		*Min*		*Max*	
	D	E	D	E	D	E	D	E	D	E
1 Imagery	4	4	3.72	3.88	1.01	1.01	1	5	1	5
2 Imagery	4	4	3.63	3.9	0.97	0.73	2	5	2	5
3 Imagery	2.5	2	2.83	2.4	1.24	1.15	1	5	1	5
4 Imagery	4	4	3.32	3.37	1.05	0.94	1	5	2	5
5 Attentional Focus	3	3	3.07	2.87	1.01	1.1	1	5	1	5
6 Immersion	3	3	2.88	3.03	1.06	1.12	1	5	1	5
7 Spatial Presence	2	2.5	2.78	2.82	1.15	1.2	1	5	1	5
8 Suspense	3	4	2.75	3.32	1.19	1.28	1	5	1	5
9-10 Emotional Involvement	4	6	4.12	5.75	1.42	1.95	2	8	2	9
11 Overall Reading Pleasure	3	3	2.57	3.25	1.05	1.05	1	4	1	5
12-13 Cognitive Involvement	6	7	5.9	6.42	1.67	1.86	3	10	2	10
14 Coherence	4	4	3.65	3.37	0.95	1.15	1	5	1	5
15-16 Ease of Cognitive Access	5	6	5	5.93	1.8	2.14	2	9	2	10
17 Atmosphere	4	4	3.2	3.65	1.13	1.12	1	5	1	5
18 Empathy	3	4	2.97	3.65	1.15	0.88	1	5	2	5

Note. Total sample size $n = 60$, *Mdn* = median, *M* = mean, *SD* = standard deviation, *Min* = minimum, *Max* = maximum, D = descriptive text, E = enactive text.

6.3. Discussion

Research reported in this chapter investigated the role of text-related factors, in particular enactment- and description-imagery type, in influencing resumption following interruptions during reading. The presented study compared two narrative texts containing enactive and descriptive style passages, with interruptions strategically interspersed. This chapter provides a novel contribution to the interrupted reading research field, addressing the unexplored question of how text style affects resumption. The study aimed to determine whether enactment-imagery, elicited by enactive text features, resulted in longer resumption–characterized by longer resumption lags and fixation durations–compared to description-imagery, following an interruption. While our hypotheses, H1 and H2, were not supported, the results revealed significantly shorter fixation durations during resumption for the enactive text, contributing to a deeper understanding of the cognitive and attentional processes underlying interrupted narrative reading.

Contrary to our expectations, our results indicated no significant effect of imagery type on resumption lags following an interruption. This suggests that the time required to resume reading may be influenced more by individual cognitive factors, such as vsWMC and familiarity with the text genre [Zermiani et al., 2024b], rather than by text features. These findings align with prior discussions (see Chapter 5) on the complexity of the resumption process, which likely draws on the interplay of reader-specific cognitive characteristics [Zermiani et al., 2024b]. Interestingly, while the effect was not statistically significant, participants resumed reading the enactive text slightly faster than the descriptive text, as indicated by the negative coefficient of the predictor as well as by the average resumption times. While in contrast with Magyari et al. [2020], this observation aligns with prior work showing faster reading times on action-laden passages, associated with motor simulation, [Mak and Willems, 2019] and sentences containing two action-related verbs [Marino et al., 2017]. Notably, the enactive text in our experiment contained a higher density of verbs compared to the descriptive text [Magyari et al., 2020], a feature strongly linked to enactment-imagery and action-oriented

cognitive processing [Kuzmičová, 2014].

When inspecting gaze behavior, however, we observed that the imagery type could significantly predict the duration of fixations during resumption lags. In other words, participants exhibited significantly shorter fixation durations when resuming enactive texts compared to descriptive ones. This finding highlights differences in how enactive text may foster cognitive processing upon resumption. Unlike description-imagery, enactive-imagery–characterized by object-directed and bodily-movement descriptions that foster greater engagement with the story [Kuzmičová, 2013, 2014]–appeared to preserve text information more effectively during interruptions. The shorter fixation durations observed upon resumption suggest that enactive-style texts offer higher ease of cognitive processing, consequently enabling swifter recovery of previously read material. These results resonate with Mak and Willems [2019], who found that motor simulation and action-related descriptions were associated with shorter gaze durations, hence faster reading time, indicating reduced processing difficulty [Rayner and McConkie, 1976, Just and Carpenter, 1980, Rayner, 1998]. Conversely, they showed that perceptual simulation, partially corresponding to description-imagery [Magyari et al., 2020], tended to prolong gaze duration, suggesting greater processing difficulty [Rayner and McConkie, 1976, Just and Carpenter, 1980, Rayner, 1998]. With object-directed and bodily-movement descriptions, enactive texts may activate enactment-imagery and action-oriented cognitive processing, facilitating the retrieval and recovery of textual information from working memory following interruptions. Specifically, the higher density of active verb forms and action-focused sentences involving interactions with objects could support cognitive re-engagement upon resumption. Such features may enhance retrieval structures in short-term working memory, thereby enabling efficient access to information stored in LT-WM.

Surprisingly, we found that participants scored significantly lower on reading comprehension for enactive texts, despite the apparent ease of resumption indicated by shorter fixation durations. This contrasts with Magyari et al. [2020], who reported no significant difference in comprehension scores for the same pair of texts. However, their design consisted of a self-paced uninter-

rupted reading task. This discrepancy may reflect the differential impact of interruptions: while the imagery-evoking mechanisms inherent to enactive texts might promote easier resumption, they may not enhance overall comprehension. The self-reported reading experience questionnaire results corroborate this interpretation, revealing no significant differences in attentional focus between the two text types. Enactive texts may thus foster resilience to interruptions without necessarily improving sustained attention or comprehension. Interestingly, while exploring whether participants' individual differences might impact the relationship between mental simulation and gaze behavior in reading, Mak and Willems [2019] showed that higher reported levels of attention had a negative effect on how strong the relationship was. Specifically, the authors postulated that readers reporting low attention were most likely experiencing more mindless reading, hence being more inclined to simulate the narrated events and making more associations while processing the text [Mak and Willems, 2019]. However, as pointed out by Mak and Willems [2019], to uncover the exact dynamics of this complex interplay between attention, mental imagery and gaze behavior, further research is needed.

The reading experience questionnaire revealed notable differences between the two texts across several factors: mental imagery–in terms of both reader's ability to generate vivid mental images of characters and settings described in the text and difficulty of generating mental images–suspense, emotional involvement, overall reading pleasure, cognitive involvement, ease of cognitive access, atmosphere and empathy. These results partially align with and extend the findings of Magyari et al. [2020] for the same pair of texts, where only the imagery-related item assessing the difficulty of generating mental images, i.e., item 3, significantly differed, with descriptive texts being more difficult to visually imagine. Likewise, our results demonstrated greater reported difficulty in conjuring visual images during the descriptive-text story, whereas the enactive-text story brought persons and places to life more vividly in participants' imagination. This suggests that imagery processing in enactive texts may require less conscious effort compared to the description-imagery style [Magyari et al., 2020]. Supporting this, significantly higher ratings for

cognitive involvement and ease of cognitive access were observed for the enactive text. Its action- and object-oriented content was generally perceived as easier to understand and follow, requiring less cognitive effort, in line with Mak and Willems [2019].

Moreover, in contrast to Magyari et al. [2020], the enactive text in this study elicited significantly higher ratings for suspense, highlighting its ability to engage readers with greater narrative tension and anticipation. Participants also reported higher emotional involvement with the enactive text, which they found more enjoyable and atmospheric. The enhanced engagement likely facilitated a deeper empathetic connection with the characters, as displayed in the higher empathy ratings. This suggests that the enactive text contributed to higher affective responses, possibly reflecting a stronger bonding between the reader and the character's experience Caracciolo [2012], Kuzmičová [2014], as well as to higher overall reading pleasure. These observations align with previous research [Oatley, 1995, Lai et al., 2015, Mak and Willems, 2019, Mak et al., 2023], which emphasized the importance of emotional resonance in mental simulation and language processing. Readers simulate the emotions of characters by drawing on their own emotional experiences, creating a reciprocal relationship between emotional engagement and the vividness of mental simulation [Oatley, 1995, Mak and Willems, 2019].

6.3.1. Implications

The findings suggest that texts eliciting enactment-imagery, characterized by action- and object-oriented descriptions, may enhance cognitive processing and facilitate the recovery of textual information during resumption after interruptions in reading. This emphasizes the critical role of task-specific features in influencing resumption processes and underscores the need for resumption frameworks to integrate such features. For instance, text style and imagery could be incorporated into the Memory for Goals framework [Altmann and Trafton, 2002, Trafton et al., 2003] as factors influencing the encoding strength and accessibility of goal representations. Task-specific characteristics, such as action-laden descriptions, may strengthen the link

between interrupted goals and working memory representations via embodied cognition mechanisms, making these goals more resistant to decay or interference during interruptions. Similarly, within the LT-WM theory [Ericsson and Kintsch, 1995, Delaney and Ericsson, 2016], text features can be viewed as part of the retrieval structure. Action-oriented verbs and motor descriptions in enactive texts may serve as stronger retrieval cues, making information more accessible during resumption. Most importantly, cognitive processes activated by text characteristics, such as mental imagery, must be carefully considered when designing interrupted reading experiments and selecting reading materials. Results derived from collections of diverse text types or short paragraphs could be confounded by variations in these features. Ensuring consistency in text characteristics might be relevant to draw more comprehensive conclusions about the cognitive processes involved in reading resumption.

Additionally, the results on emotional and empathetic response have implications for our understanding of how narrative styles influence affective engagement and mental imagery, highlighting an intricate interaction between emotion, imagery, attention and comprehension in narrative processing. Narrative texts appear to offer a rich and versatile resource for examining attention and comprehension dynamics in both uninterrupted and disrupted reading processes.

From a practical perspective, the results have implications for designing educational materials and content, particularly for use in environments prone to frequent interruptions, such as classrooms or digital learning platforms. Enactive-imagery texts, which foster swifter resumption through reduced fixation durations, may be better suited for environments where interruptions are unavoidable. Higher density of active forms, action-laden verbs and sentences involving object-directed interactions may facilitate cognitive processing during resumption, thereby enhancing the learning experience in reading contexts. However, the trade-off between cognitive ease and comprehension must be considered. While enactive texts support smoother recovery from interruptions, they may not always lead to deeper understanding of the material. Designers of instructional content might therefore need to balance the

use of enactive and descriptive styles depending on whether the goal is to facilitate interrupted reading or promote long-term retention and comprehension. These insights may also inform further development of gaze-based AALT for reading. Such technologies could tailor text styles to the learner's external environment, e.g., noisy or interruption-prone settings; internal cognitive states, e.g., learning goals or levels of motivation; and real-time behavioral indicators of attentional focus, such as eye movements. For instance, platforms could prioritize enactive-like features when users are likely to face interruptions or are experiencing distractions or mindless reading [Hutt et al., 2021, 2024], leveraging their ability to facilitate smoother cognitive recovery. Conversely, when the focus is on in-depth comprehension and retention, descriptive-like features may be more appropriate. Adaptive systems could therefore integrate a mix of both styles or enhance enactive style texts with supplemental materials, such as reflective questions or summaries to ensure comprehension is not compromised (for a similar approach leveraging reviews and previews to minimize interruption effects during reading, see Srivastava et al. [2021]), thereby enhancing metacognitive skills [Paris and Winograd, 2013, Stanton et al., 2021].

6.3.2. Limitations

Some methodological limitations in the presented study should be highlighted. While we aimed to capture the differences in mental imagery between texts through the use of subjective ratings in the reading experience questionnaire [Magyari et al., 2020], this approach lacks the precision of more direct measurements of mental imagery or simulation, such as neuroimaging techniques [Mak et al., 2023]. However, the use of subjective scales aligns with both prior work (cf. Mak and Willems [2019] for an overview) as well as efforts to maintain ecological validity in our experimental design. Unlike mobile brain imaging techniques, which still create artificial settings, subjective scales provide a less intrusive and more naturalistic method, making them better suited for assessing reading-related processes while simulating real-world contexts such as educational environments.

Additionally, interruptions were strategically inserted withing those passages that have been previously annotated for evoking enactment- or description-imagery by [Magyari et al., 2020]. Although reflecting a more naturalistic scenario, this design choice led to interruptions occurring at different points in the page, without controlling for line position, as in prior studies [Cane et al., 2012, Jo et al., 2015, Chevet et al., 2022a, Zermiani et al., 2024b]. Efforts were made to counterbalance the line position of target words across pages and texts; however, this variability may have introduced differences in resumption lags between conditions. Controlling for line position in future work could help isolate the effects of text style more precisely.

The observed moderate ICC for manually annotated resumption lags further underscores the criticalities of this method. This emphasizes the need for more reliable and automated resumption detection techniques, such as reading fixation detection algorithms [Jo et al., 2015], to reduce potential bias and improve reproducibility.

Lastly, the low conditional R^2 values for the models, particularly for fixation duration, indicate that the models explained only a small fraction of the variance in the investigated dependent variable, suggesting the influence of other variables not captured by the current design. These could include individual cognitive differences, task difficulty, or fatigue. These results also highlights the complexity of reading behaviors and the challenges of fully capturing how individuals respond to interruptions and different text styles. While the significant predictor offer insights into potential mechanisms, the limited explanatory power emphasizes caution in generalizability and the need for further research with additional variables and refined experimental designs to better understand the factors influencing resumption in interrupted reading.

6.3.3. Conclusion and Outlook

This study opens several avenues for further investigation into the relationship between text characteristics, interruption recovery, and gaze behavior during reading. Future research could extend the analysis by investigat-

ing the effect of imagery type and text features on different phases of the resumption process, following previous approaches [Cane et al., 2012, Zermiani et al., 2024a]. For instance, analyzing eye movements across pre- and post-interruption phases, with a particular focus on the corresponding text segments, could provide deeper insights into the interaction between imagery type, interruptions, and gaze behavior. Moreover, future research could incorporate a detailed linguistic analysis by extracting syntactic and semantic features from the texts. This approach could help identify specific elements that contribute to the facilitative effects of enactive text on resumption, offering a multi-layered investigation of how text characteristics impact interrupted reading.

A more in-depth analysis of the reading comprehension scores could further enhance our understanding of how individual differences influence the interrupted reading process. Exploring whether participants with lower comprehension scores exhibited distinct patterns of re-engagement with the text, or whether their scores primarily indicate a general lack of focus reflected in reading behavior, could help disentangle the cognitive factors influencing comprehension during interrupted reading of narratives.

Finally, this research could be extended to other text types and characteristics. For instance, expository texts, newspaper or scientific articles could be taken into account. Besides, different literary genres could also be explored. In this direction, responses to the reading habits questionnaire could be examined through factor analysis to identify clusters of frequently read genres, as well as frequent readers vs. non-readers. This would enable researchers to study whether genre-related reading frequency may influence resumption times, gaze behavior, and overall reading processes. Our findings also suggest that, beyond imagery, various factors such as cognitive and emotional engagement, empathy, and suspense shape the reading experience of narratives. Examining these variables in relation to resumption mechanisms could uncover their role in enhancing or diminishing resilience to interruptions. By linking text-related with reader-related characteristics, future studies could provide further insights into whether such interactions may shape the interrupted reading experience and inform further development of adaptive

learning technologies.

In conclusion, this chapter has focused and shed light on the impact of text styles eliciting enactment- vs. description-imagery on the resumption process. The presented results contribute to a deeper understanding of the cognitive mechanisms underlying interrupted reading and their interaction with text, a research domain that remains relatively unexplored. These findings emphasize the potential for future research to leverage text features to adaptive learning technologies, fostering cognitive engagement and resilience to interruptions. Integrating task-specific characteristics with individual cognitive differences would provide a more comprehensive understanding of resumption processes. This, in turn, could improve the capacity of learning technologies to tailor and optimize individual learning experiences effectively.

7. InteRead: Towards an Eye Tracking Corpus for Interrupted Reading

This chapter[1] introduces the third contribution of this thesis by presenting InteRead[2], the first publicly accessible eye tracking dataset focused on interrupted reading. Addressing the critical gap in data availability, InteRead provides a valuable resource for investigating individual differences in attentional recovery processes during reading simulating interruption-filled environments. This dataset not only includes comprehensive gaze and linguistic data but also incorporates demographic information and annotations of resumption lags, enabling detailed analysis of task resumption in narrative reading contexts.

The scarcity of publicly available datasets with comparable scope and focus has limited advancements in adaptive, gaze-based AALT and NLP applications for education, which rely on individualized gaze data to support and enhance reading engagement and assessment [Mariakakis et al., 2015, Hutt et al., 2021, Mézière et al., 2023, Southwell et al., 2023, Ganesh et al., 2023]. InteRead was developed to bridge this gap, enabling researchers to conduct investigations into how interruptions and linguistic features affect reading processes, as well as into individual variability in reading and resumption behavior, ultimately fostering the development of more personalized reading tools. This chapter therefore tackles the following overarching research

---

[1]The content of this chapter has already been published in Zermiani et al. [2024a].

