What's On Your Mind

Mental Simulation and Aesthetic Appreciation

During Literary Reading

Funding body
This research was funded by the Netherlands Organization for Scientific Research (NWO), Grant Number Vidi-276-89-007, awarded to Roel Willems

International Max Planck Research School (IMPRS) for Language Sciences
The educational component of the doctoral training was provided by the International Max Planck Research School (IMPRS) for Language Sciences. The graduate school is a joint initiative between the Max Planck Institute for Psycholinguistics and two partner institutes at Radboud University – the Centre for Language Studies, and the Donders Institute for Brain, Cognition and Behaviour. The IMPRS curriculum, which is funded by the Max Planck Society for the Advancement of Science, ensures that each member receives interdisciplinary training in the language sciences and develops a well-rounded skill set in preparation for fulfilling careers in academia and beyond. More information can be found at www.mpi.nl/imprs

The MPI series in Psycholinguistics
Initiated in 1997, the MPI series in Psycholinguistics contains doctoral theses produced at the Max Planck Institute for Psycholinguistics. Since 2013, it includes theses produced by members of the IMPRS for Language Sciences. The current listing is available at www.mpi.nl/mpi-series

ISBN: 978-94-92910-44-8
Coverdesign and artwork: Interface Communicatie Ede, www.ifcommunicatie.nl
Layout: This thesis was typeset with LaTeX 2_ε. It is based on the template created by Merel Wolf with help of Greta Kaufeld
Printed and bound by Ipskamp Drukkers, Enschede

What’s On Your Mind

Mental Simulation and Aesthetic Appreciation During Literary Reading

Proefschrift ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus prof. dr. J.H.J.M. van Krieken,
volgens besluit van het college voor promoties
in het openbaar te verdedigen op

dinsdag 29 november 2022
om 10.30 uur precies

door

Hanna Maria Louise Mak
geboren op 10 september 1991
te Leiden

Promotor:

Prof. dr. J.M. Sanders

Copromotor:

Dr. R.M. Willems

Manuscriptcommissie:

Prof. dr. W.P.M.S. Spooren (voorzitter)

Prof. dr. H.H.J. Das

Prof. dr. F. Hakemulder (Universitetet i Stavanger, Noorwegen)

Dr. L.J. Speed

Dr. M.M. Kuijpers (Universität Basel, Zwitserland)

For Anne

Contents

1 | Introduction

Many readers experience reading as a "journey" into the world of a story, where they can create a mental image of the events happening in the story, and feel the emotions of the characters in the story. This dissertation describes research into the role of mental simulation during reading, and how it is related to other reading experiences, measuring it with a combination of research methods from various disciplines. The current chapter will introduce the topics of mental simulation, reading experiences, and individual differences therein, as well as two important techniques used in this dissertation: Eye tracking and fMRI.

This Chapter Is Partly Based on

Mak, Marloes, & Willems, Roel M. (2021). Mental simulation during literary reading. In D. Kuiken & A. Jacobs (Ed.), *Handbook of Empirical Literary Studies* (pp. 63-84). Berlin, Boston: De Gruyter. DOI: 10.1515/9783110645958-004

"Outside the building, she waited for her husband to open his umbrella and then took his arm. He kept clearing his throat, as he always did when he was upset. They reached the bus-stop shelter on the other side of the street and he closed his umbrella. A few feet away, under a swaying and dripping tree, a tiny unfledged bird was helplessly twitching in a puddle."
(Nabokov, Symbols and Signs)

These four sentences from *Symbols and Signs* by Vladimir Nabokov (2003) illustrate most of what will be discussed in this dissertation. This dissertation primarily focuses on mental simulation during literary reading, building on the definition given by Barsalou (2008, p. 618):

"Simulation is the re-enactment of perceptual, motor, and introspective states acquired during experience with the world, body, and mind."

In this definition of simulation, Barsalou points out that simulation is not one of a kind. There are different kinds of simulation, stemming from different kinds of daily life experiences: Perceptual simulation, motor simulation, and the simulation of introspective states (or mental events, as I shall call them in this dissertation).

In the excerpt from Symbols and Signs, there are passages that can invite these three kinds of simulation in readers. For example, the descriptions of *"a swaying and dripping tree"* and *"a tiny unfledged bird"* that was *"helplessly twitching in a puddle"* all invite perceptual simulation, whereas *"she waited for her husband to open his umbrella and then took his arm"*, *"He kept clearing his throat"*, *"They reached the bus-stop shelter"*, and *"he closed his umbrella"* all invite motor simulation. Finally, *"as he always did when he was upset"* invites the simulation of mental events (i.e., thoughts or emotions).

Importantly, the second part of the definition of simulation as given by Barsalou states that the different kinds of simulation have one thing in common: They are all grounded in experiences with the world, body, and mind. Therefore, simulation is not simply elicited by descriptions in the text. Rather, it is invited by these descriptions, but eventually arises from an interaction between the text and the reader (i.e., their daily life experiences and personality characteristics).

The above observations result in the questions I would like to address in this dissertation:

(1) Do people use mental imagery during story reading, (2) how does this relate to the way readers experience stories, and (3) how do people differ in these respects?

These questions will be addressed using a variety of techniques, being subjective reports of reading experiences, eye tracking and fMRI. In the remainder of this introductory chapter, I will highlight these questions, and define the constructs that will be focused on in this dissertation. I will first discuss mental imagery and mental simulation, then discuss what is meant by "story reading", and what kind of reading experiences are important in this context. Next, individual differences in these processes are discussed. At the end of this chapter, I will discuss how eye tracking and fMRI can be used to answer the research questions, and provide an outline for the remainder of this dissertation.

1.1. Do People Use Mental Imagery During Story Reading?

To introduce this first research question, it is important to acknowledge that this can be seen as a question with two parts. I first ask what *mental imagery* can look like during language processing in general, and later specify this as imagining *during story reading*. To give a full background on all aspects of this question, I will unpack all these aspects individually. In the first two subsections, the general background of mental imagery and mental simulation during language processing is discussed (section 1.1.1.), followed by a discussion of the differences between the two (section 1.1.2.). To give some more background on mental simulation during literary reading, prominent theories are discussed in the next subsection (section 1.1.3.). In the final subsection, I will shift gears and focus on the "story reading" part of the question, explaining what is meant by "story" and "reading" (section 1.1.4.).

1.1.1. Mental Simulation and Mental Imagery

In order to answer the question if people use mental imagery during story reading it is important to first answer the question what mental imagery during language processing looks like in general. This has been studied in many distinct subfields of the social sciences and the humanities, under the terms of *mental simulation* and *mental imagery* (see Box 1 for a more in-depth discussion of the historical background of this term).

Box 1. Historical Background of Mental Simulation and Mental Imagery

The term mental simulation has its origins in *Simulation Theory*, a theory in the philosophy of mind which describes how people understand the mental states of others (e.g., Goldman, 2006). According to Shanton and Goldman, in mental simulation "one mental event, state or process is the re-experience of another mental event, state, or process" (Shanton & Goldman, 2010, p.528). 'Re-experience' is key in this description. Simulation theorists posit that people understand each other by reenacting their thoughts or feelings.

Historically, simulation theory has been contrasted with *Theory Theory*, the position that people reason about others in a reflective, theory-based and non-simulative manner (see Stich & Nichols, 1992). A third position to consider is *Interaction Theory*, which posits that conscious reasoning is no prerequisite for the understanding of the mental states of others (in contrast to simulation theory and theory theory); instead, people instinctively understand others by (subconsciously) mapping non-verbal cues onto their own bodies. The debate between these three positions is beyond the scope of this dissertation (but see e.g., Gallagher, 2015). Nonetheless, it is necessary to point out that it might seem that 'mental simulation', the core theme of this chapter, is directly linked to simulation theory. This does not need to be the case. The process of mental simulation during reading could be compatible with interaction theory (or a hybrid of interaction and simulation theory) equally well. From such a pluralistic standpoint it follows that the processes described in this chapter are compatible with any of the theories of social cognition (e.g., Andrews, 2008; Gallagher, 2015; Wiltshire, Lobato, McConnell, & Fiore, 2015).

Regardless, the increased interest in mental simulation in a wide variety of fields (see below) has led to stronger interest in how simulation theory can be incorporated into the philosophical and psychological theories of cognition. An intuitive starting point for the topic of this dissertation is the observation that during literary reading most readers do not only engage with a narrative in a detached, 'theorizing' manner. Instead, they experience sensations ('pictures in the head'), report feelings for a character, or think along with a fictive character.

One prominent conceptual difficulty across studies lies in the distinction between *mental simulation* and *mental imagery*, and before moving on, the contrast between these two concepts should be considered. Mental simulation and mental imagery have been approached in different ways in two important areas of research. I will now discuss each in turn.

In the study of *mental simulation* during language comprehension, researchers have presented words or sentences related to the senses to participants and observed the effects on sensory perception. An example comes from Speed and Vigliocco Speed and Vigliocco (2014) who showed that listening to sentences describing slow movement (e.g., The lion ambled to the balloon) led to slower eye movements than listening to sentences describing fast motion (e.g., The lion dashed to the balloon). In a similar vein, hearing sentences that imply a certain shape or orientation of an object primes visual recognition of that object – but only if that object is presented in the implied shape or orientation (Stanfield & Zwaan, 2001; Zwaan, Stanfield, & Yaxley, 2002). Similarly, words implying a location on a vertical axis prime perception of objects appearing in this location (Estes, Verges, & Barsalou, 2008; Ostarek & Vigliocco, 2017). Finally, visual and motor regions of the brain tend to be activated when reading action-related or sensory words (e.g., Hauk, Johnsrude, & Pulvermüller, 2004; see Willems, Labruna, D'Esposito, Ivry, & Casasanto, 2011 for an overview; see Ostarek & Huettig, 2019 for a critical note on this research). The rationale behind these studies is to show that understanding language related to actions or to the senses leads to sensorimotor activations in the brain.

In the study of *mental imagery*, researchers have looked into the neurocognitive basis of the deliberate ('conscious') creation of mental images. An early driving force for this work was the so-called 'imagery debate' (Kosslyn, 1994; Pylyshyn, 2003). An important issue in that debate was whether primary sensory and motor regions are involved in imagery the same way they are during actual perception and motor actions. In this spirit, it was found that motor imagery elicited activation in the same brain areas as motor preparation, motor control and motor execution (De Lange, Roelofs, & Toni, 2008; Jeannerod, 1994, 2001; Lotze & Halsband, 2006; Parsons et al., 1995). Similarly, for perceptual imagery, there is overlap between brain areas involved in perceptual imagery and real perception (Dijkstra, Bosch, & Gerven, 2017; Kosslyn, Ganis, & Thompson, 2001).

As mental imagery and mental simulation have been studied in different ways and in different disciplines, I will now specifically elaborate on differences between the two concepts.

1.1.2. Differences Between Mental Simulation and Mental Imagery

A striking difference between mental simulation and mental imagery is the speed at which each occurs. While reading language, mental simulation can be very fast and relatively effortless. In contrast, during deliberate mental imagery, image generation takes much longer (seconds or more) and is subjectively effortful. It has even been found that information is lost when people try to bring subconscious (or preconscious) and automatic mental simulations to awareness (Connell & Lynott, 2016). This contrast between subconscious and effortless simulation and conscious and effortful imagery makes it more likely for simulation to occur during reading, compared to imagery. The conscious, effortful nature of mental imagery would disrupt the natural flow of the reading process too much, and therefore is unlikely to occur during natural reading.

Moreover, it may seem that mental simulation and mental imagery differ in degree: perhaps what happens during reading is just an 'impoverished' or scaled down version of the image-generation that is executed during full-fledged mental imagery. However, there is reason to believe that this is not the case. It has been argued that the type of mental simulation elicited during language comprehension is qualitatively different from imagery. Troscianko (2013) makes this point on conceptual grounds. In accordance with Barsalou (2008), she argues that mental simulation during reading should not be seen as an explicit and vivid mental picture (or image) that a reader creates while reading, but rather as reactivations of motoric or sensory memories. That is, memories of previous experiences with actions and objects in the actual world determine how language is understood without the explicit need to form vivid mental pictures. Willems, Toni, Hagoort, and Casasanto (2010) provide empirical support for such a qualitatively different neural basis, and Chow et al. (2015) show that the motoric and sensory brain systems involved in language processing are indeed modulated by previous experiences. In the current dissertation, the sub-conscious simulative processes occurring during reading or listening to language will be referred to as 'mental simulation', and the term 'mental imagery' will be reserved for situations in which participants engage in deliberate and conscious mental imagery.

Having specified the differences between mental simulation and mental imagery, I will now turn to the main topic of this dissertation: (how) does mental simulation play a role in the specific context of literary reading?

1.1.3. Theories About Mental Simulation and Literary Reading

In this section, I discuss three prominent theories on how mental simulation plays a role in reading, or in literary reading specifically. I will first discuss a theory about aspects of mental simulation during literary reading (Kuzmičová, 2014), then cover the *Neurocognitive Poetics Model* (Jacobs, 2015b), and finally focus on the simulation of feelings (as opposed to limiting the discussion to the simulation of actions and perceptions; Miall & Kuiken, 2002). In the next three subsections, I will explain how the three theories differ, and how they seem to be intertwined. Importantly, these theories are not necessarily a starting point for the experiments described in this dissertation, but the results from these experiments should be interpreted in light of these three theories.

1.1.3.1. Varieties of Mental Imagery During Literary Reading

In an important theoretical contribution, Kuzmičová (2014) has suggested that mental imagery[1] during literary reading is not one of a kind, but can be experienced in a few different forms. Furthermore, which of the different forms of mental imagery is experienced at a given time during reading is dependent on both text characteristics and reader characteristics (see Table 1 for a schematic overview).

Kuzmičová calls the most basic form of mental imagery *rehearsal-imagery*. Readers experiencing this kind of imagery perceive the words in the stories they read as if they are reading them aloud (without actually articulating the words). This kind of imagery is most often triggered by longer, syntactically complex sen-

[1]Note that although Kuzmičová uses the term imagery, she defines this as the non-conscious, automatic process I have called simulation. I will also use the word imagery when discussing Kuzmičová's theories, but it is important to keep in mind that imagery as defined by Kuzmičová is by no means the same as imagery as defined by Kosslyn, Jeannerod, Parsons and others (see above).

Table 1.1.: *Schematic Overview of the Four Different Varieties of Mental Imagery (Simplified Version of the Model Proposed by Kuzmičová, 2014)*

Variety of Mental Imagery			
Verbal Domain		Referential Domain	
Rehearsal-Imagery	Speech-Imagery	Description-Imagery	Enactment-Imagery

tences, or by sentences that contain certain stylistic elements such as rhythm or alliteration (which need to be articulated to be fully appreciated).

The second level, *speech-imagery*, differs from rehearsal-imagery in that readers do not hear their own voices in their mind while reading, but rather the voices of characters in the story, as if they are witnessing their conversations. This is most often triggered by dialogues in stories and not as much by stylistic elements. Together, rehearsal-imagery and speech-imagery form the verbal domain of the mental imagery continuum.

Beyond the verbal domain, there is the referential domain, which is most closely linked to embodied cognition theories in psychology. Again, according to Kuzmičová (2014) there are two levels of imagery that comprise the referential domain. The first is called *description-imagery*, where readers form (mostly, but not only, visual) pictures of objects or situations described in a story, specifically from an observer's perspective. Description-imagery is often triggered by elaborate descriptions of how (inanimate) objects in stories look, sound, or feel. This is unlike enactment-imagery (according to Kuzmičová, the highest form of mental imagery) in which readers form mental pictures from the perspective of a character in the story, almost as if they are acting out the situations in the story. *Enactment-imagery* is triggered by concrete and imageable descriptions of the sensorimotor experiences of characters.

It could be argued that the difference between description-imagery and enactment-imagery reflects differences in viewpoint or stance (comparable to the relationship between viewpoint in narratives and identification with characters; Van Krieken, Hoeken, & Sanders, 2017). Description-imagery is experienced from a third person stance, whereas enactment-imagery is experienced from a first-person stance. Consequently, the experience of description- or enactment-imagery could be dependent on text-characteristics or contextual information encouraging a first versus third person interpretation.

It is important to underscore that Kuzmičová (2014) acknowledges that readers' experiences can also resemble an in-between form between two levels of mental imagery. Additionally, she stresses that it is not the case that a given reader can only experience one form of mental simulation. During reading, readers constantly switch between imagery modes, as a result of a continuous interplay between the text characteristics of the passages read and internal reader characteristics. Kuzmičová hypothesizes that the transition between different modes will be smooth and (almost) non-conscious within the verbal and referential domains, whereas it will be conscious when readers switch between the domains.

1.1.3.2. Neurocognitive Poetics Model

A more general theory of the cognitive processes going on during literary reading is Jacobs' Neurocognitive Poetics Model (NCPM; Jacobs, 2015b). Although this model is not specific to mental simulation during reading, as is Kuzmičová's theory, the NCPM does provide insight into the circumstances that make the occurrence of mental simulation during literary reading most likely. This theory is built on the premise that reading stories is more than just reading words on a page: if stories were processed as mere 'cold' lists of words and sentences, they would probably not elicit strong emotions (Jacobs, 2015b). Because stories challenge readers to create mental pictures during reading, readers can more easily become emotionally involved when reading stories than when reading lists of words. In the paper introducing the NCPM, (Jacobs, 2015b) argues that simulation is evoked by familiar words and phrases, high frequency words, and highly imageable words (which are typically words that also require less effort to process).

At the heart of the NCPM lies the distinction between two routes of literary reading, a fast route and a slow route. The fast route is provoked by backgrounded elements in stories, such as familiar words and phrases, high frequency words, and highly imageable words. This route evokes fluent reading through implicit processing and fiction feelings and is hypothesized to be related to immersive processes during reading. Fluent reading is considered to be automatic and subconscious, just as mental simulation during reading is considered to be automatic and subconscious. Additionally, the hypothesized link between fluent reading and immersive processes is reminiscent of the link between mental simulation and immersive processes (elaborated in the section on offline studies of simulation). Therefore, it seems probable that mental simulation plays an important role in this mode of reading.

The slow route is provoked by foregrounded elements in stories: for example, metaphors, abstract and defamiliarizing language, rhyme and rhetorical devices.

Foregrounded elements are hypothesized to evoke dysfluent reading through explicit processing and aesthetic feeling (Jacobs calls this the aesthetic trajectory). The outcome of dysfluent reading is the aesthetic appreciation of literature and poetry. Interestingly, this route is triggered by stylistic elements in stories, similar to Kuzmičová's rehearsal-imagery (Kuzmičová, 2014, see above). Although, in general, mental simulation seems to play a role in the fast route of reading, perhaps some forms of simulation (i.e., perceiving the stories as if one were reading them aloud) are actually more likely to occur in the slow route of reading.

Interestingly, Kuiken and Douglas (2017, 2018) distinguish between simulation of content related to peri-personal space versus content related to extra-personal space. They hypothesize that processing (or simulating) objects in peri-personal space (such as sensorimotor imagery) is part of the slow, foregrounded route. In contrast, content related to extra-personal space (such as visuospatial imagery) is hypothesized to be part of the backgrounded, fast route of literary reading. As (Jacobs, 2015b) does not go into detail about the involvement of different forms of mental simulation in the two routes of the NCPM, only future research will determine whether different forms of mental simulation indeed play roles in different routes of the NCPM, and, if so, which forms of simulation play roles in which routes of literary processing.

1.1.3.3. Simulating Feelings

Apart from perceptually simulating objects and situations or motorically simulating actions described in stories, it is also possible to simulate story characters' feelings. Simulating feelings elicits those feelings in readers. According to Miall and Kuiken (2002) this can happen on four levels that differ in the "depth" of these feelings. Miall and Kuiken called the first, most basic, level *evaluative feelings*. This level comprises feelings like enjoyment of a story or reading pleasure – feelings that can drive a reader to continue reading a story but do not result in a deep involvement in the story.

The second level identified by Miall and Kuiken is the level of narrative feelings. *Narrative feelings* include empathy for and sympathy with the characters in the story or feelings that are a response to specific events in the story. These feelings require a reader to step into the shoes of the story character and (to some extent) simulate the story world and feelings of the characters. This level of feelings may be elicited by description-imagery as defined by Kuzmičová (2014).

The third level of feelings elicited by stories is called *aesthetic feelings*. Aesthetic feelings are not elicited by story events but by stylistic elements in the stories. Certain metaphors, choice of words, or sentence constructions can fascinate or intrigue readers and capture their attention. In terms of the levels of mental imagery defined by Kuzmičová (2014), aesthetic feelings may be linked to the verbal domain (mainly rehearsal-imagery), which Kuzmičová claims is associated with stylistic elements of stories, such as prosody and rhythm. In terms of the NPCM (Jacobs, 2015b), this level of feelings probably results from processing via the slow route of the NCPM.

The fourth and highest level of feelings that can be elicited by literary fiction is called *self-modifying feelings* (Miall & Kuiken, 2002). At this level, a combination of perspective taking and stylistic elements elicits a deep identification with the story and story characters. Identification at the fourth level stands out from the other levels in that it is grounded in memories of feelings readers have experienced in their own lives. This is closely related to the process of "re-experiencing" proposed by Goldman and Barsalou (described at the beginning of this chapter and in Box 1). To elaborate, the power of fiction to elicit self-modifying feelings results from readers re-experiencing their past feelings, possibly through mental simulation (i.e., most probably through simulation at the level of enactment-imagery).

So far, I have focused on mental imagery and mental simulation during language processing, the differences between the mental imagery and mental simulation, and I have discussed three prominent theories on mental simulation during literary reading. In the next subsection, I will focus on the "story reading" part of the research question "Do people use mental imagery during story reading".

1.1.4. Do People Use Mental Imagery During *Story Reading*?

In this dissertation, I will focus on the role of mental simulation in *story reading*. It is therefore important to define what is meant by "story" and by "reading". According to Abrams and Harpham (2009, p. 209), a story can be defined as "a mere sequence of events in time". This sequence of events is usually structured in a narrative plot, where the story starts with an *orientation* ("setting the scene" of the story), followed by *complicating actions* that result in the *critical event* of the story (what the story is about), and ending with the *resolution* of this critical event and the *coda* that reconnects the story world with the speech situation in the "real" world (Labov, 2010; Labov & Waletzky, 1967; Sanders & Van Krieken, 2018).

The studies described in this dissertation will make use of stories from the specific subgenre of literary *short stories*, that can be defined as “a brief work of prose fiction [... that] organizes the action, thought, and dialogue of its characters into the artful pattern of a plot, directed toward particular effects on an audience” (Abrams & Harpham, 2009, p. 331). The reason for using these stories is a rather pragmatic one: these stories already exist, and are written by professional and acclaimed writers (and therefore can be expected to be of good quality). Typical of the subgenre of literary short stories as used in this dissertation is that they all follow a similar narrative plot structure: the tension builds up throughout the story, is then released in a plot twist (critical event), which is followed quite shortly by some sort of resolution and an open ending.

An important reason to study *story* reading specifically, rather than focusing on language processing in general, is that many researchers have mentioned a role for mental simulation in the processing and understanding of narratives. Narrative understanding has been proposed to be a product of the construction and updating of mental representations of story worlds based on the information given in the text, including the actions and interactions of characters (Emmott, 1997). The information derived from the text is understood within the light of readers’ real-life experiences (Emmott, 1997). This is in line with the Event-Indexing model proposed by Zwaan, Langston, and Graesser (1995), which proposes that readers understand narratives through narrative events that can be linked to their memories for similar events. These events can occur on multiple dimensions, being time, space, character, causality, and intentionality (Zwaan et al., 1995). Similarly, Toolan (2016) states that narrative understanding occurs through the interaction between repetition (experienced coherence in a text through repetition of elements in a text), situation (the anchoring of the narrative in stable locations, characters or time), and mental picturing. Mental picturing is closely related to mental simulation, being the experience of vague, relatively unstable mental images that are based in associations with knowledge or sensory memories (Toolan, 2016).

In psychological research, it is customary that researchers create their own stimuli, in order to be able to control all aspects of the stimuli and interject the variables of interest in a structured manner. Here, I choose to depart from this tradition by using existing stories as my stimuli, because I prefer stimuli with a high ecological validity (that is, stimuli that can be found “in the wild” rather than only in lab settings). It has been argued that using stimuli constructed for lab settings leads to a poorer understanding of mental simulation than using ecologically valid narratives (Willems & Jacobs, 2016). By using published literary texts of generally acclaimed authors, "natural" materials are the basis of the experiments reported in this dissertation, in which the variables of interest are employed in manner that is not too obvious or explicitly purposeful.

Finally, when talking about "reading", I am talking about reading fictional stories. Because of the methods I use in the studies I report in this dissertation (i.e., eye tracking, fMRI), I operationalize this as digital reading in Chapters 3,4, and 5. For practical reasons (I test Dutch participants), all stories are presented in Dutch.

1.2. How Does Mental Simulation Relate to the Way Readers Experience Stories?

After discussing what mental simulation during literary reading entails, I now turn to my second research question concerning other reading experiences that are theoretically related to mental simulation. One important reading experience that should be discussed in this light, concerns the experience of becoming lost in a story. Mental simulation is one aspect that plays a role in this experience (e.g., Green & Brock, 2000; Kuijpers, Hakemulder, Tan, & Doicaru, 2014). The experience of becoming lost in a story has been described as absorption (e.g., Kuijpers et al., 2014; Kuiken & Douglas, 2017), but also immersion (Ryan, 2001), transportation (Gerrig, 1993; Green & Brock, 2000), presence (Kuzmičová, 2012), or flow (Csikszentmihalyi, 1990). For the sake of clarity, I will refer the experience of becoming lost in a story as absorption for the remainder of this dissertation.

In many offline studies of mental simulation, the role of mental simulation in theories of absorption during literary reading is investigated. Absorption is an experiential state in which readers are focused on reading and on the content of what is read (Kuijpers, 2014). This process has been identified in different ways, with some of the definitions of this process emphasizing the role of mental imagery. Moreover, when readers experience absorption while reading a story, this usually includes rich and vivid mental simulation (Kuijpers et al., 2014; see also Kuiken & Douglas, 2017). Note that the same theoretical association has been proposed between transportation and mental simulation (Green & Brock, 2000).

If mental simulation indeed contributes to the experience of absorption, it could be expected that people who report experiencing absorption during reading will indeed also experience mental simulation. Individual differences in simulation as measured with eye tracking and fMRI should then be associated with individual differences in reported absorption.

Absorption, in turn, has been associated with the enjoyment of narratives, with absorbed readers reporting enjoying reading more (Busselle & Bilandzic, 2009; Green & Brock, 2000; Green, Brock, & Kaufman, 2004; Hartung, Burke, Hagoort, & Willems, 2016; Kuijpers et al., 2014). In this dissertation, I am not only interested in surface-level enjoyment of stories, but also in aesthetic evaluations on more detailed grounds. With aesthetic evaluations, I mean evaluations of the sadness, suspense or beauty of a story, for example (as suggested by Knoop, Wagner, Jacobsen, & Menninghaus, 2016). Aesthetic evaluation could be associated with absorption in multiple ways: It could be that being absorbed in a story also contributes to the aesthetic evaluation of this story. However, it is also possible that absorption is actually *distracting* from consciously evaluating stories on more detailed grounds (in line with Eekhof et al., 2021). Regardless, I view this as an important part of reading experiences, as enjoyment and appreciation of stories are an important factor in promoting reading (e.g., Mol & Jolles, 2014).

1.3. How Do People Differ in Mental Simulation?

Above, I have explored the origins of mental simulation during reading (i.e., mental simulation is based on the re-experience of personal life events and experiences), and the relationship between mental simulation and experiential states such as absorption. The final question that needs to be addressed, is whether mental simulation occurs similarly in all readers. If this is not the case, this should become visible in individual differences in the effect of mental simulation on eye movements (see Box 2) and brain activation (see Box 3).

Indeed, an fMRI study by Hartung, Hagoort, and Willems (2017) showed that people differ in their preferred kind of simulation during the reading of stories. The authors distinguished three groups of readers based on an offline self-report question. When subsequently looking at the brain activation patterns of participants in these groups during reading, they found marked differences between these groups of readers. The first group showed activation in a region in the right frontolateral pole and were called "enactors" by the authors. The second group showed activation in a network including the right inferior frontal gyrus, left postcentral gyrus, left supramarginal gyrus, and left and right posterior superior and middle temporal gyri, and were called “observers”. The third group was called “hypersimulators”; these people showed activation in both networks.

This distinction between subgroups of readers is reminiscent of the subgroups distinguished by Kozhevnikov, Kosslyn, and Shephard (2005). In a behavioral study, they classified participants as “visualizers” or “verbalizers” according to the Visualizer-Verbalizer Cognitive Style Questionnaire, and found that verbalizers scored at intermediate levels on mental imagery tasks, whereas visualizers scored high on either spatial imagery or object imagery tasks. This implies that people can be visualizers (prone to mental imagery) or verbalizers, but, interestingly, there also appeared to be individual differences within the group of visualizers. Apparently, people can be skilled in particular forms of imagery (i.e., spatial imagery or object imagery).

Another fMRI study by Nijhof and Willems (2015) looked at brain activity associated with listening to action descriptions and descriptions of mental events within literary stories. The authors found that listening to action descriptions was associated with activity in areas involved in action execution, whereas listening to descriptions of mental events was associated with activity in areas involved in mentalizing. Importantly, they found interesting individual differences, indicating that some participants were particularly responsive to action descriptions but not to descriptions of mental events, whereas others were responsive to descriptions of mental events but not to action descriptions. Together, these studies imply that there are differences between readers in the way mental simulation is part of their reading experience.

Box 2. Methodology: The Use of Eye Tracking in the Study of Mental Simulation

In order to address how eye tracking can be used in the study of mental simulation during literary reading, I will now give some background information on eye movement tracking, what is known about eye movements during reading, and why this is a useful tool for studying mental simulation.

1. How Are Eye Movements Measured?

Eye movements can be measured using an eye tracker. Eye trackers generally consist of an infrared source and a camera aimed at the eyes of participants while they are reading a text from a computer screen (see Kliegl

& Laubrock, 2017). The way eye trackers work is the following. The pupil is a hole in the middle of the iris, that allows light to pass through. The retina however, does not let light pass through. The infrared light emitted by the eye tracker's infrared source will therefore be absorbed by the iris, but reflected by the retina, resulting in a circle of reflected infrared light being recorded by the camera of the eye tracker. The eye tracker recognizes the light that has been reflected by the retina as the pupil, and can track the location of the pupil. The eye tracker is linked to the computer screen on which the participant is reading. Through calibration, the eye tracker can determine at what position on the screen a participant is fixating their eyes, based on the exact shape and position of the pupil. When recording, the eye tracker transmits a time series of locations, which can be linked to the positions of the words that were presented on the screen at each given time. Recording happens at a very fast pace, in the studies reported in this dissertation at 500 or 1000 Hz (i.e., information about pupil size and location is recorded for each one or two millisecond time period).

What is measured when readers have their eye movements tracked, is essentially which word was fixated at every moment in time, and where the eye moved next when a fixation had ended (e.g., the next word (saccades), or a word previous to the current word (regression)). Fixations and eye movements can both indicate multiple facets of cognitive processing. For example, fixation durations have been associated with attention, whereas regressive eye movements can indicate processing difficulties (e.g., Kliegl & Laubrock, 2017; Rayner, 1998).

2. What is Known About Eye Movements During Reading?

Research into eye movements during reading has discovered some general patterns. When people read, they are alternating fixations (periods of time that the eye lingers on one location) with fast saccades (jump to a new location; Kliegl & Laubrock, 2017; Rayner, 1998). During those saccades, visual input is suppressed, meaning that visual processing always takes place during the fixations (Kliegl & Laubrock, 2017). Generally, fixations last for 30 – >500 ms, with an average of 200-250 ms during (silent) reading[2], and saccades last for 10-30 ms (Kliegl & Laubrock, 2017; Rayner, 1998).

[2]Importantly, the numbers presented here are specific to silent reading. When people are reading aloud, or following the text accompanied by someone else reading the text aloud, fix-

People generally fixate every 7-9 letters, which is the span of the average saccade (Rayner, 1998). Most saccades are left-to-right meaning that they are made toward a next portion of a sentence or text. However, about 15% of saccades are called regressions: right-to-left saccades to a previous part of a line, or to a position a few lines back (Rayner, 1998). Regressions usually indicate difficulty processing a word or difficulty integrating a word into a sentence.

Although these numbers are on average a good indication of reading behavior, there is a lot of variation between individual readers and between individual words. The fixation durations and saccade span can differ between individuals due to reading proficiency, for instance. Good readers tend to show shorter fixation duration and longer saccade spans than poor readers (Rayner, 1998). Differences in fixation durations to individual words can be explained by multiple word characteristics that influence processing difficulty. In general, one can say that the more difficult it is to process a certain word, the longer the fixation duration, and the higher the probability that a regression (eye movement back to the word from a later portion of the text) is made toward this word. These effects have been found for the frequency (higher frequency words receive shorter fixations), length (longer words receive longer fixations), and predictability (highly predictable words receive shorter fixations) of words. This has been found both for currently fixated words, and for previous and following words (Kliegl, Grabner, Rolfs, & Engbert, 2004; Kliegl, Nuthmann, & Engbert, 2006). Regarding word length, it has also been found that the length of a word is indicative for the likelihood that the word will receive a fixation at all (Rayner, 1998). Shorter function words are skipped more than half of the time. Finally, Rayner and colleagues discovered an interaction between the effects of word frequency and predictability: the effect of predictability on fixation durations (i.e., highly predictable words receive shorter fixations) was more pronounced for low frequency words than for high frequency words (Rayner, Ashby, Pollatsek, & Reichle, 2004).

ations tend to be slightly longer (Rayner, 1998). Moreover, when people read incoherent text (for example, text made up from incoherent strings of letters “zzz zzz zzzzz zz zzz zzzzzz”) they show longer fixations, shorter saccades, and more frequent skipping of target string than when they read coherent text (e.g., “the man stood in the garden”; Rayner & Fischer, 1996).

3. How Can Measurements of Eye Movements Be Interpreted in the Context of Simulation?

Eye tracking measures attention (e.g., to words, passages), and processing speed. For example, increased attention to certain words or passages is associated with longer gaze duration. As mentioned, this can be seen in text features such as the lexical frequency of words. Words with a high lexical frequency, are words that occur often in day-to-day language. Because they occur frequently, these words are easier to recognize and often more predictable, warrant less attention, and are easier to process than words that infrequently occur in language. Indeed, highly frequent words are associated with shorter gaze durations (i.e., shorter reading times for frequent words; Rayner, 1998). Moreover, reading speed is not only related to the characteristics of these words themselves, but also relies on the context in which a word occurs. If a word is highly likely within its context (regardless of its frequency), this word is easier to process (and warrants less attention) than unlikely words, and is indeed associated with shorter gaze durations (Goodkind & Bicknell, 2018; Hale, 2001; Levy, 2008). When looking at simulation-inviting language in stories, this could be associated with reading speed in a similar manner. If simulation is associated with an increased processing load (because it takes time to link the meaning of words to a mental picture, for example), processing speed - and therefore reading speed - will decrease. Likewise, if passages that invite simulation attract more attention than passages that don't invite simulation, this will result in longer or more frequent fixations on those passages. However, if simulation aids in processing (for example if the mental picture formed based on the previous text renders an upcoming word more predictable), reading speed will increase.