[2]The complete dataset is available for download here: `https://doi.org/10.17605/OSF.IO/43J5F`

question:

RQ3 *How can we address data scarcity on individual differences in resumption during reading to advance the development of adaptive AALT?*

Particularly, the validation analyses reported in this chapter address the following three research questions and open up cross-disciplinary research opportunities for future studies that can leverage the capabilities of InteRead:

1. How can we leverage InteRead to investigate the impact of interruptions on gaze behavior during reading?

2. How can we leverage InteRead to investigate the impact of linguistic features on gaze behavior during interrupted reading?

3. How can we leverage InteRead to investigate individual variability in interrupted reading?

7.1. Dataset Overview

The InteRead dataset is an open-source eye tracking dataset focused on interrupted reading, with linguistic features and gaze data gathered from 50 participants performing an interrupted reading task. This dataset emerged from the data collection presented in Chapter 5. Participants engaged in three phases during data collection: a pre-test, an interrupted reading task, and a post-test.

In the pre-test, they completed demographic questions, a prior knowledge questionnaire developed by the authors, assessing familiarity with Sherlock Holmes stories and the crime fiction genre, and a SSPAN task [Unsworth et al., 2009] to measure vsWMC. The SSPAN task required participants to recall sequences of colored squares and perform symmetry judgments on grids, with scores reflecting recall accuracy and processing performance. During the interrupted reading task, participants read an excerpt from the Sherlock Holmes' story "The Adventure of the Speckled Band" [Doyle, 1892], written in Victorian British English, while their eye movements were recorded. Six pages (pages 3, 7, 11, 15, 19, 24) featured an interactive interrupting task in the form of opinion questions [Pashler et al., 2013] appearing in a

dialogue box. The interruptions, triggered by eye tracking data when participants reached specific target words, required participants to type their answer within 60 s. Finally, the post-test included four reading comprehension questions, developed by the authors, to verify participants' attentive reading and a reading experience questionnaire, assessing reading immersion and enjoyment [Kuijpers et al., 2014] and their resumption strategies [Wirzberger and Russwinkel, 2015].

The InteRead dataset therefore encompasses all recorded data and responses from each phase of the data collection, along with the reading stimuli including 5247 words over 28 pages. It also contains analysis scripts and a custom-built annotation tool, which allowed annotators to visualize gaze data following interruptions and manually select the first stable reading pattern in the pre-interruption text to derive resumption lags for each participant. The obtained resumption lags are also included in InteRead. While the methods and materials used during data collection were already detailed in Section 5.1, this section will focus on the data pre-processing procedures involved in constructing the InteRead dataset.

7.1.1. Data Pre-processing

From the initial sample size of $n = 57$, we excluded seven participants from the InteRead dataset for lacking more than 50% gaze data points on the pages including an interruption, or skipping more than two interruptions out of six, or having issue during the resumption annotation phase due to excessively noisy data. In contrast with Chapter 5, in the validation process of this dataset we did not conduct any analyses on the SSPAN task. For this reason, we did not exclude any additional participants, resulting in a final sample size of $n = 50$.

The next pre-processing step involved the extraction of gaze and linguistic features from the raw gaze data and text material.

Gaze Features Extraction

Due to fluctuations in the timestamps of the raw gaze data collected with the PyGaze package [Dalmaijer et al., 2014], we first resampled the data to ensure a consistent 1200 Hz sampling rate. We applied linear interpolation to the gaze coordinates between adjacent valid samples, omitting any samples recorded during blinks. Subsequently, we averaged the left and right eye coordinates from the raw data and detected fixation and saccade events using the REMoDNaV toolkit [Dar et al., 2020].

Following prior research [Hollenstein et al., 2022, Mézière et al., 2023], we therefore obtained relevant gaze features:

- Fixation Duration: mean duration of all fixations for a word (ms);

- First Pass Fixation: fixation duration for the first pass fixation the participant produced on a word (ms);

- Gaze Duration: sum of all first pass fixations on a word (ms);

- Fixation Count: count of fixations a participant made on a word (n);

- Non-Fixation: boolean indicator showing whether a word received fixations;

- Saccade Length: mean length of saccades (px);

- Regression Frequency: mean frequency of regressive saccades–saccades that are directed backward along the line of text (%);

- Reading Times: time required by the participant to read a page (s).

Linguistic Features Extraction

To extract linguistic features, we initially parsed the text read by the participants in Python [Van Rossum and Drake, 2009], using the widely established *spaCy*[3] library [Honnibal and Johnson, 2015]. To integrate the linguistic features with the eye tracking data, we manually aligned each parsed word to its corresponding predefined region of interest, commonly known as a bounding box, defined by its pixel coordinates. We further incorporated abstractness

---

[3]Version 3.6.1.

ratings, i.e., level of abstractness of a given word. To do this, we applied the ratings from Brysbaert et al. [2013], which indicate the extent to which a word's meaning is grounded in human perceptual experiences, on a range of 1–purely abstract to 5–purely concrete. For extracting word frequency values instead, i.e., frequency of a given word in a natural logarithmic scale, we used the SUBTLEXus database [Brysbaert and New, 2009]. Specifically, based on established evidence that logarithmic transformations of word frequency are more appropriate for analyzing word-related processing in reading [Zipf, 1935, Balota et al., 2004, Brysbaert and New, 2009], we extracted the logarithmic frequency values for each word contained in our dataset, from the related column of the SUBTLEXus database.

We thus obtained the following linguistic features:

- Sentences: average number of sentences contained in a given page;

- Part-Of-Speech Tags: grammatical tag of each word;

- Dependency Relations: syntactic dependencies of each word;

- Words: total count of words per page;

- Word Types: total count of unique words per page;

- Word Length: length of each word (in characters);

- Type-Word Ratio: proportion of types to words on a page;

- Content Words: boolean flag showing whether a word is a content word, e.g., nouns or verbs;

- Logarithmic Word Frequency: frequency of a given word in a natural logarithmic scale;

- Abstractness/Concreteness: Ratings on a range of 1–purely abstract to 5–purely concrete.

For all analyses on linguistic elements based on gaze data, we excluded fixation shorter than 100 ms, following previous approaches [Hollenstein et al., 2020, 2022]. Such brief fixations are generally assumed to provide limited meaningful information about reading processes [Just and Carpenter, 1980, Rayner, 1998, Reichle et al., 1998, Sereno and Rayner, 2003].

7.2. Dataset Validation

To validate the InteRead dataset quality, we conducted three focused analyses designed to tackle the related research questions. The analyses presented here specifically aimed to demonstrate its reliablility and applicability for studying gaze behavior during interrupted reading settings.

First, we compared eye movements associated with higher-level reading processes during the pre-interruption and post-interruption phases to determine whether interruptions caused significant differences in their means, following [Cane et al., 2012]. This analysis validated the dataset's ability to capture interruption effects on gaze behavior during reading.

Second, we investigated the influence of word-level linguistic features, specifically logarithmic word frequency and word length, on gaze behavior. These effects are well-established benchmarks in reading research since they reflect the intricate interaction of physiological, psychological, and cognitive processes involved in reading [Rayner, 1998, Kliegl et al., 2004, Sekerina et al., 2019a, Kuperman et al., 2024]. By confirming the presence of these effects in our dataset, we validated its alignment with established reading paradigms and its ability to capture core reading processes, based on linguistic features.

Lastly, we examined the individual variability in our dataset with regard to reading time and resumption time across participants. This analysis explored whether individual differences are reflected in the dataset, addressing how InteRead can be leveraged to study individual variability in interrupted reading behavior.

7.2.1. Interruption Effect on Gaze Behavior

To assess the effect of interruptions on reading behavior, we first identified two main temporal phases in the pages containing an interruption: (1) pre-interruption phase, covering the reading activity from the start of the page until the interruption occurred and (2) post-interruption phase, covering the reading activity from the end of the interruption until the end of the page. We analyzed eye movements associated with higher-level cognitive processes and semantic integration, such as reading time, fixation count, fixation dura-

tion, saccade length, and regression frequency, based on established research [Rayner, 1998, Reichle et al., 1998, Cane et al., 2012]. Our aim was to test whether these gaze metrics exhibited significant changes between the pre- and post-interruption phases. This approach was conceptually grounded in prior research showing the effect of interruptions on these specific eye movements across the temporal phases [Cane et al., 2012, Chevet et al., 2022a].

We therefore conducted a multiple paired t-test analysis [Student, 1908] with Bonferroni correction [Bonferroni, 1936], using R [R Core Team, 2021]. For the pages containing an interruption, the total number of observations ($n_{obs} = 578$) were thus assigned to pre-interruption ($n = 289$) and post-interruption ($n = 289$) groups, corresponding to the identified temporal phases. Results revealed a significant increase in reading times ($t(497.7)$ = -14.65, $p < .0001$, $d = $ -1.21), fixation counts ($t(501.12) = $ -14.16, $p < .0001$, $d = $ -1.17), and saccade lengths ($t(575.99) = $ -5.21, $p < .0001$, $d = $ -0.43) following an interruption. Specifically, as displayed in Figure 7.1, readers spent significantly more time reading, made more fixations, and exhibited longer saccades in the post-interruption phase, compared to the pre-interruption one. However, no significant change was observed for fixation duration ($t(575.97) = 0.42$, $p = 1.0$, $d = 0.03$) or regression frequency ($t(544.11) = 1.91$, $p = .281$, $d = 0.15$) across phases. Table 7.1 provides an overview of the descriptive statistics related to these results.

7.2.2. Linguistic Features Effect on Gaze Behavior

Among the extracted linguistic features, we focused on exploring the impact of logarithmic word frequency and word length on reading gaze behavior, specifically first pass fixation and gaze duration. These gaze metrics have been previously linked to lexical properties, such as word length and frequency, in the reading research literature [Schilling et al., 1998, Kliegl et al., 2004, Hollenstein et al., 2022]. The objective of this analysis was to validate whether our dataset reflected these well-established benchmark effects [Rayner, 1998, Kliegl et al., 2004, Sekerina et al., 2019a, Kuperman et al., 2024], thereby capturing fundamental linguistic influences on reading pro-
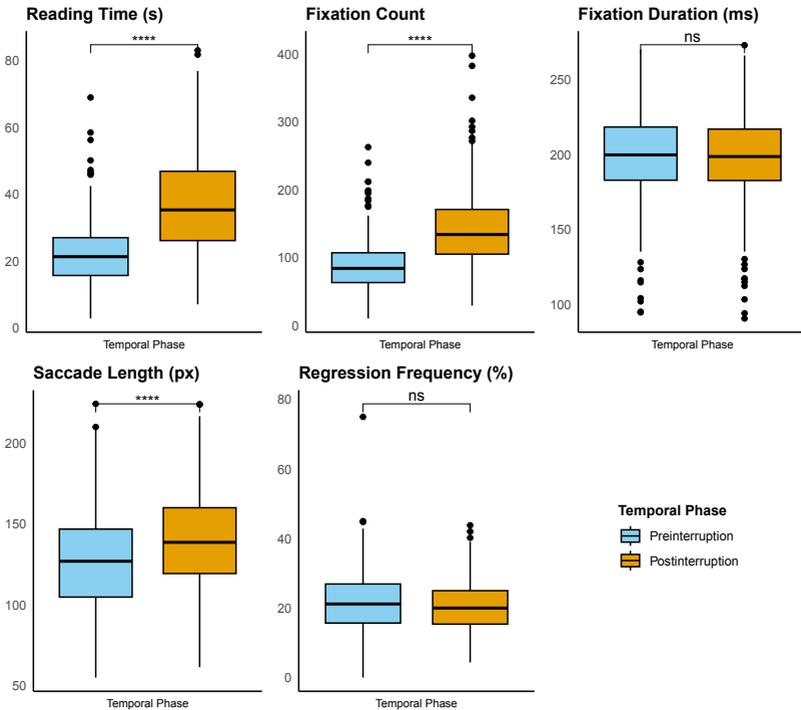
Figure 7.1.: Boxplot illustrating the distribution of reading time, fixation count, fixation duration, saccade length, and regression frequency across pre- and post-interruption phases [Zermiani et al., 2024a]. Black dots indicate outliers in each distribution. Significance between means of the two distributions is represented by stars (*) for significant differences and "ns" for non-significance.

Table 7.1.: Descriptive statistics of the inspected eye movements across the pre-interruption and post-interruption temporal phases, for pages containing an interruption.

	Pre-Interruption		Post-Interruption	
	M	*SD*	*M*	*SD*
Reading Time (s)	22.33	9.38	37.05	14.27
Fixation Count	88.89	37.19	144.82	55.91
Fixation Duration (ms)	199.13	29.02	198.10	29.21
Saccade Length (px)	127.17	29.71	140.05	29.67
Regression Frequency (%)	21.47	9.04	20.18	7.07

Note. Total sample size $n = 50$; M = mean, SD = standard deviation; s = seconds; ms = milliseconds; px = pixels; % = percentage.

cesses. In particular, we inspected whether logarithmic word frequency and word length could significantly influence gaze duration and first pass fixation in our dataset, across all pages.

To test this, we therefore used two linear mixed-effects models implemented in R [R Core Team, 2021], with the *lme4*[4] package[Bates et al., 2015]. We computed *p*-values for fixed effects with the *lmerTest*[5] package [Kuznetsova et al., 2017]. The first model had first pass fixation as the criterion variable, and logarithmic word frequency and word length as predictors. It operated on a total of 110057 observations and obtained a conditional R^2 of 0.06. The second model instead had gaze duration as the criterion variable, and again logarithmic word frequency and word length as predictors. It operated on a total of 110057 observations and obtained a conditional R^2 of 0.10. To account for variability introduced by different participants and text structures, line number, page number, and participant ID were included as random effects in both models.

Word length did not significantly predict first pass fixations (β = -0.13, CI = -0.49 – 0.23, $SE = 0.17$, $t(104300) = -0.71$, $p = .476$) but did substantially predict gaze duration (β = 12.44, CI = 11.90 – 12.97, $SE = 0.26$, $t(108300)$

---

[4]Package version 1.1.32.
[5]Package version 3.1.3.

= 45.39, $p < .0001$). In contrast, logarithmic word frequency significantly predicted both first pass fixation (β = -0.77, CI = -1.04 – -0.50, $SE = 0.14$, $t(101900)$ = -5.61, $p < .0001$) and gaze duration (β = -0.60, CI = -1.01 – -0.20, $SE = 0.20$, $t(107600)$ = -2.93, $p = .003$).

7.2.3. Individual Variability in Reading and Resumption Times

Finally, we aimed to test whether our dataset reflected individual variability in interrupted reading behavior, in particular in overall reading time and resumption time. Following a leave-one-out approach, we first compared each participant's average reading time to the mean reading time of all remaining participants ($n = 49$), using a two-sample unpaired t-test [Student, 1908], implemented in R [R Core Team, 2021]. This method allowed us to identify participants whose reading times significantly deviated from the group average. A Benjamini-Hochberg correction [Benjamini and Hochberg, 1995] was applied to the resulting p-values to control for multiple comparisons.

Out of 50 participants, 31 exhibited significant differences from the remaining group mean ($n = 49$), with 18 reading faster and 13 reading slower than the mean. Since this approach generated a large number of t-statistics, p-values, and effect sizes, we reported the average values to summarize the results more concisely. To compute these averages, we aggregated the t-statistics, p-values, and effect sizes from all comparisons separately for faster readers, slower readers, and typical readers. The average reading time among participants was significantly lower for faster readers, with an average effect of $t(1398)$ = -3.88 (p_{avg} = .007, d_{avg} = -0.74), and significantly higher for slower readers, with an average effect of $t(1398)$ = 5.06 (p_{avg} = .002, d_{avg} = 0.96). On the contrary, the remaining participants showed no significant differences in average reading time, with an average effect of $t(1398)$ = 0.26 (p_{avg} = .439, d_{avg} = 0.05), indicating a more typical reading speed within the sample. The groups were categorized into fast, moderate, and slow readers, with respective mean reading time of 37.22 s ($SD = 11.69$ s, range = 14.23 – 131.77 s), 51.60 s ($SD = 12.71$ s, range = 20.47 – 94.43 s), and 68.02 s (SD

Figure 7.2.: Boxplot illustrating individual differences in mean reading time per page (x-axis) across participants (y-axis), with average reading times arranged in ascending order. The dashed red horizontal line represents the overall mean reading time across all participants. Participants are categorized by reading speed into three groups: Fast, Moderate, and Slow, as reported in Zermiani et al. [2024a].

$= 19.11$ s, range $= 26.32 – 150.65$ s) (Figure 7.2).

To assess individual variability in resumption times, we again employed a leave-one-out approach using a two-sample unpaired t-test [Student, 1908], implemented in R [R Core Team, 2021]. For each participant, their average resumption time was compared to the mean resumption time of all remaining participants ($n = 49$). To account for the multiple comparisons conducted across participants, we applied the Benjamini-Hochberg correction [Benjamini and Hochberg, 1995].

This analysis revealed that four participants exhibited significantly longer

resumption lags compared to the group mean across our sample. Given the large number of comparisons, we summarized the results by reporting the average t-statistics, p-values, and effect sizes separately for participants with longer resumption lags and those with typical resumption lags. Specifically, the four participants indicated substantially longer resumption lags, with an average effect of $t(288) = 3.68$ ($p_{avg} = .018$, $d_{avg} = 1.63$), compared to the group mean. The remaining participants instead showed no significant deviations from the group mean, with an average effect of $t(288) = -0.28$ ($p_{avg} = .677$, $d_{avg} = -0.11$), indicating a more typical resumption time within the sample. Four participants therefore were substantially slower at resuming after interruptions, with an average time of 6.39 s ($SD = 4.05$ s, range $= 0.87 – 18.93$ s), compared to the other participants, who had an average resumption time of 2.46 s ($SD = 1.98$ s, range $= 0.26 – 12.28$ s) (Figure 7.3).

7.3. Discussion

The lack of accessible eye tracking reading datasets incorporating interruptions currently challenges advancements in key research areas. These include the development of gaze-based AALT and reading applications [D'Mello et al., 2012, Jo et al., 2015, Mariakakis et al., 2015, Srivastava et al., 2021, Hutt et al., 2021, 2024], as well as NLP-based educational tasks, such as reading comprehension prediction systems from gaze data [Malmaud et al., 2020, Mézière et al., 2023, Southwell et al., 2023]. These applications increasingly aim to adapt to individual differences and interruption-prone environments. However, development is hindered by the limited availability of datasets that capture how gaze behavior changes in reading contexts designed to simulate real-world scenarios, where interruptions are frequent [Mariakakis et al., 2015, Hutt et al., 2021, Ganesh et al., 2023].

To address this gap, this chapter introduced InteRead, a publicly available eye tracking dataset designed for interrupted reading. InteRead provides a combination of linguistic and gaze features collected from 50 participants as they engaged in reading a long fictional English text containing interruptions.
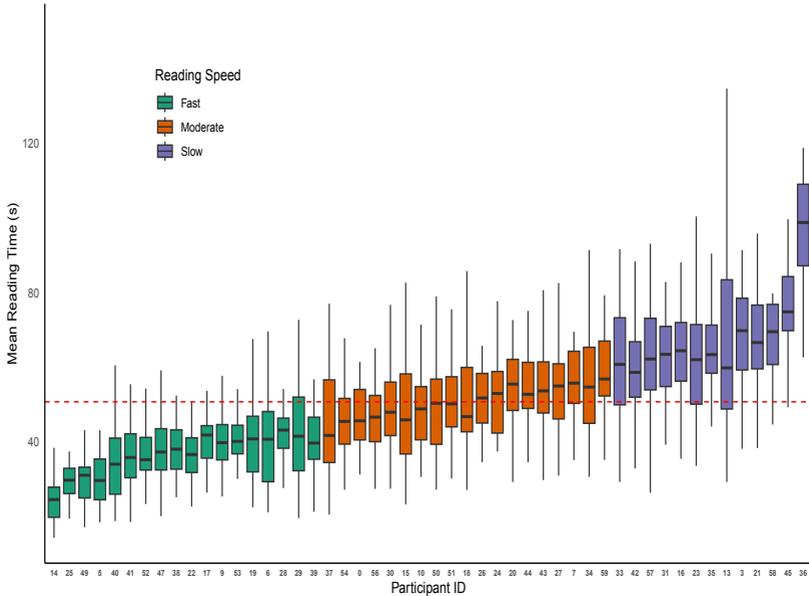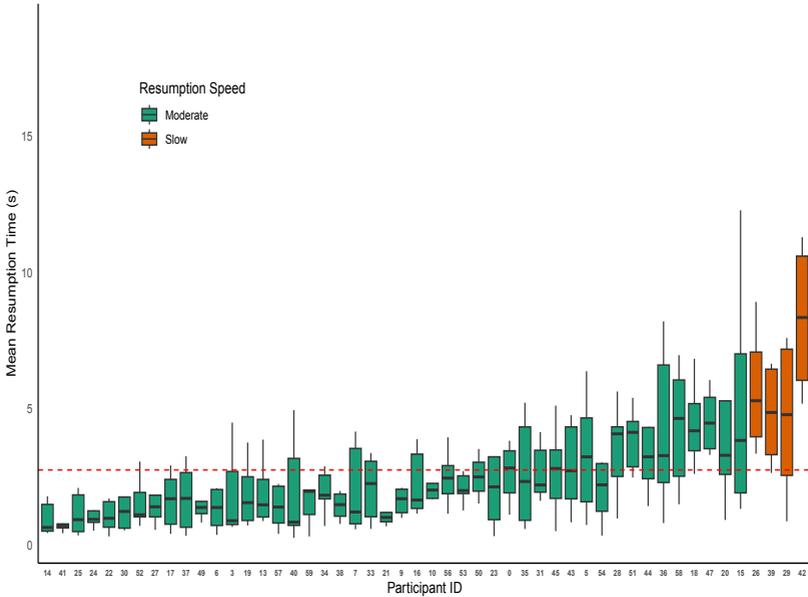
Figure 7.3.: Boxplot illustrating individual differences in mean resumption time across pages (x-axis) and participants (y-axis), with average resumption times arranged in ascending order. The dashed red horizontal line represents the overall mean resumption time across all participants. Participants are categorized by resumption speed into two groups: Moderate and Slow.

By offering detailed gaze data across pre- and post-interruption phases, InteRead has the potential to inform the development of AALT and reading applications by enabling researchers to further investigate how each reader adjusts to interruptions. Additionally, the dataset supports NLP-based educational tasks, such as predicting reading comprehension and assessment, by providing comprehensive annotations linking linguistic features to gaze metrics.

Our analyses applied interdisciplinary methodologies to examine key aspects of InteRead. The validation particularly focused on the impact of inter-

ruptions and linguistic features on gaze behavior as well as on the individual variability reflected in our dataset. These analyses addressed InteRead's validity for studying interrupted reading, providing evidence that our dataset can support future investigations into changes in gaze behavior caused by interruptions. Specifically, results revealed a significant increase in reading time, fixation count, and saccade length after interruptions. Moreover, the significant effects of word frequency on first-pass fixation and gaze duration, and of word length on gaze duration, confirmed the presence of these benchmark effects tied to lexical properties. This demonstrated that InteRead reliably captures core reading processes linked to linguistic features [Rayner, 1998, Kliegl et al., 2004, Sekerina et al., 2019a, Kuperman et al., 2024], and can be further leveraged for their study. Finally, the substantial individual variations in average reading and resumption times highlighted the dataset's capacity to support investigations on individual differences in interrupted reading behavior, offering a resource for adaptive and personalized educational tools.

7.3.1. Interruption Effect on Gaze Behavior

The first set of analyses displayed a substantial increase of reading time, fixation count and saccade length on those pages containing an interruption, for the period of time immediately following the interrupting task itself. These results are consistent with prior findings [Cane et al., 2012, Cauchard et al., 2012], suggesting that readers may revisit and make use of previously read text for reorientation and resumption before moving on to the new text, as indicated by longer saccades, higher number of fixations and longer overall time spent reading.

However, our results did not show significant changes in fixation duration or regression frequency before and after interruptions. While Cane et al. [2012] observed a significant increase in regression frequency compared to the non-interrupted condition, this analysis referred to the re-reading of the pre-interruption text after an interruption. With regard to the same phase, however, Cauchard et al. [2012] reported no significant difference in

regression frequency. When considering instead only the post-interruption text after an interruption, no significant change was reported for regression frequency [Cane et al., 2012, Cauchard et al., 2012]. Prior research also found a consistent significant effect of interruptions on fixation duration in the post-interruption phase [Cane et al., 2012, Cauchard et al., 2012]. Specifically, this effect was observed during the re-reading of previously read text, as well as during the reading of new text introduced after the interruption. However, while average fixation durations increased in the re-reading of the pre-interruption text, they decreased while reading the post-interruption text, also referred to as facilitative effect of interruptions in prior research [Cane et al., 2012, Cauchard et al., 2012]. Despite not reaching statistical significance, our results also displayed shorter fixation duration in the post-interruption phase at the descriptive level. In interrupted reading, processing of new post-interruption text may be therefore facilitated, in spite of the overall increase in reading times [Cane et al., 2012, Cauchard et al., 2012]. This facilitative effect has been explained as a consequence of re-reading the pre-interruption text, which reinforces memory traces, enabling faster integration of novel text [Oulasvirta and Saariluoma, 2006, Cane et al., 2012, Cauchard et al., 2012]. According to the LT-WM theory, skilled readers can quickly store processed information in long-term memory and retrieve it as needed, provided that retrieval cues are present in short-term working memory [Ericsson and Kintsch, 1995, Cane et al., 2012, Cauchard et al., 2012]. Upon resumption, previously read content may no longer be available in short-term memory but can be retrieved from long-term memory if the reader briefly re-reads the pre-interruption text, enhancing integration with new text once reading resumes [Cauchard et al., 2012].

The discrepancy from prior research Cane et al. [2012], Cauchard et al. [2012] may result from differences in experimental design, particularly in how the post-interruption phase was distinguished between re-reading pre-interruption content and engaging with post-interruption text. In contrast with our definition of resumption lag as the time span from the end of the interruption until the first stable reading pattern in the pre-interruption text, Cane et al. [2012] defined it as the time participants spent re-reading the

pre-interruption text upon returning to the task. Furthermore, their study employed a collection of brief paragraphs from various text types, complicating direct comparison to our data collection, which used a continuous long fictional text.

7.3.2. Linguistic Features Effect on Gaze Behavior

Findings from the second set of analyses on the influence of linguistic features on gaze behavior highlighted key mechanisms underlying reading processes across all pages, with word length and frequency effect benchmarks [Rayner, 1998, Kliegl et al., 2004, Sekerina et al., 2019a, Kuperman et al., 2024]. Specifically, word length had a significant effect on gaze duration, but not on first pass fixations. On the other hand, word frequency had a substantial effect on both first pass fixations and gaze duration. These results align with previous findings by Kliegl et al. [2004] regarding the influence of word length and frequency on eye movement during reading.

The non-significant effect of word length on first pass fixations suggests that initial fixations are not strongly influenced by the length of a given word. This may imply that readers do not immediately adjust how long they fixate a given word based on that word's length, but rather initiate their first fixation regardless of word size. However, word length significantly predicted gaze duration, indicating that readers spend more total time on longer words across multiple fixations. This supports Kliegl et al. [2004]'s observation that longer words led to increased processing time, which was reflected in the cumulative duration rather than the initial fixation. Additionally, the observed negative relationship between word frequency and both first pass fixations and gaze duration showed that more frequent words were overall processed more quickly, reducing the time spent on both initial and total fixations. This further aligns with Kliegl et al. [2004], where highly frequent, thereby more predictable, words were skipped more often or processed faster, indicating that frequency plays a key role in facilitating early lexical access and reducing reading time.

These findings contribute to the larger objective of validating InteRead's

capability to capture fundamental linguistic mechanisms, making it a valuable tool for investigating reading behavior based on linguistic and gaze data. Importantly, these results were observed across all pages, both with and without interruptions, demonstrating that InteRead can be leveraged to further study linguistic processes in both normal and interrupted reading contexts.

7.3.3. Individual Variability

The final analyses revealed significant individual differences in reading speed, suggesting that the reading process may vary across individuals. This is consistent with previous research on reading speed, indicating that slow readers tend to have a smaller perceptual span compared to fast readers [Rayner et al., 2010]. Slow readers may require more resources to process and comprehend the currently fixated word, which could limit the amount of information they can simultaneously access to the right of the currently fixated word, thereby slowing down their overall processing speed [Rayner et al., 2010].

Additionally, our results showed substantial individual differences in resumption lags within the dataset. In particular, certain participants required significantly longer times to resume reading following an interruption compared to the group mean, indicating different levels of resumption performance. These differences could stem from individual cognitive factors, as investigated in prior research across various tasks [Werner et al., 2011, Meys and Sanderson, 2013, Bai et al., 2014, Drews and Musters, 2015, Foroughi et al., 2016b, Kanaan and Moacdieh, 2021, Zermiani et al., 2024b]. For instance, participants with lower vsWMC and prior knowledge may have required significantly more time to filter irrelevant textual information and reactivate mental representations upon resumption, resulting in extended resumption times [Zermiani et al., 2024b].

7.3.4. Implications

Our results revealed substantial interruption effects on gaze behavior, indicating that InteRead offers fine-grained temporal gaze data to further inves-

tigate the specific challenges interruptions pose to reading and how readers adapt to them. Such insights could inform the development of gaze-based automatic resumption detection systems. For instance, these systems could leverage the observed variations in eye movements following interruptions to train computational models to identify resumption processes while reading, based on the annotated resumption lags as ground truth. These models could in turn be integrated in gaze-based AALT and reading applications to implement resumption detection and guide learners to re-engage with the text, through timely visual cues or metacognitive aids (see Cane et al. [2012], D'Mello et al. [2012], Mariakakis et al. [2015], Jo et al. [2015], Hutt et al. [2021], Srivastava et al. [2021]).

Additionally, these technologies could leverage the individual variability observed in InteRead to dynamically adapt to readers' differences in processing speed and recovery time. Personalized support, such as customized visual prompts or adaptive pacing, could align with each reader's unique needs, fostering improved reading outcomes and moving beyond one-size-fits-all approaches [Mariakakis et al., 2015, Hutt et al., 2021].

Our findings also demonstrated InteRead's capacity to enable detailed investigations of the interaction between linguistic features and eye movements during reading. For example, the presence of word length and frequency benchmark effects on gaze behavior confirms its quality as a resource for future psycholinguistic research. Moreover, InteRead could enhance NLP applications by integrating linguistic and gaze data into models for tasks like reading comprehension and assessment prediction. For instance, incorporating gaze data from interrupted reading pages might increase the robustness of models designed to handle interruption-prone contexts [Ganesh et al., 2023].

Fruthermore, considering the variety of extracted linguistic features, InteRead also supports comparative studies of linguistic processing during reading of uninterrupted and interrupted pages, providing insights into how readers cognitively adjust to text features after interruptions. These investigations could broaden our understanding of the interplay between attention, text and comprehension, particularly in light of the open question on the effects of interruptions on text comprehension [Chevet et al., 2022a].

7.3.5. Limitations

From an ethical perspective, InteRead's participant demographics only represents a specific subset of the global population. In addition, the reading material used in this dataset focuses exclusively on an overrepresented language, such as English, neglecting the strong effect that the specific language we speak can have on cognition [Blasi et al., 2022]. Increasing the diversity of ethnic, gender, language and cultural background representation will enhance the generalizability and relevance of our findings for interrupted reading.

Moreover, our data's generalizability should be considered within the constraint of a laboratory-based, controlled environment, which ensured high internal validity by allowing precise measurement of interruptions' effects during a task involving reading and typing. However, this setting may limit external validity, as participants were exposed to structured interruptions rather than more unpredictable interruptions encountered in real-world scenarios. Besides, head movements during typing introduced some drift in the raw gaze data. Despite these limitations, the induced interruptions provided an ecologically valid approximation of real-world disruptions.

Another methodological limitation found in our analyses relates to the highly skewed distribution of fixation durations, with a higher rate of zero-count fixations observed in the data. Whereas excluding fixation durations under 100 ms is a common approach [Hollenstein et al., 2020, 2022], prior work explored alternative methods to handle zero-count fixations, including zero-inflated models [Olkoniemi et al., 2023]. However, due to non-convergence issues when testing such models (cf. Olkoniemi et al. [2023]), we opted for standard exclusion practices. Addressing this issue remains an area for further research to optimize gaze data analysis methods.

Finally, it is noteworthy to again mention that our analyses of resumption lags relied on manual annotations, which may have introduced some degree of subjective variability. Although these annotations provided a valid measure of resumption lag during reading, they emphasize the need for automated resumption detection techniques, such as reading fixation detection algorithms [Sharmin et al., 2013, Jo et al., 2015], to enhance data reliability.

7.3.6. Conclusion and Outlook

This chapter introduced InteRead, the first fully open-source eye tracking dataset specifically designed to study interrupted reading. The reported analyses aimed to validate the dataset's scope and potential for cross-disciplinary research by exploring its utility in investigating the effects of interruptions, linguistic features, and individual variability on reading behavior. While the findings supported InteRead's scope and potential, they also emphasized several directions for future research.

Future work could delve deeper into the impact of interruptions on gaze behavior through more fine-grained analyses that compare eye movements in specific areas of interest within the text. For instance, exploring differences in eye movements directly linked to the previously read text right before vs. novel text read right after an interruption could reveal more about how readers visually processed and navigated the text, building on prior studies [Cane et al., 2012, Cauchard et al., 2012, Chevet et al., 2022a].

Additionally, InteRead offers comprehensive linguistic features. Expanding future investigations to include these features would allow for a more detailed understanding of how linguistic properties, such as parts of speech tags, syntactic dependencies, or word concreteness, may affect gaze behavior during interrupted reading. For example, InteRead could be used to examine whether specific elements, such as content words or abstract terms, are more likely to be skipped, or if specific features of Victorian British English, like complex syntax and archaic vocabulary, may be reflected in eye movements.

The dataset's detailed gaze data and resumption lag annotations make it particularly suited to further studying individual differences in interrupted reading. Future work could examine how readers showing significant differences in reading and resumption time, compared to the group mean, might display distinct gaze behaviors or interactions with linguistic features. Building on studies like Rayner et al. [2010], researchers could investigate whether faster and slower readers exhibited different gaze patterns when returning to the text after an interruption, or whether participants who experienced longer resumption lags showed specific gaze behaviors during recovery phases.