Box 3. Methodology: The Use of fMRI in the Study of Mental Simulation

In order to address how functional Magnetic Resonance Imaging (fMRI) can be used to study mental simulation during literary reading I will now give some background information on fMRI, what is known about brain activation during reading, and what brain areas are involved in literary reading.

1. What is fMRI?

As mentioned, in this dissertation fMRI will be used to measure brain activation. With fMRI, the Blood Oxygenation Level Dependent (BOLD)-response is measured (Willems & Cristia, 2017; Willems & Van Gerven, 2018). When a brain area becomes active, this results in a change in the blood oxygen level, as active areas in the brain consume more oxygen than inactive brain areas. The magnetic properties of oxygenated red blood cells differ from the magnetic properties of deoxygenated red blood cells. The scanner can pick up these changes in blood oxygenation level, and how they differ in different parts of the brain (Willems & Cristia, 2017). Pairing knowledge of the timing of events in a study with the knowledge that active brain areas consume more oxygen, helps to calculate correlations between event onset and a rise in oxygen consumption in all parts of the brain (Willems & Cristia, 2017; Willems & Van Gerven, 2018). Because of the specific designs of the studies that make use of fMRI, it has been possible to determine the processes executed by many different areas in the brain.

2. What is Known About Brain Activation During Reading?

When people are reading, they are integrating single words into sentences and binding these sentences together. There are multiple brain networks involved in the execution of this process (Hagoort, 2019). These do not only include the perisylvian areas historically linked to language processing (i.e., Broca's and Wernicke's areas), but many other parts of the temporal and parietal and frontal cortex play an important role in language processing too (Hagoort, 2019). For example, more areas of the Left Inferior Frontal cortex are involved in language than just Broca's area (BA 44 and 45). Finally, not only cortical areas have been found to be involved in language processing, but also subcortical areas such as the thalamus and basal ganglia, as well as parts of the cerebellum (Hagoort, 2019). Reading stories does not only involve brain activation in parts of the brain associated with language processing per se. When people are reading stories, they are also often making inferences about a character's thoughts and behavior, or are incorporating new parts of a story into their situation model of the story (Zwaan, 2009). Pragmatic inferencing has been linked to areas within the Theory of Mind network (such as the temporoparietal junction, medial prefrontal cortex), whereas the integration of utterances into the situation model has been linked to the inferior frontal gyrus and angular gyrus (Hagoort, 2019).

3. What Brain Areas Are Involved in Mental Simulation?

According to the embodied cognition theory, language processing is also accompanied by activation in so-called modality-specific brain areas. This means that reading about actions elicits activation in areas associated with performing these actions yourself (i.e., motor simulation), reading visual descriptions elicits activation in areas involved in vision (i.e., perceptual simulation), and reading about thoughts or emotions of characters elicits activation in areas involved in processing emotions (i.e., mentalizing). Indeed, several studies have found modality-specific brain activation during story reading. For example, reading about actions has been associated with activity in motor areas (e.g., precentral and postcentral cortex, superior temporal sulcus, cingulate cortex, supplementary motor area, middle and superior frontal gyrus, middle and superior temporal gyrus, inferior parietal lobule, intraparietal sulcus, precuneus, parahippocampal gyrus; Chow et al., 2015; Kurby & Zacks, 2013; Moody & Gennari, 2010; Nijhof & Willems, 2015), reading visual descriptions has been associated with activity in areas involved in visual processing (e.g., cuneus, lingual gyrus, fusiform gyrus and parahippocampal gyrus; Chow et al., 2015), and reading about thought and emotions has been associated with activity in the mentalizing-network (e.g., aMPFC, dMPFC, MCC, TPJ; U. Frith & Frith, 2003; Hsu, Conrad, & Jacobs, 2014; Lai, Willems, & Hagoort, 2015; Nijhof & Willems, 2015; Saxe & Kanwisher, 2003; Tamir, Bricker, Dodell-Feder, & Mitchell, 2016). Interestingly, Chow et al. (2015) have shown that the language-specific regions mentioned earlier are connected with these domain-specific regions, confirming the involvement of these domain-specific regions in language processing.

1.4. Outline of This Dissertation

In the remainder of this dissertation, I will describe the experiments I will perform in order to answer my research questions.

In Chapter 2, I describe two off-line reading experiments, in which I attempt to influence reading experiences through instructions aimed at altering mental imagery. The aim of this study is to find out whether it is possible for readers to consciously "decide" to imagine events described in stories, and how this translates to experiences such as absorption, enjoyment and aesthetic evaluations during reading. The findings from these experiments shed light on what it is that people imagine during literary reading, and how mental simulation relates to how readers experience stories.

In Chapter 3, I describe an eye tracking experiment in which participants read three literary short stories, while having their eye movements tracked. After reading, participants are asked to report on their subjective reading experiences, so that the on-line measure of eye tracking can be combined with offline measures (questionnaires). The words of the three stories are scored on simulation-eliciting content. These scores will be related to the eye movement data. This study aims to tap into the effect of mental simulation on eye movements, and whether differential effects of motor simulation, perceptual simulations, and simulation of mental events can be seen in eye movement data. In the analyses of this study, I will also pay attention to individual differences in these effects, and how these individual differences are related to absorption, enjoyment and aesthetic evaluations. Using eye tracking, this experiment sheds light on what it is that people imagine during literary reading, how people differ in mental simulation, and how mental simulation relates to how readers experience stories.

Two of the three fictional stories from Chapter 3 are reused in the experiment described in Chapter 4, where the eye tracking and questionnaire measures from the experiment described in Chapter 3 are combined with neuroimaging. The aim of this experiment is to corroborate the eye movement findings from Chapter 3 with neuroimaging data and to extend these findings by unveiling whether and how the individual differences as found in the eye tracking data are related to individual differences on a neural level. Again, I pay attention to how individual differences in mental simulation are related to absorption, enjoyment and aesthetic evaluations. Using both eye tracking and fMRI, this experiment sheds light on what it is that people imagine during literary reading, how people differ in mental simulation, and how mental simulation relates to how readers experience stories.

In Chapter 5, individual differences in the relationships between the subjective experiences of general liking, aesthetic appreciation, and narrative absorption are explored further, in a reanalysis of questionnaire data from three previous experiments. The aim of this reanalysis is to make the individual differences in these reading experiences clear, after finding strong individual variation in the association between mental simulation and reading behavior in the earlier chapters. With this experiment, I also want to pay closer attention to the open question of the role of aesthetic evaluations in literary reading, and how these evaluations relate to absorption and surface-level enjoyment. This chapter dives more deeply than the previous chapters into the questions how readers differ in how they experience stories.

In Chapter 6, the results from the experiments described in this dissertation will be discussed in light of the theories of mental simulation during reading, and their significance for scientific development will be stressed along with considerations of their methodological strengths and weaknesses.

2 | The Influence of Mental Imagery Instructions and Personality Characteristics on Reading Experiences

It is well-established that readers form mental images when reading a narrative. However, the consequences of mental imagery (i.e., the influence of mental imagery on the way people experience stories) are still unclear. Here I manipulated the amount of mental imagery that participants engaged in while reading short literary stories in two experiments. Participants received pre-reading instructions aimed at encouraging or discouraging mental imagery. After reading, participants answered questions about their reading experiences. I also measured individual trait differences that are relevant for literary reading experiences. The results from the first experiment suggests an important role of mental imagery in determining reading experiences. However, the results from the second experiment show that mental imagery is only a weak predictor of reading experiences compared to individual (trait) differences in how imaginative participants were. Moreover, the influence of mental imagery instructions did not extend to reading experiences unrelated to mental imagery. The implications of these results for the relationship between mental imagery and reading experiences are discussed.

This Chapter Is Based on

Mak, Marloes, De Vries, Clarissa, & Willems, Roel M. The influence of mental imagery instructions and personality characteristic on reading experiences. *Collabra: Psychology, 6*(1): 43. DOI: https://doi.org/10.1525/collabra.281

In Chapter 1, I described studies of mental simulation and other reading experiences, in which researchers attempt to influence reading outcomes through reading instructions aimed at altering mental simulation or mental imagery. Many studies have not been successful or only partially successful in this attempt.

This chapter describes two experiments in which I used reading instructions that were more specifically aimed at mental imagery than the reading instructions used in previous studies. Through mental imagery, these instructions were supposed to influence multiple reading experiences. The goal is to find out whether mental imagery can be manipulated in this way, or if this is not a fruitful way to approach differences in mental simulation during reading.

2.1. Introduction

It is well established that readers perceive mental images during reading (Green & Brock, 2000; Jacobs, 2015b). For instance, an eye tracking study showed that people are responsive to mental simulation-eliciting content in stories (Mak & Willems, 2019). It was found that when participants were reading action descriptions (assumed to elicit motor simulation) they sped up, whereas they slowed down when reading perceptual descriptions or mental event descriptions (assumed to elicit perceptual simulation or mentalizing, respectively).

Additionally, there is a relationship between the amount of imagery and subjective experiences during reading. Mak and Willems (2019) found that individual differences in the responsiveness to simulation-eliciting content were related to participants' subjective experiences (such as absorption and appreciation). This is only one example of work showing that mental simulation during reading is associated with absorption in and appreciation of stories (see also Green, 2004; Green & Brock, 2002; Kuijpers et al., 2014; Mol & Jolles, 2014; Weibel, Wissmath, & Mast, 2011).

Next to individual variation in amount of imagery perceived during reading, there is a number of stable (personality) characteristics that are associated with reading experiences. In the experiments I report in this chapter, I decided to study both the role of instructed mental imagery and the role of individual (trait) differences in literary reading. Below, I will discuss the relationships between (1) mental imagery and reading experiences and between (2) individual (trait) differences and reading experiences, before I (3) introduce the set-up and hypotheses of the experiments.

2.1.1. Mental Imagery and Reading Experiences

As mentioned above, people engage in mental imagery[1] when reading stories, and mental imagery is an important driver of absorption: it has been found that visualizing the story world will strengthen people's experience of absorption in a story (Green & Brock, 2002; Kuijpers et al., 2014; Kuiken & Douglas, 2017)[2]. Absorption has been defined by Kuijpers et al. (2014) as "the subjective experience of being absorbed in the story world of a narrative text" (p. 90, emphasis in the original). It describes the feeling we may have when reading a good story or book where we go beyond comprehending the meaning of the words on a page, to a captivating experience that can help us become completely involved in the stories we read. This experience has been reported on in widely diverging disciplines, and has also been defined as transportation (Gerrig, 1993; Green & Brock, 2000), narrative engagement (Busselle & Bilandzic, 2008), narrative presence (Kuzmičová, 2012), immersion (Ryan, 2001; see also Jacobs, 2015b) or flow (Csikszentmihalyi, 1990). For the sake of clarity, I will refer to this experience as absorption for the remainder of this chapter.

Absorption is a multifaceted construct that tries to capture the entirety of the subjective experience of being captivated by a good story (Kuijpers et al., 2014). It is proposed that absorption consists of multiple dimensions, being mental imagery, emotional engagement, attention, and transportation (Kuijpers et al., 2014). Mental imagery (as a dimension of absorption) is defined as a visualization of the story world, whereas emotional engagement could be seen as the emotional counterpart of mental imagery: the sympathetic and empathic feelings for the characters in the story. Attention is characterized as a heightened focus or concentration of the reader towards the story world – and as a consequence a lower concentration towards the here and now. Transportation is seen as the feeling a reader can have of being part of the story world as opposed to the real world.

[1]Mental imagery during reading is also sometimes referred to as mental simulation. Theoretically, mental simulation is a somewhat more subconscious process than mental imagery. As already explained in Chapter 1, I call the process of (more or less consciously) imagining events or perceptible elements of the story world described in a story mental imagery.

[2]Note that when looking at the literature regarding mental imagery during reading it becomes clear that people seem not just to mentally image descriptions of visual elements of story worlds, but a much more extensive range of perceptual descriptions (i.e., auditory, olfactory, proprioceptive) and motor descriptions (Kuzmičová, 2012, 2014; Mak & Willems, 2019; Nijhof & Willems, 2015). Although the exact nature of mental imagery during reading is still debated, in this chapter I do not want to focus on the content of mental imagery during reading, but instead on the act of mental imagery itself.

Together these 4 dimensions can result in an experience of complete absorption in a narrative or story world.[3]

Another connection between absorption and mental imagery, is that they have both been found to be associated with another important aspect of people's reading experiences: the enjoyment (or appreciation) of stories (Busselle & Bilandzic, 2009; Green, 2004; Green et al., 2004; Kuijpers et al., 2014; Kuiken & Douglas, 2017; Mol & Jolles, 2014; Weibel et al., 2011). As there does not seem to be strong consensus with regard to the definition of appreciation (especially between different disciplines, e.g., communication research and literature studies), in the experiments described in the current chapter I have considered both the overall enjoyment of narratives and other facets of aesthetic experiences that I believe could play a role in the enjoyment and appreciation of stories (e.g., whether a reader is emotionally moved by a story, or finds it amusing). To test this multitude of facets of aesthetic experiences, I used adjectives that many readers use to describe their aesthetic experiences while reading literature, that were obtained from a list of adjectives collected by Knoop et al. (2016). Because this list of adjectives was compiled in a "bottom-up" fashion (i.e., the adjectives are derived directly from the experiences of readers), it was assumed that these adjectives could successfully tap into multiple facets of aesthetic experiences of readers. I found it important to look at appreciation on top of absorption, because I wanted to consider a measure of reading experience that was more distantly related (but not unrelated) to mental imagery than absorption (recall that mental imagery is considered to be one of the four subcomponents of absorption).

2.1.2. Reading Instructions and Mental Imagery

In order to test the relationship between mental imagery and reading experiences, I wanted to make sure that some readers in my experiments would engage in mental imagery more than others. To this effect, I employed a method from a related but separate line of research, where it was found that instructing students to create images of what they had read impacts text comprehension (see De Koning & Van der Schoot, 2013, for an extensive overview). Apart from text comprehension, some studies have also found direct links between pre-reading instructions and reading experiences. Green and Brock (2000) found that instructing readers to judge the difficulty of a story to establish the suitability for fourth-

[3]Note however, that Kuiken and Douglas (2017) proposed that absorption is even more multidimensional, suggesting multiple types of absorption, with multiple different outcomes.

grade readers, led to lower experienced transportation (which is conceptually comparable to absorption) in comparison with readers who were instructed to pay specific attention to the story plot. The rationale of the study was that instructing participants to pay attention to the suitability of the text for fourth-graders, would lead to less absorption than instructing them to focus on the story plot. In a follow-up experiment, Green (2004) subsequently instructed participants to use relaxation strategies during reading to increase transportation. This manipulation did however not lead to more experienced transportation in these participants, when compared to participants who had received a neutral instruction. Perhaps this instruction was not associated with higher transportation because relaxation is relatively unrelated to the process of transportation. Therefore, the advice for future research was to use a pre-reading instruction focusing on a specific component of transportation (e.g., imagery; Green, 2004, p.261). Johnson, Cushman, Borden, and McCune (2013) made a successful first attempt in this direction. Instead of pre-reading instructions, they gave participants an imagery generation training, which subsequently resulted in increased transportation when reading a narrative. In the experiments described in this chapter, I take these findings as the starting point for the investigation of the influence of explicit imagery instructions on subjective reading experiences.

Participants in the experiments described in the current chapter were either encouraged or discouraged to engage in mental imagery through pre-reading instructions, after which they were asked about their subjective experiences while reading literary short stories. Even though pre-reading instructions have not been successful in all studies attempting to influence reading experiences, Tukachinsky (2014) noted in a review of these studies that the effect of pre-reading instructions seemed quite reliable. Moreover, I used an instruction specifically targeting mental imagery, which was suggested by Green (2004) as possibly more powerful in manipulating reading experiences than instructions aimed at more general processes. The purpose of this specific mental imagery instruction was to manipulate the amount of mental imagery between participants in order to establish what role mental imagery plays in subjective reading experiences.

2.1.3. Individual (Trait) Differences in Reading Experiences

As mentioned above, mental imagery is not the only factor that plays a role in reading experiences. Variation in experienced absorption and appreciation can also be due to individual differences in situational factors (e.g., stress, mood, distractions, level of energy), or more stable characteristics. For example, amount of print exposure is negatively associated with reading difficulties (Stanovich & West, 1989), positively associated with reading skills (Acheson, Wells, & MacDonald, 2008) and positively associated with language ability, school success, Theory of Mind and empathy (Brysbaert, Sui, Dirix, & Hintz, 2020). Additionally, reading habits in daily life are closely related to reading experiences: more habitual readers experience more absorption (Kuijpers, Douglas, & Kuiken, 2018) and enjoy reading more (Mol & Jolles, 2014). Furthermore, a range of personality characteristics have been found to be associated with reading experiences. Individuals reporting more need for affect (M. Appel & Richter, 2010), as well as more transportable individuals (Bilandzic & Busselle, 2011), reported experiencing more transportation while reading a story or watching a film. Similarly, Need for Cognition was found to be a predictor of transportation experienced while reading stories (Green et al., 2008) or watching films (Hall & Zwarun, 2012). Openness was positively associated with reported interest for stories (which is related to enjoyment/appreciation of stories; Fayn, Tiliopoulos, & MacCann, 2015), absorption (although indirectly, via reading habits; Kuijpers et al., 2018), and the overall likelihood that people read literature for leisure (Kraaykamp & Van Eijck, 2005; see also Schutte & Malouff, 2004). Interestingly, Malanchini et al. (2017) link differences in reading enjoyment and motivation to genetic differences.

Because of the important role of the abovementioned individual (trait) differences in reading experiences, I took these into account in the experiments reported here. Because I controlled for the role of these individual (trait) differences in reading experiences when studying the role of guided mental imagery instructions in reading experiences, I was able to draw conclusions about the role of mental imagery instructions over and above individual (trait) differences from the results of the experiments.

2.1.4. The Current Experiments

In the experiments reported in the current chapter I investigated the respective roles of mental imagery and individual differences on subjective experiences during reading. I tested in two experimental studies whether guided mental imagery instructions influenced reading experiences. In keeping with the suggestion made by Green (2004), I tested the effect of pre-reading instructions specifically focusing on mental imagery in the experiments reported in this Chapter.

In the first experiment participants read a literary short story and subsequently rated their reading experiences on several questionnaires. One group of participants was instructed to use mental imagery while reading the stories, whereas another group of participants was asked to read the stories for leisure. In the second experiment a third instruction was added, designed to distract participants from the plot of the story. This third instruction was added to control for a task confound in the first experiment, where the imagery instruction was more effortful to follow than the leisure instruction. The task in the third instruction was as effortful to complete as the task in the imagery instruction, but it was designed to distract participants from the plot of the story, and therefore to make their mental imagery less vivid. With this experiment, I wanted to test whether overall reading experiences can be modified using reading instructions focusing on one specific facet of these reading experiences and whether these reading instructions can "overrule" the influence of individual (trait) differences (e.g., reading habits, personality characteristics) on reading experiences. Based on the literature, I hypothesized that mental imagery instructions would result in more mental imagery compared to the control group (leisure readers) and therefore in more absorption and appreciation. In contrast, I hypothesized that the distracting instruction added in the second experiment would result in less mental imagery and therefore in reduced absorption and appreciation.

If specific reading instructions would indeed prove powerful in altering reading experiences these could be used to promote reading in people who do not read for leisure, which could have positive consequences for among others school success (Chiu & McBride-Chang, 2006; Mol & Jolles, 2014; Retelsdorf, Köller, & Möller, 2011), second language learning (Lao & Krashen, 2000; Lee, Schallert, & Kim, 2015; Yamashita, 2008), social cognition and empathy (e.g., Fong, Mullin, & Mar, 2013; Johnson et al., 2013; Mar & Oatley, 2008; Oatley, 2016), or persuasion (e.g., Dal Cin, Zanna, & Fong, 2004; Green & Brock, 2000). If, however, other factors (e.g., individual trait differences) are found to be a stronger driver of absorption and appreciation than imagery instructions, this would indicate that using such instructions in for instance educational settings is not an optimal intervention to increase reading pleasure. A third option would be that individual trait differences and imagery instructions interact as drivers of absorption and appreciation. In that case, this could indicate that using imagery instructions would only be useful for some individuals and that it would be necessary to find a way to determine which individuals would or would not benefit from such instructions.

2.2. Experiment 1

2.2.1. Methods

This first experiment was conducted in the context of Bachelor's theses, for which five students of Communication and Information Studies at the Radboud University in Nijmegen, The Netherlands, worked together under the supervision of the first author to conduct an experiment testing the influence of mental imagery-inducing reading instructions on reading experiences.

2.2.1.1. Participants

A total of 120 participants took part in this first experiment. To ensure that participants understood their instructions for the experiment they were asked to repeat what they had been instructed to do while reading (i.e., mental imagery versus reading for leisure) after reading the story. Due to an error during data collection, the experimental condition of 20 participants was not registered correctly. Data from these 20 participants were excluded from analysis. It was double checked that the data from the remaining 100 participants were registered correctly before moving on with the analyses. The remaining participants consisted of an experimental group (n = 45; 25 females; Age: *M(SD)* = 32 (15) years old; age range = 19–71) and a control group (n = 55; 33 females; Age: *M(SD)* = 33 (16) years old; age range = 17–82). Chi-square tests indicated that there were no significant differences between the participants in the two groups with respect to gender ($\chi^2(1) = 0.06$; $p = .81$) and educational level ($\chi^2(5) = 10.06$; $p = .07$). However, to control for any possible individual differences in age, gender or educational level (measured as the highest completed education, ranging from 1 to 6, where 6 was the highest possible level of education in the Netherlands), I considered these factors as covariates in the analyses.

Prior to the experiment, participants were informed about the procedure of the experiment. It was made clear that participation was voluntary and that participants were allowed to withdraw from the experiment at any time without need for explanation. All participants gave written informed consent in accordance with the Declaration of Helsinki. The experiment was approved by the local ethics committee.

2.2.1.2. Materials

Materials used in Experiment 1 consisted of the story that was read, the instructions participants received before reading, and the questionnaires participants filled in after reading. I will now discuss these materials in more detail

2.2.1.2.1. Story The story used in the first experiment was an existing literary short story by the acclaimed Dutch writer Rob van Essen (2014), called *De mensen die alles lieten bezorgen* (The people who had everything delivered). The story was 2988 words long and took participants about 10–15 minutes to read. The story recounted the experiences of a man who lives in an apartment building in Amsterdam. His neighbors rent out their apartment while they are on holiday for the Christmas days, and a morbidly obese British couple stays there. When the wife has a heart attack she has to be lifted out of the apartment by a firetruck, as there is no elevator in the building. The events in the story were narrated using very descriptive language and were easy to visualize.

2.2.1.2.2. Instructions Before reading the stories, participants received a reading instruction. The experimental group was instructed to "*Pay close attention to the story and read the story as you would normally read a story. Use your imagination while reading, by visualizing the surroundings described in the story and envisioning the actions of the characters. Imagine the main character standing in front of you, imagine what happens and pay close attention to what all of the characters are doing*". The control group was simply instructed to "*Pay close attention to the story and read the story as you would normally read a story*". This way both groups were instructed to read the story attentively, but the experimental group was encouraged to form a vivid mental image of the events described in the story. When participants had finished reading, they were asked to repeat their reading instruction to the experimenter. If they were unable to repeat the instruction correctly or if their answer indicated they did not use the instruction while reading the story, data for these participants were excluded from the analysis.

2.2.1.2.3. Questionnaires After reading, participants had to fill in questionnaires measuring their reading experiences (i.e., Story World Absorption, see Kuijpers et al., 2014), reading habits in daily life, and other more general information (i.e., age, level of education, and gender). Since this experiment was conducted in the context of Bachelor's theses, a couple of questionnaires were devised by the Communication and Information Studies students regarding topics not under study here (e.g., attitudes towards fast food, behavioral intentions regarding healthy eating). Since I did not have specific theoretical assumptions regarding these topics, I will not discuss these measures in the current chapter. Story world absorption was measured using the Story World Absorption Scale (SWAS; Kuijpers et al., 2014). The SWAS is a validated scale consisting of 18 items with high internal validity (Kuijpers et al., 2014) which measures 4 aspects of story world absorption on the four subscales Attention, Transportation, Emotional Engagement and Mental Imagery (e.g., *When I finished the story I was surprised to see that time had gone by so fast; I could imagine what the world in which the story took place looked like*). Participants rated each question on a 7-point scale (1 = disagree, 7 = agree). Reading habits were measured using five multiple choice questions about reading habits in everyday life, with 4 or 5 optional answers (Hartung et al., 2016; Mak & Willems, 2019; e.g., *How often do you read fiction; How many books do you read each year*). Additionally, participants were asked for their genre preference in an open-ended question, where they could list up to three preferred genres.

2.2.1.3. Procedure

Participants were recruited and tested in a quiet room at the university campus or at home. Participants were informed about the procedure and were asked for written informed consent. At the start of the experiment, participants were given one of the two possible instructions on paper (as described above). If necessary, the instruction was clarified by the experimenter. After that, participants read the story at their own pace. Both the instructions and the story were read from paper. After reading, participants were asked to fill in the questionnaires (SWAS, reading habits, general information) on the experimenter's laptop. The entire procedure took about 20–25 minutes.

2.2.1.4. Data Analysis

Data analysis was done using the *'stats'* package in R version 3.6.1 (R Core Team, 2021). I constructed a linear regression model that predicted average scores on the SWAS based on experimental group (imagery instruction contrasted with control). Gender (male contrasted with female), age, level of education, and reading habits were added to the model as general variables expected to explain additional variance. As a result, any effects of the experimental group would represent variance explained by the given instruction over and above variance explained by these demographic variables. Similar models were constructed to predict scores on the four subscales of the SWAS separately (i.e., Attention, Transportation, Emotional Engagement and Mental Imagery). All continuous predictors were centered and scaled. Variance Inflation Factors (VIFs) were calculated for all models, to check for multicollinearity between predictors. All VIFs for all models were between 1 and 2, indicating that multicollinearity was not problematic in the models and all planned predictors could be entered into the models.

2.2.2. Results

2.2.2.1. Questionnaires

To test the reliability of the used scales and subscales, Ω_t was calculated. I decided to calculate Ω as opposed to Cronbach's α, since it has been argued by several researchers that Ω_t is a more appropriate measure of reliability (e.g., Dunn, Baguley, & Brunsden, 2014; Peters, 2014; Revelle, 2014). I decided to report Ω_t as opposed to Ω_h because I assumed the constructs measured with my scale to be multidimensional (e.g., Revelle, 2014)[4]. The average scores on the Story World Absorption Scale as well as the scores on the four subscales of the SWAS all showed sufficient to excellent reliability; total SWAS (18 items), Ω_t = .95; SWAS Attention (5 items), Ω_t = .87; Transportation (5 items), Ω_t = .84; Emotional Engagement (5 items), Ω_t = .92; Mental Imagery (3 items), Ω_t = .79. Descriptive statistics for the questionnaire scores per subscale and per group are visualized in Fig. 2.1. The answers on the reading habits questionnaire were measured on a scale ranging from 1 to 5 on four of the five multiple choice questions, but from 1 to 4 on the final question. Therefore, z-scores were calculated for all questions on this questionnaire (with higher scores for more habitual readers). Overall reliability was sufficient, Ω_t = .78.

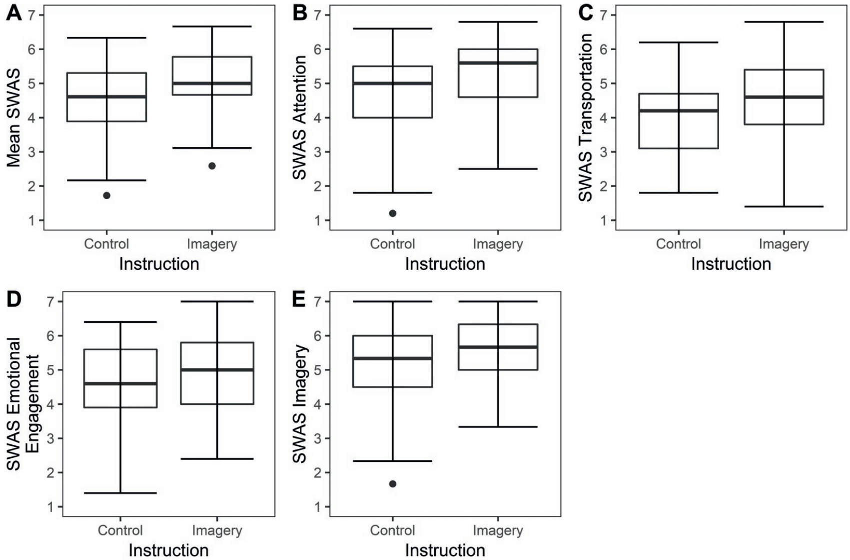
2.2.2.2. Main Analysis

The model predicting average scores on the SWAS based on gender, age, level of education, reading habits and experimental group (imagery instruction versus control; model adjusted $R^2 = 0.185$) showed that participants in the experimental group were more absorbed than participants in the control group (see Table

[4]For these same reasons, Ω_t is reported as a measure of reliability throughout this dissertation.

2.1 for all results of this model; see Fig. 2.1A). Additionally, participants who were more habitual readers, also reported more story world absorption. Finally, males reported less story world absorption than females and participants with a higher level of education were more absorbed.

Figure 2.1.: *Boxplots Depicting Scores on the SWAS, per Subscale and per Group*



To find out which aspects of story world absorption this difference between the experimental and control group stems from, similar models were built to predict the scores on the four subscales of the SWAS (Attention, Transportation, Emotional Engagement, Mental Imagery). The model predicting the scores on the Attention subscale of the SWAS (model adjusted $R^2 = 0.172$; see Table 2.1A), showed that participants in the experimental group read the story more attentively (see Fig. 2.1B). More habitual readers in daily life also reported more Attention to the story and males reported less Attention than females.

For the Transportation subscale of the SWAS (model adjusted $R^2 = 0.170$; see Table 2.1b), it was found that participants in the experimental group experienced more transportation than participants in the control group (see Fig. 2.1C). Participants with a higher level of education also reported experiencing more transportation into the story.

The model predicting scores on the Emotional Engagement subscale of the SWAS (model adjusted $R^2 = 0.118$; see Table 2.1c) revealed no differences be-

Table 2.1.: *Coefficients of the Models Predicting Absorption Based on Type of Instruction (Mental Imagery Instruction Contrasted with the Control Instruction), Taking into Account Gender (Male Contrasted with Female), Age, Self-Reported Reading Habits, and Level of Education. Significant Predictors are Marked (* p <.05, ** p <.01, *** p <.001)*

		B	*SE*	*t*-value	*p*-value
1. SWAS	(Intercept)	4.65	0.14	33.11	<.001***
	Imagery Instruction	0.58	0.18	3.18	.002**
	Reading Habits	0.32	.14	2.24	.03*
	Age	0.06	0.09	0.70	.49
	Gender (Male)	-0.43	0.18	-2.39	.02*
	Level of Education	0.21	0.09	2.30	.02*
1a. SWAS Attention	(Intercept)	4.86	0.16	29.45	<.001***
	Imagery Instruction	0.62	0.21	2.87	.005**
	Reading Habits	0.47	0.17	2.83	.006**
	Age	-0.01	0.11	-0.05	.96
	Gender (Male)	-0.44	0.21	-2.05	.04*
	Level of Education	0.20	0.11	1.83	.07
1b. SWAS Transportation	(Intercept)	3.99	0.17	23.90	<.001***
	Imagery Instruction	0.62	0.22	2.84	.005**
	Reading Habits	0.21	0.17	1.23	.22
	Age	0.02	0.11	0.17	.87
	Gender (Male)	-0.31	0.22	-1.45	.15
	Level of Education	0.43	0.11	3.95	<.001***
1c. SWAS Emotional Engagement	(Intercept)	4.77	0.18	26.25	<.001***
	Imagery Instruction	0.45	0.24	1.90	.06
	Reading Habits	0.38	0.18	2.10	.04*
	Age	0.15	0.12	1.24	.22
	Gender (Male)	-0.58	0.23	-2.46	.02*
	Level of Education	0.09	0.12	0.80	.43
1d. SWAS Mental Imagery	(Intercept)	5.22	0.17	30.01	<.001***
	Imagery Instruction	0.68	0.23	3.01	.003**
	Reading Habits	0.13	0.17	0.76	.45
	Age	0.11	0.11	1.02	.31
	Gender (Male)	-0.39	0.23	-1.72	.09
	Level of Education	0.06	0.11	0.54	.59

tween the experimental and control group with regard to emotional engagement (see Fig. 2.1D). Males were less emotionally engaged than females. Additionally, more habitual readers reported to be more emotionally engaged with the story.

In support of the effectiveness of the manipulation, participants in the experimental group reported more mental imagery than participants in the control group (see Table 2.1d; see Fig. 2.1E) on the Mental Imagery subscale of the SWAS (model adjusted $R^2 = 0.084$). None of the other tested predictors (gender, age, level of education, reading habits) were significantly related to scores on the Mental Imagery subscale of the SWAS.

2.2.3. Discussion

The results from this first experiment show that mental imagery-inducing reading instructions were associated with a stronger absorption experience, in particular a stronger attention towards the story, a stronger experience of transportation into the story world and more reported use of mental imagery (confirming the effectiveness of the manipulation). This suggests that, indeed, pre-reading instructions focusing on specific aspects of reading (such as mental imagery) can influence the way readers experience stories.

Apart from the influence of reading instructions, I found substantial individual differences in reading experiences. For instance, females reported to be more absorbed by the story. When looking at the subcomponents of absorption, it became clear that this difference between males and females was most prominent in the emotional engagement component of absorption: women were more emotionally engaged when reading the story than men. Level of education also appeared to explain some of the variation in absorption. This was mostly visible in the transportation subcomponent of absorption: participants with a higher level of education, reported experiencing more transportation into the story world. Habitual readers were also more absorbed than participants who did not read much in their own time. This was visible in their scores on the attention and emotional engagement subcomponents of absorption: more habitual readers reported more attention to the story and more emotional engagement with the story. Age was not significantly related to the absorption experience, nor to any of its subcomponents, as has also been found in previous work (Hartung, Withers, Hagoort, & Willems, 2017).

From this first experiment it could be concluded that there are indeed individual differences in reading experiences, which are related to both mental imagery-inducing reading instructions and stable individual differences (i.e., gender, level of education, reading habits). However, a few questions remain after this experiment. First and foremost, a difference in reading experiences between the experimental group and the control group could mean two things. Although it is very well possible that absorption was enhanced through induced mental imagery as a result of the mental imagery reading instruction in the experimental group, it could also be the case that elaborate reading instructions in general promote more intensive reading and as a result a stronger experience of absorption (in both reading instructions used in this experiment, participants were told to read the story attentively, but only in the imagery instruction participants were asked to vividly imagine the events happening in the story on top of reading the story attentively). That is, in this experiment, the imagery group got more elaborate (longer) instructions than the control group and this could have led to the observed differences (see De Koning & Van der Schoot, 2013, for a similar argument). Secondly, the reading instruction was aimed at mental imagery, which is part of the absorption experience. Although the effect of the reading instruction did translate to other subcomponents of absorption (attention, transportation), it would be interesting to find out if a mental imagery reading instruction could also influence other reading experiences, most notably story appreciation. Third, I did look at some general individual differences that could influence reading experiences, such as gender, age and reading experience, but I did not look at any personality characteristics that have been associated with reading experiences before. In the next experiment I have therefore considered the personality characteristics fantasy (how imaginative a person is) and perspective taking (the extent to which someone takes other peoples' perspectives in daily life).

To be able to answer these remaining questions, I ran a second experiment, in which I made the mental imagery instruction more clear (to improve compliance with the instruction) and included a third instruction, which was as elaborate as the mental imagery instruction, but its content was aimed at diverting the reader from the narrative's plot (but still encouraging thorough reading of the text; adapted from an instruction successfully used to lower experienced transportation; Green & Brock, 2000). This way, I could rule out that the effect of instruction was simply the result of more intensive reading instead of being related to the actual content of the instruction. To elaborate, I added this third instruction to test the alternative explanation for the results from experiment 1: that elaborate, detailed reading instructions in general promote more intensive reading and as a result a stronger experience of absorption. If both detailed instructions would increase absorption, this would support the alternative hypothesis that this increase in absorption is the result of more thorough reading due to the length or details in the instructions. However, if only the imagery instruction would increase absorption and the other detailed instruction would not change or even decrease absorption, I would have stronger evidence for the hypothesis that the increase in absorption after the imagery instruction is the result of the *content* of the instruction. For this third instruction (that was as detailed as the imagery instruction), I chose an instruction aimed at *decreasing* absorption since using a second detailed instruction that was aimed at *increasing* absorption would not have been helpful: if in this case both detailed instructions would increase absorption, I would still not know whether this was due to the content of the instructions, and not simply to the fact that they were detailed instructions that encouraged thorough reading.

I also included measures for story appreciation, a more thorough measure of mental imagery, and measures of personality traits I thought might play a role in reading experiences. Furthermore, to ensure more experimental control I conducted the second experiment in a more controlled environment. Finally, I decided to use two new stories to find out if the effect of reading instructions would also extend to different stories.

2.3. Experiment 2

2.3.1. Methods

In this experiment participants were divided into three groups. Apart from the group receiving an elaborate mental imagery instruction and the control group, I also included a group that received the instruction to judge whether the writing style of the story (sentence construction, word use) was suitable for teenagers of about 14 or 15 years old, who were in the lower grades of Dutch secondary education (henceforth called secondary school suitability instruction; cf. Green & Brock, 2000). Such "distraction manipulations" have in previous studies been particularly useful in manipulating transportation (for a review, see Tukachinsky, 2014). If the *length* of the imagery instruction was the reason people became more transported in the first experiment, this secondary school suitability instruction should also result in increased absorption and there should be no difference between the results for the imagery instruction and the suitability instruction. However, if the effect of the instruction was due to the content of the instruction, the imagery instruction should increase absorption but the suitability instruction should decrease absorption. This enabled me to test if an effect of instruction was the result of the actual content of the instruction or, alternatively, was simply the result of more intensive reading.

2.3.1.1. Participants

To ensure sufficient statistical power, the appropriate sample size for this experiment was calculated in G-power (Faul, Erdfelder, Buchner, & Lang, 2009) using the effect size from experiment 1. This resulted in a required sample size of approximately 120 for an estimated power of .85, divided over 3 groups. A total of 125 participants (102 females) participated in the second experiment. The data of 7 participants had to be discarded because of procedural errors (4), too much missing data (2), or because they did not have enough time to finish the experiment. The remaining 118 participants (99 females) were split up in a group receiving a mental imagery instruction ($n = 39$; 32 females; M age = 24 years old), a group receiving a secondary school suitability instruction ($n = 39$; 35 females; M age = 23 years old) and a control group ($n = 40$; 32 females; M age = 24 years old). There were no differences between the participants in the three groups with respect to gender ($\chi^2(2) = 3.07$; $p = .22$), nor were there differences between groups in age ($F(2, 233) = 0.77$, $p = .46$), reading habits (self-report: $F(2, 233) = 0.17$, $p = .84$; ART-score: $F(2, 233) = .21$, $p = .81$) or personality characteristics (IRI Fantasy: $F(2, 233) = 2.81$, $p = .06$; IRI Perspective Taking: $F(2, 233) = 0.43$, $p = .65$; see below for an extensive description of all used questionnaires). Participants were all healthy, native speakers of Dutch, without dyslexia.

Participants were recruited from the participant database of the Radboud University and received €10 or course credit for participation in this experiment. Prior to the experiment, participants were informed about the procedure of the experiment. It was made clear that participation was voluntary and that it was allowed to withdraw from the experiment at any time without need for explanation. All participants gave written informed consent in accordance with the Declaration of Helsinki and the experiment was approved by the local ethics committee.

2.3.1.2. Materials

Materials used in Experiment 2 consisted of the two stories that were read, the instructions participants received before reading, and the questionnaires participants filled in after reading. I will now discuss these materials in more detail.

2.3.1.2.1. Stories Instead of just one story, participants read two stories in the second experiment. Both were literary short stories written by acclaimed Dutch writers. The first story, *Brommer op zee* (Moped at sea), was written by Maarten Biesheuvel (1972) and was 1827 words long. It is a surrealistic story about a boy on a boat and his encounter with a man riding a moped at sea in the middle of the night. The second story, *God en de gekkenrechter* (God and the judge of the insane), was written by Adriaan van Dis (1986) and was 2026 words long. In this story, the author narrates the story of a mentally instable man who is convinced that he is God, and believes that therefore all his excrements are holy and should not be thrown away. Apart from that, he terrorizes the neighborhood, leading to his institutionalization later on in the story, after which he finally seems to realize that he was mistaken in thinking that he was God. Both stories contain many descriptions that could guide mental imagery of the stories. It took participants about 15 minutes to read each story. In the remainder of this chapter, the story *Moped on sea* will be referred to as Story A, and *God and the judge of the insane* will be referred to as Story B. Note that the stories were read in counterbalanced order: Half of the participants started with Story A and half of the participants started with Story B.

2.3.1.2.2. Instructions Before reading the stories, participants were given a reading instruction. Every participant received either a mental imagery instruction, a secondary school suitability instruction, or a neutral control instruction. After participants had read the first story, they received a short reminder of their reading instruction to urge them to also keep the instruction in mind while reading the second story. The group receiving the mental imagery instruction was told "I*n a short while, you will be reading a short story. During reading, try to vividly imagine the events happening in the story. Vividly imagine what you see, hear, feel or smell. For example, envision the characters and places described in the story, imagine what the conversations and environmental sounds sound like, what the odors smell like, how the physical experiences of the characters feel*". The group receiving the secondary school suitability instruction was told "*In a short while, you will be reading a short story. Your job is to make sure the text is suitable for students in the lower grades of secondary school, of about 14 or 15 years old. The content of the story is not important, please pay attention to the writing style: the sentence constructions and the word use of the author of the story. Try to focus on these two aspects while reading the story. Determine whether the word use and sentence constructions are of a suitable level for students in the lower grades of secondary school*". The control group, who received a short, neutral instruction, was told "*In a short while, you will be reading a short story. Please read this story the way you would usually read a story for your leisure*". To make sure all participants understood their reading instruction, they were presented with a short excerpt from a different story (which was stylistically comparable to the two experimental stories). Participants had to apply the reading instruction while reading this fragment and were afterwards asked to check with themselves whether they indeed applied the reading instructions while reading. After participants had practiced the instruction on the example fragment the instructions were repeated and participants were told to start reading the stories. Participants were reminded that they were allowed to read at their own pace and did not have to hurry. As mentioned previously, participants were given another reminder of the reading instruction just before they started reading the second story. The motivation behind this reminder was that participants had to fill in some questionnaires about their reading experience after reading the first story. Therefore, I wanted to make sure they would still remember the instruction while reading the second story. Although participants received different instructions, they all practiced their instruction on the same fragment and received reminders of their instruction at the same moments. This way, I tried to make sure that all three groups would read the stories equally attentively.

2.3.1.2.3. Questionnaires Just as in experiment 1, I asked participants to fill in questionnaires (discussed in detail below) regarding their reading experience after reading each story. However, this time I did not only use the Story World Absorption Scale to measure reading experience, but also a questionnaire measuring story appreciation and a questionnaire measuring the vividness of the imagery experienced during reading (for an overview of all questionnaires used in this experiment, see Fig. 2.2 and Table 2.2). After reading both stories and filling in the story-related questionnaires, participants were asked to fill in some additional questionnaires measuring reading habits in daily life, personality characteristics, story comprehension and more general information (i.e., age and gender). Level of education was not considered in this experiment as nearly all of the participants were university students and therefore no claims could be made about the role of level of education in reading experiences from this experiment. I will discuss the questionnaires that were not used in experiment 1 in more detail below (see section 2.2.1.2.3. for a detailed discussion of the other questionnaires).

As mentioned in the introduction of this chapter, I looked at appreciation of the stories. I measured this with the Appreciation Questionnaire, which is previously described in Mak and Willems (2019) and consists of a general score of story liking (*How did you like the story*; 1 = It was very bad, 7 = It was very good) and twelve adjectives (e.g., [*did you find the story*] *Beautiful, ... Ominous*) that could be used to describe the stories (adapted from Knoop et al., 2016). These adjectives are taken from a list of adjectives that were most often used by people to describe their opinion of poetry and which are also used to describe

Figure 2.2.: *Schematic Overview of the Procedure of Experiment 2*

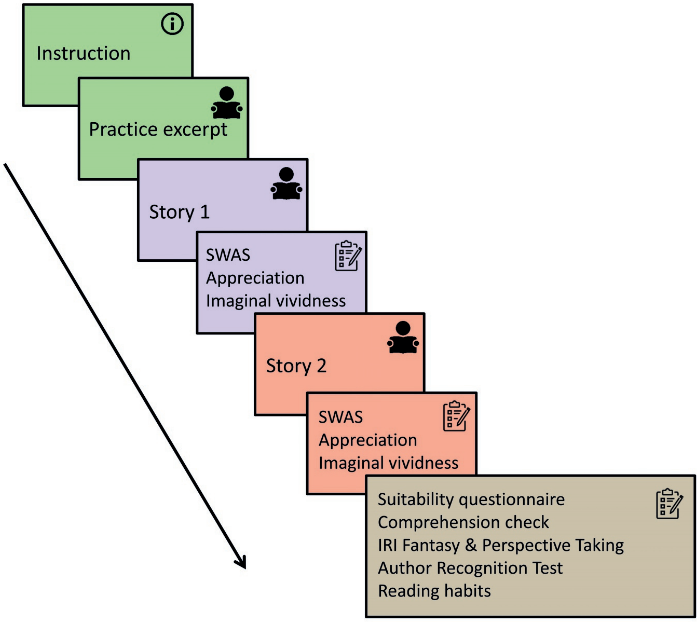


Table 2.2.: *Hierarchy of the Dependent Variables, Including Measuring Instruments (and Subscales) Used in Experiment 2*

Construct	Measuring Instrument	Variable of Main Interest	Subscales (If Applicable)
Absorption	Story World Absorption Scale (SWAS)	1. Mean SWAS	1a. SWAS Attention 1b. SWAS Transportation 1c. SWAS Emotional Engagement 1d. SWAS Mental Imagery
Vividness of Mental Imagery	Imaginal Vividness Scale (IVS)	2. Mean IVS	2a. IVS Character 2b. IVS Setting
Appreciation	Appreciation Questionnaire	3a. General Appreciation 3b. Interest 3c. Emotional Response 3d. Amusement 3e. Suspense	

aesthetic appeal in the domain of literature (Knoop et al., 2016). In the original scale by Mak and Willems (2019), a thirteenth adjective (*[did you find the story] Entertaining*) was used, but the Dutch word used for "entertaining" (in Dutch, "onderhoudend") appeared to be unknown to some of the participants, and was therefore removed from the questionnaire in the present experiment. Finally, 6 questions are asked regarding the enjoyment of the story (from Kuijpers et al., 2014; e.g., *I was constantly curious about how the story would end; I thought the* *story was written well*). Participants rated both the adjectives and the questions regarding enjoyment on a 7-point scale (1 = disagree, 7 = agree).

Vividness of experienced imagery was measured using a slightly adapted version of the Imaginal Vividness Scale (IVS; Fialho, n.d.), which is partly based on the Literary Response Questionnaire (Miall & Kuiken, 1995) and partly on a series of in-depth interviews with readers. This questionnaire was used because it is a more elaborate measure of imagery than the imagery subscale of the SWAS (which consists of only three items; and which is mainly focused on visual aspects of mental simulation, instead of the multisensory mental simulation I wanted to investigate, as explained in the introduction). The IVS as used in this experiment consisted of a total of 15 items divided over two subscales: Character (7 items, e.g., *While reading this story I could see the events happening in the story through the eyes of the main character; While reading the story I could almost feel the physical experiences of the characters in my own body*) and Setting (8 items, e.g., *While reading this story I often saw the described places so clearly, that it almost was as if I was there; I sometimes had auditory experiences (for example, hearing sounds) as if I was present in the world of the story*). This captured the quality of the imagery experienced by the participants in more detail.

After reading both stories and finishing the story-related questionnaires all participants answered a suitability questionnaire asking about the suitability of the text for 14–15-year-olds and a comprehension check. The questions on the suitability questionnaire were asked in such a way, that the questionnaire would not feel "out of the blue" for participants in the control or imagery groups.[5] The suitability questionnaire was simply a follow-up on the secondary school suitability instruction, where participants had to answer three questions about the suitability of the word use, sentence constructions, and the general suitability of the story for students in the lower grades of secondary school, of about 14 or 15 years old. The comprehension check consisted of three multiple choice questions per story (3 possible answers per question), that should have been possible to answer correctly for people who had read the stories with normal attention. Participants who answered two or more questions of the comprehension check incorrectly for one or both stories, were excluded from analysis.

[5] Some participants still found these questions "out of the blue". However, the majority of the participants did not report being surprised by these questions. Furthermore, this questionnaire was only presented at the end of the experiment, after all the questionnaires regarding reading experiences had already been filled in, and will therefore not have affected answers on the questionnaires that were of interest for the analyses.

Finally, participants were asked to report their reading habits and some personality characteristics. Reading habits were measured using the same questionnaire as used in the first experiment. Additionally, as an implicit measure of reading habits, participants completed the Author Recognition Test (ART; Stanovich & West, 1989; Dutch adaptation reported in Koopman, 2015), consisting of 42 names (30 real authors and 12 foils), where they had to indicate who they thought were genuine authors.

Personality characteristics were measured using the Fantasy and Perspective Taking subscales of the Interpersonal Reactivity Index (IRI; Davis, 1980; Dutch translation adapted from De Corte et al., 2007) on a 7-point scale (e.g., Becoming extremely involved in a good book or movie is somewhat rare for me; When I'm upset at someone, I usually try to "put myself in his shoes" for a while). The Fantasy subscale (which is conceptually related but not identical to the Transportability Scale; Dal Cin et al., 2004) measures the extent to which someone gets mentally very involved in the stories they encounter, to the point at which they imagine themselves being part of the story. The Perspective Taking subscale measures the extent to which someone is able to take another person's perspective in daily life.

2.3.1.3. Procedure

Participants were tested in small groups in lecture rooms at the university campus. One or two experimenters were always present to make sure the participants did not disturb each other and (if necessary) to answer questions. Before the start of the experiment, participants were informed about the procedure and were asked for written informed consent. At the start of the experiment, participants were given one of the three instructions (as described above; see Fig. 2.2 for a schematic overview of the procedure of this experiment). After having read the reading instruction and having practiced the instruction on an excerpt from an unrelated story not used in the remainder of the experiment, they started reading the first story. After reading, participants filled in the Story World Absorption Scale, the Appreciation Questionnaire, and the Imaginal Vividness scale. When they had finished, the reading instruction was repeated and participants read the second story. After finishing reading the second story participants completed the SWAS, Appreciation Questionnaire and IVS again, followed by the remaining questionnaires. The stories were read in counterbalanced order. Both the instructions and the stories were read from paper and the questionnaires were completed as paper and pencil tests. Participants were allowed to read the stories and fill in the questionnaires at their own pace. The entire procedure took about 40 minutes.

2.3.1.4. Data Analysis

Data analysis was done using the package *'lme4'* (Bates, Mächler, Bolker, & Walker, 2015) in R version 3.5.1 (R Core Team, 2021). I constructed a linear mixed effects regression model that predicted average scores on the SWAS based on experimental group (mental imagery instruction and secondary school suitability instruction contrasted with control). Random intercepts were allowed per participant.[6] Story (Story B contrasted with Story A), gender (male contrasted with female), age, self-reported reading habits, ART-score, and the Fantasy and Perspective taking subscales of the IRI were added to the model as general variables expected to explain additional variance. As a result, any effects of the experimental group would represent variance explained by the given instruction over and above variance explained by any story effects, demographic variables and other important individual difference measures. *P*-values were estimated using the "lmerTest" package (Kuznetsova, Brockhoff, & Christensen, 2017).

An initial effect of Experimental Group was calculated by comparing a base model (a model containing all predictors except for Experimental Group) with the full model, using an ANOVA. If this indicated a significant effect of experimental group, post-hoc pairwise comparisons (with Tukey HSD adjustment for multiple comparisons) between all three groups were made using the 'emmeans' package in R (Lenth, 2020). All continuous predictors were centered and scaled. Variance Inflation Factors (VIFs) were calculated for all models to check for multicollinearity. All VIFs for all models were between 1 and 2, indicating that multicollinearity was not problematic in the models and all planned predictors could be entered into the models.

Similar models were constructed to predict scores on the other variables of main interest (i.e., the average scores on the IVS as a more elaborate measure of mental imagery, scores on the different components of appreciation; see Fig. 2.2). Additionally, I also ran analyses for the four subscales of the SWAS separately (i.e., Attention, Transportation, Emotional Engagement and Mental Imagery), and the subscales of the IVS. I analyzed the data at the level of these subscales for two reasons. Firstly, the subscales of the SWAS and IVS are arguably

[6] I did not have enough observations in my dataset to support any random slopes (let alone a maximal model with random slopes for all predictor variables). Therefore, I decided that it was most appropriate to use a random intercept only model.

built up from subscales measuring diverging sub-constructs. Therefore, analyzing the subscales separately may provide additional information with regard to the processes underlying possible effects. The second reason for analyzing the data at the level of subscales in addition to the average scores was to find out whether the null-effect for Instruction I found when looking at the variables of main interest (i.e., the average scores on the SWAS and the IVS), would also be visible within all of the subscales.

2.3.2. Results

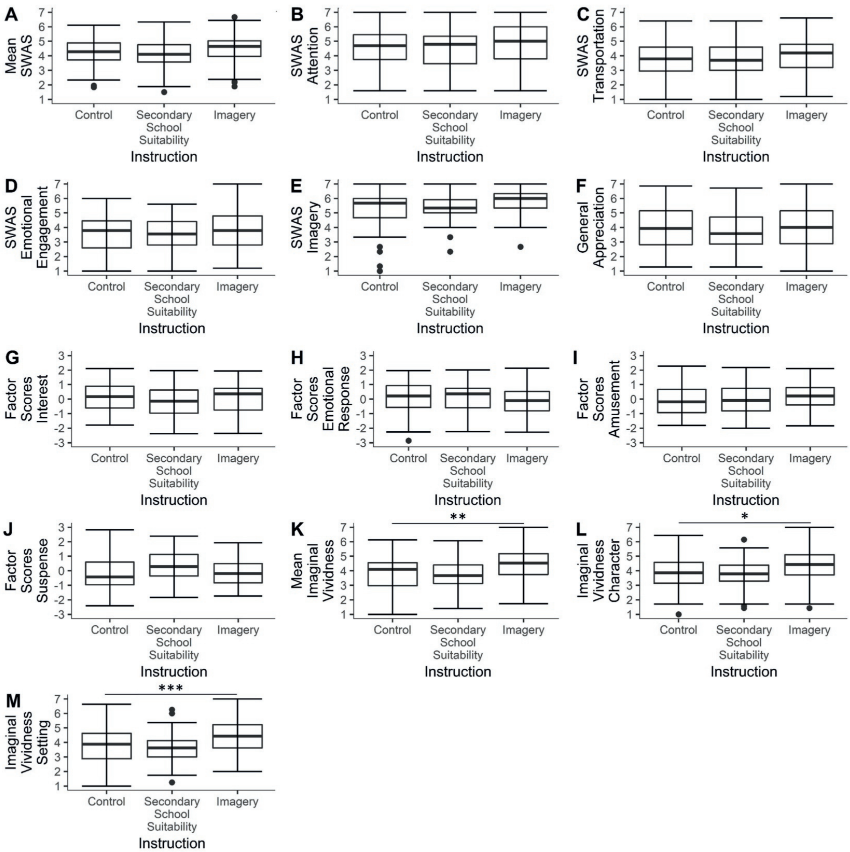
2.3.2.1. Questionnaires

The overall scores on the Story World Absorption Scale as well as the scores on the four subscales of the SWAS all showed good to excellent reliability; total SWAS (18 items), $\Omega_t = .95$; Attention (5 items), $\Omega_t = .91$; Transportation (5 items), $\Omega_t = .89$; Emotional Engagement (5 items), $\Omega_t = .92$; Mental Imagery (3 items), $\Omega_t = .77$. Descriptive statistics for the questionnaire scores per subscale and per group are visualized in Fig. 2.3.

The Appreciation Questionnaire was divided into two parts for the analysis. The first part, consisting of twelve adjectives that could be used to describe the stories, was analyzed using a principal components analysis (PCA) with oblique rotation (direct oblimin). The Kaiser-Meyer-Olkin measure was good, KMO = .86, indicating that the sampling adequacy for this analysis was good (all KMO values for individual items >.72). Bartlett's test of sphericity showed that there was sufficient correlation between items, $\chi^2(66) = 1558.99$, $p < .001$. An initial analysis showed that three components had eigenvalues over 1 (Kaiser's criterion), but a model with three components did not fit the data well enough (fit = .93). Therefore, in the final model four components were retained (fit = .95). This model explained 71% of the variance. The first component contained items measuring the evoked interest in the story (beautiful, boring (–), interesting, captivating, gripping, suspenseful), the second component contained items measuring the emotional response to the story (gripping, tragic, sad), the third component contained items measuring the suspense elicited by the story (ominous, suspenseful), and the fourth component contained items measuring the amusement elicited by the story (funny, witty, special).[7] The structure and pat-

[7] Note that the items "gripping" and "suspenseful" load strongly (above .40) on more than one component. This is due to the nature of PCA as an unsupervised dimension reduction method. Items loading strongly on a component are considered "typical" items for the component, and can be used for the interpretation of the components. However, every item will load on every

Figure 2.3.: *Boxplots Depicting Scores on all Questionnaires and Subscales, Displayed per Group. Asterisks indicate significant differences between groups on the post-hoc comparisons, * p < .05, ** p < .01, *** p < .001*



tern matrices for the factor loadings after rotation can be seen in Table 2.3. Factor scores per participant and story were used in the subsequent analyses for the constructs Evoked Interest, Emotional Response, Suspense, and Amusement.

The second part of the appreciation questionnaire consisted of a general score of story liking, and six questions regarding the enjoyment of the story, $\Omega_t = .95$

component, and the loadings of all items (both items strongly associated with the components and items weakly associated with the components) are taken into account when calculating the component or factor scores of all components. The strong association of the items "gripping" and "suspenseful" with more than one component therefore does not have important consequences for the calculation of the component scores, but only for the theoretical interpretation of the components.

Table 2.3.: *Pattern Matrix for the PCA of the 12 Adjectives on the Appreciation Questionnaire (N = 236). Factor Loadings over .40 Appear in Bold*

	Pattern Matrix			
	Evoked Interest	Emotional Response	Amusement	Suspense
Beautiful	**0.81**	0.09	0.01	-0.16
Boring	**-0.81**	-0.04	-0.08	0.17
Gripping	**0.53**	**0.43**	-0.03	0.14
Funny	0.02	-0.05	**0.89**	-0.12
Interesting	**0.71**	0.06	0.20	0.10
Ominous	-0.20	0.23	0.11	0.83
Sad	0.08	**0.85**	-0.05	-0.04
Suspenseful	**0.50**	-0.18	-0.11	**0.66**
Tragic	0.01	**0.85**	0.01	0.07
Witty	0.11	0.01	0.73	0.20
Captivating	**0.75**	0.03	0.17	0.18
Special	0.33	0.10	**0.40**	0.20

(7 items). The answers on these questions were collapsed into a mean score for General Appreciation.

The average scores on the Imaginal Vividness scale, as well as the scores on the two subscales, all showed sufficient to excellent reliability; total Imaginal Vividness (15 items), $\Omega_t = .93$; Character subscale (7 items), $\Omega_t = .86$; Setting subscale (8 items), $\Omega_t = .89$.

Reading experience was measured both directly using a reading habits questionnaire, and indirectly using the Author Recognition Test (ART). Because answers on different items of the reading habits questionnaire were measured on different scales, z-scores were calculated for all questions on this questionnaire (higher values indicating more habitual readers). Overall reliability was sufficient, $\Omega_t = .82$. The scores on the ART were slightly positively skewed ($M = 7.46$, $SD = 4.03$, median = 7.00, IQR = 5.00–10.00) with higher values indicating more (literary) reading experience.

Reliability of both subscales of the Interpersonal Reactivity Index was sufficient; Fantasy subscale (7 items), $\Omega_t = .84$, and Perspective Taking subscale (7 items), $\Omega_t = .84$.

2.3.2.2. Main Analysis

2.3.2.2.1. Story World Absorption The model predicting average scores on the SWAS based on story, gender, age, ART-score, reading habits, IRI Fantasy, IRI Perspective Taking and experimental group showed no differences in SWAS scores between the three experimental groups (see Fig. 2.3A; see Table 2.4.1 for all results of this model). Interestingly, participants with higher scores on the Fantasy subscale of the IRI reported more story world absorption (see Fig. 2.4A). To find out which aspects of story world absorption this relationship between IRI fantasy and scores on the SWAS stems from, similar models were constructed to predict the scores on the four subscales of the SWAS (see Fig. 2.4 for a visual representation of the relationship between scores on the Fantasy subscale of the IRI and the tested reading experiences).

Table 2.4.: *Coefficients of the Models Predicting Reading Experiences (Absorption, Vividness of Mental Imagery, Story Appreciation) Based on Type of Instruction (Mental Imagery Instruction and Secondary School Suitability Instruction Contrasted with the Control Instruction), Taking into Account Story (Story B Contrasted with Story A), Gender (Male Contrasted with Female), Age, Self-Reported Reading Habits, ART-Score, and the Fantasy and Perspective Taking Subscales of the IRI. Significant Predictors Are Marked (* p <.05, ** p <.01, *** p <.001)*

		B	*SE*	*df*	*t*-value	*p*-value
1. SWAS	(Intercept)	4.22	0.13	156.82	32.37	<.001***
	Imagery Instruction	0.17	0.16	118.00	1.01	.32
	School Instruction	-0.12	0.16	118.00	-0.71	.48
	IRI Fantasy	0.39	0.08	118.00	5.12	<.001***
	IRI Perspective Taking	-0.04	0.07	118.00	-0.60	.55
	Reading Habits	0.09	0.13	118.00	0.70	.49
	ART-Score	-0.09	0.08	118.00	-1.09	.28
	Age	0.06	0.08	118.00	0.77	.44
	Gender (Male)	0.30	0.19	118.00	1.60	.11
	Story (B)	0.03	0.10	118.00	0.28	.78
1a. SWAS Attention	(Intercept)	4.42	0.17	149.38	25.89	<.001***
	Imagery Instruction	0.10	0.22	118.00	0.48	.64
	School Instruction	-0.22	0.22	118.00	-1.00	.32
	IRI Fantasy	0.32	0.10	118.00	3.13	.002**
	IRI Perspective Taking	-0.02	0.09	118.00	-0.22	0.83
	Reading Habits	0.31	0.17	118.00	1.81	.07
	ART-Score	-0.13	0.11	118.00	-1.14	.26
	Age	0.03	0.10	118.00	0.32	.75
	Gender (Male)	0.28	0.25	118.00	1.12	.27
	Story (B)	0.47	0.12	118.00	3.95	<.001***
1b. SWAS	(Intercept)	3.65	0.16	147.16	22.12	<.001***

Transpor-tation	Imagery Instruction	0.16	0.21	118.00	0.74	.46
	School Instruction	-0.07	0.21	118.00	-0.34	.74
	IRI Fantasy	0.49	0.10	118.00	4.98	<.001***
	IRI Perspective Taking	-0.09	0.09	118.00	-1.00	.32
	Reading Habits	0.13	0.17	118.00	0.79	.43
	ART-Score	-0.14	0.11	118.00	-1.33	.19
	Age	0.11	0.10	118.00	1.10	.28
	Gender (Male)	0.45	0.25	118.00	1.85	.07
	Story (B)	0.15	0.11	118.00	1.39	0.17
1c. SWAS Emotional Engagement	(Intercept)	4.08	0.16	165.90	25.11	<.001***
	Imagery Instruction	0.17	0.20	118.00	0.86	.39
	School Instruction	-0.07	0.20	118.00	-0.38	.71
	IRI Fantasy	0.38	0.09	118.00	4.02	<.001***
	IRI Perspective Taking	-0.03	0.08	118.00	-0.32	.75
	Reading Habits	-0.06	0.16	118.00	-0.39	.70
	ART-Score	-0.08	0.10	118.00	-0.78	.44
	Age	0.01	0.09	118.00	0.07	.94
	Gender (Male)	0.22	0.23	118.00	0.95	.35
	Story (B)	-0.89	0.14	118.00	-6.55	<.001***
1d. SWAS Mental Imagery	(Intercept)	5.08	0.12	158.85	41.03	<.001***
	Imagery Instruction	0.27	0.16	118.00	1.75	0.08
	School Instruction	-0.09	0.15	118.00	-0.60	0.55
	IRI Fantasy	0.38	0.07	118.00	5.20	<.001***
	IRI Perspective Taking	-0.02	0.06	118.00	-0.30	.76
	Reading Habits	-0.09	0.12	118.00	-0.77	.44
	ART-Score	0.04	0.08	118.00	0.50	.62
	Age	0.11	0.07	118.00	1.50	.14
	Gender (Male)	0.23	0.18	118.00	1.29	.20
	Story (B)	0.62	0.10	118.00	6.39	<.001***
2. IVS	(Intercept)	3.82	0.14	143.53	27.73	<.001***
	Imagery Instruction	0.35	0.18	118.00	1.97	.051
	School Instruction	-0.29	0.18	118.00	-1.63	.11
	IRI Fantasy	0.47	0.08	118.00	5.69	<.001***
	IRI Perspective Taking	-0.08	0.07	118.00	-1.06	.29
	Reading Habits	0.12	0.14	118.00	0.86	.39
	ART-Score	-0.11	0.09	118.00	-1.15	.25
	Age	0.07	0.08	118.00	0.91	.36
	Gender (Male)	-0.22	0.21	118.00	-1.09	.28
	Story (B)	0.33	0.09	118.00	3.77	<.001***
2a. IVS Character	(Intercept)	3.94	0.15	145.97	26.31	<.001***
	Imagery Instruction	0.30	0.19	118.00	1.53	.13
	School Instruction	-0.22	0.19	118.00	-1.15	0.25
	IRI Fantasy	0.45	0.09	118.00	5.02	<.001***

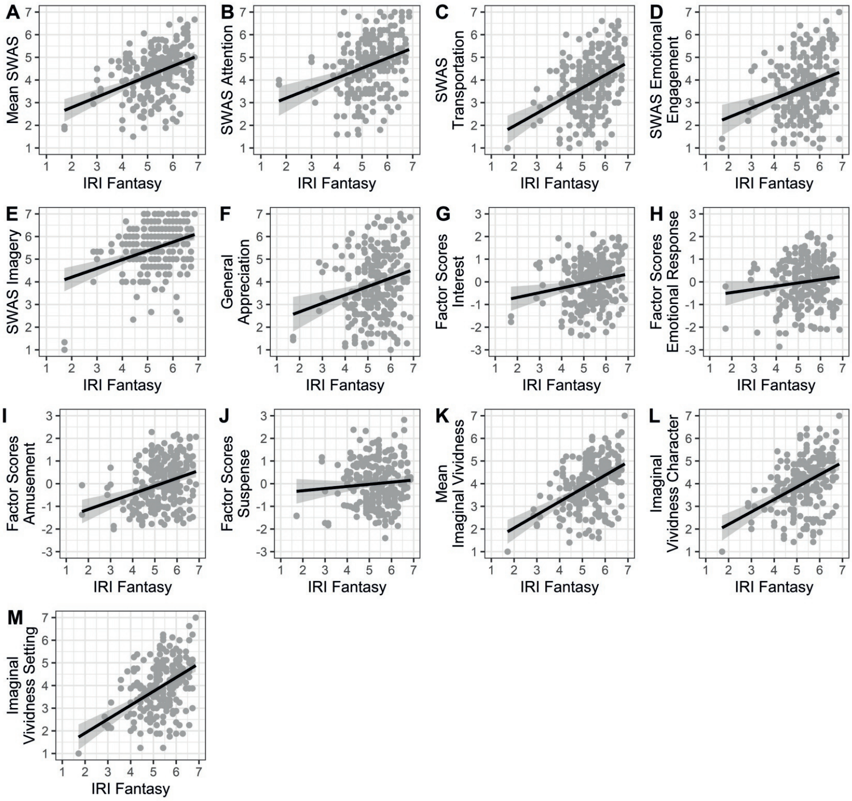
	IRI Perspective Taking	-0.10	0.08	118.00	-1.23	.22
	Reading Habits	0.07	0.15	118.00	0.46	.65
	ART-Score	-0.13	0.10	118.00	-1.37	.17
	Age	0.03	0.09	118.00	0.33	.75
	Gender (Male)	-0.29	0.22	118.00	-1.30	.20
	Story (B)	0.21	0.10	118.00	2.12	.04*
2b. IVS Setting	(Intercept)	3.73	0.14	145.24	26.15	<.001***
	Imagery Instruction	0.40	0.18	118.00	2.15	.03*
	School Instruction	-0.35	0.18	118.00	-1.93	.06
	IRI Fantasy	0.49	0.09	118.00	5.75	<.001***
	IRI Perspective Taking	-0.06	0.08	118.00	-0.78	.44
	Reading Habits	0.17	0.14	118.00	1.15	0.25
	ART-Score	-0.08	0.09	118.00	-0.86	.39
	Age	0.11	0.08	118.00	1.35	.18
	Gender (Male)	-0.17	0.21	118.00	-0.79	.43
	Story (B)	0.43	0.09	118.00	4.62	<.001***
3a. General Appreciation	(Intercept)	3.55	0.19	162.42	18.74	<.001***
	Imagery Instruction	-0.14	0.24	118.00	-0.60	.55
	School Instruction	-0.22	0.23	118.00	-0.94	.35
	IRI Fantasy	0.36	0.11	118.00	3.23	.002**
	IRI Perspective Taking	-0.03	0.10	118.00	-0.32	.75
	Reading Habits	0.07	0.19	118.00	0.37	.71
	ART-Score	-0.05	0.12	118.00	-0.44	.66
	Age	0.13	0.11	118.00	1.18	.24
	Gender (Male)	0.36	0.27	118.00	1.30	.19
	Story (B)	0.84	0.15	118.00	5.51	<.001***
3b. Factor Scores Interest	(Intercept)	-0.11	0.13	166.17	-0.84	.40
	Imagery Instruction	-0.19	0.16	117.67	-1.17	.24
	School Instruction	-0.32	0.16	118.67	-1.97	.052
	IRI Fantasy	0.18	0.08	117.78	2.39	.02*
	IRI Perspective Taking	-0.02	0.07	118.39	-0.30	.77
	Reading Habits	0.15	0.13	117.72	1.16	.25
	ART-Score	-0.06	0.08	117.67	-0.74	.46
	Age	0.09	0.07	117.79	1.14	.26
	Gender (Male)	0.29	0.19	120.55	1.51	.13
	Story (B)	0.46	0.11	117.98	4.12	<.001***
3c. Factor Scores Emotional Response	(Intercept)	-0.10	0.14	115.01	-0.69	.49
	Imagery Instruction	-0.28	0.18	116.44	-1.58	.12
	School Instruction	-0.03	0.18	117.38	-0.17	.86
	IRI Fantasy	0.15	0.08	116.55	1.85	.07
	IRI Perspective Taking	0.02	0.07	117.12	0.28	.78
	Reading Habits	-0.02	0.14	116.49	-0.15	.88
	ART-Score	0.04	0.09	116.44	0.40	.69

	Age	-0.004	0.08	116.56	-0.05	.96
	Gender (Male)	0.28	0.21	119.15	1.34	.18
	Story (B)	0.30	0.11	116.61	2.80	.006**
3d. Factor Scores Amusement	(Intercept)	-0.38	0.13	149.28	-2.89	.004**
	Imagery Instruction	0.17	0.17	117.76	1.01	.31
	School Instruction	0.02	0.17	118.63	0.10	.92
	IRI Fantasy	0.29	0.08	117.86	3.60	<.001***
	IRI Perspective Taking	-0.05	0.07	118.39	-0.78	.43
	Reading Habits	-0.02	0.13	117.80	-0.13	.90
	ART-Score	0.04	0.09	117.76	0.51	.61
	Age	-0.16	0.08	117.86	-2.08	.04*
	Gender (Male)	0.31	0.20	120.27	1.56	.12
	Story (B)	0.55	0.09	117.80	5.90	<.001***
3e. Factor Scores Suspense	(Intercept)	-0.36	0.14	142.78	-2.59	.01*
	Imagery Instruction	0.004	0.18	117.35	0.02	.98
	School Instruction	0.44	0.18	118.14	2.45	.02*
	IRI Fantasy	0.11	0.09	117.44	1.27	.21
	IRI Perspective Taking	-0.11	0.07	117.92	-1.52	.13
	Reading Habits	-0.14	0.14	117.39	-0.95	.34
	ART-Score	-0.01	0.09	117.35	-0.09	.93
	Age	-0.12	0.08	117.45	-1.39	.17
	Gender (Male)	0.01	0.21	119.63	0.03	.97
	Story (B)	0.44	0.09	117.26	4.97	<.001***

The positive relationship between scores on the Fantasy subscale of the IRI and scores on the Story World Absorption Scale was visible on all subscales of the SWAS (see Fig. 2.4B–E; see Table 2.4.1a–d). On the Attention and Mental Imagery subscales, story effects became visible; after reading Story B participants reported higher attention to the story world and higher Mental Imagery. In contrast, participants reported lower emotional engagement with Story B (see 2.4.1a–d). No other significant associations between the predictors and any of the subscales of the SWAS were found.