Moreover, linking these participants' answers on the reading comprehension questionnaire to specific areas of interest or words in the text could deepen our understanding on the interplay of attention, linguistic elements and comprehension.

While InteRead lays the groundwork for eye tracking datasets on interrupted reading research, its scope could be enriched by including datasets in additional languages, especially underrepresented ones. Applying similar methodologies across languages would broaden the dataset's reach and allow for cross-linguistic comparisons, toward the creation of a corpus for the study of interrupted reading through eye tracking.

In conclusion, InteRead enables investigations into how interruptions influence eye movements, how linguistic complexity shapes gaze behavior, and how individual traits may affect reading and resumption processes. These insights hold meaningful promise for bridging the gaps across disciplines and advancing adaptive gaze-based educational technologies and reading tools, particularly those aimed at accommodating diverse reading behaviors.

8. Conclusion

Attention is a cornerstone of effective learning, especially in rapidly chang-
ing educational environments, where multitasking has been taking over sus-
tained focus. Being a foundational skill for the entire educational process,
understanding attention dynamics during reading and how to counteract the
effects of interruptions and distractions is highly relevant to enhance knowl-
edge acquisition and learning outcomes. As digital technologies increasingly
dominate educational contexts, the challenges of maintaining focus among
frequent interruptions have been receiving a growing research interest. In
this regard, special efforts are being made to develop educational technologies
capable of mitigating the disruptive effects of interruptions during learning,
particularly by leveraging eye tracking data to track learners' attention and
intervene to redirect the focus on the learning material in case of distraction.
These advancements have the potential to foster the learning experience and
equip students with the skills needed to thrive in a digitally-driven education.

However, there has been a lack of research on what influencing factors are
at play in the process of resumption during reading, leading to a scarcity
of related data. While previous research has focused on the properties of
interruptions, the factors that influence learners' efficiency in managing in-
terruptions remain largely unexplored. This gap poses a critical limitation for
further development of current gaze-based AALT and reading applications,
which mostly adopt one-size-fits-all designs, thereby placing less emphasis
on learners' individual differences and task-specific elements. To bridge this
existing gap, this thesis aimed to tackle the question of what factors can
influence resumption time following interruptions during reading, hence pro-
viding the scientific community with additional data to advance research on
interrupted reading. The thesis particularly focused on narrative texts as

reading material, given their established yet under-explored potential to engage readers through their structure and linguistic features [Stein and Glenn, 1979, Graesser and Clark, 1985, Emmott et al., 2006, Kukkonen, 2020, Mar et al., 2021, Kukkonen and Baumbach, 2022].

A key result from Chapter 5 consisted in the significant interaction between vsWMC and prior knowledge as predictor of resumption lags after interruptions while reading a detective story. This offers a novel insight into what individual differences in cognitive factors can have an impact on the time required to resume reading a text. Specifically, the related study was the first to assess vsWMC and prior knowledge respectively through a psychometric test and a customized questionnaire, in the context of interrupted narrative reading. Another key finding from Chapter 6 related to how the style of a narrative text, in particular enactive-style descriptions prompting enactment-imagery, could significantly reduce fixation durations during the resumption process. This insight from Chapter 6 provides an additional novel contribution to the field, by comparing for the first time two narrative texts during interrupted reading, addressing the gap of what text-specific features may influence resumption process, with a focus on mental imagery. Additionally, the InteRead dataset, introduced in Chapter 7, offered a final key contribution, with validation analyses showing substantial effects of interruptions, linguistic features, and individual variability on gaze and reading behavior. This dataset aimed to tackle the existing need for experimental data specifically designed to study resumption mechanisms and related individual differences during reading. Although previous studies already investigated this phenomenon, InteRead is the first fully open-source dataset to include gaze and linguistic features, along with resumption lag annotations, from a long real-word narrative text containing interruptions, thereby opening up new opportunities for cross-disciplinary research.

8.1. Broader Impact

The key findings presented in this thesis have practical implications for teachers aiming to support students in managing interruptions during reading,

particularly in digital learning environments filled with distractions. First, the interaction between vsWMC and prior knowledge emphasizes the importance of recognizing individual cognitive differences among students. Teachers could benefit from tailoring reading activities and trainings students' assessment on these factors. For instance, teachers may implement visualization strategies, which have been shown to also improve reading comprehension [De Koning and van der Schoot, 2013], or incorporate more visualizations in reading materials for students with lower vsWMC. Besides, providing context-building exercises might be beneficial for students with limited prior knowledge on the learning material. Another key finding was that enactive-style descriptions, which prompted enactment-imagery, could reduce fixation durations during resumption. This highlights the potential of using action-oriented styles in learning material to facilitate cognitive processing and smoother transitions after interruptions. Teachers might incorporate such features into their materials, especially in environments where interruptions are frequent, to maintain engagement and focus. This result further underscores how supporting students in building non-linguistic representations of text might promote text comprehension, hence learning outcomes (cf. De Koning and van der Schoot [2013] for a collection of visualization strategies). Lastly, the development of the InteRead dataset emphasizes the importance of fostering collaboration and dialogue between researchers and educators. By translating insights from eye tracking data into actionable and concrete strategies, researchers can promote a mutual exchange of knowledge, helping educators better understand how reading behaviors relate to student needs. This dialogue can further inform the joint development of materials, tools, and strategies that support students more effectively, ultimately enhancing learning outcomes in real-world educational settings.

These insights have also meaningful implications for advancing research and development of gaze-based AALT. The interaction between vsWMC and prior knowledge as predictor of resumption lags underscores the crucial need for adaptive learning technologies to integrate learners' individual differences in their design, avoiding one-size-fits-all approaches. AALT could better support learners in handling interruptions by tailoring interventions

based on students' vsWMC and domain-specific knowledge. For instance, visually highlighting the last fixated word before an interruption, integrating visuo-spatial information in the forms of visualizations in the text, or offering previews and reviews upon resumption might support only students with lower vsWMC and prior knowledge. As prior research efforts already led to the design of similar solutions for different reading technologies and applications [Cane et al., 2012, Jo et al., 2015, Mariakakis et al., 2015, Hutt et al., 2021, Srivastava et al., 2021], insights from this thesis could help researchers to extend prior solutions into more adaptive and comprehensive systems. Similarly, the observed impact of narrative styles on reducing fixation durations during resumption have implications for leveraging text-specific features to enhance cognitive processing in interrupted reading scenarios. Gaze-based AALT could be designed to dynamically adapt not only to individual differences, but also to the inherent characteristics of the learning material, shaping texts that are both engaging and easier to comprehend. Most importantly, the open-access InteRead dataset provides a foundation for interdisciplinary research on the interplay of interruptions, linguistic features, and individual variability in reading behavior. By integrating gaze and linguistic data related to the resumption process into the development pipeline, gaze-based AALT can evolve into more personalized and context-aware tools, improving their ability to foster resilience to distractions and support diverse learning needs.

8.2. Limitations

Despite the outlined contributions, this thesis also presents several limitations to be acknowledged. Firstly, the recruitment criteria excluded individuals with attention and reading difficulties, and limited the sample to native speakers or those with high language proficiency. While these restrictions ensured a homogeneous group for controlled experimentation, they limit the generalizability of the findings to populations with diverse cognitive or linguistic backgrounds.

Second, the choice of texts was restricted to narrative materials. While

narratives were selected due to their potential to reveal insights into attentional mechanisms and their limited investigation in educational research, this choice may limit the applicability of findings to other text types, such as expository or instructional texts, which require different cognitive skill sets.

A further limitation concerns the absence of a standardized assessment of reading comprehension. While this choice was made to specifically focus on resumption processes, the lack of a robust comprehension measure poses a challenge in generalizing the findings to potential learning outcomes. Lastly, the scope of investigated influencing factors was relatively narrow, encompassing only a small subset of reader-related and text-related factors. This focused approach was necessary to maintain control in the experimental design but leaves room for expanding the scope in future studies to include a wider range of individual differences and text features.

8.3. Future Directions

Building on the discussed limitations, numerous avenues for future research emerge from this thesis, which would enhance the generalizability, applicability, and depth of understanding in the study of interrupted reading and its implications for learning technologies and educational practices. Firstly, expanding the participant pool to include neurodivergent populations, such as individuals challenged by maintaining their attentional focus and processing textual stimuli, would offer critical insights into how neurodevelopmental conditions beyond the cognitive majority influence resumption processes. This could not only improve theoretical understanding but also inform the design of supportive strategies or tools tailored to enhance interruption resilience for neurodivergent groups. For instance, it might be beneficial to explore the effect of the interaction between vsWMC and prior knowledge on resumption in relation to attention disorders, particularly in light of recent findings linking individual variability in executive function to learning performance during distracting multimedia lessons [Lawson and Mayer, 2024]. Similarly, a deeper examination of the linguistic features of enactive and descriptive texts could be conducted, focusing on their effects on populations with read-

ing difficulties to assess whether their experience and processing of such texts differ. Furthermore, the participant pool could also be expanded to include speakers of different languages or individuals with varying proficiency levels in the same language. This would enable researchers to explore how linguistic and cognitive factors related to a specific language may influence interrupted reading behavior, given how language can shape our cognitive systems [Majid et al., 2004]. For instance, comparing native and non-native speakers of a language could reveal differences in how linguistic complexity and familiarity may affect the resumption process.

Second, future research could replicate and extend these studies using expository texts, which are central to educational settings. Comparing different types of expository texts through the combined lenses of psycholinguistics and educational science could help identify textual features that promote resilience to interruptions. For narrative texts, exploring a wider array of genres or works from different historical periods could uncover how different variables associated with the narrative reading experience interact with resumption processes. Such studies, although complex in design [Wolfe and Woodwyk, 2010], could pinpoint specific textual triggers that either facilitate or hinder resumption, as already outlined in prior research with regard to mind wandering and linguistic attractors and distractors [Kukkonen and Baumbach, 2022].

A third direction involves integrating robust reading comprehension measures into studies of interrupted reading. Previous research leveraging eye tracking data to model reading comprehension have either used assessments from established inventories [Mézière et al., 2023], or custom-designed tests [Southwell et al., 2023]. In the context of interrupted reading instead, custom-designed tests are more common, often encompassing several text comprehension and recognition questions [Glanzer et al., 1984, Oulasvirta and Saariluoma, 2006, Cane et al., 2012, Foroughi et al., 2015, Chevet et al., 2022a]. Both methods present limitations. Established tests often require the use of their specific materials, whereas long comprehension tests might extend experimental times, increasing participant fatigue and introducing additional confounding variables. In light of the conflicting results regarding the im-

pact of interruptions on reading comprehension (see Chevet et al. [2022a] for an overview), future research could aim to find a trade-off by incorporating both text comprehension and recognition questions while carefully managing experimental length to avoid fatigue effects.

As direct extensions of the presented findings, additional individual cognitive factors could be included in future analyses of reading resumption, such as reading habits, genre preferences, and reading proficiency. Moreover, an in-depth investigation into the factors associated with significant differences between enactive and descriptive texts–such as suspense, emotional and cognitive engagement, reading pleasure, ease of cognitive access, atmosphere, and empathy–could shed light on whether the experiences evoked by these factors aid the resumption process. Finally, constructing and sharing new emerging datasets including additional languages could contribute to a robust and comprehensive corpus of interrupted reading, facilitating cross-linguistic comparisons and interdisciplinary research.

In summary, this thesis highlights how interruptions in reading reflect the intricate interplay of reader-related cognitive factors and text-specific factors, offering insights into how these aspects can potentially guide and support resumption. The reported findings are especially relevant in light of the observed shift from a deeper attention toward rapid attentional shifts among multiple tasks, also referred to as hyper attention by Hayles [2007]. As she concluded: *"Whether inclined toward deep or hyper attention, toward one side or another of the generational divide separating print from digital culture, we cannot afford to ignore the frustrating, zesty, and intriguing ways in which the two cognitive modes interact."* (Hayles [2007]: 197–198). Over 15 years later, this thesis underscores the need for further research on the process of resumption during reading, to inform adaptive technologies that can better accommodate the evolving ways we engage with text, in the interruption-prone digital age.

A. Supplementary material

Interrupting Task: Opinion Questions

This section provides the 16 opinion questions used as interrupting task, in the study presented in Chapter 6, with eight interruptions per text. Within the descriptive text, interruptions were placed on page 3, 5, 6, 8, 9, 15, 16, 17. Within the enactive text, interruptions were placed on page: 1, 2, 3, 5, 8, 9, 15, 16. English translations of the questions are included in brackets.

Descriptive Text

1. Sind Sie der Meinung, dass die Regierungen den Einsatz von künstlicher Intelligenz regulieren sollten? Warum oder warum nicht? (Do you think governments should regulate the use of artificial intelligence? Why or why not?)

2. Was denken Sie über die zunehmende Automatisierung von Arbeitsplätzen und deren mögliche Auswirkung auf die Arbeitslosigkeit? (How do you feel about the increasing automation of jobs and its potential impact on unemployment?)

3. Wie denken Sie über die Darstellung von Diversität in der Medien- und Unterhaltungsindustrie? (What are your thoughts on the representation of diversity in the media and entertainment industry?)

4. Welches sind Ihrer Meinung nach die dringlichsten Probleme, mit denen das Bildungswesen heutzutage konfrontiert ist? (In your opinion, what are the most pressing issues facing education today?)

5. Wie würden Sie das Problem der Stigmatisierung der psychischen Gesundheit in der Gesellschaft angehen? (How would you address the challenges of mental health stigma in society?)

6. Befürworten Sie die Idee von Pflichtimpfungen oder lehnen Sie diese ab? Warum? (Do you support or oppose the idea of mandatory vaccinations? Why?)

7. Welche Maßnahmen würden Sie vorschlagen, um Gleichstellung und Inklusion am Arbeitsplatz zu fördern? (What measures would you propose to promote equality and inclusivity in the workplace?)

8. Wie sollte die Gesellschaft Ihrer Meinung nach die Rechte einer Person auf Privatsphäre mit der Notwendigkeit einer verstärkten Überwachung zur Gewährleistung der öffentlichen Sicherheit in Einklang bringen? (How do you think society should balance individual privacy rights with the need for enhanced surveillance to ensure public safety?)

Enactive Text

1. Wie würden Sie mit einer Situation umgehen, in der Sie Zeuge oder Zeugin davon werden, wie jemand in der Öffentlichkeit misshandelt wird? (How would you handle a situation where you witness someone being mistreated in public?)

2. Was denken Sie über die Rolle der sozialen Medien zur Bildung der öffentlichen Meinung? (What are your thoughts on the role of social media in shaping public opinion?)

3. Wie würden Sie am Arbeitsplatz mit einem/einer Mitarbeitenden umgehen, der/die ständig die Anerkennung für Ihre Ideen an sich reißt? (In a workplace, how would you address a colleague who consistently takes credit for your ideas?)

4. Vertreten Sie das Konzept eines allgemeinen Grundeinkommens? Warum oder warum nicht? (Do you believe in the concept of a universal basic income? Why or why not?)

5. Wie denken Sie über die Auswirkungen der Technologie auf die Privatsphäre im digitalen Zeitalter? (How do you feel about the impact of technology on privacy in the digital age?)

6. Welche Maßnahmen sollten Ihrer Meinung nach ergriffen werden, um den Klimawandel auf globaler Ebene zu bekämpfen? (What measures do you think should be taken to address climate change on a global scale?)

7. Wie würden Sie ein Gespräch mit einer Person angehen, die ganz andere politische Ansichten vertritt als Sie? (How would you approach a conversation with someone who holds drastically different political views than you?)

8. Welche Rolle spielt Ihrer Meinung nach Einfühlungsvermögen für eine effektive Führung? (What role do you think empathy plays in effective leadership?)

Bibliography

Noureldin Mohamed Abdelaal and Amal Saleh Sase. Relationship between prior knowledge and reading comprehension. *Advances in Language and Literary Studies*, 5:125–131, 2014. doi: 10.7575/aiac.alls.v.5n.6p.125.

Lidia Altamura, Ladislao Salmerón, and Yvonne Kammerer. Instant messaging multitasking while reading: A pilot eye-tracking study. In *2022 Symposium on Eye Tracking Research and Applications*, ETRA '22, pages 1–6. Association for Computing Machinery, 2022. ISBN 9781450392525. doi: 10.1145/3517031.3529237.

Erik M. Altmann and J. Gregory Trafton. Memory for goals: An activation-based model. *Cognitive Science*, 26:39–83, 2002. doi: 10.1016/S0364-0213(01)00058-1.

Erik M. Altmann and J. Gregory Trafton. Task interruption: Resumption lag and the role of cues. *Proceedings of the 26th Annual Conference of the Cognitive Science Society*, pages 43–48, 2004.

Erik M. Altmann and J. Gregory Trafton. Timecourse of recovery from task interruption: Data and a model. *Psychonomic Bulletin & Review*, 14: 1079–1084, 2007. ISSN 1531-5320. doi: 10.3758/BF03193094.