2.3.2.2.2. Vividness of Mental Imagery To investigate the effect of mental imagery instructions on reported mental imagery more thoroughly, I also tested differences between groups in mental imagery as reported on the IVS (see Fig. 2.3K; see Table 2.4.2). The results on this questionnaire also indicate whether participants complied with the mental imagery instructions. A comparison between the base model and the full model suggested a significant effect of group on scores on the IVS ($F(2, 118) = 6.49$, $p = .002$), but post-hoc comparison of

Figure 2.4.: *Relationship Between Scores on the Fantasy Subscale of the IRI and the Tested Aspects of the Reading Experience*



the two experimental groups and the control group showed no notable differences (Mental Imagery Instruction vs. Control: $B = -0.35$, $SE = 0.19$, $df = 128$, $t = -1.90$, $p = .14$; Secondary School Suitability instruction vs. Control: $B = 0.29$, $SE = 0.18$, $df = 128$, $t = 1.56$, $p = .27$). However, the group receiving the Mental Imagery instruction did report significantly more vivid imagery than the group receiving the Secondary School Suitability instruction ($B = 0.64$, $SE = 0.19$, $df = 128$, $t = 3.46$, $p = .002$). IRI Fantasy was positively related to the vividness of mental imagery (see Fig. 2.4K). After reading Story B, participants reported more vivid imagery. To find out if the effect of instruction was perhaps only visible on one of the two subscales and to find out which aspects of imaginal vividness the relationship between IRI fantasy and imaginal vividness stems from, these analyses were repeated for the individual subscales of the IVS.

The Character subscale of the IVS showed similar results for the relationship between instructions and imaginal vividness as were found on the overall scale: An initial comparison between the base model and the full model suggested a significant effect of type of instruction ($F(2, 118) = 3.61$, $p = .03$; see Fig. 2.3L), but post-hoc comparison of the two experimental groups and the control group did not reveal any statistically significant differences (Mental Imagery instruction vs. Control: $B = –0.30$, $SE = 0.20$, $df = 128$, $t = –1.47$, $p = .31$; Secondary School Suitability instruction vs. Control: $B = 0.22$, $SE = 0.20$, $df = 128$, $t = 1.10$, $p = .51$). The group receiving the Mental Imagery instruction did score higher on the Character subscale of the IVS than the group receiving the Secondary School Suitability instruction ($B = 0.51$, $SE = 0.20$, $df = 128$, $t = 2.58$, $p = .03$).

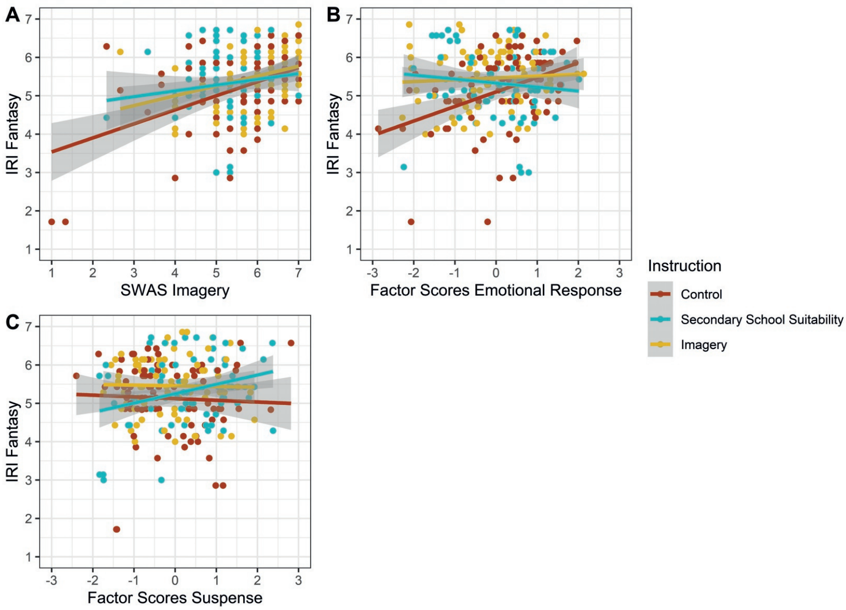
The same pattern was revealed for the Setting subscale of the IVS: an initial comparison between the base model and the full model suggested a significant effect of type of instruction ($F(2, 118) = 8.35$, $p < .001$; see Fig. 2.3M), but post-hoc comparisons of the two experimental groups and the control group did not reveal any statistically significant differences (Mental Imagery instruction vs. Control: $B = –0.40$, $SE = 0.19$, $df = 128$, $t = –2.07$, $p = .10$; Secondary School Suitability instruction vs. Control: $B = 0.35$, $SE = 0.19$, $df = 128$, $t = 1.86$, $p = .16$). Again, the group receiving the Mental Imagery instruction scored higher on the Setting subscale of the IVS than the group receiving the Secondary School Suitability instruction ($B = 0.75$, $SE = 0.19$, $df = 128$, $t = 3.93$, $p < .001$). On both subscales of the Imaginal Vividness Scale, I found a positive relationship between scores on the Fantasy subscale of the IRI and imaginal vividness (see Fig. 2.4L&M; See Table 2.4.2a–b). Similarly, differences between stories were found for both subscales of the IVS. After reading Story B, participants reported more vivid imagery of the characters in the story and of the settings described in the story. The results on the Imaginal Vividness Scale suggest that the reading instructions indeed influenced the experienced vividness of mental imagery, with respect to both the characters in the story and the environment described in the stories. The imagery instruction was associated with more vivid mental imagery than the secondary school suitability instruction. This suggests that participants are able to follow these instructions while reading, and that they indeed target mental imagery, as intended. Apart from instructed mental imagery, I also found a significant role for the personality trait Fantasy in the experienced vividness of mental imagery.

2.3.2.2.3. Appreciation To test whether mental imagery instructions would also have an impact on the appreciation of stories, I tested differences between groups in general appreciation and story appreciation as reported on the four components of the Appreciation Questionnaire (see Table 2.4.3a–e). Initial comparisons between the base models and the full models showed that the reading instructions influenced only the experienced Suspense ($F(2, 117.90) = 3.95$, $p = .02$; see Fig. 2.3J), although post-hoc comparisons of the three groups failed to reach statistical significance (Mental Imagery instruction vs. Control: $B = –0.004$, $SE = 0.19$, $df = 127$, $t = –0.02$, $p = .9998$; Secondary School Suitability instruction vs. Control: $B = –0.44$, $SE = 0.19$, $df = 128$, $t = –2.35$, $p = .052$; Mental Imagery Instruction vs. Secondary School Suitability instruction: $B = –0.44$, $SE = 0.19$, $df = 128$, $t = –2.32$, $p = .057$; Note however, that – although not statistically significant – this suggests that the participants receiving the secondary school suitability instructions experienced the stories they read as being somewhat more ominous and suspenseful, opposite to my expectations). The reading instructions did not have an effect on the four other aspects of appreciation. Comparable to the findings for the SWAS, participants scoring higher on the Fantasy subscale of the IRI appreciated the stories they read more (General Appreciation; see Fig. 2.4F; see Table 2.4.3a). Similarly, a positive association was found between IRI Fantasy and factor scores for Evoked Interest (see Fig. 2.4G; see Table 2.4.3b) and the factor scores for Amusement (see Fig. 2.4I; see Table 2.4.3d). Additionally, there was a negative association between age and factor scores for Amusement: older participants reported finding the stories less funny, witty or special (see Table 2.4.3d). Differences between stories were found for General Appreciation, Evoked Interest, Emotional Response, Amusement, and Suspense (see Table 2.3.3a–e). Story B was generally appreciated more, evoked more interest, elicited a stronger emotional response, was considered more Amusing, and more Suspenseful.

2.3.2.2.4. Interaction Between Instruction Condition and IRI Fantasy To test whether there was an interaction between instruction condition and IRI Fantasy, I performed exploratory analyses in which I included an interaction term between IRI Fantasy and instruction in the models. There was a significant interaction between IRI Fantasy and the Secondary School Suitability instruction for three of the thirteen tested dependent variables (SWAS Mental Imagery: $B = –0.36$, $SE = 0.14$, $df = 118.00$, $t = –2.91$, $p = .004$; Emotional Response: $B = –0.47$, $SE = 0.16$, $df = 116.5$, $t = –2.91$, $p = .004$; Suspense: $B = 0.38$, $SE =$

0.17, $df = 117.40$, $t = 2.28$, $p = .02$). From the visualization of these interactions in Fig. 2.5 can be seen that for Mental Imagery and Emotional Response, the relationship between IRI Fantasy and the dependent variable is attenuated when participants have to read with a reading instruction in mind (and mostly so if this is the Secondary School Suitability instruction). Oppositely, the relationship between IRI Fantasy and Suspense seems only present in participants who received the Secondary School Suitability instruction, but not in the other groups. Although these results are interesting in themselves, I have to be careful with interpreting them, as this interaction only appears for a few of the dependent variables and does not follow a highly consistent pattern.

Figure 2.5.: *Interaction Between IRI Fantasy and Instruction Group, for SWAS Mental Imagery Scores (5A), and Factor Scores for Emotional Response (5B) and Suspense (5C)*



2.3.3. Discussion

The aim of this second experiment was to replicate the findings of the first experiment in a more controlled setting, with additional stories, and while considering an extra set of control variables (most importantly aiming at personality characteristics that might influence reading experiences). As can be seen from the results of this experiment, reading instructions only played a very minor role in defining reading experiences.[8] Although it is visible that, in particular, the Mental Imagery instruction did influence the reading experiences directly involving mental imagery – suggesting that the instruction was indeed successful in influencing mental imagery –, the effect of this instruction did not translate to other reading experiences.[9] The most notable statistically significant finding in the experiments described in this chapter was that the Fantasy subscale of the Interpersonal Reactivity Index appeared to be positively associated with all aspects of participants' reading experiences. Even though it may be possible to influence certain reading experiences through reading instructions, personality characteristics appear to be much more important in determining people's reading experiences. As was described above, the Fantasy subscale measures the extent to which someone has the tendency to get mentally involved in the stories they encounter by imagining themselves being part of the story or by trying to empathize with characters in the story. Together, the questions on this subscale of the IRI give an impression of the amount of "fantasy" with which participants experience fiction on a day-to-day basis. Because of the theoretical relationship between this personality characteristic and reading experiences, it

[8]A possible concern with the analyses of these data would be that the lack of effect of reading instructions is due to the other predictors (The Fantasy and Perspective Taking subscales of the IRI, scores on the Author Recognition Test) absorbing so much variance that effects of reading instructions would not become visible. To check this, I ran reduced models that are comparable to the models used in Experiment 1 (without extra predictors, but including a random intercept for Participant and a predictor for Story, to control for participant and story effects). Although there were some minor changes in the effect of reading instruction on some of the dependent variables (i.e., SWAS Mental Imagery, Mean Imaginal Vividness, the Setting subscale of the Imaginal Vividness scale, and the Suspense component of the Appreciation Questionnaire), this concerned more pronounced effects rather than effects I failed to find with the models that did include the additional measures. As the latter models were significantly better than the reduced models (based on AIC, BIC and LogLikelihood), I chose to report only the results for the complete models.

[9]Note that it is possible that the findings with regard to the vividness of mental imagery are due to experimental demand, as the mental imagery instruction specifically asked participants to increase their mental imagery. However, I think this not the most likely explanation of these results. If the imagery findings would be entirely due to the experimental demand, I would have expected that both the secondary school instruction group and the control group would differ from the mental imagery instruction group with respect to reported vividness of mental imagery (since neither the secondary school suitability instruction nor the control instruction mentioned mental imagery). In contrast, the situation is that the secondary school suitability instruction lowered reported mental imagery somewhat with respect to the control group, and the mental imagery instruction increased reported mental imagery somewhat, resulting in a significant difference between the secondary school suitability group and the mental imagery instruction group (with the control group being somewhere in between). Although it is possible that the decrease in mental imagery in the secondary school suitability group is coincidental, and the increase in mental imagery in the imagery instruction group is due to experimental demand, it seems more likely that both instructions had a (small) influence on mental imagery – but in opposite directions. Importantly, the general conclusions remain the same, regardless of which explanation of the results is true.

is interesting to find that this personality characteristic is indeed positively associated with reading experiences across the board (note that this personality characteristic is not just associated with absorption or mental imagery, but also with several aspects of how participants appreciated the stories). A question of causality remains, however: future studies will have to determine whether those who engage more in imagery during reading, as a consequence like the stories they read more, or whether people who enjoy stories gradually become more imaginative as a result of reading (comparable to the question of causality in the study of the relationship between reading and theory of mind: do more empathic people read more, or does reading result in more empathy? See Panero et al., 2016; Samur, Tops, & Koole, 2018).

Finally, it is interesting to find that the influences of gender and reading habits on reading experiences as found in the first experiment reported in this chapter were not significantly associated with reading experiences in the second experiment. Just like the influence of reading instructions, the effects of gender and reading habits do not seem to be as important as personality characteristics in determining reading experiences (although note that the variation in gender in this second experiment was far from balanced: in the second experiment only 20% of participants were male, compared to 50% in experiment 1. Therefore, no strong conclusions about the effects of gender can be drawn from the results of experiment 2). Interestingly, age was negatively associated with amusement: older participants found the stories less funny, witty or special. However, it should be noted that the majority of participants in this experiment were university students of about 21 or 22 years of age, and this effect might be due to a couple of outliers (only 7 participants were older than 30 years of age, with 3 of them being 55 years or older). Nevertheless, a previous study with a larger variation in age between participants (and a large number of participants being between 50 and 75 years old) showed that older participants rated the stories they read as less literary and less beautiful than younger participants did Hartung, Withers, et al. (2017). The results reported in this chapter further showed a few differences between the two stories with respect to the reading experiences they elicited. As I did not have any hypotheses regarding how the stories would differ, I will not interpret the results regarding the differences between the stories.

2.4. General Discussion

In this chapter I investigated the relationship between mental imagery and reading experiences. In particular, I was interested in the act of mental imagery during reading and whether differences between people in the extent to which they engaged in mental imagery was related to their reading experiences. To make sure that participants differed in the extent to which they engaged in mental imagery, they received reading instructions in which they were instructed to envision the stories as much as possible, to read as if they were reading for leisure, or to focus on surface characteristics of the stories (word use and sentence construction; only in experiment 2). Apart from mental imagery, I was interested in the role of stable or trait-like individual differences (such as reading habits, gender, age, education, and personality characteristics) in determining reading experiences.

Although experiment 1 suggested that mental imagery instructions, as well as level of education, gender, and reading habits, played a significant role in determining reading experiences, experiment 2 showed that after controlling for personality characteristics (in particular “fantasy”) and adding an extra control condition, this association between mental imagery instructions, gender, reading habits, and reading experiences disappeared for a large part. This suggests that, besides all other aspects involved in reading experiences, these experiences are most strongly influenced by personality characteristics, such as readers’ proneness to “fantasy”. Fantasy has been suggested to be one of the aspects underlying the “Openness” personality characteristic (Fayn et al., 2015), a characteristic that has also been found to be associated with reading experiences and reading habits in other studies (Kraaykamp & Van Eijck, 2005; Kuijpers et al., 2018; Schutte & Malouff, 2004).

Mental imagery was mainly found to be related to mental imagery-related reading experiences, and not as strongly to other reading experiences. The reason that this relationship was not found to be very strong in the experiments described in this chapter, could be that there is a difference between (explicit) mental imagery and (implicit) mental simulation (see Jacobs & Willems, 2018). Perhaps the explicit mental imagery the participants were instructed to perform in these experiments was too different from the implicit mental simulation elicited by stories during naturalistic reading, and was therefore relatively unrelated to reading experiences. This could also explain why submitting participants to a more implicit mental imagery training before reading, did prove effective in increasing experienced transportation (Johnson et al., 2013).

Moreover, the interactions between the effects of fantasy and reading instructions on some of the tested reading experiences in experiment 2 even suggest that pre-reading instructions might in fact negatively influence naturalistic processes during reading. For both mental imagery (as reported on the SWAS) and for the emotional response to the story, it was found that the positive relationship between fantasy and these reading experiences, was attenuated in readers who received pre-reading instructions compared to readers in the control group (regardless of the content of the instructions). Therefore, it could be possible that having to remember and execute instructions during reading interferes with reading experiences as they would normally occur. However, the interactions that were found were not present for all aspects of reading experiences, and these analyses were highly exploratory, so further research should indicate whether this is indeed the case. However, when studying subjective reading experiences, it seems wise to only study naturalistic reading instead of trying to influence reading experiences using pre-reading instructions.

Another explanation for the weak association between mental imagery and reading experiences in the experiments in this chapter could be that readers differ greatly in the form of mental imagery they prefer during reading. Kuzmičová (2014) suggests four possible forms of mental imagery during literary reading: Enactment-imagery (where readers imagine themselves executing the actions described in the story), description-imagery (where readers visually imagine the objects and scenes described in the story), speech-imagery (where readers imagine hearing the narrator tell the story) and rehearsal-imagery (where readers imagine reading the story out lout). Kuzmičová suggests readers differ in the form of imagery they perform during reading (and this can also differ from one story to the next within a given reader). Perhaps the instructions given in this experiment did not match the preferred form of imagery of some (or all) of the readers, resulting in weak effects of the mental imagery instruction on reading experiences.

A different possibility would be that mental imagery just doesn't play a role in people's ability to become involved in a story. However, previous research has shown relationships between imagery and absorption, transportation, and appreciation or enjoyment of narratives, which does not fit with the proposal that imagery is unimportant in story involvement (Green, 2004; Green & Brock, 2002; Kuijpers et al., 2014; Mol & Jolles, 2014; Weibel et al., 2011). Another possibility would be that people are unable to perform mental imagery "on command". However, the fact that reading instructions were successful at inducing or reducing mental imagery in my participants contradicts this claim.

Overall, it seems that the use of mental imagery-inducing reading instructions does have a small influence on people's reading experiences. However, this effect pales into insignificance compared to the effect of personality characteristics. Perhaps a single reading instruction is insufficient for altering reading experiences: to really enhance reading experiences, readers will have to be trained intensively to read in a different way. For instance, Janssen, Braaksma, and Couzijn (2009) found that students' appreciation for stories increased after an intervention where they were allowed to come up with their own questions about the stories (as opposed to answering a teacher's questions). Perhaps a comparable intervention, but instead aimed at mental imagery, might have a stronger influence on reading experiences than a single instruction, and may perhaps prove more powerful in overcoming personality characteristics (but see De Koning & Van der Schoot, 2013).

3 | Mental Simulation During Literary Reading: Individual Differences Revealed with Eye Tracking

People engage in simulation when reading literary narratives. In the experiment reported in this chapter, I tried to pinpoint how different kinds of simulation (perceptual and motor simulation, mentalizing) affect reading behavior. Eye tracking (gaze durations, regression probability) and questionnaire data were collected from 102 participants, who read three literary short stories. In a pre-test, 90 additional participants indicated which parts of the stories were high in one of the three kinds of simulation-eliciting content. The results show that motor simulation reduces gaze duration (faster reading), whereas perceptual simulation and mentalizing increase gaze duration (slower reading). Individual differences in the effect of simulation on gaze duration were found, which were related to individual differences in aspects of story world absorption and story appreciation. These findings suggest fundamental differences between different kinds of simulation and confirm the role of simulation in absorption and appreciation.

This Chapter Is Based on

Mak, Marloes, & Willems, Roel M. Mental simulation during literary reading: Individual differences revealed with eye-tracking. *Language, Cognition and Neuroscience, 34*(4), 511-535. DOI: 10.1080/23273798.2018.1552007

Data Collected for this Chapter Have Been Published Open Access in

Mak, M., & Willems, R. M. (2021). Eyelit: Eye Movement and Reader Response Data During Literary Reading. *Journal of Open Humanities Data, 7,* 1–6. https://doi.org/10.5334/johd.49

Mak, H.M.L., & Willems, R.M. (2021). *Eyelit: Eye-movement and reader response data during literary reading.* DANS-EASY [Dataset]. doi: 10.17026/dans-zqk-zmqs

In Chapter 2, I described experiments in which I tried to study the role of mental imagery during reading by introducing group differences in mental imagery through reading instructions. This appeared to be difficult for two potential reasons. First, most of the time, readers probably do not use explicit mental imagery during natural literary reading. Rather, it is more likely that they use more implicit mental simulation instead. Secondly, the results from the experiments in Chapter 2 pointed into the possibility of individual differences in the role of mental simulation, for example through personality characteristics.

In Chapter 3 I describe an experiment in which I try to measure mental simulation as it occurs during reading by means of eye tracking. This way I hoped to be able to pick up the subtle influence of mental imagery, and its fluctuations between individual readers (as no one reader is the same as other readers or experiences stories in the same way as other readers).

3.1. Introduction

When people read stories, they sometimes vividly imagine the events occurring in the stories and in the story world in which these events are happening. The process underlying this vivid imagination has been called mental simulation. One result of mental simulation is that readers get the feeling that they are part of the story they are reading. Consequently, literary stories can take a strong grip on readers, although the strength of this "grip" can vary widely from one story to the next, and from one reader to the next. This sense of grip has been described in the literature as absorption[1] (Gerrig, 1993; Green & Brock, 2000; Jacobs & Willems, 2018; Kuijpers et al., 2014; Kuzmičová, 2012). In this experiment I will focus on mental simulation as an important driver of absorption. I distinguish between three kinds of mental simulation, and I have a particular focus on identifying individual differences in simulation.

Mental simulation has been defined as "*… the re-enactment of perceptual, motor, and introspective states acquired during experience with the world, body, and mind.*" (Barsalou, 2008, p. 618). Importantly, the definition of Barsalou suggests that mental simulation is not one of a kind. Indeed, from theoretical (Barsalou, 2008; and see Shanton & Goldman, 2010, for a review of Simulation Theory;

[1] Many researchers have tried to capture the experience of becoming part of a story, resulting in constructs such as immersion (Ryan, 2001; see also Jacobs, 2015b), absorption (e.g., Kuijpers et al., 2014), transportation (Gerrig, 1993; Green & Brock, 2000) or presence (Kuzmičová, 2012). For the sake of clarity, I will refer to this experience as absorption for the remainder of this chapter.

see also section 1.1.1.) as well as empirical work (discussed below), it is known that mental simulation should be subdivided in different kinds of simulation. Language users are capable of simulating perceptual and motor events on the one hand, and mental processes of others on the other hand (also referred to as mentalizing; e.g., Goldman, 2012). The effects of these three kinds of simulation on language processing have so far been studied mostly separately from each other (see Jacobs & Willems, 2018). In this chapter, I investigate them in one experiment. This enables me to disentangle the effects of different kinds of simulation on literary reading behavior. An additional advantage of the experiment described in this chapter is that I use narratives as stimuli. I choose to investigate mental simulation in the context of narratives because narratives allow readers to construct a much richer mental story world as compared to single sentences or paragraphs that are sometimes used in research.

3.1.1. Empirical Background

The aforementioned three kinds of simulation have been extensively studied in different subdisciplines of cognitive science. Perceptual and motor simulation (sometimes called 'sensorimotor simulation') have been studied in the tradition of embodied cognition. According to Zwaan (2009)), there is an important link between situation models, simulation and grounding in perception and action. When people form situation models (for example when they encounter narratives), the events and event nodes within these situation models are grounded in perception and action through (sensorimotor) simulation.

There is some evidence to suggest that readers indeed form perceptual mental images when reading language. It has, for example, been found that reading implicit descriptions of shape (e.g., "the ranger saw the eagle in the nest" vs. "the ranger saw the eagle in the sky") or orientation (e.g., "John put the pencil in the drawer" vs. "John put the pencil in the cup") primes subsequent visual perception of the described object in the implied orientation, both in adults and children (e.g., Engelen, Bouwmeester, Bruin, & Zwaan, 2011; Stanfield & Zwaan, 2001; Zwaan et al., 2002). Reading words that imply a certain location on a vertical axis similarly primes perception of pictures of semantically related objects appearing in the implied location (e.g., "sky" primes the detection of a picture of a cloud if this picture is presented in the top half of the screen; Ostarek & Vigliocco, 2017; see also Estes et al., 2008). Additionally, reading descriptions of biological motion has been associated with activation in a motion processing area (i.e., middle temporal gyrus; Deen & McCarthy, 2010; Samur, Lai, Hagoort,

& Willems, 2015). Similar associations between reading auditory descriptions and activation in multiple areas involved in auditory processing (Kurby & Zacks, 2013) and between reading vivid visual descriptions and connectivity between different areas in the visual processing system (Chow et al., 2015) have also been found.

Similarly, motor simulation has been found to play a role in language processing. Movements and actions implied in sentences were found to prime related actions when they had to be executed after reading these sentences. This could happen directly via verbs (Glenberg & Kaschak, 2002), but also more indirectly through the context presented within sentences (Bergen & Wheeler, 2010; Foroni & Semin, 2013). Additionally, in several neuroimaging studies, it has been found that action words (Hauk et al., 2004), sentences (Tettamanti et al., 2005), and even complete passages describing actions (Kurby & Zacks, 2013) all elicit activation in areas in the (pre)motor cortex. An association has also been found between reading vivid descriptions of actions and connectivity between different areas in the motor cortex (Chow et al., 2015). However, motoric language processing does not always elicit activation in the same way and in the same areas; for a more elaborate review of the task effects at hand and the precise role of the motor cortex in action language processing see for example Kemmerer (2015) or Willems and Casasanto (2011).

In sum, evidence suggests that perceptual and motor simulation occur during language understanding. Participants show behavioral or neural indices of the involvement of brain areas involved in perception and action when comprehending language that is related to perception and action. At the same time, it seems that such activation does not invariably occur when readers encounter perceptuo-motor language (see e.g., Willems & Francken, 2012, for discussion).

Another kind of simulation under study here is the simulation of introspective states, sometimes referred to as “mentalizing” (also related to Theory of Mind, see C. Frith & Frith, 1999; U. Frith & Frith, 2003; Goldman, 2012). When people are mentalizing, they are attributing mental states (thoughts, emotions, intentions) to other people. In this process, they may link perceptions of these people to pre-existing social knowledge. Research has shown that people engage in mentalizing both consciously and unconsciously (i.e., explicit and implicit mentalizing, see Apperly & Butterfill, 2009), and that both processes (at least in part) make use of similar brain areas (U. Frith & Frith, 2003; Van Overwalle & Vandekerckhove, 2013), and are similarly reflected in behavioral data (Nijhof, Brass, Bardi, & Wiersema, 2016), although the extent to which they overlap neurally as well as conceptually is a matter of ongoing debate (e.g., Kovács, Kühn, Gergely, Csibra, & Brass, 2014). Understanding the beliefs, intentions and thoughts of fictional characters is vital to the experience of being in a fiction world, and it is fair to say that mentalizing is an important aspect of literary reading (see, e.g., Bruner, 1986; Burke, 2011; Hartung et al., 2016; Oatley, 2012; Van Krieken et al., 2017).

The involvement of mentalizing in narrative reading would imply that readers attribute mental states to characters and link the actions of these characters to the knowledge they have gained about these characters over the course of the story. Indeed, Filik and Leuthold (2013) found that if the subsequent information about the actions of a character in their study was incongruent with the personality or beliefs of this character, people exhibited N400 responses to the critical words and a higher number of regressions away from these words as well as longer regression path reading times associated with these words, as if they were interpreting semantic incongruity. Additionally, reading mentalizing-eliciting content in a narrative has been associated with brain activation in areas involved in social cognition and mentalizing (Hsu et al., 2014; Nijhof & Willems, 2015; Tamir et al., 2016).

3.1.2. Relationship Between Kinds of Simulation During Narrative Reading

Although all three kinds of simulation seem to be involved in reading, findings from a few studies hint at the possibility that they can be involved in different ways. In an experiment tapping into the role of mentalizing in language processing, Wallentin, Simonsen, and Nielsen (2013) presented participants with a short story and asked them to indicate the level of intensity they experienced while reading the different passages of the story. Subsequently, these intensity ratings were linked to participants' empathy scores. The researchers found that the level of reported intensity was highest in passages describing (fear-induced) action, but that the reported intensity in these passages was not related to empathy. In contrast, they did find a correlation between intensity rating and participants' empathy scores in mentalizing-eliciting passages (describing social interactions), even though the level of intensity reported for these passages was not particularly high. Together, these findings suggest differences in the processing of mentalizing-eliciting and motor simulation-eliciting passages in a story. In a different experiment, a comparable dissociation was found between mentalizing and motor simulation (Nijhof & Willems, 2015). Nijhof and Willems found that mentalizing- and motor simulation-eliciting descriptions in stories activated brain areas involved in mentalizing and action execution, respectively. Interestingly, there was a negative correlation between the effects of both kinds of descriptions, implying that individual participants could prefer mentalizing over motor simulation, or vice versa. Together, these two studies suggest that different kinds of simulation might have differential effects on reading behavior (in general or within participants), but the precise relationship between the effects of different kinds of simulation remains unclear.

Both on theoretical as well as on empirical grounds there is good reason to expect that perceptual, motor, and mental state simulation play a role in narrative understanding. In the experiment described in the current chapter, I aimed to investigate these kinds of simulation and how they are interrelated within one experiment.

3.1.3. Current Experiment

Most of the studies described above looked at the relationship between one or two kinds of simulation and language processing, but the tasks used are divergent, and none of the above studies tried to pinpoint the differential influence of all different kinds of simulation on literary reading behavior. In the experiment described in this chapter, I tried to disentangle the individual roles of perceptual simulation, motor simulation and mentalizing in reading behavior, as measured using eye tracking. As described above, there is reason to believe that the different kinds of simulation have different effects during narrative reading. I presented literary narratives to participants while tracking their eye movements, to find out whether perceptual descriptions, motor descriptions and mental event descriptions (as identified in the stories in a pre-test) were differentially related to gaze duration and the probability of regressing back to a word. The rationale for using eye tracking as a method of choice was that if mental simulation is a time-sensitive cognitive process (as suggested by reaction time studies, see e.g., Fischer & Zwaan, 2008, for an overview), increased simulation should be detectable in gaze durations to passages in the text that are thought to elicit simulation. I hence predicted that mental simulation would increase gaze durations (i.e., slower reading).

Apart from the general effects of different kinds of simulation on reading behavior, I was interested in the question whether all people show these effects in the same or a similar way, or if they show individual differences in their responses. Previous research has suggested sizeable individual differences in how much readers engage in mental simulation (e.g., Altmann, Bohrn, Lubrich, Menninghaus, & Jacobs, 2014; Chow et al., 2015; Hartung, Hagoort, & Willems, 2017; Hsu et al., 2014; Nijhof & Willems, 2015). In this chapter's experiment I linked individual differences in gaze duration to passages high in simulation-eliciting content to absorption and appreciation for the stories. Previous behavioral (questionnaire) research suggests that simulation influences story world absorption, and that absorption correlates with appreciation (Busselle & Bilandzic, 2009; Green & Brock, 2000; Green et al., 2004; Hartung et al., 2016; Kuijpers et al., 2014). Here I tried to replicate these earlier findings and importantly investigated whether the relationship between mental simulation and absorption / appreciation is exclusive to one of the kinds of simulation or not. In order to investigate individual differences effectively, I collected data from a relatively large sample (N=102).

Additionally, I linked individual differences in simulation to individual differences in personality traits or characteristics that have been found to be related to absorption, such as fiction reading habits (Mar, Oatley, Hirsh, Paz, & Peterson, 2006), the tendency to get transported (see Green & Donahue, 2009) and perspective taking (Mar & Oatley, 2008; see Mumper & Gerrig, 2017 for a meta-analysis). As simulation is related to absorption, I expected that simulation would also be related to these traits. Any association between individual differences in simulation and individual differences in one or more of these personality traits or characteristics might give an indication as to why people seem to differ so much in their experiences during reading.

3.2. Methods

The experiment described in the current chapter was pre-registered in the Open Science Framework (`osf.io/qgx26`).

3.2.1. Participants

I recruited 109 participants (85 females) from the participant database of the Radboud University. All participants were native speakers of Dutch, and had normal or corrected to normal vision. Based on poor quality of the eye tracking data or insufficient performance on a comprehension check, data for seven participants were rejected. Of these participants, four were female. The mean age of the remaining participants (N=102) was 23 years (range 18 – 40).

Participants received €15 or course credit for their participation in the experiment. Prior to the experiment, participants were informed about the procedure of the experiment. It was made clear that participation was voluntary and that it was allowed to withdraw from the experiment at any time without need for explanation. All participants gave written informed consent in accordance with the Declaration of Helsinki. The experiment was approved by the local ethics committee.

3.2.2. Materials

Three existing Dutch short stories (see Table 3.1) were presented to the participants. The stories were selected based on length, the presence of simulation-eliciting content, and the probability that the stories would be unknown to the target group (to ensure that all participants would read the stories for the first time). All stories were written by acclaimed writers (who all have received literary awards for their work) and have been published by literary publishers. Stories A and B are written by contemporary Dutch writers, and story C was translated from American English to Dutch. This story was taken from a professional and published translation. The stories were on average around 2600 words (2143, 2659, and 2988 words), and each story took around 10-15 minutes to read. A pre-test (see below) confirmed that all stories contained simulation-eliciting passages, indicating that all stories contained passages (or sentences or clusters of words) that were likely to elicit motor simulation, perceptual simulation or mentalizing. All participants read all three stories (in counterbalanced order). None of the participants reported having read any of the three stories before.

3.2.3. Simulation Scoring Pre-Test

For a pre-test, 90 participants were recruited from the same participant database. These participants did not take part in the main part of the experiment (i.e., the eye tracking session). All participants read all three stories (in counterbalanced order), and were asked to focus on one of the three kinds of simulation. They were instructed to underline all the words, sentences, or passages that they considered to be part of either of three possible types of simulation-eliciting content: perceptual descriptions, motor descriptions, or descriptions of mental

Table 3.1.: *Title, Author, Year of Publication and Word Count of the Experimental Stories.*

	Title		Author	Year of Publication	Word Count
Story A	De Mensen die Alles Lieten Bezorgen	*(The People that Had Everything Delivered)*	Rob van Essen	2014	2988
Story B	De Chinese Bruiloft	*(The Chinese Wedding)*	Sanneke van Hassel	2012	2659
Story C	Signalen en Symbolen	*(Symbols and Signs)*	Vladimir Nabokov	1948/2003	2143

events (e.g., thoughts, feelings, opinions) that revealed what was going on in the mind of a character in the story. Perceptual descriptions were defined as "things that are perceivable with the senses", motor descriptions as "concrete acts or actions performed by a person or object", and mental event descriptions as "explicit descriptions of the thoughts, feelings and opinions of a character" and/or "reflection by a character on his own or someone else's thoughts, feelings or behavior". In addition to the definitions of the types of descriptions, participants were presented with a short explanation of what was meant by these definitions, including a couple of example sentences derived from different stories. I hence used subjective ratings for the occurrence of simulation-eliciting content in a story. Several previous studies have similarly used subjective ratings as a proxy for simulation Nijhof & Willems, 2015, imagery (Kurby & Zacks, 2013), or other variables affecting the reading process, such as foregrounding (Van den Hoven, Hartung, Burke, & Willems, 2016).

I asked participants to underline all words or passages in each of the three stories that they considered to be part of one of the three types of simulation-eliciting content, resulting in scores between 0 and 30 for every word in each of the three stories, for all three types of simulation-eliciting content: a score of 0 if none of the participants had underlined it and a score of 30 if every participant had underlined it (with higher scores theoretically resulting in a higher probability this word would be mentally simulated). The underlining of perceptual descriptions was performed by 24 females and six males, the underlining of motor descriptions was performed by 22 females and eight males, and mental event descriptions were underlined by 19 females, ten males and one unspecified.

The distribution of scores per kind of simulation is shown in Fig. 3.1, and the average number of times the words were underlined per kind of simulation and per story can be seen in Table 3.2. Importantly, 2968 words, 1952 words and 3555 words were underlined by none of the pre-test participants for perceptual, motor and mental events descriptions, respectively. This indicated considerable consensus between participants about whether words were part of a description. The number of times a word was underlined was similar for all percentiles of sentence length (with the last word set at 100% and the middle word at 50%) for all types of underlining (see Appendix A1-3). This means that underlinings did not occur systematically more at the beginning or end of the sentence.

Figure 3.1.: *Distribution of the Number of Times Words Were Underlined for Mental Event Descriptions, Motor Descriptions and Perceptual Descriptions. 4235 out of 7790 Words Were Underlined At Least Once for Mental Event Descriptions, 5838 Words Were Underlined At Least Once for Motor Descriptions and 4822 Words Were Underlined At Least Once for Perceptual Descriptions*

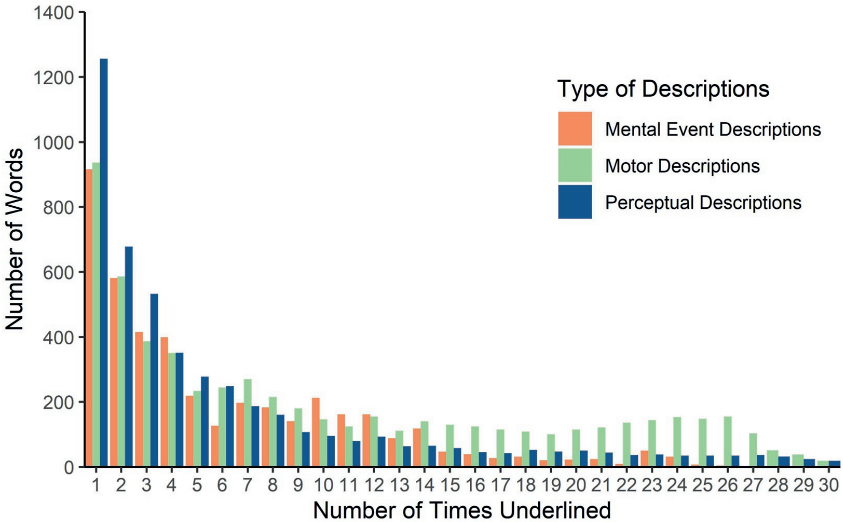


Table 3.2.: *Descriptive Statistics of the Underlined Words: Mean, Standard Deviation and Maximal Observed Value of the Number of Times Words Were Underlined for Each Type of Description in Each Story (the Maximal Possible Value of the Number of Times Words Were Underlined Is, in All Instances, 30)*

	Perceptual Descriptions		Motor Descriptions		Descriptions of Mental Events	
	M (SD)	Max. Observed Value	*M (SD)*	Max. Observed Value	*M (SD)*	Max. Observed Value
Story A	5.69 (6.31)	30	9.88 (8.62)	30	5.21 (4.57)	25
Story B	6.08 (6.56)	30	11.95 (8.98)	30	5.54 (4.68)	23
Story C	7.27 (7.75)	30	8.00 (7.47)	29	7.32 (6.51)	27

3.2.4. Apparatus

For eye movement data collection, a monocular desktop-mounted EyeLink1000-plus eye tracking system was used. Data were recorded with a sampling rate of 500Hz. Head movements were minimized using a head stabilizer. This allowed me to ensure all participants were seated at 108 cm from the screen (i.e., distance from the eye to the bottom of the screen).

3.2.5. Stimulus Presentation

The stimuli were presented using SR Research's Experiment Builder software (SR Research, Ottowa, Canada), on a BenQ XL 2420T 24" LED screen. The experiment was presented at a resolution of 1024 x 768 (32 bits per pixel). The stories were divided into 30 sections each, that were presented to the participants one at a time. These sections resembled the author's original division of the story into paragraphs as much as possible. For presentation of the sections, minimum margins of 120 pixels were used on all sides. They were presented as black letters on a white background, in a 15-point Calisto MT font, corresponding to an on-screen size of 4 mm high for letters such as "m", 6 mm high for capital letters and letters such as "h", and 8 mm high for letters such as "j". Between different lines on a page, there was 24 mm white space. The Experiment Builder software automatically defined interest areas for all words. There was no space between interest areas; the boundaries of the interest areas were centered between horizontally and vertically adjacent words.

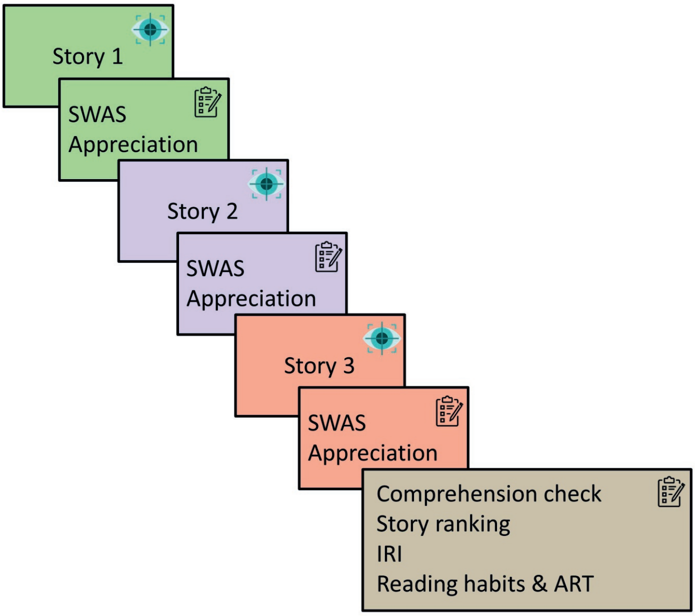
3.2.6. Procedure

At the beginning of the experiment, participants were instructed to move as little as possible, but to read as naturally as possible, the way they would read a story outside of the laboratory. Since the eye tracking data were collected monocularly, the dominant eye was tracked. To identify the dominant eye, participants performed an eye dominance test. In seven participants it was not possible to track the dominant eye, so in these participants the non-dominant eye was tracked. The stories were presented in a sound proof booth. There was no time restriction, participants could proceed to the next section of a story by pressing the space bar as soon as they had finished reading the current section. In addition to the eye tracking part of the experiment, questionnaires (see below) were presented as paper and pencil tests outside of the booth, to enable participants to take a break from the computer screen.

The stories were presented in counterbalanced order while maintaining the overall gender balance within each order. At the beginning of each story, participants performed a 9-point calibration, and after every five sections a drift check was performed. During 1000ms before the next section appeared, participants fixated on a fixation cross at the point of the screen the first character of the text would appear.

The order in which participants completed the different parts of the experiment (story readings, questionnaires) is depicted in Fig. 3.2. Participants first read a story, and filled in questionnaires (Story World Absorption Scale, appreciation questionnaire; see for more detailed information about the questionnaires section 3.2.7.) about that specific story. They repeated this for Story 2 and 3. After finishing reading the third story (and answering the two questionnaires), they were presented with the comprehension check for all stories. After that, the additional questionnaires (story ranking, reading habits, Interpersonal Reactivity Index and Author Recognition Test) were presented.

Figure 3.2.: *Graphic Depiction of the Time Line of the Procedure of the Experiment. The Stories Were Presented in Counterbalanced Order*



3.2.7. Additional Measures

All used questionnaires can be found in Appendix B. After reading each story, participants filled out a sim\subulation and an appreciation questionnaire. The simulation questionnaire consisted of the story world absorption scale (SWAS; Kuijpers et al., 2014; e.g., *When I finished the story I was surprised to see that time had gone by so fast; I could imagine what the world in which the story took place looked like*), complemented with six additional questions (partly based on items originally designed by Kuijpers et al., 2014) regarding the experience of different kinds of simulation (mainly perceptual and motor simulation, e.g., *I could see the events in the story happening as if I could see through the eyes of the main character; I could easily depicture the characters in the story;* See Appendix B1 for a list of the questions added to the SWAS, all other used questionnaires can be found in Appendix B2-8). The SWAS is a validated scale consisting of 18 items with high internal validity (Kuijpers et al., 2014), which measures four dimensions of story world absorption via the subscales Attention, Transportation, Emotional Engagement and Mental Imagery. Participants rated each question on a 7-point scale (1 = disagree, 7 = agree).

The appreciation questionnaire consisted of a general score of story liking (How did you like the story; 1 = It was very bad, 7 = It was very good) and thirteen adjectives (e.g., *[did you find the story] Entertaining, ... Ominous*) that could be used to describe the stories (adapted from Knoop et al., 2016). These adjectives were taken from a list of adjectives that people often use to describe their opinion of poetry, and which are also used to describe aesthetic appeal in the domain of literature (Knoop et al., 2016). Finally, six questions were asked regarding the enjoyment of the story (from Kuijpers et al., 2014; e.g., *I was constantly curious about how the story would end; I thought the story was written well*). Participants rated both the adjectives and the questions regarding enjoyment on a 7-point scale (1 = disagree, 7 = agree).

At the end of the experiment, participants completed additional, more general, questionnaires. First, participants were presented with a comprehension check, consisting of three multiple choice questions per story with four possible answers per question, that should have been possible to answer correctly after reading the stories with normal attention (example question, *Why did Jeffrey and Rita leave the flat?*). Subsequently, they were asked to rank the stories from most appreciated to least appreciated, and they were asked to indicate whether they had read the story before. Next, they answered six questions about their reading habits in everyday life, choosing from four or five optional answers (adapted from Hartung et al., 2016; e.g., *How often do you read fiction; How often do you read non-fiction; How many books do you read each year*), and filled out the Fantasy and Perspective Taking subscales of the Interpersonal Reactivity Index (IRI; Davis, 1980; Dutch translation adapted from De Corte et al., 2007) on a 7-point scale (e.g., *Becoming extremely involved in a good book or movie is somewhat rare for me; When I'm upset at someone, I usually try to "put myself in his shoes" for a while*). The Fantasy subscale measures the extent to which someone gets mentally very involved in the stories they encounter, to the point where they imagine themselves being part of the story. The Perspective Taking subscale measures the extent to which someone is able to take someone else's perspective in daily life. Finally, as an implicit measure of reading habits, participants completed the Author Recognition Test (ART; Stanovich & West, 1989; Dutch adaptation reported in Koopman, 2015), consisting of 42 names (30 real authors and 12 foils), of which they had to indicate who they thought were genuine authors.

3.2.8. Eye Movement Data Pre-Processing

Before data analysis, all fixations were checked for all sections of all stories. This was done to make sure that they did not drift off so much that they entered a different interest area, thus corrupting the data. If necessary, they were manually aligned using SR Research's EyeLink Data Viewer. If this was impossible, because fixations did not lie on clear lines (corresponding to the lines on the pages presented on the screen), individual sections were excluded. If more than six sections of one story had to be excluded (more than 20% of the data for that story), data for entire story-readings were excluded in order to reduce noise in the data. If the entire story-reading had to be excluded for more than one story in the same participant, all data for this participant was excluded. For 62 participants, no sections of any of the stories had to be removed. For 40 participants at least one section had to be removed. For Story A, at least one but no more than six sections had to be removed for nine participants (on average 1.56 sections). For Story B, at least one but no more than six sections had to be removed from the analysis for 14 participants (on average 2.14 sections). For Story C, at least one but no more than six sections had to be removed for 21 participants (on average 2.05 sections). For four participants, the number of excluded sections exceeded six, resulting in the exclusion of one story-reading for this participant. This happened twice for Story A, and twice for Story C, resulting in the exclusion of the data for one story-reading for four participants. In total, eye movement data pre-processing resulted in the loss of 2.26% of the total amount of data.

3.2.9. Comprehension Check

Seventy-four participants answered all multiple-choice questions (four answer-options) in the comprehension check correctly. Participants were allowed to answer one question per story incorrectly. If participants answered more than one question incorrectly for a given story, it was concluded that they had not paid sufficient attention to this story, and data for this story-reading was excluded for these participants. If the entire story-reading had to be excluded for more than one story, all data for this participant was excluded. Eight participants answered more than one question incorrectly for one of the three stories. This was the case four times for Story B and four times for Story C, resulting in the exclusion of the data for one story-reading for eight participants (an additional loss of 2.42% of the total amount of data).

3.2.10. Data Analysis: Step 1

In the first part of the analysis, I used the *"lme4"* package in R (Bates et al., 2015; R version 3.5.1) to analyze the data with a linear mixed effects regression model that predicted gaze duration for each individual word (i.e., the total duration of all fixations on a word the first time that word is read) by simulation-eliciting content (as the effect of interest), with lexical frequency, word length and surprisal value as covariates (see Fig. 3.3A; values for all predictors were at the word-level). *P*-values were estimated using the "lmerTest" package (Kuznetsova et al., 2017). I controlled for lexical frequency, word length and surprisal value, because previous studies have shown that high frequency words are associated with shorter gaze durations than low frequency words (see Rayner, 1998 for a review), longer words with longer gaze durations than shorter words (e.g., Rayner & Fischer, 1996; Rayner, Sereno, & Raney, 1996), and words that are more likely to occur given their context (low surprisal value) with shorter gaze durations than unlikely words (Goodkind & Bicknell, 2018; Hale, 2001; Levy, 2008). The effects of the different types of descriptive words (motor descriptions, perceptual descriptions and descriptions of mental events) were allowed to vary per story per participant (i.e., different intercepts and slopes were allowed for stories A, B, and C for each participant[2]). This resulted in a total of

[2]As a result of the nature of the random effect structure, random effects were calculated for (1 | story:subject) – which has 294 levels (i.e., 294 individual subject and story combinations). Although it can be argued that in the current experiment story is not necessarily nested in subject, crossed random effects (where random effects are calculated for (1 | subject) and (1 | story), but not for the interaction between the two) would mean that random effects would be calculated for (1 | story) – which has only 3 levels (namely, three stories were used in this experiment).

294 different coefficients[3] for each predictor (102 participants times three stories, minus single story-readings of four participants based on insufficient quality of the eye tracking data and of eight participants based on poor performance on the comprehension check). Data for the first word of each slide were excluded, as previous research has shown that fixations on these words are disproportionately long, due to the after effect of the fixation cross (Van den Hoven et al., 2016).

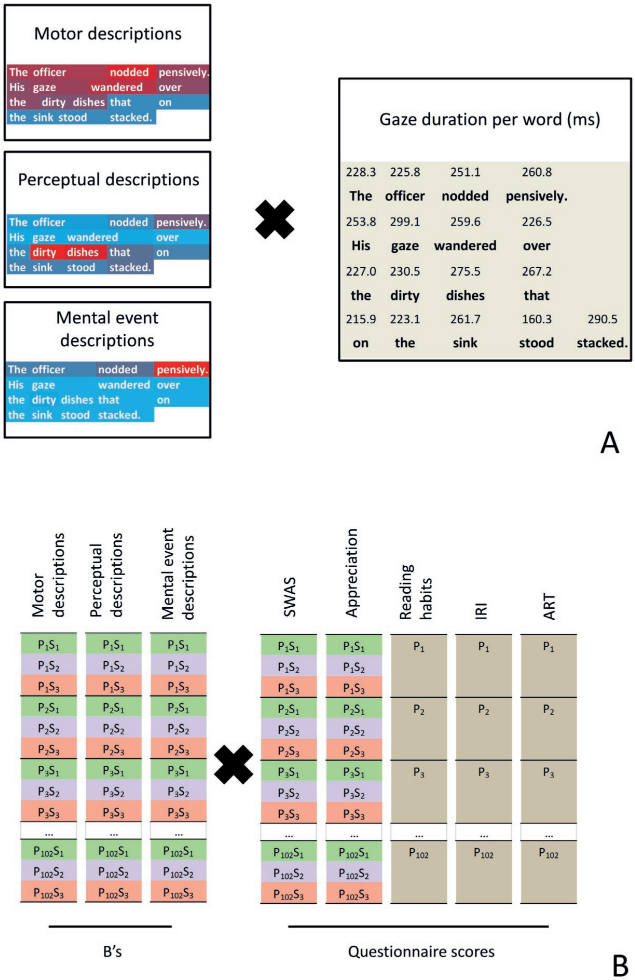
Lexical frequency was derived from the SUBTLEX-NL database and consisted of the logarithm of the frequency with which a word appeared in the database (Keuleers, Brysbaert, & New, 2010). Word length was determined by counting the number of characters for each word. Surprisal value was derived from perplexity, calculated using a 3-gram model trained by SRILM on 1 million sentences from the NLCOW2012 corpus. Perplexity was equal to 10 to the power of negative surprisal. Words for which one of the covariates was unknown were excluded from the analysis (resulting in the loss of another 3.20% of the total amount of data).

A model of eye movements during reading, the E-Z Reader model (Reichle, Pollatsek, Fisher, & Rayner, 1998; see Reichle, Warren, & McConnell, 2009 for version 10 of this model), predicts spill-over effects in reading behavior, when considering reading behavior on the word level (these spill-over effects have been previously described by Mitchell, 1984; Rayner, 1998; see Reichle & Drieghe, 2015 for an account on how E-Z reader can be used to explain spill-over effects). The E-Z Reader model assumes that words are processed serially, meaning that the processing of one word has to be completed before processing of the next word can be started. However, after a first stage of initial processing (which is based on the "familiarity" of the word), a saccade can already be made towards the next word. As a result, deeper processing (based on the meaning of the word) and integration of the word into a sentence representation can actually take place while the gaze has already shifted towards the next word, resulting in spill-over effects (the effect of a variable on the processing speed at word n is reflected in the gaze duration towards word $n + 1$). The effect of simulation-eliciting content on gaze duration that I would like to unveil would not primarily

As it is advised to only calculate random effects for variables with more than 5 or 6 levels, this approach would, statistically speaking, not have been favorable.

[3]These coefficients, derived from the random slopes in the model, are strictly speaking Best Linear Unbiased Predictions (or BLUPs), as opposed to the population-level coefficients derived from the model summary (Best Linear Unbiased Estimates, or BLUEs). However, using such terminology in the text of this chapter, might make it unnecessarily complex, therefore the BLUPs will be referred to as "coefficients".

Figure 3.3.: *A. In the First Step of the Analysis, Scores from the Underlining Pre-Test for Motor, Perceptual and Mental Event Descriptions per Word (with Red Words Having a High Score in the Underlining Pre-Test, that is They Score Highly on that Type of Description) Were Linked to Gaze Duration Scores (or to Scores Indicating Whether or Not a Regression Back to a Word Had Been Made) per Word. B. In the Second Step of the Analysis, the Coefficients for the Relationships Between Motor, Perceptual and Mental Event Descriptions and Gaze Duration/Regression Probability (per Participant per Story) Were Linked to the Questionnaire Data (per Participant; for the SWAS and Appreciation Questionnaires per Participant per Story). The Rationale of this Second Analysis Was to See if Individual Differences (as Discovered in the Questionnaire Data) Were Related to Individual Differences in the Way Reading Behavior Was Affected by the Different Kinds of Simulation, as Established in Step 1 of the Analysis*



lie in the processing of the familiarity of the word being processed, but rather in the processing of the meaning of the word, and would therefore be expected to be reflected by the gaze duration in the spill-over region (as later processing of words is reflected by spill-over effects; Rayner et al., 2004). To account for these spill-over effects, the scores for all predictors included in the model (i.e., lexical frequency, word length, surprisal value, motor descriptions, perceptual descriptions and mental event descriptions) were taken from the previous word (all predictors thus taken to predict gaze duration at the next word in the story instead of the word they were derived from; comparable to approaches reported by, among others, Calvo & Meseguer, 2002; Frisson, Koole, Hughes, Olson, & Wheeldon, 2014; Kliegl et al., 2006; Rayner et al., 2004; Schroyens, Vitu, Brysbaert, & D'Ydewalle, 1999). However, please note that after analyzing the data at the level of the target word (i.e., if predictors derived from the current word were included instead of predictors derived from the previous word), results remained highly similar: Statistically significant effects were found, with the same direction for the three types of descriptions as in the spill-over analysis.

I constructed a generalized linear mixed effects regression model to predict the probability of regressing into an interest area (i.e., word). Regressions are right-to-left eye movements, indicating a difficulty in the processing of a previous portion of the text (see Rayner, 1998). Regressions into an interest area (or regressions back) are an indicator of effects on later processing (Rayner et al., 2004). It would be interesting to find out if words high in simulation-eliciting content would be easier or more difficult to process than other parts of the text. Hence, I repeated the statistical analysis as described above for this dependent variable. In the model, the probability of regressing back to a word was predicted by simulation-eliciting content, with lexical frequency, word length and surprisal value as covariates. Again, the effects of the different types of descriptions were allowed to vary per story per participant, resulting in a total of 294 different coefficients for each predictor. In this model, the predictors included in the model were derived from the current (target) word, as I did not expect a spill-over effect for regressions back to a word. The two models described above predicted the effect of different types of simulation-eliciting content on reading behavior, per participant and per story.

I decided to look at gaze duration, as this has been found to be a good measure of difficulty of processing of a word (Just & Carpenter, 1980; Rayner, 1998). Another possibility would have been to look at first fixation duration. However, first fixation duration and gaze duration are not independent of one another:

first fixation durations are always part of, and often equal to, gaze durations, making it statistically undesirable to perform analyses on both variables (Kliegl & Laubrock, 2017). Therefore, I chose to use gaze durations in my analyses, as this is "considered [to be] the upper bound of early processing" in reading research (Kliegl & Laubrock, 2017, p. 77). Apart from gaze duration, I looked at the probability of regressing back to a word, as a measure of difficulty in the processing of a previous portion of the text or of incorporating a word into a mental representation of a sentence (Rayner et al., 2004). Readers have been found to be quite accurate in making a saccade back to the word with which they have trouble integrating, indicating that the word on which the eye lands after a regression is usually the word they found difficult to process (Rayner, 1998). Because of the strong co-dependency of different eye tracking measures, I decided to choose the measures of interest prior to the experiment to avoid the pitfall of looking at too many different measures and subsequently reporting "spuriously significant results" (Kliegl & Laubrock, 2017, p. 78).

3.2.11. Data Analysis: Step 2

In the second part of the analysis, I wanted to link individual differences in the relationships between the different kinds of simulation and gaze duration to individual differences in absorption, appreciation, perspective taking-ability, and reading habits. The purpose of this analysis was to see whether self-report measures of reading experiences such as transportation, mental imagery, appreciation, etc. that have been used in previous studies would be associated with simulation as measured using eye tracking. In order to test this, I derived the slopes per participant per story for the relationships between the different kinds of simulation and gaze duration / regression probability (i.e., 294 different coefficients for motor, perceptual and mental event descriptions) from the predictions of both models from the first part of the analysis, and investigated how these were related to absorption and appreciation. In this part of the analysis, I constructed three models, each predicting the coefficients of one of the three types of descriptions (per participant and per story), by the questionnaire scores per participant (and per story for the simulation and appreciation questionnaires), allowing for random intercepts per participant (see Fig. 3.3B).

3.3. Results

3.3.1. Questionnaires

3.3.1.1. SWAS

The four subscales of the Story World Absorption Scale all showed good or excellent reliability; Attention (5 items), Ω_t = .92; Transportation (5 items), Ω_t = .90; Emotional Engagement (6 items), Ω_t = .94, Mental Imagery (8 items), Ω_t = .93. Descriptive statistics per subscale and per story are given in Table 3.3.

Table 3.3.: *Mean, Standard Deviation, Median and Interquartile Range for the Scores on the SWAS Subscales, per Story*

Subscale		*M (SD)*	Median	IQR
Attention	Story A	5.286 (0.863)	5.400	4.800 – 6.000
	Story B	4.308 (1.191)	4.400	3.600 – 5.150
	Story C	3.825 (1.178)	3.700	3.000 – 4.650
Transportation	Story A	4.304 (1.071)	4.400	3.600 – 5.000
	Story B	3.515 (1.182)	3.600	2.800 – 4.350
	Story C	3.075 (1.117)	3.200	2.200 – 3.800
Emotional Engagement	Story A	4.680 (0.979)	4.833	4.125 – 5.333
	Story B	3.920 (1.189)	4.000	3.167 – 4.667
	Story C	3.299 (1.140)	3.333	2.333 – 4.042
Mental Imagery	Story A	5.516 (0.732)	5.625	5.000 – 6.000
	Story B	4.688 (0.932)	4.875	4.125 – 5.250
	Story C	4.099 (1.107)	4.312	3.344 – 4.875

3.3.1.2. Appreciation Questionnaire

The Appreciation Questionnaire was divided into two parts for the analysis. The first part, consisting of thirteen adjectives that could be used to describe the stories, was analyzed using a principal components analysis (PCA) with oblique rotation (direct oblimin). Using the Kaiser-Meyer-Olkin measure, it was determined that the sampling adequacy for this analysis was good, KMO = .87 (all KMO values for individual items > .75). There was sufficient correlation between items, as indicated by Bartlett's test of sphericity, χ^2 (78) = 2061.961, p < .001. An initial analysis showed that two components had eigenvalues over 1 (Kaiser's criterion). However, a model with two components did not fit the data well enough (fit based upon off diagonal values). Therefore, in the final model three components were retained. This model explained 68% of the variance. The first component contained items measuring the evoked interest in the story (beautiful, boring (-), captivating, interesting), the second component contained items measuring the emotional response to the story (sad, tragic, ominous, gripping, suspenseful), and the third component contained items measuring the positive affect elicited by the story (witty, funny, entertaining, special). The structure and pattern matrices for the factor loadings after rotation can be seen in Table 3.4. Factor scores per participant and story were used in the subsequent analyses.

Table 3.4.: *Summary of the Principal Components Analysis Results for the 13 Adjectives on the Appreciation Questionnaire (N = 294). Factor Loadings Over .40 Appear in Bold*

	Structure Matrix			Pattern Matrix		
	Evoked Interest	Emotional Response	Positive Affect	Evoked Interest	Emotional Response	Positive Affect
Beautiful	**0.813**	0.159	0.364	**0.910**	-0.058	-0.148
Boring	**-0.840**	-0.092	**-0.516**	**-0.842**	0.110	-0.043
Gripping	**0.566**	**0.731**	0.260	0.395	**0.636**	0.033
Entertaining	**0.490**	0.159	**0.749**	0.052	0.140	0.719
Funny	0.387	-0.311	**0.801**	0.018	-0.322	0.793
Interesting	**0.815**	0.323	**0.526**	**0.706**	0.153	0.128
Ominous	0.170	**0.774**	0.053	-0.067	**0.790**	0.085
Sad	0.076	**0.864**	-0.155	-0.065	**0.880**	-0.127
Suspenseful	**0.562**	**0.490**	**0.437**	0.321	**0.410**	0.253
Tragic	0.165	**0.864**	-0.032	-0.033	**0.872**	-0.021
Witty	**0.505**	0.054	**0.851**	0.024	0.041	0.837
Captivating	**0.887**	0.234	**0.674**	**0.723**	0.058	0.267
Special	**0.548**	0.377	**0.601**	0.195	0.326	0.488

A second part of the questionnaire consisted of a general score of story liking, and six questions regarding the enjoyment of the story, $\Omega_t = .96$ (7 items). The answers on these questions were collapsed into a mean score for General Appreciation. These General Appreciation scores turned out to be highly correlated with the Evoked Interest factor scores, $r_s = .846$, $p < .001$. To prevent multicollinearity, it was decided to use only the Evoked Interest factor score in further analyses as an indicator of evoked interest/general appreciation (EI/GA).

3.3.1.3. Top 3 Questionnaire

After reading all three stories, participants were asked to rank the stories from most appreciated to least appreciated. Most participants preferred story A, ranked story B as second best and story C as least appreciated (see Table 3.5). Note that participants read all stories in counterbalanced order, this preference was not an order effect.

Table 3.5.: *Percentage of Times Each Story Was Ranked as Most Appreciated, Intermediate or Least Appreciated (n=101)*

	Most Appreciated (%)	Intermediate (%)	Least Appreciated (%)
Story A	83.168	8.911	7.921
Story B	13.861	59.406	26.733
Story C	2.970	31.683	65.347

3.3.1.4. Reading Habits and Author Recognition Test

Reading habits were measured both directly using a reading habits questionnaire, and indirectly using the Author Recognition Test (ART). Because answers on the reading habits questionnaire were measured on a scale ranging from 1-5 on four of the five questions, but from 1-4 on the final question, z-scores were calculated for all questions on this questionnaire (higher values indicate more habitual readers). Overall reliability was sufficient if the question about non-fiction reading was excluded, $\Omega_t = .78$. The scores on the ART were positively skewed ($M = 7.324$, $SD = 4.695$, median = 6.000, IQR = 4.000 – 9.000) with higher values indicating more habitual readers.

3.3.1.5. Interpersonal Reactivity Index

Two subscales of the Interpersonal Reactivity Index were administered: The Fantasy subscale ($M = 5.134$, $SD = 0.861$, median = 5.167, IQR = 4.667 – 5.667) and the Perspective Taking subscale ($M = 5.059$, $SD = 0.917$, median = 5.143, IQR = 4.571 – 5.571). The Perspective Taking subscale was sufficiently reliable, $\Omega_t = .83$, and the Fantasy subscale was reliable if the item about daydreaming (i.e., *I daydream and fantasize, with some regularity, about things that might happen to me*) was dropped, $\Omega_t = .88$.

3.3.2. Eye Tracking Data

3.3.2.1. Gaze Duration

For the full model summaries, see Appendix C. A Linear Mixed Effects Regression model was created, that predicted gaze duration by the number of times the previous word was underlined for motor descriptions, perceptual descriptions and descriptions of mental events, controlling for lexical frequency, number of characters and surprisal value, and allowing random slopes per story per participant for the three different types of descriptions. All predictors were centered and scaled, to improve model fit. Variance Inflation Factors (VIFs) were calculated for this model, to check for multicollinearity (VIFs were calculated using the function reported on `https://github.com/aufrank/R-hacks/blob/master/mer-utils.R`). All VIFs were below five, the VIFs for the underlining-scores were all around one. This indicates that multicollinearity was not problematic in the models and all planned predictors were entered into the models.

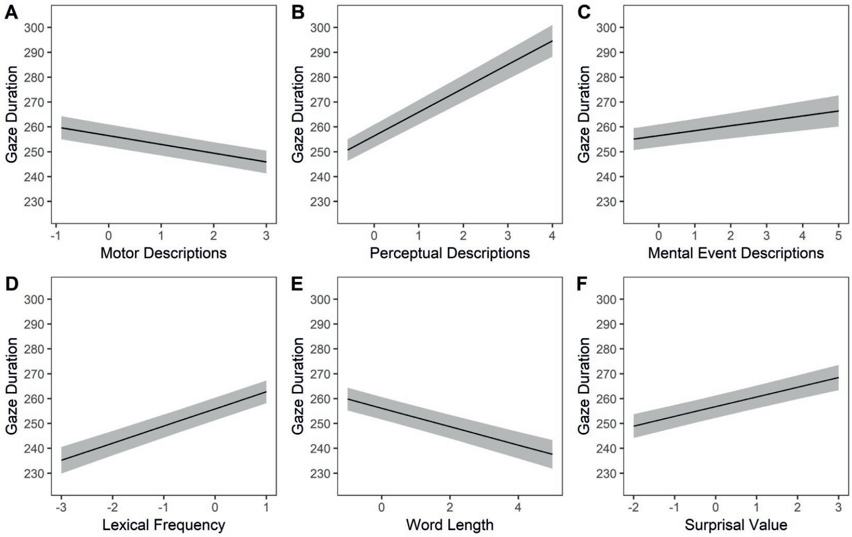
As can be seen in Table 3.6 and Fig. 3.4, motor descriptions were associated with a decrease in gaze duration (increased reading speed), whereas perceptual descriptions and descriptions of mental events were associated with an increase in gaze duration (decreased reading speed). More frequent or more unexpected (as reflected by a high surprisal value) words were read slower than infrequent or more expected words, as reflected by an increase in gaze duration towards frequent words and unexpected words. Longer words (words consisting of more characters) were associated with a decrease in gaze duration, reflecting increased reading speed. The effects of lexical frequency and word length were reversed to the effects that would have been expected based on previous research (see Data Analysis: Step 1). It should be noted that these effects were as expected in the analysis for the target word (in that analysis more frequent words were associated with a decrease in gaze duration (faster reading) and longer words with an increase in gaze duration (slower reading); for an overview of the results of the analysis for the target word, see Appendix D; for an elaborate discussion of these results, see the Discussion section). As can be seen in Fig. 3.5, the associations between all types of descriptions and gaze duration varied between participants (and between different stories within participants; more detailed figures per participant can be found in Appendix E1-3).