Erik M. Altmann and J. Gregory Trafton. Brief lags in interrupted sequential performance: Evaluating a model and model evaluation method. *International Journal of Human-Computer Studies*, 79:51–65, 2015. doi: 10.1016/j.ijhcs.2014.12.007.

Ulrike Altmann, Isabel C. Bohrn, Oliver Lubrich, Winfried Menninghaus, and Arthur M. Jacobs. The power of emotional valence—from cognitive to affective processes in reading. *Frontiers in Human Neuroscience*, 6, 2012. ISSN 1662-5161. doi: 10.3389/fnhum.2012.00192.

APA American Psychological Association. *Ethical Principles of Psychologists*

Bibliography

and Code of Conduct. Including 2010 and 2016 Amendments. American Psychological Association, Washington, DC, 2020. URL `https://www.apa.org/ethics/code`.

John R. Anderson. *The Architecture of Cognition*. Harvard University Press, Cambridge, MA, USA, 1983.

John R. Anderson and Christian Lebiere. *The Atomic Components of Thought*. Lawrence Erlbaum Associates Publishers, 1998.

Richard C. Anderson. Allocation of attention during reading. *Advances in Psychology*, 8:292–305, 1982. doi: 10.1016/S0166-4115(08)62699-0.

Richard C. Anderson. Role of the reader's schema in comprehension, learning, and memory. In Richard C. Anderson, J. Osborn, and R. J. Tierney, editors, *Learning to Read in American Schools*, pages 243–257. Lawrence Erlbaum Associates, 1984. ISBN 9780898592191.

Markus Appel, Erik Koch, Margrit Schreier, and Norbert Groeben. Aspekte des leseerlebens: Skalenentwicklung. *Zeitschrift für Medienpsychologie*, 14: 149–154, 2002. doi: 10.1026//1617-6383.14.4.149.

Richard C. Atkinson and Richard M Shiffrin. The control of short-term memory. *Sci Am*, 225(2):82–90, 1971. doi: 10.1038/scientificamerican0871-82.

Alan Baddeley. Working memory. *Science*, 255(5044):556–559, 1992. ISSN 00368075, 10959203. doi: 10.1126/science.1736359.

Hao Bai, Winston E. Jones, Jarrod Moss, and Stephanie M. Doane. Relating individual differences in cognitive ability and strategy consistency to interruption recovery during multitasking. *Learning and Individual Differences*, 35:22–33, 2014. ISSN 1041-6080. doi: 10.1016/j.lindif.2014.07.002.

David Balota, Michael Cortese, Susan Sergent-Marshall, Daniel Spieler, and Melvin Yap. Visual word recognition of single-syllable words. *Journal of Experimental Psychology. General*, 133:283–316, 2004. doi: 10.1037/0096-3445.133.2.283.

Maria Barrett and Nora Hollenstein. Sequence labelling and sequence classification with gaze: Novel uses of eye-tracking data for natural language processing. *Language and Linguistics Compass*, 14:1–16, 2020. doi: 10.1111/lnc3.12396.

Frederic C. Bartlett. *Remembering: A study in experimental and social psy-*

chology. Cambridge University Press, 1932.

Douglas Bates, Martin Mächler, Ben Bolker, and Steve Walker. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67:1–48, 2015. doi: 10.18637/jss.v067.i01.

Yoav Benjamini and Yosef Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)*, 57(1):289–300, 1995. doi: 10.1111/j.2517-6161.1995.tb02031.x.

Sandra Berkowitz and B.M. Taylor. The effects of text type and familiarity on the nature of information recalled by readers. In Michael L. Kamil, editor, *Directions in reading: Research and instruction*, pages 157–161. National Reading Conference, Washington, DC, 1981.

Yevgeni Berzak, Chie Nakamura, Amelia Smith, Emily Weng, Boris Katz, Suzanne Flynn, and Roger Levy. Celer: A 365-participant corpus of eye movements in l1 and l2 english reading. *Open mind : Discoveries in Cognitive Science*, 6(11):41–50, 2022. doi: 10.1162/opmi_a_00054.

Damián E. Blasi, Joseph Henrich, Evangelia Adamou, David Kemmerer, and Asifa Majid. Over-reliance on english hinders cognitive science. *Trends in Cognitive Sciences*, 26(12):1153–1170, 2022.

Carlo Bonferroni. Teoria statistica delle classi e calcolo delle probabilita. *Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commericiali di Firenze*, 8:3–62, 1936.

Laura L. Bowman, Laura E. Levine, Bradley M. Waite, and Michael Gendron. Can students really multitask? an experimental study of instant messaging while reading. *Computers & Education*, 54:927–931, 2010. ISSN 0360-1315. doi: 10.1016/j.compedu.2009.09.024.

Marc Brysbaert and Boris New. Moving beyond kucera and francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for american english. *Behavior Research Methods*, 41(4):977–990, 2009. doi: 10.3758/BRM.41.4.977.

Marc Brysbaert, Amy Warriner, and Victor Kuperman. Concreteness ratings for 40 thousand generally known english word lemmas. *Behavior Research Methods*, 46:904–911, 10 2013. doi: 10.3758/s13428-013-0403-5.

Bibliography

Kirsten R. Butcher. The multimedia principle. In Richard E. Mayer, editor, *The Cambridge Handbook of Multimedia Learning*, pages 174–205. Cambridge University Press New York, NY, 2014. doi: 10.1017/CBO9781139547369.010.

Alejandro J. Cambronero-Delgadillo, Sarah Jasmin Nachtnebel, Christof Körner, Iain D. Gilchrist, and Margit Höfler. Interruption in visual search: a systematic review. *Frontiers in Psychology*, 15, 2024. ISSN 1664-1078. doi: 10.3389/fpsyg.2024.1384441.

James E. Cane, Fabrice Cauchard, and Ulrich W. Weger. The time-course of recovery from interruption during reading: Eye movement evidence for the role of interruption lag and spatial memory. *Quarterly Journal of Experimental Psychology*, 65:1397–1413, 2012. ISSN 1747-0218. doi: 10.1080/17470218.2012.656666.

Marco Caracciolo. Fictional Consciousnesses: A Reader's Manual. *Style*, 46:42–65, 03 2012. ISSN 0039-4238. doi: 10.5325/style.46.1.42.

Fabrice Cauchard, James Cane, and Ulrich Weger. Influence of background speech and music in interrupted reading: An eye-tracking study. *Applied Cognitive Psychology*, 26(3):381–390, 2012. doi: 10.1002/acp.1837.

Guillaume Chevet, Thierry Baccino, Lucas Marlot, Annie Vinter, and Véronique Drai-Zerbib. Effects of interruption on eye movements and comprehension during reading on digital devices. *Learning and Instruction*, 80:101565, 2022a. ISSN 0959-4752. doi: 10.1016/j.learninstruc.2021.101565.

Guillaume Chevet, Thierry Baccino, Annie Vinter, and Véronique Drai-Zerbib. What breaks the flow of reading? a study on characteristics of attentional disruption during digital reading. *Frontiers in Psychology*, 13:987964, 2022b. doi: 10.3389/fpsyg.2022.987964.

Guillaume Chevet, Thierry Baccino, Annie Vinter, and Véronique Drai-Zerbib. Effect of prior knowledge on re-reading behavior after an interruption and text comprehension. In *European Conference on Eye Movements*, 2022c. doi: 10.13140/RG.2.2.24342.83523.

Kit Cho, Jeanette Altarriba, and Maximilian Popiel. Mental juggling: When does multitasking impair reading comprehension? *The Journal of general psychology*, 142:90–105, 2015. doi: 10.1080/00221309.2014.1003029.

Ruth C Clark and Richard E Mayer. *e-Learning and the Science of Instruction: Proven Guidelines for Consumers and Designers of Multimedia Learning*. John Wiley & Sons, 2016. doi: 10.1002/9781119239086.

Fiona Collins. The use of traditional storytelling in education to the learning of literacy skills. *Early Child Development and Care*, 152(1):77–108, 1999. doi: 10.1080/0300443991520106.

Cristina Conati, Vincent Aleven, and Antonija Mitrovic. Eye-tracking for student modelling in intelligent tutoring systems. *Design Recommendations for Intelligent Tutoring Systems*, 1:227–236, 2013.

Uschi Cop, Nicolas Dirix, Denis Drieghe, and Wouter Duyck. Presenting GECO: An eyetracking corpus of monolingual and bilingual sentence reading. *Behavior Research Methods*, 49:602–615, 2017. doi: 10.3758/s13428-016-0734-0.

Nelson Cowan. Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system. *Psychological Bulletin*, 104(2):163–191, 1988. doi: 10.1037/0033-2909.104.2.163.

Nelson Cowan. The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, 24(1): 87–114, 2001. doi: 10.1017/s0140525x01003922.

Edwin S. Dalmaijer, Sebastiaan Mathôt, and Stefan Van der Stigchel. PyGaze: An open-source, cross-platform toolbox for minimal-effort programming of eyetracking experiments. *Behavior Research Methods*, 46: 913–921, 2014. ISSN 1554-3528. doi: 10.3758/s13428-013-0422-2.

Meredyth Daneman and Patricia A. Carpenter. Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, 19(4):450–466, 1980. ISSN 0022-5371. doi: https://doi.org/10.1016/S0022-5371(80)90312-6.

Amedeo D'Angiulli and Adam Reeves. Probing vividness increment through imagery interruption. *Imagination, Cognition and Personality*, 23(1):63–78, 2003. doi: 10.2190/FD86-RWDH-FLX2-ULRG.

Asim H. Dar, Adina S. Wagner, and Michael Hanke. Remodnav: Robust eye-movement classification for dynamic stimulation. *bioRxiv*, 53:399–414,

Bibliography

2020. doi: 10.1101/619254. URL https://www.biorxiv.org/content/early/2020/03/19/619254.

Lydia Davis. *The End of the Story*. Farrar, Straus, and Giroux, 1994. ISBN 0374148317.

Björn De Koning and Menno van der Schoot. Becoming part of the story! refueling the interest in visualization strategies for reading comprehension. *Educational Psychology Review*, 25(2):261–287, 2013. doi: 10.1007/s10648-013-9222-6.

Peter Delaney and Karl Ericsson. Long-term working memory and transient storage in reading comprehension: What is the evidence? comment on foroughi, werner, barragán, and boehm-davis (2015). *Journal of Experimental Psychology: General*, 145:1406–1409, 2016. doi: 10.1037/xge0000181.

Peter F. Delaney. The role of long-term working memory and template theory in contemporary expertise research. *Journal of Expertise*, 1(3):155–161, 2018. URL https://www.journalofexpertise.org.

Pablo Delgado and Ladislao Salmerón. The inattentive on-screen reading: Reading medium affects attention and reading comprehension under time pressure. *Learning and Instruction*, 71:101396, 2021. doi: 10.1016/j.learninstruc.2020.101396.

Arthur C. Doyle. Adventures of Sherlock Holmes. VIII.–The Adventure of the Speckled Band. In George Newnes, editor, *The Strand Magazine*, volume 3, pages 142–157. George Newnes Ltd, London, UK, 1892. URL https://archive.org/details/StrandMagazine14.

Frank A. Drews and Adrian Musters. Individual differences in interrupted task performance: One size does not fit all. *International Journal of Human-Computer Studies*, 79:97–105, 2015. ISSN 1095-9300. doi: 10.1016/j.ijhcs.2015.01.003.

Sidney D'Mello. Gaze-based attention-aware cyberlearning technologies: Learning in the age of emerging technologies. In Thomas D. Parsons, Lin Lin, and Deborah Cockerham, editors, *Mind, Brain and Technology*, pages 87–105. Springer International Publishing, 2019. ISBN 978-3-030-02630-1. doi: 10.1007/978-3-030-02631-8_6.

Sidney D'Mello and Arthur C. Graesser. Dynamics of affective states during

complex learning. *Learning and Instruction*, 22(2):145–157, 2012. ISSN 0959-4752. doi: 10.1016/j.learninstruc.2011.10.001.

Sidney D'Mello, Andrew Olney, Claire Williams, and Patrick Hays. Gaze tutor: A gaze-reactive intelligent tutoring system. *International Journal of Human-Computer Studies*, 70:377–398, 2012. doi: 10.1016/j.ijhcs.2012. 01.004.

Sidney D'Mello, Kristopher Kopp, Robert Earl Bixler, and Nigel Bosch. Attending to Attention: Detecting and combating mind wandering during computerized reading. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '16, pages 1661–1669. Association for Computing Machinery, 2016. ISBN 9781450340823. doi: 10.1145/2851581.2892329.

Sidney D'Mello, Caitlin Mills, Robert Bixler, and Nigel Bosch. Zone out no more: Mitigating mind wandering during computerized reading. In *2017 International Conference on Educational Data Mining*, EDM 2017, 2017.

Oliver Eberle, Stephanie Brandl, Jonas Pilot, and Anders Søgaard. Do transformer models show similar attention patterns to task-specific human gaze? In Smaranda Muresan, Preslav Nakov, and Aline Villavicencio, editors, *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 4295–4309, Dublin, Ireland, 2022. Association for Computational Linguistics. doi: 10.18653/v1/2022.acl-long.296. URL `https://aclanthology.org/2022. acl-long.296`.

Educational Testing Service. Official gre practice test. Copyright ©, 2024. URL `https://www.ets.org/gre`. [Accessed: 2024-11-30].

Amy M. Elleman and Eric L. Oslund. Reading comprehension research: Implications for practice and policy. *Policy Insights from the Behavioral and Brain Sciences*, 6:3–11, 2019. doi: 10.1177/2372732218816339.

Catherine Emmott, Anthony J. Sanford, and Lorna I. Morrow. Capturing the attention of readers? stylistic and psychological perspectives on the use and effect of text fragmentation in narratives. *jlse*, 35(1):1–30, 2006. ISSN 0341-7638, 1613-3838. doi: 10.1515/JLS.2006.001.

K. Anders Ericsson and Walter Kintsch. Long-term working memory. *Psy-*

chological Review, 102:211–245, 1995. ISSN 1939-1471. doi: 10.1037/0033-295X.102.2.211.

Beth Fischer and Murray Glanzer. Short-term storage and the processing of cohesion during reading. *The Quarterly Journal of Experimental Psychology Section A*, 38(3):431–460, 1986. doi: 10.1080/14640748608401607.

Martin H. Fischer and Rolf A. Zwaan. Embodied language: A review of the role of the motor system in language comprehension. *Quarterly Journal of Experimental Psychology*, 61(6):825–850, 2008. doi: 10.1080/17470210701623605.

Cyrus K. Foroughi, Nicole E. Werner, Daniela Barragán, and Deborah A. Boehm-Davis. Interruptions disrupt reading comprehension. *Journal of Experimental Psychology: General*, 144:704–709, 2015. ISSN 1939-2222. doi: 10.1037/xge0000074.

Cyrus K. Foroughi, Daniela Barragán, and Deborah A. Boehm-Davis. Interrupted reading and working memory capacity. *Journal of Applied Research in Memory and Cognition*, 5:395–400, 2016a. ISSN 2211-3681. doi: 10.1016/j.jarmac.2016.02.002.

Cyrus K. Foroughi, Nicole E. Werner, Ryan McKendrick, David M. Cades, and Deborah A. Boehm-Davis. Individual differences in working-memory capacity and task resumption following interruptions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 42:1480–1488, 2016b. ISSN 1939-1285. doi: 10.1037/xlm0000251.

Jeffrey L. Foster, Zach Shipstead, Tyler L. Harrison, Kenny L. Hicks, Thomas S. Redick, and Randall W. Engle. Shortened complex span tasks can reliably measure working memory capacity. *Memory & Cognition*, 43:226–236, 2015. ISSN 1532-5946. doi: 10.3758/s13421-014-0461-7.

Annie Fox, Jonathan Rosen, and Mary Crawford. Distractions, distractions: Does instant messaging affect college students' performance on a concurrent reading comprehension task? *Cyberpsychology & Behavior: The Impact of the Internet, Multimedia and Virtual Reality on Behavior and Society*, 12:51–3, 2008. doi: 10.1089/cpb.2008.0107.

Stefan Frank, Irene Monsalve, Robin Thompson, and Gabriella Vigliocco. Reading time data for evaluating broad-coverage models of english sen-

tence processing. *Behavior Research Methods*, 45:1182–1190, 2013. doi: 10.3758/s13428-012-0313-y.

Kathleen Marie Gallagher. In search of a theoretical basis for storytelling in education research: Story as method. *International Journal of Research & Method in Education*, 34(1):49–61, 2011. doi: 10.1080/1743727X.2011.552308.

Vittorio Gallese and Hannah Wojciehowski. How stories make us feel: Toward an embodied narratology. *California Italian Studies*, 2, 2011. doi: 10.5070/C321008974.

Matthias Gamer, Jim Lemon, and Ian Fellows Puspendra Singh. *irr: Various Coefficients of Interrater Reliability and Agreement*, 2019. R package version 0.84.1.

Ananya Ganesh, Jie Cao, E. Margaret Perkoff, Rosy Southwell, Martha Palmer, and Katharina Kann. Mind the gap between the application track and the real world. In Anna Rogers, Jordan Boyd-Graber, and Naoaki Okazaki, editors, *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 1833–1842, Toronto, Canada, July 2023. Association for Computational Linguistics. doi: 10.18653/v1/2023.acl-short.156.

Vanessa E. Ghosh and Asaf Gilboa. What is a memory schema? a historical perspective on current neuroscience literature. *Neuropsychologia*, 53:104–114, 2014. ISSN 0028-3932. doi: 10.1016/j.neuropsychologia.2013.11.010.

Murray Glanzer and Suzanne Donnenwerth Nolan. Memory mechanisms in text comprehension. In Gordon H. Bower, editor, *Psychology of Learning and Motivation*, volume 20, pages 275–317. Academic Press, 1986. doi: https://doi.org/10.1016/S0079-7421(08)60021-3.

Murray Glanzer, David Dorfman, and Barbara Kaplan. Short-term storage in the processing of text. *Journal of Verbal Learning and Verbal Behavior*, 20 (6):656–670, 1981. ISSN 0022-5371. doi: 10.1016/S0022-5371(81)90229-2.

Murray Glanzer, Beth Fischer, and David Dorfman. Short-term storage in reading. *Journal of Verbal Learning and Verbal Behavior*, 23:467–486, 1984. ISSN 0022-5371. doi: 10.1016/S0022-5371(84)90300-1.

Arthur M. Glenberg and Peter Kruley. Pictures and anaphora: Evidence for

independent processes. *Memory & Cognition*, 20(5):461–471, 1992. ISSN 0090-502X. doi: 10.3758/BF03199579.

Kevin A. Gluck, John R. Anderson, and Scott A. Douglass. Broader bandwidth in student modeling: What if ITS were "eye"TS? In Gilles Gauthier, Claude Frasson, and Kurt VanLehn, editors, *Intelligent Tutoring Systems*, ITS 2000, pages 504–513, 2000. ISBN 978-3-540-45108-2. doi: 10.1007/3-540-45108-0_54.

Fernand Gobet. Long-term working memory: A computational implementation for chess expertise. In *Proceedings of the Third International Conference on Cognitive Modelling*, pages 142–149. Universal Press, 01 2000.

Arthur C. Graesser and Leslie F. Clark. The generation of knowledge based inferences during narrative comprehension. In Gert Rickheit and Hans Strohner, editors, *Inferences in Text Processing*, volume 29, pages 53–94. North-Holland, Amsterdam, 1985. doi: 10.1016/S0166-4115(08)62732-6.

Arthur C. Graesser, Jonathan M. Golding, and Debra L. Long. Narrative representation and comprehension. In *Handbook of reading research, Vol. 2.*, pages 171–205. Lawrence Erlbaum Associates, Inc, Hillsdale, NJ, US, 1991.

Arthur C. Graesser, Keith K. Millis, and Rolf A. Zwaan. Discourse comprehension. *Annual Review of Psychology*, 48(1):163–189, 1997. doi: 10.1146/annurev.psych.48.1.163.

Peter Green and Catriona MacLeod. SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, 7:493–498, 11 2015. doi: 10.1111/2041-210X.12504.

Thor Grünbaum. Action between plot and discourse. *Semiotica*, 2007(165): 295–314, 2007. doi: 10.1515/SEM.2007.045.

Abid Haleem, Mohd Javaid, Mohd Asim Qadri, and Rajiv Suman. Understanding the role of digital technologies in education: A review. *Sustainable Operations and Computers*, 3:275–285, 2022. ISSN 2666-4127. doi: 10.1016/j.susoc.2022.05.004.

Shannon Harp and Richard E. Mayer. How seductive details do their damage: A theory of cognitive interest in science learning. *Journal of Educational Psychology*, 90:414–434, 1998. doi: 10.1037/0022-0663.90.3.414.

Joellen T. Hartley. Reader and text variables as determinants of discourse memory in adulthood. *Psychology and Aging*, 1:150–158, 1986. doi: 10. 1037/0882-7974.1.2.150.

Franziska Hartung, Michael Burke, Peter Hagoort, and Roel Willems. Taking perspective: Personal pronouns affect experiential aspects of literary reading. *PLOS ONE*, 11:e0154732, 2016. doi: 10.1371/journal.pone.0154732.

Katherine Hayles. Hyper and deep attention: The generational divide in cognitive modes. *Profession*, 2007:187–199, 2007. doi: 10.1632/prof.2007. 2007.1.187.

Patricia Hirsch, Luca Moretti, Sibel Askin, and Iring Koch. Examining the cognitive processes underlying resumption costs in task-interruption contexts: Decay or inhibition of suspended task goals? *Memory & Cognition*, 52(2):271–284, 2024. doi: 10.3758/s13421-023-01458-8.

Helen M. Hodgetts and Dylan M. Jones. Interruption of the tower of london task: Support for a goal-activation approach. *Journal of Experimental Psychology: General*, 135:103–15, 2006a. doi: 10.1037/0096-3445.135.1. 103.

Helen M. Hodgetts and Dylan M. Jones. Contextual cues aid recovery from interruption: The role of associative activation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32:1120–1132, 2006b. ISSN 1939-1285. doi: 10.1037/0278-7393.32.5.1120.

Patrick Colm Hogan. *The Mind and its Stories: Narrative Universals and Human Emotion*. Studies in Emotion and Social Interaction. Cambridge University Press, 2003.

Nora Hollenstein, Marius Troendle, Ce Zhang, and Nicolas Langer. ZuCo 2.0: A dataset of physiological recordings during natural reading and annotation. In Nicoletta Calzolari, Frédéric Béchet, Philippe Blache, Khalid Choukri, Christopher Cieri, Thierry Declerck, Sara Goggi, Hitoshi Isahara, Bente Maegaard, Joseph Mariani, Hélène Mazo, Asuncion Moreno, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 138–146, Marseille, France, 2020. European Language Resources Association. ISBN 979-10-95546-34-4. URL https://aclanthology.org/2020.lrec-1.18.

Bibliography

Nora Hollenstein, Maria Barrett, and Marina Björnsdóttir. The copenhagen corpus of eye tracking recordings from natural reading of Danish texts. In Nicoletta Calzolari, Frédéric Béchet, Philippe Blache, Khalid Choukri, Christopher Cieri, Thierry Declerck, Sara Goggi, Hitoshi Isahara, Bente Maegaard, Joseph Mariani, Hélène Mazo, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 1712–1720, Marseille, France, 2022. European Language Resources Association. URL `https://aclanthology.org/2022.lrec-1.182`.

Matthew Honnibal and Mark Johnson. An Improved Non-monotonic Transition System for Dependency Parsing. In Lluís Màrquez, Chris Callison-Burch, and Jian Su, editors, *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 1373–1378, Lisbon, Portugal, September 2015. Association for Computational Linguistics. doi: 10.18653/v1/D15-1162. URL `https://aclanthology.org/D15-1162`.

Samar Husain, Shravan Vasishth, and Narayanan Srinivasan. Integration and prediction difficulty in Hindi sentence comprehension: Evidence from an eye-tracking corpus. *Journal of Eye Movement Research*, 8:1–12, 2015. doi: 10.16910/jemr.8.2.3.

Stephen Hutt, Caitlin Mills, Shelby White, Patrick J. Donnelly, and Sidney D'Mello. The Eyes Have It: Gaze-based detection of mind wandering during learning with an intelligent tutoring system. In *Proceedings of the 9th International Conference on Educational Data Mining*, EDM 2016, pages 86–93. International Educational Data Mining Society, 2016. URL `https://api.semanticscholar.org/CorpusID:16029341`.

Stephen Hutt, Caitlin Mills, Nigel Bosch, Kristina Krasich, James Brockmole, and Sidney D'Mello. "Out of the Fr-Eye-ing Pan": Towards gaze-based models of attention during learning with technology in the classroom. In *Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization*, UMAP '17, page 94–103, New York, NY, USA, 2017. Association for Computing Machinery. ISBN 9781450346351. doi: 10.1145/3079628.3079669.

Stephen Hutt, Kristina Krasich, James R. Brockmole, and Sidney D'Mello.

Breaking out of the lab: Mitigating mind wandering with gaze-based attention-aware technology in classrooms. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, pages 1–14, New York, NY, USA, 2021. Association for Computing Machinery. ISBN 9781450380966. doi: 10.1145/3411764.3445269.

Stephen Hutt, Aaron Wong, Alexandra Papoutsaki, Ryan S. Baker, Joshua I. Gold, and Caitlin Mills. Webcam-based eye tracking to detect mind wandering and comprehension errors. *Behavior Research Methods*, 56(1): 1–17, January 2024. ISSN 1554-351X. doi: 10.3758/s13428-022-02040-x. Publisher Copyright: © 2022, The Psychonomic Society, Inc.

Julie R. Irwin and Gary H. McClelland. Negative consequences of dichotomizing continuous predictor variables. *Journal of Marketing Research*, 40(3): 366–371, 2003. doi: 10.1509/jmkr.40.3.366.19237.

Arthur M. Jacobs. Neurocognitive poetics: methods and models for investigating the neuronal and cognitive-affective bases of literature reception. *Frontiers in Human Neuroscience*, 9, 2015a. ISSN 1662-5161. doi: 10.3389/fnhum.2015.00186.

Arthur M. Jacobs. Towards a neurocognitive poetics model of literary reading. In Roel M. Willems, editor, *Cognitive Neuroscience of Natural Language Use*, pages 135–159. Cambridge University Press, 2015b.

Arthur M. Jacobs and Roel Willems. The fictive brain: Neurocognitive correlates of engagement in literature. *Review of General Psychology*, 22(2): 1–27, 2018. doi: 10.1037/gpr0000106.

Lena A. Jäger, Thomas Kern, and Patrick Haller. Potsdam textbook corpus (potec), January 2021.

Elspeth Jajdelska, Christopher Butler, Stephen Kelly, Allan Mcneill, and Katie Overy. Crying, moving, and keeping it whole: What makes literary description vivid? *Poetics Today*, 31:433–463, 2010. doi: 10.2307/25835325.

Jaemin Jo, Bohyoung Kim, and Jinwook Seo. EyeBookmark: Assisting recovery from interruption during reading. In *Proceedings of the 2015 CHI Conference on Human Factors in Computing Systems*, CHI '15, pages 2963–2966. Association for Computing Machinery, 2015. ISBN 978-1-4503-

3145-6. doi: 10.1145/2702123.2702340.

Susan Joslyn and Earl Hunt. Evaluating individual differences in response to time-pressure situations. *Journal of Experimental Psychology: Applied*, 4:16, 1998. doi: 10.1037/1076-898X.4.1.16.

Justin A. Jungé, Timothy F. Brady, and Marvin M. Chun. The contents of perceptual hypotheses: Evidence from rapid resumption of interrupted visual search. *Attention, Perception, & Psychophysics*, 71:681–689, 2009. doi: 10.3758/APP.71.4.681.

Marcel A. Just and Patricia A. Carpenter. A theory of reading: from eye fixations to comprehension. *Psychological review*, 87:329, 1980. doi: 10. 1037/0033-295X.87.4.329.

Dina Kanaan and Nadine Marie Moacdieh. Eye tracking to evaluate the effects of interruptions and workload in a complex task. *Human Factors*, 64:1168–1180, 2021. ISSN 0018-7208. doi: 10.1177/0018720821990487.

Michael J. Kane, Zach Hambrick, Stephen Tuholski, Oliver Wilhelm, Tabitha Payne, and Randall Engle. The generality of working memory capacity: A latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology. General*, 133:189–217, 2004. doi: 10.1037/0096-3445.133.2.189.

Byeongwoo Kang. How the COVID-19 pandemic is reshaping the education service. *The Future of Service Post-COVID-19 Pandemic, Volume 1: Rapid Adoption of Digital Service Technology*, pages 15–36, 2021. doi: 10.1007/978-981-33-4126-5_2.

Alan Kennedy, Robin Hill, and Joël Pynte. The Dundee Corpus. In *Proceedings of the 12th European Conference on Eye Movement*, ECEM 2003, 2003.

Walter Kintsch. The role of knowledge in discourse comprehension: A construction-integration model. In G.E. Stelmach and P.A. Vroon, editors, *Text and Text Processing*, volume 79 of *Advances in Psychology*, pages 107–153. North-Holland, 1991. doi: 10.1016/S0166-4115(08)61551-4.

Walter Kintsch. *Comprehension: A paradigm for cognition*. Cambridge University Press, 1998.

Walter Kintsch and Teun A. van Dijk. Toward a model of text comprehension

and production. *Psychological Review*, 85:363–394, 1978. doi: 10.1037/ 0033-295X.85.5.363.

Sigrid Klerke, Yoav Goldberg, and Anders Søgaard. Improving sentence compression by learning to predict gaze. In Kevin Knight, Ani Nenkova, and Owen Rambow, editors, *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1528–1533, San Diego, California, 2016. Association for Computational Linguistics. doi: 10.18653/v1/N16-1179. URL https://aclanthology.org/N16-1179.

Reinhold Kliegl, Ellen Grabner, Martin Rolfs, and Ralf Engbert. Length, frequency, and predictability effects of words on eye movements in reading. *European Journal of Cognitive Psychology - EUR J COGN PSYCHOL*, 16 (1-2):262–284, 2004. doi: 10.1080/09541440340000213.

Terry K. Koo and Mae Y. Li. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, 15:155–63, 2016. doi: 10.1016/j.jcm.2016.02.012.

Iuliia Kotseruba, Oscar J. Avella Gonzalez, and John K. Tsotsos. A review of 40 years of cognitive architecture research: Focus on perception, attention, learning and applications. *arXiv preprint arXiv:1610.08602*, pages 1–74, 2016. doi: 10.48550/arXiv.1610.08602.

Izabela Krejtz, Agnieszka Szarkowska, and Maria Łogińska. Reading function and content words in subtitled videos. *The Journal of Deaf Studies and Deaf Education*, 21:222–232, 2016. ISSN 1081-4159. doi: 10.1093/ deafed/env061. URL https://www.ncbi.nlm.nih.gov/pmc/articles/ PMC4886319/.

Moniek M. Kuijpers, Frank Hakemulder, E. S. Tan, and Miruna M. Doicaru. Exploring absorbing reading experiences: Developing and validating a self-report scale to measure story world absorption. *Scientific Study of Literature*, 4(1):89–122, 2014. doi: 10.1075/ssol.4.1.05kui.

Moniek M. Kuijpers, Shawn Douglas, and Don Kuiken. Capturing the Ways We Read. Introducing the Reading Habits Questionnaire. *Anglistik*, 31(1): 53–69, 2020. doi: 10.33675/ANGL/2020/1/6.

Karin Kukkonen. *Probability Designs: Literature and Predictive Processing.*

Oxford University Press, New York, 03 2020. ISBN 9780190050955. doi: 10.1093/oso/9780190050955.001.0001.

Karin Kukkonen and Sibylle Baumbach. Mind-wandering and attention in literature. *Diegesis. Interdisciplinary E-Journal for Narrative Research/Interdisziplinäres E-Journal für Erzählforschung*, 11(2):1–17, 2022. URL https://www.diegesis.uni-wuppertal.de/index.php/diegesis/article/download/456/628.

Leah Kumle, Melissa L.-H. Võ, and Dejan Draschkow. Estimating power in (generalized) linear mixed models: An open introduction and tutorial in R. *Behavior Research Methods*, 53:1–16, 05 2021. doi: 10.3758/s13428-021-01546-0.

Kai Kunze, Yuzuko Utsumi, Yuki Shiga, Koichi Kise, and Andreas Bulling. I know what you are reading: recognition of document types using mobile eye tracking. In *Proceedings of the 2013 International Symposium on Wearable Computers*, ISWC '13, page 113–116, New York, NY, USA, 2013. Association for Computing Machinery. ISBN 9781450321273. doi: 10.1145/2493988.2494354.

Victor Kuperman, Noam Siegelman, Sascha Schroeder, Cengiz Acartürk, Svetlana Alexeeva, Simona Amenta, Raymond Bertram, Rolando Bonandrini, Marc Brysbaert, Daria Chernova, et al. Text reading in English as a second language: Evidence from the Multilingual Eye-Movements Corpus. *Studies in Second Language Acquisition*, 45(1):3–37, 2023. doi: 10.1017/S0272263121000954.

Victor Kuperman, Sascha Schroeder, and Daniil Gnetov. Word length and frequency effects on text reading are highly similar in 12 alphabetic languages. *Journal of Memory and Language*, 135:104497, 2024. ISSN 0749-596X. doi: 10.1016/j.jml.2023.104497.

Anežka Kuzmičová. Presence in the reading of literary narrative: A case for motor enactment. *Semiotica*, 2012(189):23–48, 2012. doi: doi:10.1515/semi.2011.071.

Anežka Kuzmičová. *Mental imagery in the experience of literary narrative: Views from embodied cognition*. PhD thesis, Stockholm University, 2013.

Anežka Kuzmičová. Literary narrative and mental imagery: A view from

embodied cognition. *Style*, 48(3):275–293, 2014. doi: 10.5325/style.48.3. 275.

Alexandra Kuznetsova, Per B. Brockhoff, and Rune H. B. Christensen. lmertest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13):1–26, 2017. doi: 10.18637/jss.v082.i13.

Katherine Labonté and François Vachon. Resuming a dynamic task following increasingly long interruptions: The role of working memory and reconstruction. *Frontiers in Psychology*, 12:659451, 2021. doi: 10.3389/fpsyg. 2021.659451.