At a closer look, interesting individual differences in the association between simulation and gaze duration are visible (see Fig. 3.5). For motor descriptions, the response of participants was rather homogeneous: nearly all participants showed decreased gaze duration after reading motor descriptions. Comparably, all participants showed an increase in gaze duration after reading perceptual descriptions. The association between mental events and gaze duration, however, seemed more variable between (and within, see Appendix E3) participants, sometimes being associated with an increase and sometimes with a decrease in gaze duration (see Fig. 3.5). Even though on average there was a decrease in

Table 3.6.: *Coefficients of the Model Predicting Gaze Duration by Motor Descriptions, Perceptual Descriptions and Descriptions of Mental Events, Taking into Account the Influence of Lexical Frequency, Number of Characters and Surprisal Value (all Predictors Taken from the Previous Word)*

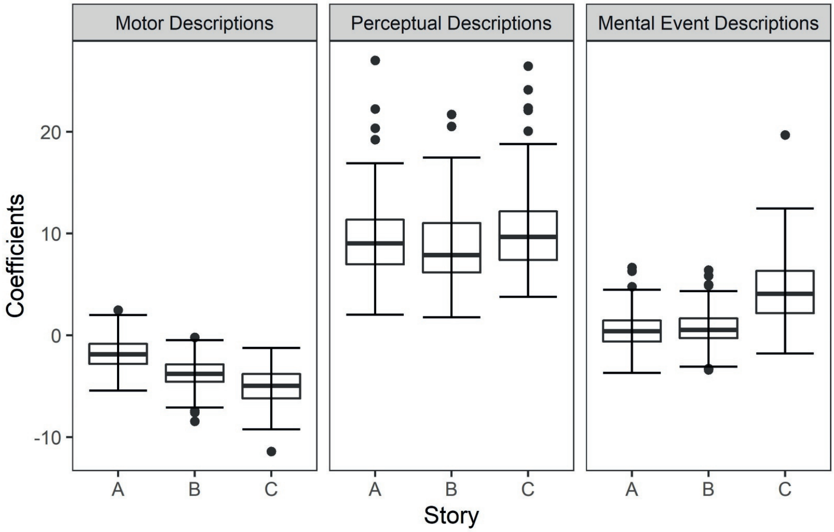
	B	*SE*	*df*	*t*-value	*p*-value
(Intercept)	255.500	2.295	294.1	111.317	<.001
Motor Descriptions	-3.525	0.286	274.5	-12.316	<.001
Perceptual Descriptions	9.555	0.354	318.3	26.977	<.001
Mental Event Descriptions	1.980	0.321	326.2	6.165	<.001
Lexical Frequency	6.884	0.462	449200	14.908	<.001
Number of Characters	-3.707	0.354	464900	-10.463	<.001
Surprisal Value	3.912	0.387	462500	10.097	<.001

Figure 3.4.: *Effects Plots for the Predictors of Gaze Duration in the Spillover Area. Note that All Predictors are Centered and Scaled. Gaze Duration Is Given in Milliseconds, the Grey Areas Indicate the 95-Percent Confidence Intervals*



reading speed for mental event descriptions, this was not always the case on the individual level. In fact, a considerable number of participants showed an increased instead of a decreased reading speed when reading mental event descriptions, in particular for stories A and B. When comparing the coefficients for the relationships of the different types of descriptions with gaze duration, a significant negative correlation between the coefficients for the associations of motor descriptions and mental event descriptions with gaze duration appeared,

Figure 3.5.: *Range of Coefficients Across Participants for the Relationships Between Different Types of Descriptions and Gaze Duration, Depicted per Story and per Type of Description*



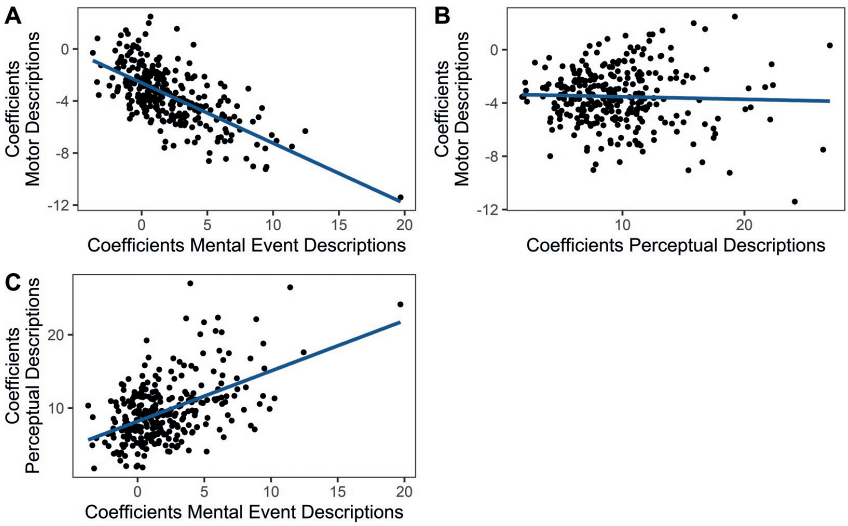
r_s = -.622, $p <. 001$.[4] This indicated that participants showing a strong negative relationship between motor descriptions and gaze duration, showed a strong positive relationship between mental event descriptions and gaze duration, and vice versa (see Fig. 3.6A). In contrast, no correlation was found between the coefficients for the associations of motor descriptions and perceptual descriptions with gaze duration, r_s = -.003, p = .964 (see Fig. 3.6B). There was a significant positive correlation between the coefficients for the associations of perceptual descriptions and mental event descriptions with gaze duration, r_s = .403, $p < .001$, indicating that participants showing a strong positive relationship between perceptual descriptions and gaze duration also showed a strong positive relationship between mental event descriptions and gaze duration (see Fig. 3.6C).

3.3.2.2. Individual Differences: Gaze Duration

To test whether self-report measures of reading experiences were associated with simulation as measured using eye tracking, coefficients for the associations of the three types of descriptions with gaze duration were derived per story per participant (total number of coefficients = 294). I created three new models in

[4]Note that this analysis was exploratory, and was not part of the pre-registration of this experiment.

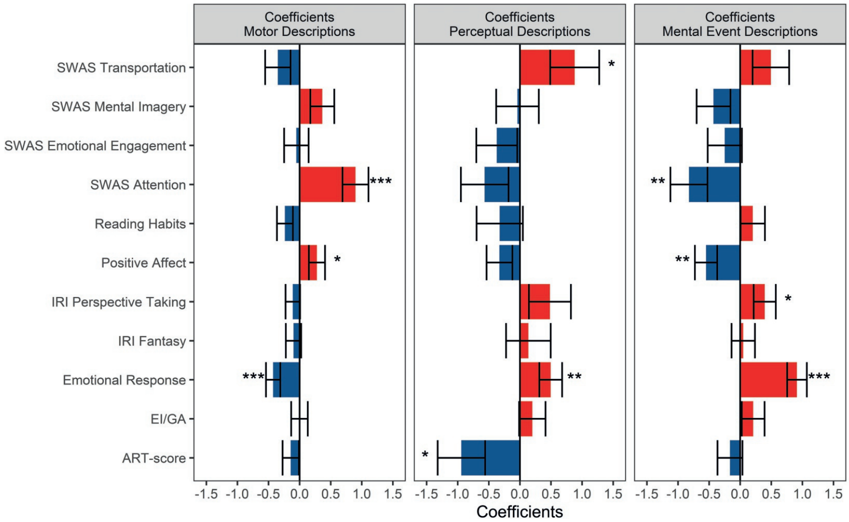
Figure 3.6.: *Correlation Between the Coefficients of the Relationships of (A) Motor Descriptions and Mental Event Descriptions with Gaze Duration, (B) Motor Descriptions and Perceptual Descriptions with Gaze Duration, and (C) of Perceptual and Mental Event Descriptions with Gaze Duration*



which I tried to explain individual differences in the strength of these associations by scores on the four subscales of the SWAS, the three factor scores for appreciation, the Fantasy and Perspective Taking subscales of the Interpersonal Reactivity Index and the direct and indirect measures of reading habits (i.e., the reading habits questionnaire and Author Recognition Test). Random intercepts were allowed per participant. Again, all predictors were centered and scaled, to improve model fit. VIFs were below five for all predictors (indicating that multicollinearity was not problematic in the models) and all planned predictors were entered into the models. An overview of the results of the three models can be seen in Fig. 3.7.

3.3.2.2.1. Motor Simulation As can be seen in Table 3.7 and Fig. 3.8, the Attention subscale of the SWAS, the evoked emotional response, and the positive affect elicited by the story were significantly associated with the strength of the relationship between motor descriptions and gaze duration. Attention and positive affect were positively associated with the strength of the relationship between motor descriptions and gaze duration, implying that this relationship (faster reading of motor descriptions) was attenuated (motor descriptions were read relatively slower) in people who reported higher attention towards the story

Figure 3.7.: *Coefficients per Predictor for Each of the Three Models Predicting the Strength of the Relationships Between the Three Types of Descriptions and Gaze Duration by Individual Differences Measured with Questionnaires. Negative Coefficients Appear in Light Grey, Positive Coefficients in Dark Grey. Error Bars Indicate Standard Errors. EI/GA = Evoked Interest/General Appreciation. Significant Predictors Are Marked (* p < .05, ** p < .01, *** p < .001)*



world or higher positive affect after reading a story. Evoked emotional response was negatively associated, implying that people who reported experiencing a high level of emotion while reading a story, read motor descriptions even faster (the strength of the relationship between motor descriptions and gaze duration was increased).

3.3.2.2.2. Perceptual Simulation Table 3.8 and Fig. 3.9 show that the Transportation subscale of the SWAS was significantly associated with the strength of the relationship between perceptual descriptions and gaze duration, as well as the emotional response evoked by the story and the ART-score (indirect measure of reading habits). Reading habits were negatively associated with the strength of the relationship between perceptual descriptions and gaze duration, implying that in people reporting being more habitual readers, the relationship between perceptual descriptions and gaze duration (slower reading of perceptual descriptions) was attenuated (perceptual descriptions were read relatively faster). Transportation and the evoked emotional response, however, were positively associated: Individuals reporting a higher level of transportation or experienced emotion while reading a story, read perceptual content even slower (the

Table 3.7.: *Coefficients of the Model Predicting the Effect of Motor Descriptions on Gaze Duration by Scores on the Questionnaires. Significant Predictors Are Marked (* p <.05, ** p <.01, *** p <.001)*

	B	*SE*	*df*	*t*-value	*p*-value
(Intercept)	-3.530	0.110	90.95	-32.238	<.001***
SWAS Attention	0.899	0.209	227.43	4.300	<.001***
SWAS Transportation	-0.350	0.204	199.36	-1.714	.088
SWAS Emotion	-0.053	0.197	275.44	-0.268	.789
SWAS Mental Imagery	0.365	0.192	244.87	1.900	.059
Evoked Interest/General Appreciation	-0.002	0.133	290.22	-0.013	.990
Emotional Response	-0.426	0.115	290.89	-3.709	<.001***
Positive Affect	0.279	0.130	290.19	2.15	.032*
IRI Fantasy	-0.098	0.124	95.76	-0.793	.430
IRI Perspective Taking	-0.109	0.116	94.81	-0.937	.351
ART-Score	-0.142	0.131	96.70	-1.082	.282
Reading Habits	-0.236	0.128	95.65	-1.841	.069

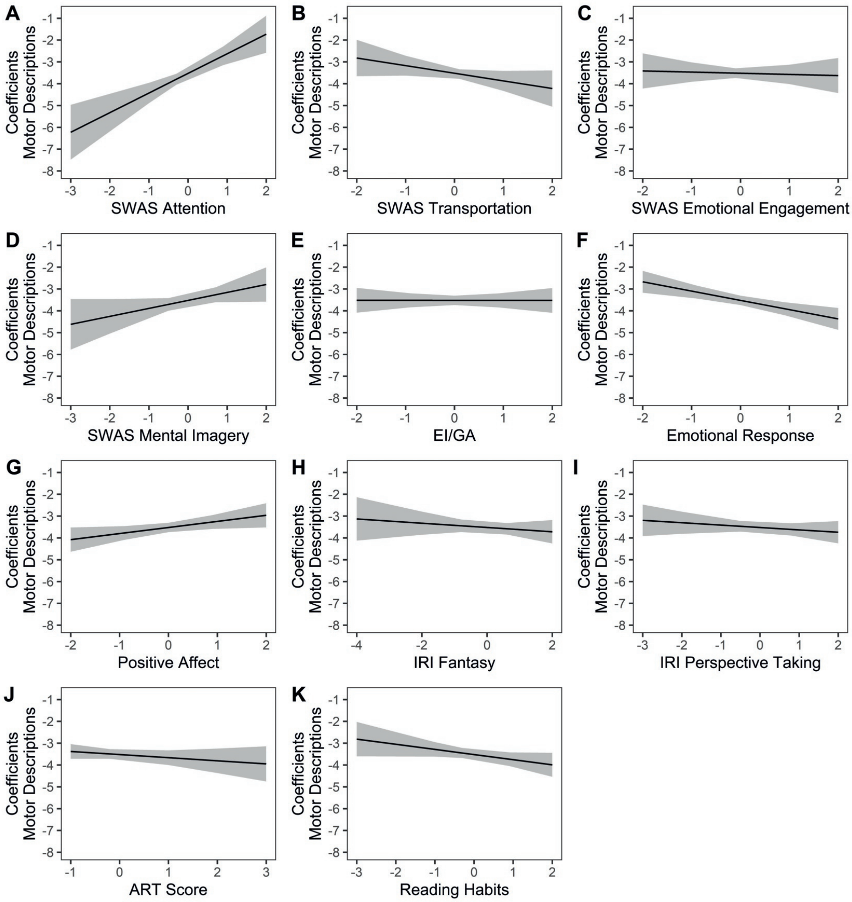
strength of the relationship between perceptual descriptions and gaze duration was increased).

Table 3.8.: *Coefficients of the Model Predicting the Effect of Perceptual Descriptions on Gaze Duration by Scores on the Questionnaires. Significant Predictors Are Marked (* p <.05, ** p <.01, *** p <.001)*

	B	*SE*	*df*	*t*-value	*p*-value
(Intercept)	9.552	0.323	100.84	29.596	<.001***
SWAS Attention	-0.566	0.382	275.56	-1.481	.140
SWAS Transportation	0.881	0.394	287.24	2.237	.026*
SWAS Emotion	-0.372	0.329	250.04	-1.129	.260
SWAS Mental Imagery	-0.039	0.343	270.23	-0.114	.910
Evoked Interest/General Appreciation	0.200	0.212	237.45	0.941	.348
Emotional Response	0.496	0.183	237.33	2.712	.007**
Positive Affect	-0.329	0.208	239.25	-1.578	.116
IRI Fantasy	0.136	0.359	103.57	0.381	.704
IRI Perspective Taking	0.483	0.338	104.48	1.430	.156
ART-Score	-0.942	0.382	103.19	-2.468	.015*
Reading Habits	-0.325	0.372	102.90	-0.872	.385

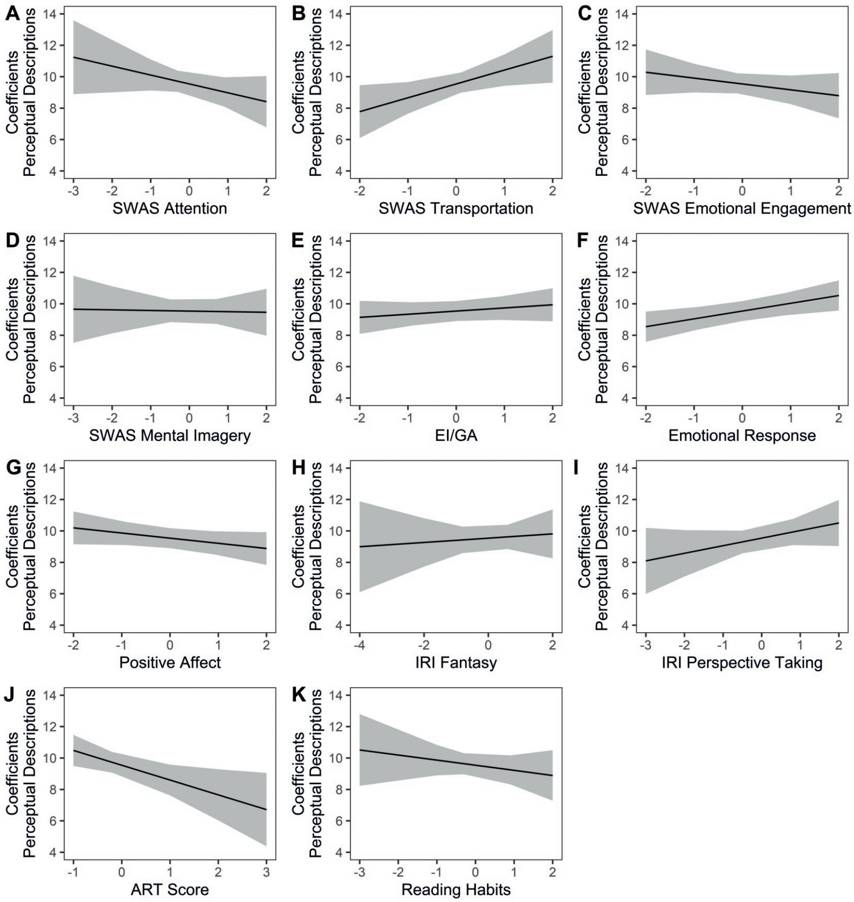
3.3.2.2.3. Mentalizing The strength of the relationship between descriptions of mental events and gaze duration was associated with scores on the Attention subscale of the SWAS, as well as the evoked emotional response and the positive affect elicited by the story, and the perspective taking-ability of the participants (see Table 3.9 and Fig. 3.10). Attention and elicited positive affect both had a negative association with the strength of the relationship between mental event descriptions and gaze duration: in readers who reported higher

Figure 3.8.: *Effects Plots for the Predictors of the Strength of the Relationship Between Motor Descriptions and Gaze Duration. Note that All Predictors Are Centered and Scaled. The Grey Areas Indicate the 95-Percent Confidence Intervals*



attention or higher positive affect, the strength of this relationship was attenuated (resulting in some participants even showing an increased instead of decreased reading speed when reading mental event descriptions). In contrast, evoked emotional response and participants' perspective taking abilities had a positive association with the strength of the relationship between mental event descriptions and gaze duration. The strength of this relationship was increased in participants who reported experiencing a high level of emotion while reading a story or who reported often considering other people's perspectives.

Figure 3.9.: *Effects Plots for the Predictors of the Strength of the Relationship Between Perceptual Descriptions and Gaze Duration. Note that all Predictors Are Centered and Scaled. The Grey Areas Indicate the 95-Percent Confidence Intervals*



3.3.2.3. Regression Probability

For the full model summaries, see Appendix C. A Generalized Linear Mixed Effects model was created, that predicted the probability of regressing back to a word by the number of times this word was underlined for motor descriptions, perceptual descriptions and descriptions of mental events, controlling for lexical frequency, number of characters and surprisal value, and allowing random intercepts and slopes for underlining-scores per story per participant. Again, all predictors were centered and scaled, to improve model fit. All VIFs for this model were below five, the VIFs for the underlining-scores were all close to one.

Table 3.9.: *Coefficients of the Model Predicting the Effect of Descriptions of Mental Events on Gaze Duration by Scores on the Questionnaires. Significant Predictors Are Marked (* p <.05, ** p <.01, *** p <.001)*

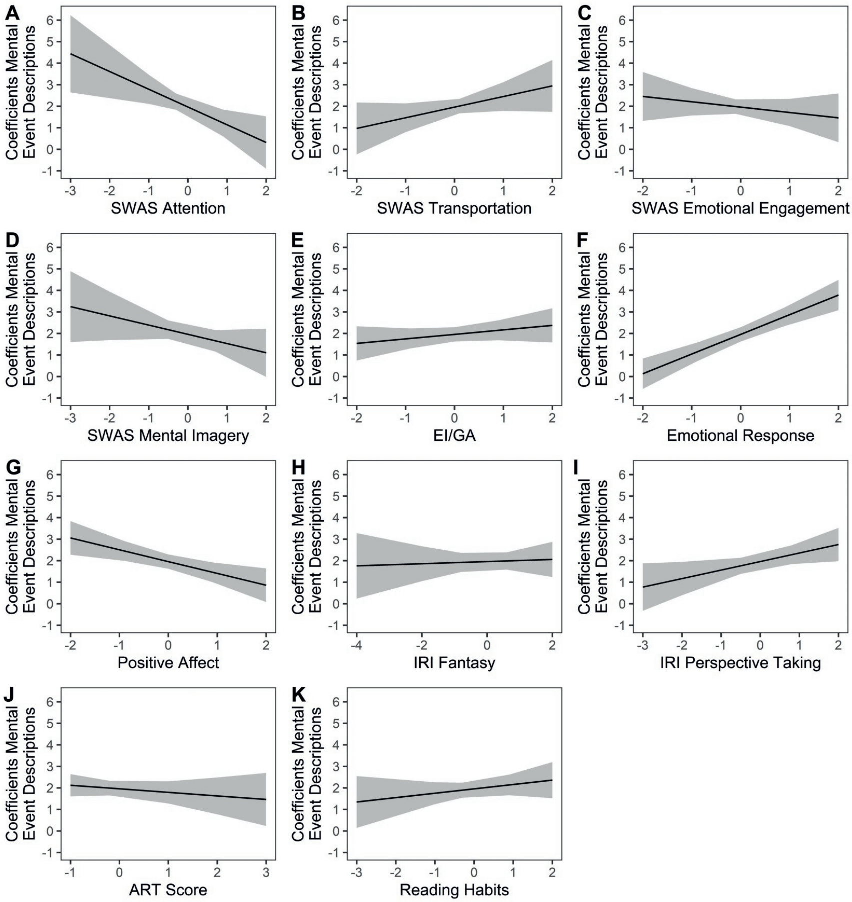
	B	*SE*	*df*	*t*-value	*p*-value
(Intercept)	1.969	0.168	100.68	11.734	<.001***
SWAS Attention	-0.824	0.298	260.89	-2.766	.006**
SWAS Transportation	0.494	0.294	235.54	1.679	.095
SWAS Emotion	-0.248	0.275	289.63	-0.902	.368
SWAS Mental Imagery	-0.429	0.272	272.57	-1.576	.116
Evoked Interest/General Appreciation	0.209	0.183	288.95	1.140	.255
Emotional Response	0.913	0.158	287.99	5.778	<.001***
Positive Affect	-0.549	0.179	289.39	-3.067	.002**
IRI Fantasy	0.049	0.189	105.39	0.262	.794
IRI Perspective Taking	0.395	0.178	104.99	2.221	.028*
ART-Score	-0.164	0.201	105.95	-0.820	.414
Reading Habits	0.203	0.195	105.01	1.037	.302

As can be seen in Table 3.10 and Fig. 3.11, motor descriptions, perceptual descriptions and descriptions of mental events were all associated with a decrease in the probability of regressing back to a word. More frequent or more unexpected (as reflected by a high surprisal value) words were more likely to be looked back to than infrequent or more expected words, as reflected by an increase in the probability of regressing back to frequent words and unexpected words. Longer words (words consisting of more characters) were associated with a decrease in the probability of regressing back to that word.

Table 3.10.: *Coefficients of the Model Predicting the Probability of Regressing Back to a Word by Motor Descriptions, Perceptual Descriptions and Descriptions of Mental Events, Taking into Account the Influence of Lexical Frequency, Number of Characters and Surprisal Value*

	B	*SE*	*z*-value	*p*-value
(Intercept)	-1.203	0.028	-43.075	<.001
Motor Descriptions	-0.070	0.004	-16.959	<.001
Perceptual Descriptions	-0.033	0.004	-7.869	<.001
Mental Event Descriptions	-0.043	0.004	-9.767	<.001
Lexical Frequency	0.180	0.008	22.894	<.001
Number of Characters	-0.209	0.006	-36.109	<.001
Surprisal Value	0.199	0.007	30.139	<.001

Figure 3.10.: *Effects Plots for the Predictors of the Strength of the Relationship Between Mental Event Descriptions and Gaze Duration. Note that all Predictors Are Centered and Scaled. The Grey Areas Indicate the 95-Percent Confidence Intervals*



3.3.2.4. Individual Differences: Regression Probability

The relationships between the three types of descriptions and the regression probability did not vary much between participants (and between different stories within participants; see also Fig. 3.12). As a result, no notable associations were found between any of the measures of individual differences (i.e., SWAS, appreciation, IRI, and measures of reading habits) and the strength of the relationship between any type of description and regression probability.

Figure 3.11.: *Effects Plots for the Predictors of the Probability of Regressing Back to a Word. Note that All Predictors Are Centered and Scaled. The Grey Areas Indicate the 95-Percent Confidence Intervals*

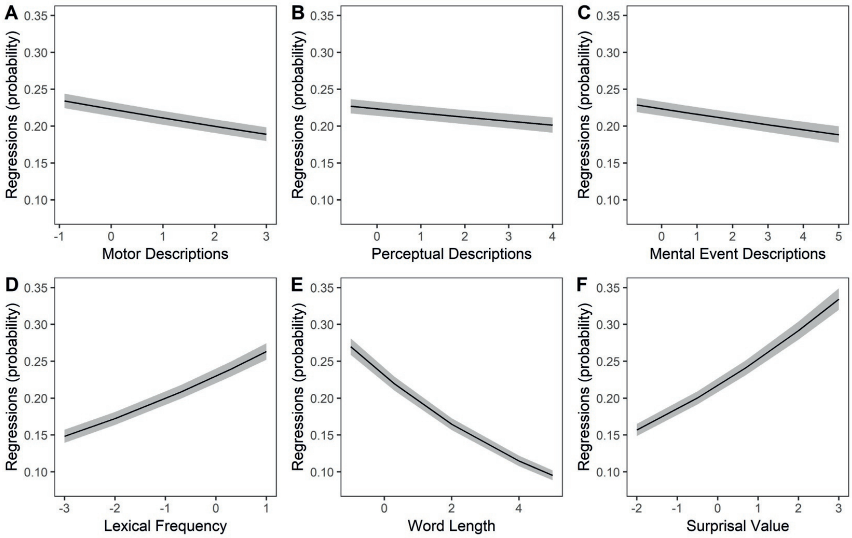
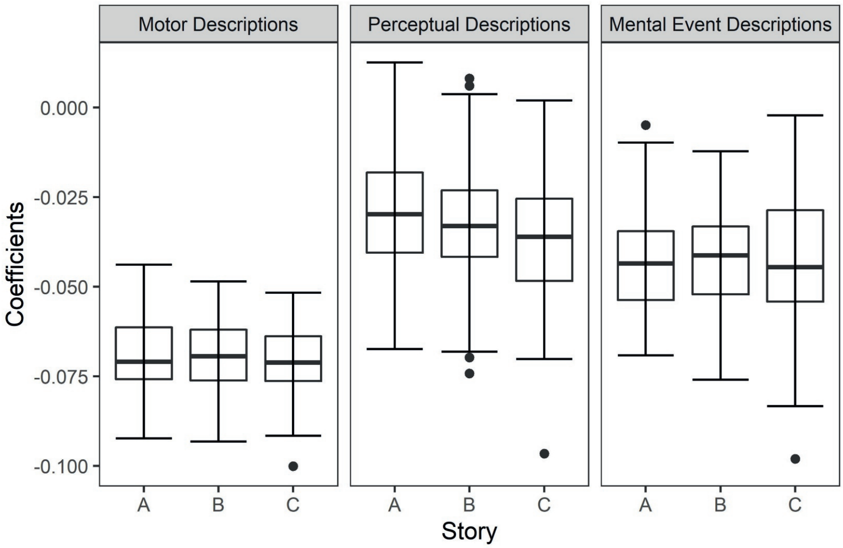


Figure 3.12.: *Range of Coefficients Across Participants for the Relationships Between Different Types of Descriptions and Regression Probability, Depicted per Story and per Type of Description*



3.4. Discussion

In the experiment described in this chapter, I found associations of motor descriptions, perceptual descriptions and mental event descriptions with gaze duration. Nijhof and Willems (2015) found that descriptions of action content and mentalizing content elicited motor simulation and mentalizing, as shown in fMRI data. For more examples of the validity of subjective ratings as a proxy for the involvement of simulation and embodiment in the processing of words and sentences, see also Kurby and Zacks (2013; who used subjective ratings of imagery modalities to assess imagery elicited by stories); and Binder et al. (2016) and Fernandino et al. (2015; who both used subjective ratings of words on several semantic components to predict brain activation patterns as a response to these words). Therefore, the motor descriptions in the current experiment are assumed to elicit motor simulation, the perceptual descriptions are assumed to elicit perceptual simulation, and the mental event descriptions (comparable to mentalizing content) are assumed to elicit mentalizing.

I found that motor simulation was associated with shorter gaze duration (faster reading), whereas perceptual simulation and mentalizing were associated with longer gaze duration (slower reading). Possibly, the processes of perceptual simulation and mentalizing are rather demanding and time-consuming, thus prolonging gaze durations (slowing down reading speed; cf. the idea that gaze duration is indicative of the ease or difficulty of processing, see Just & Carpenter, 1980). Apparently, motor simulation works differently, as this had the opposite relationship with gaze durations: people read passages richer in descriptions of actions relatively fast. Because I looked at the degree to which a passage described action, this is compatible with the findings of Marino, Borghi, Buccino, and Riggio (2017), who found that people reacted to sentences containing two verbs describing actions (e.g., “grasp and use”) faster than to sentences containing two verbs describing observational acts (e.g., “look at and stare”). This finding suggests that sentence processing is faster for sentences that are more action-laden, which fits my findings of shortened gaze duration as a function of the degree to which words are considered action descriptions (and consequently the degree to which they are likely to elicit motor simulation).

Importantly, a difference between the simulation of action language and mentalizing has already been found by Wallentin et al. (2013). Their findings led these authors to claim that the processes underlying action simulation and mentalizing rely on different cognitive systems, which fits with the finding that the relationships of motor simulation, perceptual simulation and mentalizing with reading behavior are essentially different from each other.

Apart from these differences between the three kinds of simulation on the group level, there was also quite some individual variation in the strength of the relationship between simulation and reading speed. Interestingly, the relationship of mentalizing with reading behavior (i.e., gaze duration) correlated negatively with the relationship of motor simulation with gaze duration. This means that participants who read faster when encountering motor-related content, read slower when encountering mentalizing-related content. The group results from the current experiment suggest that faster reading of motor-related content is an indicator of increased motor simulation (see above). If this is indeed the case, the present negative correlation is best interpreted as a sign that those who engage in motor simulation also tend to engage in mentalizing (which is characterized by slower reading). Comparably, the relationship of perceptual simulation with reading behavior (i.e., slower reading) correlated positively with the relationship of mentalizing with reading behavior, indicating that participants that engage in perceptual simulation also tend to engage in mentalizing. Note however that this is a different conclusion than based on an earlier fMRI study using a similar approach. That is, Nijhof and Willems (2015) observed a negative correlation between activation levels in motor areas (in reaction to motor-related content) and medial prefrontal cortex (in reaction to mentalizing content), suggesting that participants do not engage both in motor simulation and mentalizing. It is difficult to find conclusive evidence for or against these scenarios in the present data and it will be a task of future research to investigate individual differences in cortical activation levels during different kinds of mental simulation in a larger sample than has been done to date.

In the analysis, I controlled for a number of factors known to influence reading behavior. Higher surprisal value (indicating lower expectancy of a word given its context) was associated with prolonged gaze duration (slower reading) in the spill-over region (i.e., gaze durations on words occurring after unexpected words were relatively long). Interestingly, lexical frequency was associated with prolonged gaze duration (i.e., gaze durations on words occurring after frequent words were relatively long) and word length (number of characters) was associated with shortened gaze duration (i.e., gaze durations on words occurring after longer words were relatively short) in the spill-over region. These effects are surprising, as frequent words are generally associated with faster reading and longer words with slower reading. It is important to note that, when analyzing the data for the target area, the results were as expected (i.e., frequent words were read relatively fast, longer words were read relatively slow), contrary to some of my results for the spill-over area. Interestingly, a reversed effect for word length in the spill-over area has been found before (although in a slightly different context; Pollatsek, Juhasz, Reichle, Machacek, & Rayner, 2008). Pollatsek et al. (2008) propose in the E-Z Reader 10 model, that in some instances, the eye fixation may already have shifted towards word $n + 1$, even though the meaning of word n has not been fully integrated into the sentence representation (see also Sereno & Rayner, 2003). This way, word $n + 1$ will be fixated slightly longer (as first word n needs to be integrated and subsequently word $n + 1$ still needs to be processed and integrated). Perhaps short gaze durations on short words do not always allow integration to be fully completed before the gaze is shifted towards the next word. This could explain why in the experiment reported here short words were associated with longer gaze duration in the spill-over region. The paradoxical effect of lexical frequency can be explained in a similar fashion: short gaze durations on frequent words may not always allow for full integration of a word into a sentence representation before the shift to the next word is made. Consequently, gaze durations in the spill-over region may be prolonged for frequent words: in the spill-over region, integration of word n (the frequent word) still has to be completed before word $n + 1$ can be processed.

As I controlled for lexical frequency, word length and surprisal value in the analyses, it is unlikely that differences in these characteristics between motor descriptions and other parts of the stories caused the negative association between motor simulation and gaze duration. Moreover, when looking at the distribution of the data from the pre-test (see Appendix A1-3), it becomes visible that motor descriptions occurred comparably often in all parts of the sentences. The same was true for perceptual descriptions and mental event descriptions. As a result, it is unlikely that the relationship between motor descriptions and gaze duration merely reflects an effect of position in the sentence. Apparently, motor descriptions are processed differently from the rest of the text (and differently from other types of descriptions), suggesting that motor descriptions might actually be easier to process than other parts of the text (as they are associated with faster reading). How this exactly relates to motor simulation, and whether motor simulation is indeed "easier" than other types of simulation, still remains to be seen.

In the second stage of the analyses, I investigated whether individual differences in the strength of the relationship between simulation and gaze duration could be linked to individual differences in absorption, appreciation for the stories, reading habits, and interpersonal reactivity (fantasy and perspective taking). I found that answers on the Attention subscale of the SWAS were negatively related to the strength of the relationships of motor simulation and mentalizing with gaze duration: A high level of attention towards a story was associated with a relatively weak association between motor simulation / mentalizing and gaze duration. Interestingly, the negative relationship between the strength of the relationship of motor simulation and mentalizing with reading and attention is somewhat reminiscent of the attenuation of the effect of lexical and linguistic variables on reading during mindless reading (see Reichle, Reineberg, & Schooler, 2010). It seems that people's experiences while reading texts influence the effects of a number of variables that influence normal reading. However, attention and mindless reading seem to be opposite one another: it is likely that participants engaging in mindless reading will report low attention to the stories they read. As a consequence of this mindless reading, participants may be more prone to simulate the events in the stories they read, perhaps because of a more associative reading style. This could explain why I found low attention to be associated with more simulation. When and how this exactly works is a question that will need to be much more thoroughly investigated.