Vicky Tzuyin Lai, Roel M. Willems, and Peter Hagoort. Feel between the lines: Implied emotion in sentence comprehension. *Journal of Cognitive Neuroscience*, 27(8):1528–1541, 2015. ISSN 0898-929X. doi: 10.1162/ jocn_a_00798.

Susan S Lanser. *The narrative act: Point of view in prose fiction*. Princeton University Press, Princeton, New Jersey, 1981.

Alyssa P. Lawson and Richard E. Mayer. Role of individual differences in executive function for learning from distracting multimedia lessons. *Journal of Educational Computing Research*, 62(3):756–784, 2024. doi: 10.1177/07356331231215752.

Claudia Leopold. The imagination principle in multimedia learning. In Richard E. Mayer and LoganEditors Fiorella, editors, *The Cambridge Handbook of Multimedia Learning*, Cambridge Handbooks in Psychology, pages 370–380. Cambridge University Press, 2021.

Claudia Leopold and Richard Mayer. An imagination effect in learning from scientific text. *Journal of Educational Psychology*, 107:47–63, 2015. doi: 10.1037/a0037142.

Yulan Liang, Michelle L. Reyes, and John D. Lee. Real-time detection of driver cognitive distraction using support vector machines. *IEEE Transactions on Intelligent Transportation Systems*, 8(2):340–350, 2007. doi: 10.1109/TITS.2007.895298.

Yinan Liu, Bong Gee Jang, and Zaline Roy-Campbell. Optimum input mode in the modality and redundancy principles for university ESL students' multimedia learning. *Computers & Education*, 127:190–200, 2018. doi:

Bibliography

10.1016/j.compedu.2018.08.025.

Ziming Liu. Reading in the age of digital distraction. *Journal of Documentation*, 78(6):1201–1212, 2022. doi: 10.1108/JD-07-2021-0130.

Alejandro Lleras, Ronald A. Rensink, and James T. Enns. Rapid resumption of interrupted visual search: New insights on the interaction between vision and memory. *Psychological Science*, 16:684–688, 2005. doi: 10.1111/j. 1467-9280.2005.01596.x.

Robert Logie. The seven ages of working memory. In J. T. E. Richardson, R W Engle, L Hasher, R. H. Logie, E. R. Stoltzfus, and R. T. Zacks, editors, *Working Memory and Human Cognition*, pages 31–59. Oxford University Press, 1996. ISBN 9780195101003.

Cristina Loi, Moniek Kuijpers, Astrid Ensslin, and Gerhard Lauer. Paths to transformation across contemporary reading practices: The role of motivations and genre preferences. *Psychology of Aesthetics, Creativity, and the Arts*, 2023. doi: 10.1037/aca0000622.

Robert F. Lorch. Integration of topic and subordinate information during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19(5):1071–1081, 1993.

Jana Lüdtke, Burkhard Meyer-Sickendieck, and Arthur M. Jacobs. Immersing in the stillness of an early morning: testing the mood empathy hypothesis of poetry reception. *Psychology of Aesthetics, Creativity, and the Arts*, 8(3):363–377, 2014. doi: 10.1037/a0036826.

Steven Luke and Kiel Christianson. The Provo Corpus: A large eye-tracking corpus with predictability norms. *Behavior Research Methods*, 50:826–833, 2018. doi: 10.3758/s13428-017-0908-4.

Lilla Magyari, Anne Mangen, Anežka Kuzmičová, Arthur Jacobs, and Jana Lüdtke. Eye movements and mental imagery during reading of literary texts in different narrative styles. *Journal of Eye Movement Research*, 13: 3, 2020. doi: 10.16910/jemr.13.3.3.

Asifa Majid, Melissa Bowerman, Sotaro Kita, Daniel B.M. Haun, and Stephen C. Levinson. Can language restructure cognition? the case for space. *Trends in Cognitive Sciences*, 8(3):108–114, 2004. ISSN 1364-6613. doi: https://doi.org/10.1016/j.tics.2004.01.003.

Marloes Mak and Roel M. Willems. Mental simulation during literary reading: Individual differences revealed with eye-tracking. *Language, Cognition and Neuroscience*, 34(4):511–535, 2019. doi: 10.1080/23273798.2018. 1552007.

Marloes Mak and Roel M. Willems. Mental simulation during literary reading. In Donald Kuiken and Arthur M. Jacobs, editors, *Handbook of Empirical Literary Studies*, pages 63–84. De Gruyter, Berlin, Boston, 2021. ISBN 9783110645958. doi: 10.1515/9783110645958-004.

Marloes Mak, Myrthe Faber, and Roel M. Willems. Different kinds of simulation during literary reading: Insights from a combined fmri and eye-tracking study. *Cortex*, 162:115–135, 2023. ISSN 0010-9452. doi: 10.1016/j.cortex.2023.01.014.

Jonathan Malmaud, Roger Levy, and Yevgeni Berzak. Bridging information-seeking human gaze and machine reading comprehension. In *Proceedings of the 24th Conference on Computational Natural Language Learning*, pages 142–152, 2020. doi: 10.18653/v1/2020.conll-1.11.

Thomas Mann. Buddenbrooks. In Gerhard Köpf, editor, *ICD-10 Literarisch*, pages 255–263. VS Verlag für Sozialwissenschaften, Wiesbaden, 2006. ISBN 978-3-8350-9667-7. doi: 10.1007/978-3-8350-9667-7_20.

Raymond Mar, Jingyuan Li, Anh Nguyen, and Cindy Ta. Memory and comprehension of narrative versus expository texts: A meta-analysis. *Psychonomic Bulletin & Review*, 28:732–749, 2021. doi: 10.3758/ s13423-020-01853-1.

Alexander Mariakakis, Mayank Goel, Md Tanvir Islam Aumi, Shwetak N. Patel, and Jacob O. Wobbrock. Switchback: Using focus and saccade tracking to guide users' attention for mobile task resumption. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, page 2953–2962, New York, NY, USA, 2015. Association for Computing Machinery. ISBN 9781450331456. doi: 10.1145/2702123.2702539.

Barbara F. Marino, Anna M. Borghi, Giovanni Buccino, and Lucia Riggio. Chained activation of the motor system during language understanding. *Frontiers in Psychology*, 8, 2017. ISSN 1664-1078. doi: 10.3389/fpsyg. 2017.00199.

Bibliography

James H. Mathewson. Visual-spatial thinking: An aspect of science overlooked by educators. *Science Education*, 83(1):33–54, 1999. doi: 10.1002/(SICI)1098-237X(199901)83:1<33::AID-SCE2>3.0.CO;2-Z.

Sandeep Mathias, Diptesh Kanojia, Abhijit Mishra, and Pushpak Bhattacharya. A survey on using gaze behaviour for natural language processing. In Christian Bessiere, editor, *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI-20*, pages 4907–4913. International Joint Conferences on Artificial Intelligence Organization, 2020. doi: 10.24963/ijcai.2020/683.

Patricia D. Mautone and Richard E. Mayer. Signaling as a cognitive guide in multimedia learning. *Journal of Educational Psychology*, 93:377–389, 2001. doi: 10.1037/0022-0663.93.2.377.

Richard E. Mayer. Multimedia learning: Are we asking the right questions? *Educational Psychologist*, 32:1–19, 1997. doi: 10.1207/S15326985EP3201_1.

Richard E. Mayer. Cognitive theory of multimedia learning. In Richard E. Mayer, editor, *The Cambridge Handbook of Multimedia Learning*, Cambridge Handbooks in Psychology, chapter 3, pages 31–48. Cambridge University Press, 2005. doi: 10.1017/CBO9780511816819.004.

Richard E. Mayer. *Multimedia Learning*. Cambridge University Press, 2 edition, 2009.

Richard E. Mayer and Logan Fiorella. 12 principles for reducing extraneous processing in multimedia learning: Coherence, signaling, redundancy, spatial contiguity, and temporal contiguity principles. In Richard E. Mayer, editor, *The Cambridge Handbook of Multimedia Learning*, volume 279, pages 279–315. Cambridge University Press New York, NY, 2014. doi: 10.1017/CBO9781139547369.015.

Richard E. Mayer and Roxana Moreno. Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist*, 38(1):43–52, 2003. doi: 10.1207/S15326985EP3801_6.

Richard E. Mayer and Celeste Pilegard. Principles for managing essential processing in multimedia learning: Segmenting, pretraining, and modality principles. In Richard E. Mayer, editor, *The Cambridge Handbook of Mul-*

timedia Learning, pages 169–182. Cambridge University Press New York, NY, 2014. doi: 10.1017/CBO9781139547369.016.

Kathryn McCarthy and Danielle S. McNamara. The multidimensional knowledge in text comprehension framework. *Educational Psychologist*, 56:1–19, 2021. doi: 10.1080/00461520.2021.1872379.

Kathryn McCarthy, Tricia Guerrero, Kevin Kent, Laura Allen, Danielle S. McNamara, Szu-Fu Chao, Jonathan Steinberg, Tenaha O'Reilly, and John Sabatini. Comprehension in a scenario-based assessment: Domain and topic-specific background knowledge. *Discourse Processes*, 55:1–15, 2018. doi: 10.1080/0163853X.2018.1460159.

Danielle S. McNamara and W. Kintsch. Working memory in text comprehension: Interrupting difficult text. In *Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society*, pages 1–20. Routledge, 1996.

Danielle S. McNamara, Yasuhiro Ozuru, and Randy G. Floyd. Comprehension challenges in the fourth grade: The roles of text cohesion, text genre, and readers' prior knowledge. *International Electronic Journal of Elementary Education*, 4(1):229–257, 2011. ISSN 1307-9298. URL https://www.iejee.com/index.php/IEJEE/article/view/222.

Hilkka L. Meys and Penelope M. Sanderson. The effect of individual differences on how people handle interruptions. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 57:868–872, 2013. ISSN 2169-5067. doi: 10.1177/1541931213571188.

Diane Mézière, Lili Yu, Erik D. Reichle, Titus von der Malsburg, and Genevieve M McArthur. Using eye-tracking measures to predict reading comprehension. *Reading Research Quarterly*, 58:425–449, 2023. doi: 10.1002/rrq.498.

Caitlin Mills, Julie Gregg, Robert Bixler, and Sidney D'Mello. Eye-Mind reader: an intelligent reading interface that promotes long-term comprehension by detecting and responding to mind wandering. *Human–Computer Interaction*, 36:306–332, 2021. doi: 10.1080/07370024.2020.1716762.

Abhijit Mishra, Diptesh Kanojia, and Pushpak Bhattacharyya. Predicting readers' sarcasm understandability by modeling gaze behavior. In *Proceed-*

ings of the 2016 AAAI Conference on Artificial Intelligence*, AAAI '16', pages 3747—-3753. AAAI Press, 2016. doi: 10.5555/3016387.3016431.

Christopher Monk, J. Gregory Trafton, and Deborah Boehm-Davis. The effect of interruption duration and demand on resuming suspended goals. *Journal of Experimental Psychology. Applied*, 14:299–313, 2008. doi: 10. 1037/a0014402.

Roxana Moreno and Richard E. Mayer. Engaging students in active learning: The case for personalized multimedia messages. *Journal of Educational Psychology*, 92(4):724–733, 2000. doi: 10.1037/0022-0663.92.4.724.

Roxana Moreno and Richard E. Mayer. Interactive multimodal learning environments. *Educ Psychol Rev*, 19:309–326, 2007. doi: 10.1007/ s10648-007-9047-2.

Mary Muir and Cristina Conati. An analysis of attention to student – adaptive hints in an educational game. In Stefano A. Cerri, William J. Clancey, Giorgos Papadourakis, and Kitty Panourgia, editors, *Intelligent Tutoring Systems*, pages 112–122, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg. ISBN 978-3-642-30950-2. doi: 10.1007/978-3-642-30950-2_15.

Jan Mukařovský. Standard language and poetic language. In Paul L. Garvin, editor, *A Prague School Reader on Esthetics, Literary Structure, and Style*, pages 17–30. Georgetown University Press, Washington, D.C., 1964. URL `https://hdl.handle.net/11222.digilib/131565`.

Duygu Mutlu-Bayraktar, Veysel Cosgun, and Tugba Altan. Cognitive load in multimedia learning environments: A systematic review. *Computers & Education*, 141:103618, 2019. ISSN 0360-1315. doi: https://doi.org/10. 1016/j.compedu.2019.103618.

Keith Oatley. A taxonomy of literary response and a theory of identification in fictional narrative. *Poetics*, 23:53–74, 01 1995. doi: 10.1016/ 0304-422X(94)P4296-S.

Henri Olkoniemi, Eerika Johander, and Johanna K. Kaakinen. The role of look-backs in the processing of written sarcasm. *Memory & Cognition*, 47: 87–105, 2018. doi: 10.3758/s13421-018-0852-2.

Henri Olkoniemi, Sohvi Halonen, Penny M. Pexman, and Tuomo Häikiö. Children's processing of written irony: An eye-tracking study. *Cogni-*

tion, 238:105508, 2023. ISSN 0010-0277. doi: https://doi.org/10.1016/j. cognition.2023.105508.

OpenAI. Chatgpt (version 3.5). [Large language model], 2023. URL `https: //chat.openai.com/chat`.

Antti Oulasvirta and Pertti Saariluoma. Surviving task interruptions: Investigating the implications of long-term working memory theory. *International Journal of Human-Computer Studies*, 64:941–961, 2006. ISSN 1071-5819. doi: 10.1016/j.ijhcs.2006.04.006.

Allan Paivio. *Mental Representations: A dual coding approach*. Oxford University Press, 1990. ISBN 9780195066661. doi: 10.1093/acprof: oso/9780195066661.001.0001.

Jinger Pan, Ming Yan, Eike M. Richter, Hua Shu, and Reinhold Kliegl. The Beijing Sentence Corpus: A Chinese sentence corpus with eye movement data and predictability norms. *Behavior Research Methods*, 54:1989–2000, 2022. doi: 10.3758/s13428-021-01730-2.

Scott G. Paris and Peter Winograd. How metacognition can promote academic learning and instruction. In *Dimensions of Thinking and Cognitive Instruction*, pages 15–51. Routledge, 1 edition, 2013. doi: 10.4324/ 9780203771686.

Harold Pashler, Sean H. K. Kang, and Renita Y. Ip. Does multitasking impair studying? depends on timing. *Applied Cognitive Psychology*, 27: 593–599, 2013. ISSN 1099-0720. doi: 10.1002/acp.2919.

Diane Pecher and Rolf Zwann. *Grounding Cognition: The Role of Perception and Action in Memory, Language, and Thinking*. Cambridge University Press, 2005. ISBN 9780521834643. doi: 10.1017/CBO9780511499968.

Jonathan Peirce, Jeremy R. Gray, Sol Simpson, Michael MacAskill, Richard Höchenberger, Hiroyuki Sogo, Erik Kastman, and Jonas Kristoffer Lindeløv. PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, 51:195–203, 2019. ISSN 1554-3528. doi: 10.3758/ s13428-018-01193-y.

Leslie Gross Portney and Mary P. Watkins. *Foundations of clinical research: applications to practice*, volume 892. Pearson/Prentice Hall Upper Saddle River, NJ, 2009.

Bibliography

R Core Team. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria, 2021. URL `https://www.R-project.org/`.

Raj M. Ratwani and J. Gregory Trafton. Spatial memory guides task resumption. *Visual Cognition*, 16:1001–1010, 2008. ISSN 1350-6285. doi: 10.1080/13506280802025791.

Keith Rayner. Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124:372–422, 1998. doi: 10.1037/0033-2909.124.3.372.

Keith Rayner and George W. McConkie. What guides a reader's eye movements? *Vision Research*, 16:829–837, 1976. ISSN 0042-6989. doi: 10.1016/0042-6989(76)90143-7.

Keith Rayner and Erik D. Reichle. Models of the reading process. *WIREs Cognitive Science*, 1(6):787–799, 2010. doi: 10.1002/wcs.68.

Keith Rayner, Sara C Sereno, Robin K Morris, A Rene Schmauder, and Charles Clifton Jr. Eye movements and on-line language comprehension processes. *Language and Cognitive Processes*, 4:SI21–SI49, 1989. doi: 10.1080/01690968908406362.

Keith Rayner, Kathryn Chace, Timothy Slattery, and Jane Ashby. Eye movements as reflections of comprehension processes in reading. *Scientific Studies of Reading*, 10:241–255, 2006. doi: 10.1207/s1532799xssr1003_3.

Keith Rayner, Timothy Slattery, and Nathalie Bélanger. Eye movements, the perceptual span, and reading speed. *Psychonomic Bulletin & Review*, 17:834–839, 12 2010. doi: 10.3758/PBR.17.6.834.

Thomas S. Redick, James M. Broadway, Matt E. Meier, Princy S. Kuriakose, Nash Unsworth, Michael J. Kane, and Randall W. Engle. Measuring working memory capacity with automated complex span tasks. *European Journal of Psychological Assessment*, 28:164–171, 2012. ISSN 2151-2426. doi: 10.1027/1015-5759/a000123.