The emotional response evoked by the stories, on the other hand, was positively related to the strength of the relationships between all kinds of simulation and gaze duration. Participants who found the stories more sad, tragic, ominous, gripping, and suspenseful showed relatively strong relationships between simulation and gaze duration. This can be interpreted as evidence that participants who were moved by the stories they read were more prone to mentally simulate the events happening in the story (or the other way around: participants who are more prone to simulate the events happening in a story are more moved by the stories). This is an extension of what Oatley (1995) suggested about simulation and emotion. With simulation, he meant identification with characters and the simulation of the emotions of the characters in a story, which is reminiscent of the concept of mentalizing (simulation of mental events, such as thoughts and emotions) I use here. He suggested that readers simulate the emotions of a character by tapping into their own emotional experience. This implies that when readers report more emotions elicited by a story, this results from a more vivid simulation of the emotions described in the story. This explains why in the experiment reported here the strength of the relationship of mentalizing (and to a lesser extent motor and perceptual simulation) with gaze duration was larger in participants reporting a higher emotional response to the stories.

Answers on the Transportation subscale of the SWAS were positively related to the strength of the relationship between perceptual simulation and gaze duration. Participants experiencing higher levels of transportation into the story world also showed a stronger relationship between perceptual simulation and gaze duration. This suggests a role for simulation in transportation, which is supportive of the different theories surrounding transportation and absorption described in the introduction (Green & Brock, 2000; Kuijpers et al., 2014), stating that simulation is an important part of absorption.

Higher perspective taking-abilities were associated with a stronger relationship between mentalizing and gaze duration. It is interesting to note that perspective taking-abilities are only associated with mentalizing, and not with motor or perceptual simulation. The degree to which participants engage in mentalizing seems to be specifically associated with perspective taking-abilities. Indeed, the Perspective Taking subscale of the IRI correlates with measures of empathy and EQ (Davis, 1983; De Corte et al., 2007). Moreover, empathy and sympathy (of which perspective taking as measured by the IRI is one component; Davis, 1980) are important aspects of mentalizing (see Miall & Kuiken, 2002). The close relationship between mentalizing and perspective taking from a theoretical standpoint, and the association between strength of the relationship of mental event descriptions with gaze duration and individual differences in perspective taking found in the experiment described in this chapter, together confirm that mental event descriptions indeed elicit mentalizing. This finding opens up the possibility of using reading behavior (in the sense of gaze durations towards mentalizing-eliciting aspects of a story) as an implicit indicator of social perspective taking abilities.

Reading habits were negatively associated with the strength of the relationship between perceptual simulation and gaze duration (meaning that more habitual readers read perceptual descriptions relatively fast). Interestingly, a comparable result was found in a combined analysis of several eye tracking datasets including the present dataset (Eekhof et al., 2021). It was observed that participants showing the weakest relationships between word characteristics (such as lexical frequency) and gaze durations, reported being relatively avid readers. This suggests that more avid readers are more "detached" from low-level word characteristics. The results reported in this chapter suggest that, to some extent, the same can be said about simulation-eliciting content: more avid readers seem to be less influenced by this kind of content.

In addition to the relationship between simulation and gaze duration, I looked at the relationships of motor simulation, perceptual simulation and mentalizing with the probability of regressing back to a word. I found that highly descriptive (and thus simulation-inducing) words were slightly (but significantly) less likely to be looked back to. This suggests that these words are easier to process than words in the remainder of the stories (see Rayner, 1998). In contrast to the findings regarding gaze duration, no notable differences in the strength of the relationships between the number of regressions and the three kinds of simulation were found, both on the group level and within participants.

The amount of individual variation in the extent to which simulation was associated with gaze duration while reading literary stories, confirmed that mental simulation is not evoked equally in all people. This accounts for the differences between participants in experienced transportation, of which simulation is an underlying process (e.g., Green & Brock, 2000; Kuijpers et al., 2014). Because individual differences in transportation have been found to correlate with story appreciation (Busselle & Bilandzic, 2009; Green et al., 2004; Hartung et al., 2016; Kuijpers et al., 2014), it would be interesting to find out whether there is a direct link between simulation (as an underlying process of transportation) and appreciation. Interestingly though, in the experiment reported here, individual differences in simulation were not directly associated with individual differences in general measures of story appreciation. However, when looking at a more indirect measure of appreciation (using adjectives describing the stories, cf. Knoop et al., 2016), individual variation in scores on this measure could be linked to individual variation in simulation. Perhaps the more direct measures of appreciation were correlated too highly with measures of transportation, and could therefore not explain enough individual variance. In any case, the connection between individual differences in simulation and the individual differences in story appreciation deserves more attention in future research.

3.4.1. Conclusion

In conclusion, I found that motor simulation, perceptual simulation and mentalizing are differentially associated with gaze duration in literary reading. Consequently, it is important not to assume that all kinds of simulation have a similar effect on reading behavior, but to take the individual effects of the different kinds of simulation into consideration. If we do not take this into consideration, but instead study mental simulation in general (or just combine motor simulation and perceptual simulation into sensorimotor simulation), we will overlook the differential effects of the different kinds of simulation (or even be unable to find any results, because of opposite effects of different kinds of simulation on language processing).

Apart from these differential associations between the three kinds of simulation and gaze duration, I found that individual differences in simulation were related to aspects of story world absorption and of story appreciation. I showed that simulation is related to absorption, and that there is some evidence for a direct connection between simulation and appreciation (a connection which has so far only been found between absorption and appreciation). Future research should delve deeper into the precise mechanisms underlying these relations.

4 | Different Kinds of Simulation During Literary Reading: Insights from a Combined fMRI and Eye Tracking Study

Mental simulation is an important aspect of narrative reading. In a previous study, I found that gaze durations are differentially impacted by different kinds of mental simulation. Motor simulation, perceptual simulation, and mentalizing as elicited by literary short stories influenced eye movements in distinguishable ways (Mak & Willems, 2019). In the current study, I investigated the existence of a common neural locus for these different kinds of simulation. I additionally investigated whether individual differences during reading, as indexed by eye movements, are reflected in domain-specific activations in the brain. I found a variety of brain areas activated by simulation-eliciting content, both modality-specific brain areas and a general simulation area. Individual variation in percent signal change in activated areas was related to measures of story appreciation as well as personal characteristics (i.e., transportability, perspective taking). Taken together, these findings suggest that mental simulation is supported by both domain-specific processes grounded in previous experiences, and by the neural mechanisms that underlie higher-order language processing (e.g., situation model building, event indexing, integration).

This Chapter Is Based on

Mak, Marloes, Faber, Myrthe, & Willems, Roel M. Different kinds of simulation during literary reading: Insights from a combined fMRI and eye-tracking study (submitted)

In Chapter 3, I described an eye tracking experiment in which participants read three literary short stories while their eye movements were being tracked. I found that descriptive language, hypothesized to elicit mental simulation, was associated with changes in reading behavior. Motor descriptions were associated with faster reading, whereas perceptual and mental event descriptions both were associated with slower reading. The strength of this association between descriptive language and reading speed was dependent on the reader. In some, reading behavior was much more susceptible to descriptive language than in others. These individual differences were associated with subjective reading experiences, such as absorption and appreciation.

In Chapter 4, I describe a combined fMRI and eye tracking experiment with a very similar design to the experiment reported in Chapter 3. Again, participants read literary stories while their eye movements are being tracked, but this time the reading takes place inside an MRI scanner. In the study described in this chapter, I aim to answer the question by what neural mechanisms simulation is supported and if I can uncover similar individual differences on the neural level as I did in the eye tracking data.

4.1. Introduction

Many readers experience mental simulation while they read. Mental simulation has been defined as, *"the reenactment of perceptual, motor, and introspective states acquired during experience with the world, body, and mind"* (Barsalou, 2008, p. 618). This definition implies the presence of different kinds of simulation: perceptual simulation, motor simulation, and the simulation of introspective states (more commonly called "mentalizing"). Research has shown that, indeed, there is evidence for a difference between these three kinds of simulation in language processing. For example, in a study in which participants listened to audio-narratives, Nijhof and Willems (2015) found that motor simulation and mentalizing activated different brain areas. Moreover, the second part of Barsalou's definition states that simulation is *"the reenactment of ... states acquired during experience with the world, body and mind"*. As each person's experiences with the world, body and mind are different from everyone else's experiences, this grounding of mental simulation in previous experiences implies that individuals will experience mental simulation in different ways. Indeed, individual differences in the connectivity between areas in both the visual and motor cortices were found to be related to personal experience in a study which looked at the relationship between the understanding of narrative events (i.e., actions, visual descriptions of scenes) and functional connectivity in the brain (Chow et al., 2015). Additionally, in the aforementioned study by Nijhof and Willems (2015), there was a negative correlation between the levels of activation in motor areas and mentalizing areas, indicating that some readers are more prone to motor simulation as opposed to mentalizing, and the other way around. In the previous chapter I found that the three different kinds of mental simulation as described by Barsalou indeed have qualitatively different effects on eye movements during reading. Here I ask whether these different kinds of simulation share an overlapping neural locus, and / or are domain-specific processes. I examine if and how mental simulation during literary reading is visible on the neural level and if this differs between different kinds of simulation. Secondly, I attempt to find out whether there are individual differences in brain activation levels in the brain regions implicated in simulation.

4.1.1. Differences Between Motor, Perceptual and Mental Simulation: Eye Movements

As described above, in the previous chapter I found differences in reading behavior between perceptual simulation, motor simulation, and mentalizing (see Mak & Willems, 2019). Reading behavior was studied in an eye tracking experiment in which participants read literary short stories. It was found that motor simulation reduced gaze duration (faster reading), whereas perceptual simulation and mentalizing increased gaze duration (slower reading). Additionally, individual differences in the effect of simulation on gaze duration were found, which were most striking in the case of mentalizing: although on average mentalizing increased gaze duration, there was a sizeable number of participants for whom mentalizing actually decreased gaze duration at the individual level. These individual differences in simulation were related to aspects of story world absorption and story appreciation. For example, the more attention someone paid to the story, the less their gaze behavior was affected by mental simulation. In contrast, the higher someone's emotional response to the story, the more their gaze behavior was affected by mental simulation.

These findings show that different kinds of mental simulation during narrative reading exist, and that people differ in how much they engage in either kind of simulation. In the current chapter, I investigate which brain areas are sensitive to mental simulation and how the strength of activation in these brain areas is associated with measures of subjective reading experiences such as absorption and appreciation. In doing so, I aim to replicate and extend the findings of Mak and Willems (2019) in the context of a combined eye tracking and fMRI experiment.

4.1.2. Hypotheses

The changes in gaze duration in reaction to the different kinds of simulation found by Mak and Willems (2019) gave rise to the question what these changes in gaze duration reflect on a neural level. There are three possible answers to this question. The first possibility is that the different types of simulation are all represented in the same brain area, for example an area involved in the general process of constructing a coherent representation of the content of narratives, or *situation model building* (e.g., Martín-Loeches, Casado, Hernández-Tamames, & Álvarez Linera, 2008; see also Smirnov et al., 2014). Candidate areas for such a process would be the posterior cingulate cortex, anterior cingulate cortex, angular gyrus, precuneus, insula, dorsolateral and medial prefrontal cortex, and the superior frontal gyrus (with the first three being found to be involved in situation model building in multiple studies; see Hartung, Wang, Mak, Willems, & Chatterjee, 2021; Hasson, Egidi, Marelli, & Willems, 2018; Kurby & Zacks, 2008; Speer, Reynolds, Swallow, & Zacks, 2009; Speer, Zacks, & Reynolds, 2007; see also Hagoort, 2019).

Another possibility is that the different kinds of simulation do not all activate the same brain area, but activate different areas, reflecting the different modalities of simulation. Motor simulation would then activate motor areas (e.g., precentral and postcentral cortex, superior temporal sulcus, cingulate cortex, supplementary motor area, middle and superior frontal gyrus, middle and superior temporal gyrus, inferior parietal lobule, intraparietal sulcus, precuneus, parahippocampal gyrus; Chow et al., 2015; Kurby & Zacks, 2013; Moody & Gennari, 2010; Nijhof & Willems, 2015), perceptual simulation would activate areas involved in the processing of perceptual information (e.g., posterior temporal gyrus including posterior superior temporal sulcus and middle temporal gyrus and hMT for motion simulation (Deen & McCarthy, 2010; Samur et al., 2015); left superior temporal gyrus, bilateral superior temporal sulcus, but also perisylvan language-related regions for auditory simulation (Kurby & Zacks, 2013); and cuneus, lingual gyrus, fusiform gyrus and parahippocampal gyrus for visual simulation, (Chow et al., 2015)), and mentalizing would activate the mentalizing-network (e.g., aMPFC, dMPFC, MCC, TPJ; U. Frith & Frith, 2003; Hsu et al., 2014; Lai et al., 2015; Nijhof & Willems, 2015; Saxe & Kanwisher, 2003; Tamir et al., 2016).

A final possibility would be that both a common "situation-model building" simulation area, *and* modality-specific brain areas responding to the different kinds of simulation. Evidence for both options has been found before (as has been reported above), but to my knowledge no previous studies have investigated the possibility of both common simulation areas *and* modality specific brain areas for all three kinds of simulation.

4.1.3. Individual Differences in the Effect of Simulation

A second, more exploratory, question that I would like to answer in this study, concerns possible variation between participants in the strength of the brain activation associated with simulation. Individual differences in the relationship between simulation and gaze behavior were found to be related to differences in subjective reading experiences, notably appreciation and absorption (Mak & Willems, 2019). If there is a common mechanism by which simulation is associated with subjective experiences, a similar result may be found for neuroimaging data. For example, individual differences in the strength of activation in simulation-sensitive brain areas may be correlated with individual differences in subjective reading experiences. In the current experiment, I will therefore analyze if individual differences in the strength of the brain activation associated with simulation (operationalized as the individual percent signal change in areas associated with simulation) are associated with subjective reading experiences (story world absorption, story appreciation), reading habits, and certain personal characteristics (i.e., empathy and transportability). In this context, story world absorption refers to an experiential state in which readers are feeling as if they are "lost in a story" (Kuijpers, 2014).

4.1.4. Current Experiment

In the current experiment, participants read two Dutch literary short stories while they were in an MRI scanner and simultaneously had their eye movements tracked. This allowed me to link eye tracking data to neuroimaging data, within participants. I was interested in the responses of participants at the word level: I measured fixation duration and brain activation as a response to the number of times the fixated words were underlined for being part of motor descriptions, perceptual descriptions, and mental event descriptions. The scoring (underlining) of motor, perceptual, and mental event descriptions was acquired in a separate pre-test with different participants (see section 4.2.1.).

The stories were presented visually, and participants were asked to read the stories at their own pace. After reading the two stories, participants performed four localizer tasks. With these tasks, I intended to localize regions of interest for this study. Directly after each story (while still in the scanner), participants completed questionnaires regarding their experience related to the story they just read (story world absorption, story appreciation). After scanning, questionnaires regarding reading habits in daily life (directly and indirectly measured), and personal characteristics questionnaires regarding empathy and transportability were administered.

4.2. Methods

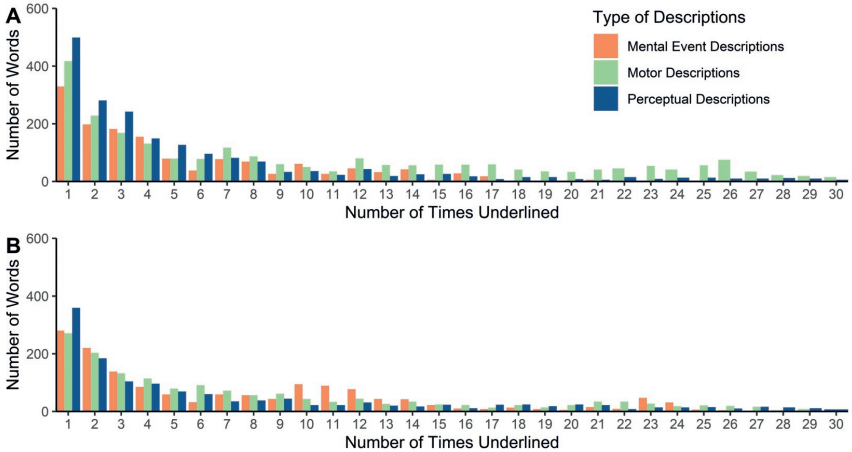
4.2.1. Pre-test

In a pre-test described in detail in Mak and Willems (2019), all words in the two stories used in the current experiment were rated by 30 participants on whether these words were part of a motor description, a perceptual description, or a mental event description. A total of 90 participants took part in the pre-test: each type of description was rated by a different group of 30 participants. Motor descriptions are defined as "concrete acts or actions performed by a person or object," such as *"They reached the bus-stop shelter"* (Story B: Symbols & Signs). Perceptual descriptions are "things that are perceivable with the senses," such as *"A tiny unfledged bird"* (Story B: Symbols & Signs). Mental event descriptions are "explicit descriptions of the thoughts, feelings and opinions of a character" and/or "reflection[s] by a character on his own or someone else's thoughts, feelings or behavior". For example, *"She thought of the recurrent waves of pain"* (Story B: Symbols & Signs). It was counted how many participants underlined each word for each of the three types of description. This resulted in scores ranging from 0-30 per word, per type of description. These scores were taken as regressors for the fMRI and eye tracking data analyses. The number of descriptive words per story per type of description can be seen in Fig. 4.1.

4.2.2. Participants

Forty participants took part in the current experiment (16 male). Participants were between 18 and 43 years old ($M = 24.61$, $SD = 5.22$). A power analysis

Figure 4.1.: *Number of Times Words Were Underlined for Mental Event, Motor and Perceptual Descriptions in Stories (A) The People Who Had Everything Delivered and (B) Symbols and Signs. Note: In Story A, 1562 Words, 659 Words and 1070 Words Were Underlined by None of the Pre-Test Participants for Mental Event, Motor and Perceptual Descriptions, Respectively. In Story B, 646, 586 and 788 Words Were Underlined by None of the Pre-Test Participants for Mental Event, Motor and Perceptual Descriptions, Respectively*



based on data from Mak and Willems (2019) showed that power would be above .8 to capture small effects, in a study with 40 participants, reading two stories, with minimally 400 descriptive words per story per type of description (power analysis based on code by Jobe, 2009). I tested participants from the participant database of the Radboud University. Participants had no dyslexia, and had normal vision or vision correction of maximally +4 or -4 (vision correction in the scanner was done with contact lenses or MR compatible glasses that were attached to the head coil). Other exclusion criteria were epilepsy, claustrophobia, pregnancy, brain surgery, or non-removable metal in or on the body. Participants gave informed consent prior to the study and were allowed to withdraw their consent at any point throughout the experiment, in accordance with the declaration of Helsinki. This experiment was approved under the ethical approval of the ethical committee CMO Arnhem/Nijmegen (CMO 2014/288; version 2).

4.2.3. Materials

4.2.3.1. Stories

Two existing Dutch short stories were presented to the participants (also used in Mak & Willems, 2019). Story selection was based on the length of the stories, the presence of descriptive content, and the probability that the stories were unknown to the participants (in the study of Mak and Willems, none of the participants reported having read one of these stories before). Both stories are written by acclaimed writers and have been published by literary publishers. One story (*De mensen die alles lieten bezorgen [The people who had everything delivered]*; henceforth Story A) is written by the contemporary Dutch writer Rob van Essen (2014), and the other story (*Signalen en Symbolen*; henceforth Story B) is a professional and published translation (American English to Dutch) of *Symbols and Signs* by Vladimir Nabokov (translation in: Nabokov, 2003). The stories are 2988 and 2143 words long, respectively, and take around 10 minutes to read (Story A: $M = 10.08$, $SD = 3.01$; Story B: $M = 9.70$, $SD = 2.94$). All participants read both stories, in counterbalanced order.

4.2.3.2. Questionnaires

After reading the stories, participants completed a set of questionnaires. The questionnaires measuring reading experiences were filled out twice, directly after reading each story. The rest of the questionnaires were completed at the end of the experiment (see also *Procedure* below).

4.2.3.2.1. Reading Experiences Reading experiences (i.e., story world absorption, story appreciation) were measured using the two following questionnaires. Story world absorption was measured with the Story World Absorption Scale (SWAS; Kuijpers et al., 2014; e.g., *When I finished the story I was surprised to see that time had gone by so fast; I could imagine what the world in which the story took place looked like*), complemented with six additional questions (partly based on items originally designed by Kuijpers et al., 2014) more specifically aimed at measuring the experience of different kinds of simulation (mainly perceptual and motor simulation, e.g., *I could see the events in the story happening as if I could see through the eyes of the main character; I could easily depicture the characters in the story*). The SWAS is a validated scale consisting of 18 items with high internal validity (Kuijpers et al., 2014), which measures 4 dimensions of story world absorption via the subscales Attention, Transportation, Emotional Engagement and Mental Imagery. Participants rated each question on a 7-point scale (1 = disagree, 7 = agree). Story appreciation was measured with a questionnaire consisting of a general score of story liking (*How did you like the story*; 1 = It was very bad, 7 = It was very good) and thirteen adjectives (e.g., *[did you find the story] Entertaining, ... Ominous*) that can be used to describe the stories (adapted from Knoop et al., 2016). These adjectives were taken from a list of adjectives that were found to be most often used by people to describe their opinion of poetry, and which can also be used to describe aesthetic appeal in the domain of literature (Knoop et al., 2016; Mak, Faber, & Willems, 2022). Finally, 6 questions were asked regarding the enjoyment of the story (from Kuijpers et al., 2014; e.g., *I was constantly curious about how the story would end; I thought the story was written well*). Participants rated both the adjectives and the questions regarding enjoyment on a 7-point scale (1 = disagree, 7 = agree). Both of these questionnaires were also used in the previous eye tracking experiment.

4.2.3.2.2. Comprehension Check Story comprehension was measured using a comprehension check, consisting of 3 multiple choice questions per story with 4 possible answers per question, that should have been possible to answer correctly for people who read the stories with normal attention (example question, *Why did Jeffrey and Rita leave the flat?*). Additionally, participants were asked to indicate whether they had read any of the stories before.

4.2.3.2.3. Reading Habits Reading habits were assessed both directly and indirectly. The direct measure consisted of a reading habits questionnaire, containing six questions regarding participants' reading habits in everyday life, for each of which participants had to select one of five optional answers (adapted from Hartung et al., 2016; e.g., *How often do you read fiction; How often do you read non-fiction; How many books do you read each year*). The indirect measure of reading habits was the Author Recognition Test (ART; Stanovich & West, 1989; Dutch adaptation reported in Koopman, 2015), consisting of 42 names (30 real authors and 12 foils), where participants had to indicate who they thought were genuine authors.

4.2.3.2.4. Personal Characteristics (Empathy, Transportability) To measure personal characteristics, such as transportability and empathy, participants filled out the Fantasy and Perspective Taking subscales of the Interpersonal Reactivity Index (IRI; Davis, 1980; Dutch translation adapted from De Corte et al., 2007) on a 7-point scale (e.g., *Becoming extremely involved in a good book or movie is somewhat rare for me; When I'm upset at someone, I usually try to "put myself in his shoes" for a while*). The Fantasy subscale measures the extent to which someone tends to get mentally involved in the stories they encounter, to the point at which they imagine themselves being part of the story (transportability).

The Perspective Taking subscale measures the extent to which someone is able to take someone else's perspective in daily life.

4.2.4. Procedure

Participants first read the two stories in the MRI scanner, while their eye movements were being tracked. Stories were presented in counterbalanced order. Participants were instructed to read the stories the way they would also read for their own leisure. There was no additional task, and participants were able to proceed through the stories at their own pace. To proceed through the pages in the story, participants pressed a button with their right index finger when they finished reading a page. Both stories were divided into 30 pages. After each story, participants were allowed to take a short break from reading, to fill in the SWAS and appreciation questionnaire about the story they just read (while remaining inside the MRI scanner). After reading the two stories, participants performed four localizer tasks, that were not included in the analyses. After the localizer tasks, participants left the MRI scanner, and completed the final questionnaires in a separate booth. First, participants answered the comprehension check questions about the two stories, after which the questions regarding reading habits and personal characteristics were asked.

4.2.5. Stimulus Presentation

Stimuli were presented page by page on a projection screen (see `www.macada-innovision.nl`) at the end of the bore, using a EIKI LC - XL100 beamer with a native resolution of 1024x768, with Presentation software (NBS, Berkeley, California). Participants could view the screen via a mirror (`https://www.pgo-online.com/intl/katalog/cold-mirrors.html`) mounted on the head coil. Pages consisted of maximally eight triple spaced lines. The distance between the mirror (110x100mm) and the projection screen (369x277mm) was 855 mm, and the distance between the mirror and the eye about 100mm (depending on how high a participant's head lay in the head coil).

4.2.6. Eye Movement Data Acquisition and Pre-Processing

An MR compatible ceiling mounted Eyelink 1000 eye tracker (SR Research, Ottawa, Canada), with a sampling rate of 1000 Hz was used for eye movement data acquisition during scanning. The eye tracker records infrared light reflected by the eyes, via a mirror attached to the head coil. The eye tracker was calibrated and calibration was validated before the presentation of each story.

Using SR Research's Eyelink Data Viewer, all fixations were checked before data analysis, and, if necessary, manually aligned. If this was impossible, because data were too noisy, data were excluded on a page-by-page basis. If too many pages had to be excluded within a participant, all data for this participant were excluded. This resulted in the exclusion of all data for three participants, the exclusion of one story for five participants (for one of these participants this was due to tracker malfunction rather than poor data quality), and in the exclusion of one to five pages in six participants (14 pages in total in these six participants). This amounts to a total of 14.33% of data loss based on eye tracking issues. After preprocessing, data for 37 participants were retained (full data for 27 participants, rejection of one story for five participants, rejection of a small portion of data for another five participants).

If entire story readings (or entire participants) needed to be removed due to poor eye tracking quality, the fMRI data for these story readings were also discarded (as I needed the eye tracking data to be able to analyze the fMRI data). In the cases where only one to five pages of eye tracking data needed to be removed, I did not discard the fMRI data for these participants. After preprocessing, I was able to use all fMRI data (two stories) for 32 participants, and fMRI data for one story for five participants. To be able to still analyze the fMRI data for the pages of which eye tracking data were discarded in the five participants for whom a small portion of the eye tracking data was rejected, I needed to impute the eye fixations for these data. To stay as close as possible to the participant's natural reading behavior, I modelled the onset and duration of the fixations, but imputed the mean value of the word characteristics I wanted to model as the weights of these fixations. This way the discarded part of the data would have no influence on the results of the analyses.

4.2.7. fMRI Data Acquisition and Pre-Processing

Data were collected at the Donders Centre for Cognitive Neuroimaging in Nijmegen, The Netherlands. fMRI data were acquired using a 3T MAGNETOM PrismaFit MR scanner (Siemens AG, Healthcare sector, Erlangen, Germany) with a 64-channel head-coil. Functional (TR = 1000ms, TE = 34ms, flip angle = 60°, Field of View = 210mm, voxel size = 2.0x2.0x2.0mm, number of slices = 66, Multi-band acceleration factor = 6, multi-slice mode = interleaved, echo spacing = 0.62ms) and anatomical (Magnetization Prepared Rapid Acquisition Gradient

Echo, voxel size = 1.0x1.0x1.0mm) images were acquired in one session lasting about 60 to 90 minutes, depending on the participants' reading speed.

Preprocessing was carried out using FEAT (version 6.00) in FSL (FMRIB's Software Library, `www.fmrib.ox.ac.uk/fsl`). The first ten or eleven volumes (ten or eleven seconds) were discarded (depending on the task programming) to allow for magnetic field saturation. Using FLIRT (Jenkinson, Bannister, Brady, & Smith, 2002; Jenkinson & Smith, 2001), functional images were registered to the high-resolution structural images (using Rigid-Body Transformation (6 DOF) and Boundary-Based Registration; Greve & Fischl, 2009) after non-brain tissue was removed using BET (Smith, 2002). Motion correction was performed using MCFLIRT (Jenkinson et al., 2002), and values for the framewise displacement (average of rotation and translation parameter differences, using weighted scaling; Power, Barnes, Snyder, Schlaggar, & Petersen, 2012) as calculated using FSLMotionOutliers were saved as a confound EV for the first level analyses. High resolution structural images were registered to standard (MNI152 template, 2x2x2mm) space using FNIRT (Andersson, Jenkinson, & Smith, 2007a, 2007b) nonlinear registration ($\geq$ 12 DOF). Spatial smoothing was performed using SUSAN noise reduction (Smith & Brady, 1997) with a 5mm FWHM Gaussian kernel. Grand-mean intensity normalization of the entire 4D dataset was done by a single multiplicative factor. High pass temporal filtering was applied using Gaussian-weighted least-squares straight line fitting, with sigma=45.0s.

fMRI data preprocessing resulted in the additional exclusion of one story for one participant (1.25% data loss in addition to the data loss due to poor eye tracking quality). Note that there was a lot of overlap between the quality of the eye tracking and fMRI data: participants who move much during scanning, tend to have both poor eye tracking and poor fMRI data.

4.2.8. Data Analysis

4.2.8.1. Eye Tracking Data

The eye tracking data were analyzed in a similar way as the eye tracking data in Mak and Willems (2019) to make direct comparison of the eye tracking results possible. I analyzed how motor descriptions, perceptual descriptions and mental event descriptions related to gaze duration, while controlling for lexical frequency, word length and surprisal value as regressors of no interest, and allowing for random slopes and intercepts for the three types of descriptions over the interaction between subject and story to allow for individual variation between subjects and stories. As in the analyses in Mak and Willems (2019), I used the values for the descriptions and word characteristics of the previous word, which allowed me to look in the spillover regions. I analyzed this with a Bayesian Multilevel Model using the package *brms* (Bürkner, 2017, 2018) and *Stan* (Stan Development Team, 2020) in *R* version 4.0.3 (R Core Team, 2021). The rationale for calculating a Bayesian multilevel model as opposed to a "classical" frequentist model, was that Bayesian models are more flexible and more capable of fitting complex models (e.g., Bürkner, 2018; Nalborczyk, Batailler, Lœvenbruck, Vilain, & Bürkner, 2019). Additionally, the analyses of the fMRI data were also done within a Bayesian framework (Beckmann, Jenkinson, & Smith, 2003; Woolrich, Behrens, Beckmann, Jenkinson, & Smith, 2004; Woolrich et al., 2009). Rather intuitively, Bayesian multilevel models calculate the range of the most probable values of each parameter, a 95% Credible Interval. If this Credible Interval does not cross zero for a given parameter, this indicates a 95% certainty that the true value of this parameter is distinguishable from zero.

In the model, I used weakly informative, normally-distributed priors with a mean of 0 and a standard deviation of 10 for all fixed effects. These priors are considered relatively conservative (McElreath, 2016). For the population-level intercept I used an informative, normally-distributed prior with a mean of 250 and a standard deviation of 50, since gaze durations are generally between 200ms and 300ms long on average. As variance can only be positive, weakly regularizing, half-cauchy priors with a mean of 0 and a standard deviation of 1 were used for the variance of the random effects as well as the overall variance (as suggested by Gelman, 2006; McElreath, 2016). The Gelman-Rubin diagnostic (*Rhat*) was 1.0 for all parameters (except for the intercept, for which it was 1.01), indicating that the model had converged.

4.2.8.2. fMRI Data

The fMRI-data were analyzed using a fixation-based analysis (comparable to an event-related analysis; see Richlan et al., 2014). In this analysis, the onset of a fixation was seen as the event onset, and the duration of the fixation as the event duration. These fixation events were then convolved with the HRF. From the eye tracking data, I extracted the fixation (event) onsets and durations per word (which were determined automatically by SR Research's default parsing algorithm), to determine which word was looked at, at any given time during reading. Data analyses were performed in Feat (FMRI Expert Analysis Tool) Version 6.00, part of FSL (FMRIB's Software Library, `www.fmrib.ox.ac.uk/fsl`).

Time-series statistical analysis was carried out using FILM with local autocorrelation correction (Woolrich, Ripley, Brady, & Smith, 2001). For the first level analysis, I ran a GLM per participant (per run, one analysis for each story) where I modelled the onset and duration of each fixation, weighted by the scores for motor descriptions, perceptual descriptions, mental event descriptions, and the first principal component of lexical frequency, word length and surprisal value of the word that was fixated (to control for these word characteristics). This way I determined which brain areas respond specifically to either of the three types of descriptions in stories, while controlling for differences in word characteristics.

For statistical inference I contrasted each type of description with the other types of descriptions, to find the activation that was *specific* to that type of description (i.e., weighted contrasts [1 -.5 -.5] for motor > perceptual and mental event, perceptual > motor and mental event, mental event > motor and perceptual). Additionally, I contrasted each type of description with baseline (contrasts [1 0 0], [0 1 0] and [0 0 1]), and visualized which areas were commonly activated by all three types of descriptions (as a conjunction analysis). The z-statistic images resulting from the contrasts were thresholded using clusters determined by $z > 3.1$ and a (corrected) cluster significance threshold of $p = .05$. The cluster threshold was determined based on the theory of Gaussian Random Fields (Worsley, 2001). This effectively controls the multiple comparisons problem introduced by the massive univariate approach taken at a family-wise error rate of $p < .05$ (Worsley, 2001).

As participants each read two stories, in two separate runs, I first aggregated the results for the two stories at the participant level using a standard weighted fixed effects model in FLAME (FMRIB Local Analysis of Mixed Effects). In this model, the variances from the first level analysis were used as the fixed effect error variances, and the random effects variance was forced to zero (Beckmann et al., 2003; Woolrich, 2008; Woolrich et al., 2004). The output from the fixed effects models per participant was used as input for the second level analysis. The second level analysis was performed using FLAME (FMRIB Local Analysis of Mixed Effects) stage 1 with automatic outlier detection, which estimates between-subject random effects using MCMC (Beckmann et al., 2003; Woolrich, 2008; Woolrich et al., 2004).

As a final, exploratory analysis, I investigated how individual differences in brain activation in response to the different types of descriptions are related to the experience of narrative reading as well as individual difference measures. I did this to find out whether brain activation due to simulation-eliciting content occurs equally or differently across individuals and to find out whether any individual differences could be explained by reading experiences or personal characteristics. The analysis was done by first extracting the percent signal change (per participant and per story, from the first level analyses) in five or six regions of interest for each of the three types of descriptions. I selected regions of interest that had been significantly activated by the three types of descriptions on the group level, and that I believed would be good candidates for finding individual differences in simulation. I derived the regions of interest from the results of the group analysis: I extracted the by-participant by-story percent signal change from areas that were found to be commonly activated by these descriptions. I then built models to predict percent signal change in each area, by scores on the Story World Absorption Scale, the appreciation questionnaire, the Fantasy and Perspective taking subscales of the Interpersonal Reactivity Index, the Author Recognition Test, and the questions about reading habits (see heading "Questionnaires" below for more information on these questionnaires). I built separate models for Story World Absorption and for appreciation, to make sure that any conceptual overlap between the two would not skew the results. I analyzed this with Bayesian Multilevel Models using the package *brms* (Bürkner, 2017, 2018) and *Stan* (Stan Development Team, 2020) in R version 4.0.3 (R Core Team, 2021). In the models, I used weakly informative, normally-distributed priors with a mean of 0 and a standard deviation of 1 for all fixed effects. For the population-level intercept I used a weakly informative, normally-distributed prior with a mean of 0 and a standard deviation of 10. These priors are considered relatively conservative (McElreath, 2016). As variance can only be positive, weakly regularizing, half-cauchy priors with a mean of 0 and a standard deviation of 1 were used for the variance of the random effects as well as the overall variance (as suggested by Gelman, 2006; McElreath, 2016). In all models, the Gelman-Rubin diagnostic (*Rhat*) was 1.0 for all parameters, indicating that the models had converged.