Erik D Reichle, Alexander Pollatsek, Donald L Fisher, and Keith Rayner. Toward a model of eye movement control in reading. *Psychological Review*, 105:125, 1998. doi: 10.1037/0033-295x.105.1.125.

Günter Daniel Rey. A review of research and a meta-analysis of the seductive

detail effect. *Educational Research Review*, 7(3):216–237, 2012. ISSN 1747-938X. doi: 10.1016/j.edurev.2012.05.003.

Ralph E. Reynolds and Richard C. Anderson. Influence of questions on the allocation of attention during reading. *Journal of Educational Psychology*, 74(5):623–632, 1982. doi: 10.1037/0022-0663.74.5.623.

Lauren Richmond, Lois Burnett, Alexandra Morrison, and Hunter Ball. Performance on the processing portion of complex working memory span tasks is related to working memory capacity estimates. *Behavior Research Methods*, 54:780–794, 2022. doi: 10.3758/s13428-021-01645-y.

Juliane Richter, Katharina Scheiter, and Alexander Eitel. Signaling text-picture relations in multimedia learning: A comprehensive meta-analysis. *Educational Research Review*, 17:19–36, 2016. ISSN 1747-938X. doi: 10.1016/j.edurev.2015.12.003.

Rui Rodrigues and Margarida Silva. Emotional design in multimedia learning: Systematic review. In Nuno Martins and Daniel Brandão, editors, *Advances in Design and Digital Communication II*, pages 223–234, Cham, 2022. Springer International Publishing. ISBN 978-3-030-89735-2. doi: 10.1007/978-3-030-89735-2_19.

Patrick Royston, Douglas G. Altman, and Willi Sauerbrei. Dichotomizing continuous predictors in multiple regression: a bad idea. *Statistics in Medicine*, 25(1):127–141, 2006. doi: 10.1002/sim.2331.

Rufat Rzayev, Paweł W Woźniak, Tilman Dingler, and Niels Henze. Reading on smart glasses: The effect of text position, presentation type and walking. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, pages 1–9, 2018. doi: 10.1145/3173574.3173619.

Charles J. Rzepka. *Detective Fiction*, volume 6. Polity Press, Cambridge, UK, 2005.

Mark Sadoski. An exploratory study of the relationships between reported imagery and the comprehension and recall of a story. *Reading Research Quarterly*, 19(1):110–123, 1983. URL http://www.jstor.com/stable/747341.

Mark Sadoski, Ernest T. Goetz, Jr. Arturo Olivarez, Sharon Lee, and Nancy M. Roberts. Imagination in story reading: The role of imagery,

Bibliography

verbal recall, story analysis, and processing levels. *Journal of Reading Behavior*, 22(1):55–70, 1990. doi: 10.1080/10862969009547694.

Molood Sadat Safavi, Samar Husain, and Shravan Vasishth. Dependency resolution difficulty increases with distance in persian separable complex predicates: Evidence for expectation and memory-based accounts. *Frontiers in Psychology*, 7:403, 2016. doi: 10.3389/fpsyg.2016.00403.

Alison J. S. Sanford, Anthony J. Sanford, Jo Molle, and Catherine Emmott. Shallow processing and attention capture in written and spoken discourse. *Discourse Processes*, 42:109–130, 2006. ISSN 0163-853X. doi: 10.1207/s15326950dp4202_2.

Anthony J. Sanford and Catherine Emmott. *Attention in text: foregrounding and rhetorical focussing*, pages 72–102. Cambridge University Press, 2012. doi: 10.1017/CBO9781139084321.005.

Elaine Scarry. *Dreaming by the Book*. Farrar, Straus and Giroux, New Work, 1999.

Hildur Schilling, Keith Rayner, and James Chumbley. Comparing naming, lexical decision, and eye fixation times: Word frequency effects and individual differences. *Memory & Cognition*, 26:1270–1281, 12 1998. doi: 10.3758/BF03201199.

Darryl W. Schneider and Peter Dixon. Visuospatial cues for reinstating mental models in working memory during interrupted reading. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, 63(3):161–172, 2009. doi: 10.1037/a0014867.

Wolfgang Schnotz. Towards an integrated view of learning from text and visual displays. *Educational Psychology Review*, 14:101–120, 2002. doi: 10.1023/A:1013136727916.

Teresa Schurer, Bertram Opitz, and Torsten Schubert. Working memory capacity but not prior knowledge impact on readers' attention and text comprehension. *Frontiers in Education*, 5, 2020. ISSN 2504-284X. doi: 10.3389/feduc.2020.00026.

Thomas L. Seamster, Richard E. Redding, John R. Cannon, Joan M. Ryder, and Janine A. Purcell. Cognitive task analysis of expertise in air traffic control. *The International Journal of Aviation Psychology*, 3(4):257–283,

1993. doi: 10.1207/s15327108ijap0304_2.

Irina Sekerina, Anna Laurinavichyute, Svetlana Alexeeva, Kristina Bag-
dasaryan, and Reinhold Kliegl. Russian sentence corpus: Benchmark mea-
sures of eye movements in reading in Russian. *Behavior Research Methods*,
51:1161–1178, 06 2019a. doi: 10.3758/s13428-018-1051-6.

Irina Sekerina, Anna Laurinavichyute, Svetlana Alexeeva, Kristina Bag-
dasaryan, and Reinhold Kliegl. Russian sentence corpus: Benchmark mea-
sures of eye movements in reading in Russian. *Behavior Research Methods*,
51:1161–1178, 2019b. doi: 10.3758/s13428-018-1051-6.

Sara C. Sereno and Keith Rayner. Measuring word recognition in reading:
eye movements and event-related potentials. *Trends in Cognitive Sciences*,
7(11):489–493, 2003. ISSN 1364-6613. doi: 10.1016/j.tics.2003.09.010.

Priti Shah and Akira Miyake. The separability of working memory resources
for spatial thinking and language processing: An individual differences
approach. *Journal of Experimental Psychology. General*, 125:4–27, 1996.
doi: 10.1037//0096-3445.125.1.4.

Selina Sharmin, Oleg Špakov, and Kari-Jouko Räihä. Reading on-screen
text with gaze-based auto-scrolling. In *Proceedings of the 2013 Con-
ference on Eye Tracking South Africa*, ETSA '13, pages 24–31. Asso-
ciation for Computing Machinery, 2013. ISBN 978-1-4503-2110-5. doi:
10.1145/2509315.2509319.

Y. Jeremy Shen and Yuhong V. Jiang. Interrupted visual searches reveal
volatile search memory. *Journal of Experimental Psychology: Human Per-
ception and Performance*, 32:1208–20, 10 2006. doi: 10.1037/0096-1523.
32.5.1208.

John L. Sibert, Mehmet Gokturk, and Robert A. Lavine. The reading assis-
tant: Eye gaze triggered auditory prompting for reading remediation. In
*Proceedings of the 13th Annual ACM Symposium on User Interface Soft-
ware and Technology*, UIST '00, pages 101–107. Association for Computing
Machinery, 2000. ISBN 1581132123. doi: 10.1145/354401.354418.

Noam Siegelman, Sascha Schroeder, Cengiz Acartürk, Hee-Don Ahn, Svet-
lana Alexeeva, Simona Amenta, Raymond Bertram, Rolando Bonan-
drini, Marc Brysbaert, Daria Chernova, et al. Expanding horizons of

cross-linguistic research on reading: The multilingual eye-movement corpus (MECO). *Behavior Research Methods*, 54(6):2843–2863, 2022. doi: 10.3758/s13428-021-01772-6.

Reid Smith, Pamela Snow, Tanya Serry, and Lorraine Hammond. The role of background knowledge in reading comprehension: A critical review. *Reading Psychology*, 42:214–240, 2021. ISSN 0270-2711, 1521-0685. doi: 10.1080/02702711.2021.1888348.

Ekta Sood, Simon Tannert, Diego Frassinelli, Andreas Bulling, and Ngoc Thang Vu. Interpreting attention models with human visual attention in machine reading comprehension. In *Proc. ACL SIGNLL Conference on Computational Natural Language Learning*, CoNLL, pages 12–25. Association for Computational Linguistics, 2020a. doi: 10.18653/v1/P17.

Ekta Sood, Simon Tannert, Philipp Mueller, and Andreas Bulling. Improving natural language processing tasks with human gaze-guided neural attention. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 6327–6341. Curran Associates, Inc., 2020b. URL https://proceedings.neurips.cc/paper_files/paper/2020/file/460191c72f67e90150a093b4585e7eb4-Paper.pdf.

Ekta Sood, Fabian Kögel, Florian Strohm, Prajit Dhar, and Andreas Bulling. VQA-MHUG: A gaze dataset to study multimodal neural attention in VQA. In *Proc. ACL SIGNLL Conference on Computational Natural Language Learning*, CoNLL, pages 27–43. Association for Computational Linguistics, 2021. doi: 10.18653/v1/2021.conll-1.3.

Rosy Southwell, Caitlin Mills, Megan Caruso, and Sidney D'Mello. Gaze-based predictive models of deep reading comprehension. *User Modeling and User-Adapted Interaction*, 33(3):687–725, 2023. doi: 10.1007/s11257-022-09346-7.

Namrata Srivastava, Rajiv Jain, Jennifer Healey, Zoya Bylinskii, and Tilman Dingler. Mitigating the effects of reading interruptions by providing reviews and previews. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, number 229 in CHI EA '21, pages 1–6. Association for Computing Machinery, 2021. ISBN 978-1-4503-8095-9.

doi: 10.1145/3411763.3451610.

Julie Dangremond Stanton, Amanda J. Sebesta, and John Dunlosky. Fostering metacognition to support student learning and performance. *CBE–Life Sciences Education*, 20(2):1–7, 2021. doi: 10.1187/cbe.20-12-0289.

Ben Steichen, Michael M. A. Wu, Dereck Toker, Cristina Conati, and Giuseppe Carenini. Te,te,hi,hi: Eye gaze sequence analysis for informing user-adaptive information visualizations. In Vania Dimitrova, Tsvi Kuflik, David Chin, Francesco Ricci, Peter Dolog, and Geert-Jan Houben, editors, *User Modeling, Adaptation, and Personalization*, pages 183–194, Cham, 2014. Springer International Publishing. ISBN 978-3-319-08786-3. doi: 10.1007/978-3-319-08786-3_16.

Nancy Stein and Christine Glenn. An analysis of story comprehension in elementary school children: A test of a schema. In Roy O. Freedle, editor, *New Directions in Discourse Processing*, volume 2, pages 53–120. Ablex, 1979.

Student. The probable error of a mean. *Biometrika*, 6(1):1–25, 1908. ISSN 00063444, 14643510. URL `http://www.jstor.org/stable/2331554`.

John Sweller. Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2):257–285, 1988. doi: 10.1207/s15516709cog1202_4.

John Sweller. Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction*, 4(4):295–312, 1994. ISSN 0959-4752. doi: 10.1016/0959-4752(94)90003-5.

John Sweller. Evolution of human cognitive architecture. In B. Ross, editor, *The psychology of learning and motivationAcademic Press*, pages 215–266. Academic Press, San Diego, 2003.

John Sweller. Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational Psychology Review*, 22(2):123–138, 2010. doi: 10.1007/s10648-010-9128-5.

Michelle Taub and Roger Azevedo. How does prior knowledge influence eye fixations and sequences of cognitive and metacognitive srl processes during learning with an intelligent tutoring system? *International Journal of Artificial Intelligence in Education*, 29:1–28, 2019. doi: 10.1007/

Bibliography

s40593-018-0165-4.

The College Board. Official sat practice test. Copyright ©, 2024. URL `https://collegereadiness.collegeboard.org`. [Accessed: 2024-11-30].

J. Gregory Trafton and Christopher A. Monk. Task interruptions. *Reviews of Human Factors and Ergonomics*, 3(1):111–126, 2007. doi: 10.1518/155723408X299852.

J. Gregory Trafton, Erik M. Altmann, Derek P. Brock, and Farilee E. Mintz. Preparing to resume an interrupted task: Effects of prospective goal encoding and retrospective rehearsal. *International Journal of Human-Computer Studies*, 58:583–603, 2003. ISSN 1095-9300. doi: 10.1016/S1071-5819(03)00023-5.

J. Gregory Trafton, Erik M. Altmann, and Derek Brock. Huh, what was i doing? how people use environmental cues after an interruption. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 49: 468–472, 2005. doi: 10.1177/154193120504900354.

Phuoc Tran, Rogelio Carrillo, and Kaveri Subrahmanyam. Effects of online multitasking on reading comprehension of expository text. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, 7:1–18, 2013. ISSN 1802-7962. doi: 10.5817/CP2013-3-2.

Eunjoon Um, Jan L. Plass, Elizabeth O. Hayward, and Bruce D. Homer. Emotional design in multimedia learning. *Journal of Educational Psychology*, 104(2):485, 2011. doi: 10.1037/a0026609.

Nash Unsworth, Richard Heitz, Chad Schrock, and Randall Engle. An automated version of the operation span task. *Behavior Research Methods*, 37:498–505, 2005. doi: 10.3758/BF03192720.

Nash Unsworth, Thomas S. Redick, Richard P. Heitz, James M. Broadway, and Randall W. Engle. Complex working memory span tasks and higher-order cognition: A latent-variable analysis of the relationship between processing and storage. *Memory*, 17:635–654, 2009. doi: 10.1080/09658210902998047.

Inge van de Ven. Attentional modulation in literary reading: A theoretical-empirical framework. *Orbis Litterarum*, 79:1–18, 11 2023. doi: 10.1111/oli.12431.

Martial Van der Linden. The relationships between working memory and long-term memory. *Comptes Rendus de l'Académie des Sciences. Série III, Sciences de la Vie*, 321:175–177, 02 1998. doi: 10.1016/S0764-4469(97)89818-6.

Teun Adrianus Van Dijk and Walter Kintsch. *Strategies of Discourse Comprehension*. Academic Press New York, 1983.

Guido Van Rossum and Fred L. Drake. *Python 3 Reference Manual*. CreateSpace, Scotts Valley, CA, 2009. ISBN 1441412697. doi: 10.5555/1593511.

Peter Vorderer, Werner Wirth, Feliz Gouveia, Timo Saari, Lutz Jäncke, Saskia Böcking, Holger Schramm, Andre Gysbers, Tilo Hartmann, Christoph Klimmt, Jari Laarni, Niklas Ravaja, Ana Sacau, Thomas Baumgartner, and Petra Jäncke. Mec spatial presence questionnaire (mec-spq, english and german version): Short documentation and instructions for application. *Report to the European Community, Project Presence: MEC (IST-2001-37661)*, 2004. doi: 10.13140/RG.2.2.26232.42249.

David Wechsler. Wechsler adult intelligence scale–fourth edition (wais-iv) [database record]. *APA PsycTests*, 2008. doi: 10.1037/t15169-000.

Nicole E. Werner, David M. Cades, Deborah A. Boehm-Davis, Jessica Chang, Hibah Khan, and Gia Thi. What Makes Us Resilient to Interruptions? Understanding the Role of Individual Differences in Resumption. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 55, pages 296–300, 2011. doi: 10.1177/1071181311551062.

Frank Wilcoxon. Individual comparisons by ranking methods. *Biometrics Bulletin*, 1(6):80–83, 1945. doi: 10.2307/3001968.

Maria Wirzberger and Nele Russwinkel. Modeling interruption and resumption in a smartphone task: An ACT-R approach. *i-com*, 14(2):147–154, 2015. ISSN 2196-6826. doi: 10.1515/icom-2015-0033.

Maria Wirzberger, Jelmer Borst, Josef Krems, and Günter Daniel Rey. Memory-related cognitive load effects in an interrupted learning task: A model-based explanation. *Trends in Neuroscience and Education*, 20:100139, 2020. doi: 10.1016/j.tine.2020.100139.

Michael B.W. Wolfe and Joshua M. Woodwyk. Processing and memory of information presented in narrative or expository texts. *British*

Bibliography

Journal of Educational Psychology, 80(3):341–362, 2010. doi: 10.1348/ 000709910X485700.

Mustafa Yildiz and Ezgi Çetinkaya. The relationship between good readers' attention, reading fluency and reading comprehension. *Universal Journal of Educational Research*, 5:366–371, 2017. doi: 10.13189/ujer.2017.050309.

Francesca Zermiani, Prajit Dhar, Ekta Sood, Fabian Kögel, Andreas Bulling, and Maria Wirzberger. InteRead: An Eye Tracking Dataset of Interrupted Reading. In Nicoletta Calzolari, Min-Yen Kan, Veronique Hoste, Alessandro Lenci, Sakriani Sakti, and Nianwen Xue, editors, *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 9154–9169, Torino, Italia, 2024a. ELRA and ICCL. URL `https://aclanthology.org/ 2024.lrec-main.802`.

Francesca Zermiani, Prajit Dhar, Florian Strohm, Sibylle Baumbach, Andreas Bulling, and Maria Wirzberger. Individual differences in visuospatial working memory capacity and prior knowledge during interrupted reading. *Frontiers in Cognition*, 3, 2024b. ISSN 2813-4532. doi: 10.3389/fcogn.2024.1434642.

George Kingsley Zipf. *The Psycho-Biology of Language: An Introduction to Dynamic Philology*. Houghton Mifflin, 1935.