4.3. Results

4.3.1. Behavioral Results

4.3.1.1. Questionnaires

4.3.1.1.1. SWAS The Story World Absorption Scale showed excellent reliability, $\Omega_t = .96$ (Story A: $M = 4.81$, $SD = 0.79$; Story B: $M = 3.55$, $SD = 0.83$).

4.3.1.1.2. Appreciation In order to reduce the 13 adjectives to a smaller number of components consisting of highly similar adjectives, I conducted a principal component analysis (PCA) with oblique rotation (direct oblimin) on the 13 appreciation adjectives (cf. Mak & Willems, 2019), using the package *psych* (Revelle, 2020) in *R* version 4.0.3 (R Core Team, 2021).

The Kaiser-Meyer-Olkin measure (KMO) was .81 (all KMO values for individual items > .62), indicating good sampling adequacy for this analysis. Bartlett's test of sphericity showed sufficient correlation between items, χ^2 (55) = 1879.28, $p < .001$. Based on the scree-plot in combination with the eigenvalues found in an initial analysis (Kaiser's method) and the model fit (fit based upon off diagonal values), it was decided to retain five components in the final model. This model explained 77% of the variance.

The first component that I found corresponded to Interest (consisting of items Suspenseful, Interesting, Captivating, Gripping, and Boring (-)); the second component to Sadness (Tragic, Sad); the third component to Special (Special); the fourth component to Positive Affect (Witty, Beautiful); and the final component to Ominous (Ominous, Funny (-), and Entertaining (-)). The pattern matrix for the factor loadings after rotation can be found in Table 4.1.

Table 4.1.: *Pattern Matrix for the PCA of the 11 Adjectives on the Appreciation Questionnaire (N = 703). Factor Loadings over .40 Appear in Bold*

	Pattern Matrix				
	Interest	Sadness	Special	Positive Affect	Ominous
Beautiful	0.31		0.33	**0.45**	**0.40**
Boring	**-0.54**	-0.11	-0.19	-0.27	0.25
Gripping	**0.59**	0.28	0.17		0.20
Entertaining	0.38	-0.11		0.34	**-0.49**
Funny	0.17	-0.16		0.36	**-0.59**
Interesting	**0.76**	0.17	0.11		0.14
Ominous					**0.89**
Sad	-0.17	**0.89**			
Suspenseful	**0.91**	-0.13	-0.11	-0.23	
Tragic	0.16	**0.91**		-0.11	
Witty			-0.10	**0.93**	
Captivating	**0.73**			0.28	-0.17
Special			**0.99**	-0.10	

4.3.1.1.3. Comprehension Check On the comprehension check for story A (three multiple choice questions, with four answer options each), three participants made one mistake. All other participants got all three questions correct. For story B, ten participants made one mistake, and three participants made two mistakes. Seven out of the ten participants who made one mistake and all three participants who made two mistakes, answered the second question incorrectly (this question was answered incorrectly 25% of the time). Therefore, I decided not to reject participants based on this question. Hence, none of the data was rejected based on the comprehension check for either of the stories.

4.3.1.1.4. Reading habits Answers on the reading habits questionnaire were measured on a scale ranging from 1 to 5 on four of the five questions, but from 1 to 4 on the final question. Therefore, z-scores were calculated for all questions on this questionnaire (higher values indicating more reading experience). Overall reliability was good if the question about non-fiction reading was excluded, Ω_t = .83.

4.3.1.1.5. ART The scores on the ART were slightly positively skewed (M = 6.28, SD = 2.89, median = 6.00, IQR = 4.75–7.00) with higher values indicating more (literary) reading experience.

4.3.1.1.6. IRI Scores on the two subscales of the Interpersonal Reactivity Index were analyzed separately, as they measure different constructs. The Fantasy subscale (M = 5.05, SD = 0.92) showed good reliability, Ω_t = .89, as did the Perspective Taking subscale (M = 4.95, SD = 0.75), Ω_t = .86.

4.3.1.2. Eye Tracking Data

The model predicting the gaze durations on individual words, by the values of motor descriptions, perceptual descriptions, mental event descriptions, lexical frequency, word length and surprisal value of the previous word (to account for spillover effects, cf. Mak & Willems, 2019), showed that motor descriptions were associated with shorter gaze durations on the next word (i.e., faster reading; see Table 4.2, Fig. 4.2B). Perceptual and mental event descriptions were associated with longer gaze durations on the next word (i.e., slower reading; see Table 4.2, Fig. 4.2C-D). These results nicely replicate the results in Mak and Willems (2019), indicating that eye movements could reliably be tracked inside the fMRI scanner. Unlike the results in Mak and Willems (2019), the studied word characteristics (lexical frequency, word length and surprisal value) did not reliably influence gaze durations in the spillover region (see Table 4.2, Fig. 4.2E-F).

Looking at the individual variation in the relationship between the three types of descriptions and gaze durations, I saw that these relationships varied between all story*subject combinations (the standard deviation of the slope of Motor Descriptions = 3.84 [CI: 2.43-5.36] across subject*story combinations; the standard deviation of the slope of Perceptual Descriptions = 5.75 [CI: 4.39-7.30] across subject*story combinations; the standard deviation of the slope of Mental Event Descriptions = 4.38 [CI: 2.96-5.91] across subject*story combinations; see Fig. 4.3).

Table 4.2.: *Posterior Distributions (Mean, SE, 95% Credible Interval) of the Associations between the Three Types of Descriptions and Word Characteristics and Gaze Durations*

	Estimate (Mean)	Est. Error	Lower Bound (95% CI)	Upper Bound (95% CI)
(Intercept)	280.00	6.10	267.80	291.83
Motor Descriptions	-2.58	0.73	-4.05	-1.16
Perceptual Descriptions	9.60	0.88	7.87	11.29
Mental Event Descriptions	2.69	0.78	1.11	4.19
Lexical Frequency	1.45	1.14	-0.76	3.70
Word Length	-0.37	0.87	-2.09	1.31
Surprisal Value	1.64	0.96	-0.23	3.51

4.3.2. Neuroimaging Results

All localizations reported here are based on the *Harvard-Oxford Cortical Structural Atlas*, *Harvard-Oxford Subcortical Structural Atlas*, or the *Cerebellar Atlas in MNI152 space after normalization with FNIRT*.

4.3.2.1. Activations Specific to Each of the Three Types of Descriptions

Activations specific to each of the three types of descriptions were clusters of voxels that were reliably activated to one type of description but not to the two other types of descriptions. Peak coordinates for the activation clusters can be found in Table 4.3, for each of the types of description.

4.3.2.1.1. Motor Descriptions Fig. 4.4 visualizes the main results for the contrast for motor descriptions versus perceptual and mentalizing descriptions. Motor descriptions were associated with activation bilaterally in the precuneus, cingulate gyrus, angular gyrus, superior lateral occipital cortex, superior frontal

Figure 4.2.: *Posterior Distributions of the Fixed Effects of the Relationships Between the Three Types of Descriptions and Word Characteristics and Gaze Durations*

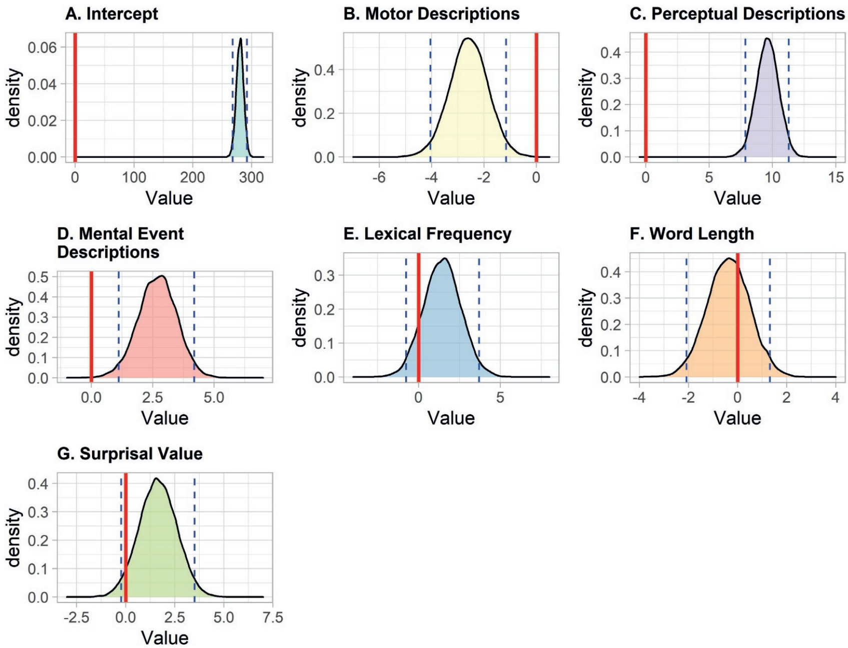
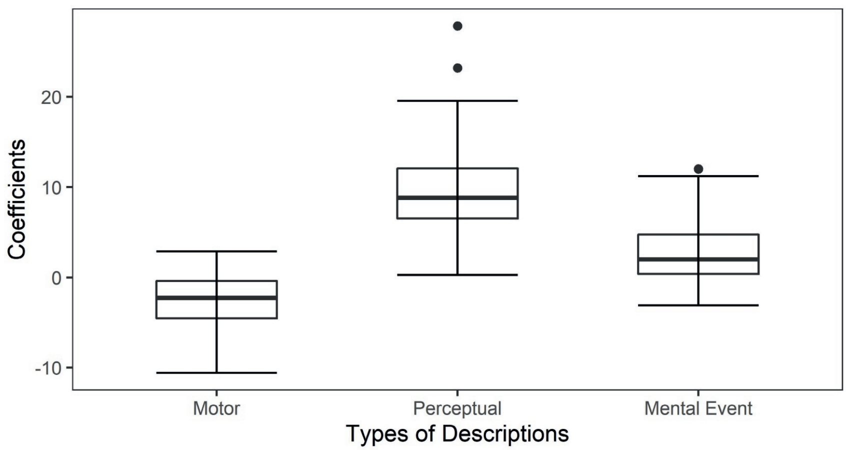


Figure 4.3.: *Estimated Means of Participants' Individual Posterior Distributions for the Relationships Between the Types of Descriptions and Gaze Durations*



gyrus and middle frontal gyrus. Additionally, motor descriptions were associated with activation in the left subcallosal cortex, left lingual gyrus, left posterior parahippocampal gyrus, and multiple areas in the left cerebellum (i.e., Left Crus I&II, Left VI and Left IX). Finally, activation was visible in the right paracingulate gyrus, right frontal pole, and right frontal medial cortex.

Table 4.3.: *Clusters of Regions Activated Specifically for Words Belonging to Motor Descriptions, Perceptual Description and Mental Event Descriptions in the Stories. Volume is Given for the Entire Cluster, with the First Structure in that Cluster*

		Peak MNI Coordinates				
Region	L/R	x	y	z	*Z*-Max	Volume (Voxels)
Motor Descriptions						
Precuneus Cortex	R	6	-60	34	7.1	9408
Precuneus Cortex	L	-8	-58	18	7.03	
Cingulate Gyrus	L	-2	-42	42	6.59	
Cingulate Gyrus	R	14	-50	38	6.37	
Angular Gyrus	R	54	-56	16	6.69	5094
Lateral Occipital Cortex (Superior Division)	R	48	-64	30	6.08	
Lateral Occipital Cortex (Superior Division)	L	-40	-72	38	6.96	1424
Angular Gyrus	L	-40	-56	14	5.61	
Superior Frontal Gyrus	R	20	28	46	5.93	3797
Middle Frontal Gyrus	R	26	32	44	5.58	
Paracingulate Gyrus	R	6	52	10	5.49	
Subcallosal Cortex	L	-10	28	-10	5.2	
Frontal Pole	R	24	44	48	5.18	
Frontal Medial Cortex	R	4	54	-8	5.01	
Superior Frontal Gyrus	L	-20	24	44	5.86	712
Middle Frontal Gyrus	L	-32	4	66	4.08	
Middle Frontal Gyrus	R	44	14	36	4.4	310
Lingual Gyrus	L	-26	-42	-8	4.86	220
Parahippocampal Gyrus (Posterior Division)	L	-22	-40	-12	4.82	

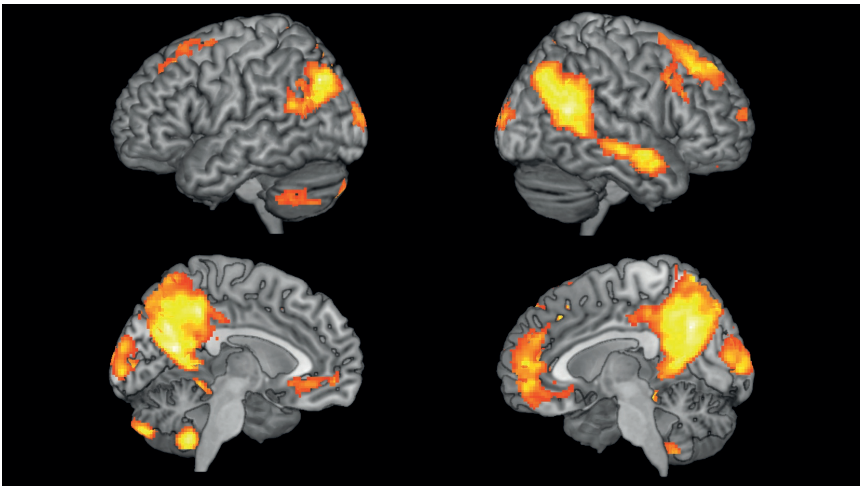
Cerebellum: Left Crus II	L	-10	-80	-40	5.74	595
Cerebellum: Left VI	L	-14	-68	-30	4.36	
Cerebellum: Left Crus I	L	-26	-74	-36	4.09	
Cerebellum: Left IX	L	-8	-50	-46	6.22	352
Cerebellum: Right IX	R	8	-50	-52	4.48	
Cerebellum: Left Crus II	L	-50	-52	-46	3.96	190
Cerebellum: Left Crus I	L	-50	-60	-40	3.43	
Perceptual Descriptions						
Temporal Fusiform Cortex (Posterior Division)	L	-36	-34	-18	7.2	3629
Temporal Occipital Fusiform Cortex	L	-42	-52	-18	7.14	
Parahippocampal Gyrus (Anterior Division)	L	-32	-4	-32	6.77	
Amygdala	L	-18	0	-20	5.36	
Inferior Temporal Gyrus (Temporooccipital Part)	L	-50	-56	-12	6.42	
Temporal Pole	L	-30	2	-32	5.71	
Temporal Pole	R	26	8	-40	5.67	1016
Temporal Fusiform Cortex (Anterior Division)	R	36	-6	-34	5.55	
Inferior Temporal Gyrus (Temporooccipital Part)	R	48	-50	-18	5.45	
Parahippocampal Gyrus (Anterior Division)	R	30	0	-34	4.94	
Temporal Fusiform Cortex (Posterior Division)	R	36	-32	-24	4.92	
Amygdala	R	18	-4	-18	5.57	235
Frontal Pole	R	50	38	12	5.49	481
Inferior Frontal Gyrus (Pars Triangularis)	R	44	30	16	5.27	

Precentral Gyrus	L	-44	4	24	6.27	468
Inferior Frontal Gyrus (Pars Opercularis)	L	-46	8	24	6.19	
Precentral Gyrus	R	40	4	26	4.81	147
Inferior Frontal Gyrus (Pars Triangularis)	L	-44	34	14	4.54	345
Frontal Pole	R	30	34	-10	4.73	239
Frontal Orbital Cortex	R	24	28	-12	4.7	
Frontal Orbital Cortex	L	-24	30	-12	4.61	236
Frontal Pole	L	-28	36	-20	3.4	
Lateral Occipital Cortex (Superior Division)	L	-28	-74	34	5.76	450
Lateral Occipital Cortex (Superior Division)	R	36	-58	48	5.24	214
Angular Gyrus	R	34	-54	42	4.2	
Mental Event Descriptions						
Cerebellum: Right Crus I	R	30	-82	-32	7.63	11753
Lingual Gyrus	L	-6	-72	-4	6.74	
Lingual Gyrus	R	2	-72	-2	6.48	
Cuneal Cortex	R	18	-74	28	4.95	
Cuneal Cortex	L	-8	-82	34	4.06	
Lateral Occipital Cortex (Superior Division)	R	26	-86	26	5.11	
Lateral Occipital Cortex (Superior Division)	L	-20	-85	22	4.21	
Cerebellum: Right Crus II	R	30	-76	-42	6.26	
Superior Frontal Gyrus	L	-4	22	64	7.37	5763
Frontal Pole	L	-6	48	50	5.99	
Inferior Frontal Gyrus (Pars Opercularis)	L	-54	20	0	7.5	4328

Middle Temporal Gyrus (Anterior Division)	L	-52	0	-34	6.44	
Inferior Frontal Gyrus (Pars Triangularis	L	-52	22	-6	6.34	
Temporal Pole	L	-52	16	-10	6.25	
Middle Temporal Gyrus (Posterior Division)	L	-50	-38	-4	5.02	2051
Superior Temporal Gyrus (Posterior Division)	L	-66	-24	0	4.95	
Angular Gyrus	L	-48	-50	30	4.76	
Parietal Operculum Cortex	L	-38	-34	18	4.45	
Temporal Pole	R	54	20	-12	4.71	1859
Planum Temporale	R	64	-18	12	4.64	
Parietal Operculum Cortex	R	54	-28	22	4.18	
Temporal Pole	R	50	8	-42	4.6	188
Middle Temporal Gyrus (Posterior Division)	R	60	-10	-28	3.63	
Middle Temporal Gyrus (Anterior Division)	R	54	-4	-32	3.6	
Cingulate Gyrus (Anterior Division)	L	0	-14	38	5.33	150
Cingulate Gyrus (Anterior Division)	R	8	-16	32	3.41	
Cerebellum: Right IX	R	6	-54	-40	4.41	123
Cerebellum: Vermis VIIIa	R	4	-68	-38	4.22	

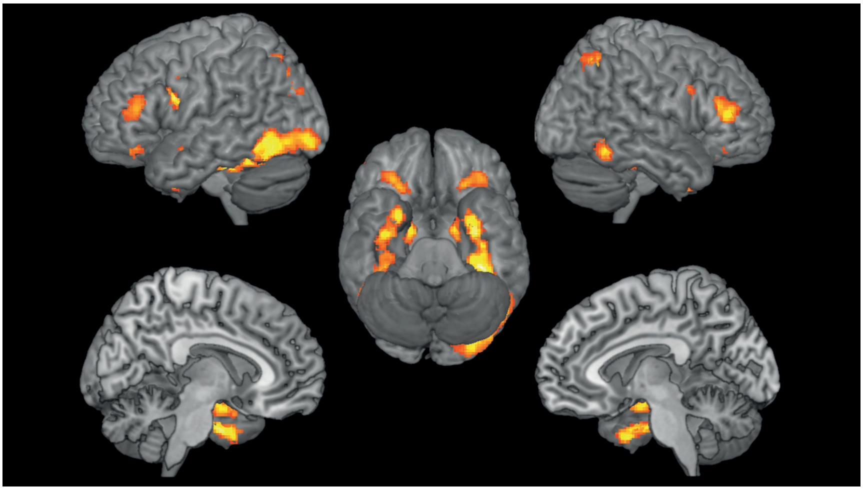
4.3.2.1.2. Perceptual Descriptions Fig. 4.5 visualizes the main results for the contrast for perceptual descriptions versus motor and mentalizing descriptions. Perceptual descriptions were associated with activation bilaterally in the posterior temporal fusiform cortex, anterior parahippocampal gyrus, the amygdalae, the temporooccipital part of the inferior temporal gyrus, the temporal pole, frontal pole, inferior frontal gyrus (pars triangularis), precentral gyrus, su-

Figure 4.4.: *Visualization of Activations Specific to Motor Descriptions (Contrast Motor Versus Perceptual & Mentalizing [1 -.5 -.5]). The Left Panel Represents the Left Hemisphere (Lateral and Medial), the Right Panel Represents the Right Hemisphere (Lateral and Medial)*



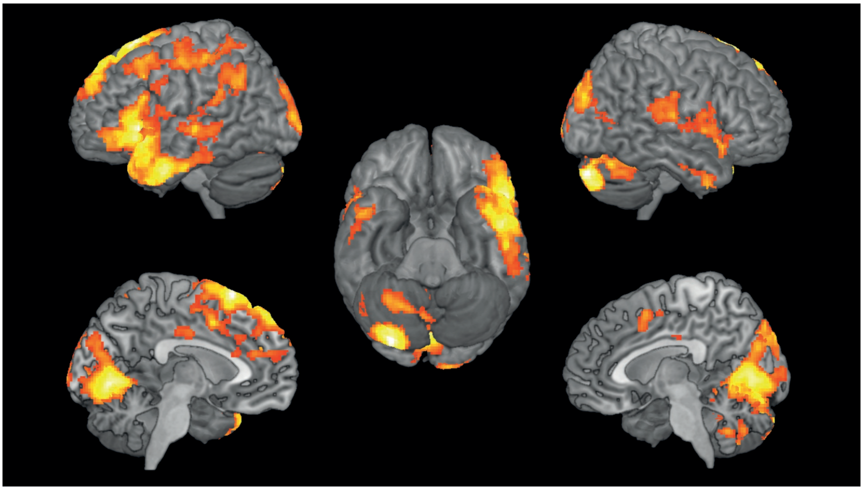
perior lateral occipital cortex, and the frontal orbital cortex. Additional left lateralized activation was seen in the left temporal occipital fusiform cortex, and in the pars opercularis of the left inferior frontal gyrus. Right lateralized activation was seen in the right angular gyrus.

Figure 4.5.: *Visualization of Activations Specific to Perceptual Descriptions (Contrast Perceptual Versus Motor & Mentalizing [-.5 1 -.5]). The Left panel Represents the Left Hemisphere (Lateral and Medial), the Right Panel Represents the Right Hemisphere (Lateral and Medial), the Middle Panel Represents the Inferior View (Bilateral)*



4.3.2.1.3. Mental Event Descriptions Fig. 4.6 visualizes the main results for the contrast for mental event descriptions versus perceptual and motor descriptions. Mental event descriptions were associated with activation bilaterally in the lingual gyrus, cuneal cortex, superior lateral occipital cortex, anterior and posterior middle temporal gyrus, temporal pole, parietal operculum cortex, and anterior cingulate cortex. Left lateralized activation was visible in the left superior frontal gyrus, left frontal pole, left inferior frontal gyrus (pars opercularis and pars triangularis), left angular gyrus, and left posterior superior temporal gyrus. Right lateralized activation was seen in the right planum temporale, and in multiple areas of the right cerebellum (i.e., Right Crus I&II, Right IX, and Vermis VIIIa).

Figure 4.6.: *Visualization of Activations Specific to Mental Event Descriptions (Contrast Mentalizing Versus Motor & Perceptual [-.5 -.5 1]). The Left Panel Represents the Left Hemisphere (Lateral and Medial), the Right Panel Represents the Right Hemisphere (Lateral and Medial), the Middle Panel Represents the Inferior View (Bilateral)*



4.3.2.2. Shared Activation of All Three Types of Descriptions

Shared activation of all three types of descriptions (i.e., the overlap between the contrasts of each type of description with the implicit baseline; conjunction analysis) was found in the left anterior supramarginal gyrus (see Fig. 4.7; see Table 4.4 for the peak coordinates). No other reliable clusters of activation were visible in all three contrasts.

Figure 4.7.: *Visualization of the overlap of the activations to the three types of descriptions. Blue indicates activations to motor descriptions (contrast Motor versus Baseline [1 0 0]), green indicates activations to perceptual descriptions (contrast Perceptual versus Baseline [0 1 0]), and red indicates activations to mental event descriptions (contrast Mentalizing versus Baseline [0 0 1]). The left panel represents the left hemisphere (lateral and medial), the right panel represents the right hemisphere (lateral and medial)*

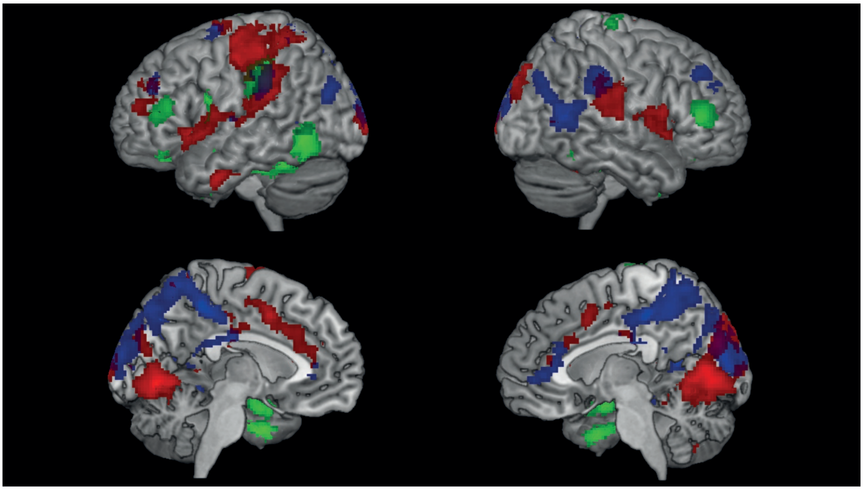


Table 4.4.: *Regions Generally Activated for All Three Types of Descriptions (Motor, Perceptual and Mental Events) Compared to the Implicit Baseline*

		Peak MNI Coordinates				
Region	L/R	x	y	z	*Z*-Max	Volume (Voxels)
Supramarginal Gyrus (Anterior Division)	L	-62	-34	36	5.08	152

4.3.2.3. Individual Differences: Percent Signal Change Analysis

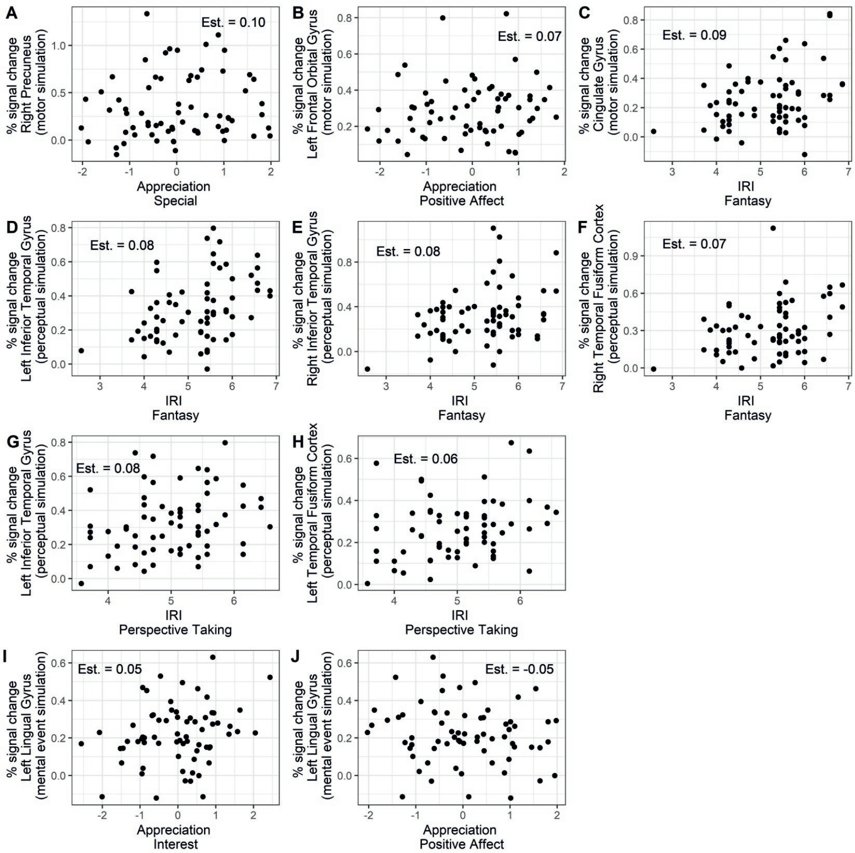
To find out whether brain activation due to simulation-eliciting content occurs equally or differently across individuals and to find out whether any individual differences could be explained by reading experiences or personal characteristics, I calculated spherical regions of interest (10mm diagonally around a peak coordinate) in regions that had been significantly activated by the three types of descriptions on the group level, and that I believed would be good candidates for finding individual differences in simulation. For the motor contrast, these regions included the left supramarginal gyrus, the left frontal orbital gyrus, the left and right precuneus, and the cingulate gyrus. For the perceptual contrast, the regions of interest were the left supramarginal gyrus, left and right inferior temporal gyrus, left and right temporal fusiform cortex, and the right parahippocampal gyrus. For the mental event contrast, I selected the left supramarginal gyrus, left and right insula, left and right lingual gyrus, and the left temporal pole as regions of interest.

After creating ROI maps of these regions of interest in MNI space, I converted the maps to each participant's native space, in order to be able to extract data on percent signal change per participant from the first level analyses. Percent signal change data was extracted per participant and per story using Featquery. I decided to analyze the 90th percentile of the percent signal change, meaning that within the ROIs I looked at the percent signal change in response to the three types of descriptions, in the 10% of voxels in which this percent signal change was highest in each participant in each story. I chose to look at the 90th percentile of the percent signal change, to make my findings more robust to outliers than if I would have looked in the 95th percentile.

I then created models predicting percent signal change in each of the ROIs, by mean absorption (per participant and story), the five principal components of appreciation (per participant and story), and scores on the Fantasy and Perspective Taking subscales of the Interpersonal Reactivity Index, on the Author Recognition Test, and for daily life reading habits (per participant). I will report all statistically reliable results here (i.e., the instances where questionnaire scores were reliably associated with the percent signal change; see Fig. 4.8), and provide tables with all results for all ROIs in Appendix F.

4.3.2.3.1. Motor Descriptions The results of the models predicting percent signal change for the five ROIs for the motor contrast showed that there was a positive association between percent signal change and the Special component of appreciation in the right precuneus (Estimate = 0.10; 95% CI = 0.01; 0.18), showing that experiencing a story as special is associated with higher percent signal change in this brain region. In the left frontal orbital gyrus, I found a positive association between percent signal change and the Positive Affect component of appreciation (Estimate = 0.07, 95% CI = 0.03; 0.11), meaning that participants who experienced stories as witty or beautiful showed a larger percent signal change in the frontal orbital gyrus. Finally, for the anterior cingulate gyrus, I found a positive association between percent signal change and the Fantasy subscale of the IRI (Estimate = 0.09, 95% CI = 0.03; 0.14), indicating that people who are prone to getting mentally involved in stories showed higher percent signal change in the cingulate gyrus.

Figure 4.8.: *Scatter Plots and Estimated Slopes for the Relationships Between Percent Signal Change and Components of Appreciation or Subscales of the Interpersonal Reactivity Index. Plots A-C Are Relationships Found for Motor Simulation, Plots D-H Are Relationships Found for Perceptual Simulation, and Plots I and J Are Relationships Found for Mental Event Simulation*



4.3.2.3.2. Perceptual Descriptions The results of the models predicting percent signal change for the five ROIs for the perceptual contrast showed that there was a positive association between percent signal change and the Fantasy subscale of the IRI, in both the left (Estimate = 0.08, 95% CI = 0.02; 0.13) and right (Estimate = 0.08, 95% CI = 0.01; 0.15) inferior temporal gyrus, and in the right temporal fusiform cortex (Estimate = 0.07, 95% CI = 0.00; 0.13), showing that people who are prone to getting mentally involved in stories show higher percent signal change in response to perceptual descriptions in these three areas. Additionally, there was a positive association between percent signal change and the Perspective Taking subscale of the IRI in the left inferior temporal gyrus (Estimate = 0.08, 95% CI = 0.02; 0.14) and in the left temporal fusiform cortex (Estimate = 0.06, 95% CI = 0.01; 0.10), indicating that people who more often take other people's perspectives in daily life show a higher percent signal change in these areas.

4.3.2.3.3. Mental Event Descriptions The results of the models predicting percent signal change for the five ROIs for the mental event contrast showed that there was a positive association between percent signal change and the Interest component of appreciation in the left lingual gyrus (Estimate = 0.05, 95% CI = 0.00; 0.09), meaning that percent signal change in the left lingual gyrus as a response to mental event descriptions was higher in participants who found the stories more suspenseful, interesting, captivating and gripping. In contrast, there was a negative association between percent signal change and the Positive Affect component of appreciation in the left lingual gyrus (Estimate = -0.05, 95% CI = -0.09; 0.01), indicating that participants who experienced the stories as witty and beautiful showed a lower percent signal change in this area.

4.4. Discussion

In this study, combined eye tracking and fMRI scanning was used to study narrative reading. Participants read two literary short stories in the scanner while their eye movements were being measured so that each participant's individual eye movements could be linked to their brain activity. With this study, I examined whether different kinds of simulation were associated with different patterns of brain activation and / or if there was one general area in the brain that can be associated with all three kinds of simulation under study. Secondly, I attempted to find out whether individual differences in reading experiences can be linked to differences in activation levels in specific brain regions.

The analysis of the eye tracking data in the current study showed similar effects of perceptual simulation, motor simulation and mentalizing on gaze durations as were found in the previous chapter. Motor simulation was associated with shorter gaze durations (faster reading), whereas perceptual simulation and mentalizing were associated with longer gaze durations (slower reading), with effect sizes that were highly similar to the previously found effect sizes (see Mak & Willems, 2019). Apart from replicating the results reported in Chapter 3, the eye tracking results from the current study also show that eye tracking inside the MRI scanner yields the same results as eye tracking within a standard eye tracking lab setting, which is encouraging for future studies attempting to find the relationship between eye movements and brain activation during reading.

4.4.1. Differences in the Effect of the Three Kinds of Simulation on Brain Activation

The first question that I attempted to answer with the current experiment was whether different kinds of simulation were associated with different patterns of brain activation, or, alternatively, if there was one general area in the brain that can be associated with all three kinds of simulation under study. In this study, I found both modality-specific activations and a brain area that was activated in response to all kinds of simulation. I will first review the patterns of brain activation that were specific to motor simulation, perceptual simulation and mentalizing, before discussing a possible domain-general "simulation area".

In line with my expectations, I found that motor descriptions were associated with activity in the cingulate and paracingulate cortex, precuneus, parahippocampal gyrus, superior frontal gyrus and middle frontal gyrus. These areas have been observed to be activated in studies of motor simulation before (Kurby & Zacks, 2013; Moody & Gennari, 2010; Nijhof & Willems, 2015).

Furthermore, a range of brain areas that have been associated with inferencing, event segmentation, and situation model building were activated when people read motor descriptions. For example, the angular gyrus, subcallosal cortex and frontal medial cortex have all been associated with event segmentation and situation model updating, together with the modality-specific motor areas mentioned above (Kurby & Zacks, 2008; Speer et al., 2009, 2007). Perhaps inferencing, event segmentation and situation model building is mainly done through interpretations of actions, something that fiction authors make use of when they decide to "show, don't tell" about their characters minds and intentions. This could also explain why readers speed up while reading motor descriptions: if people are used to inferring other people's minds and intentions based on their actions, this might make motor simulation a relatively automatic, fast process compared to other kinds of simulation (i.e., perceptual and mental event simulation).

For perceptual simulation, I found several modality-specific areas that I had expected to find based on previous work. Areas in the fusiform and parahippocampal gyri have previously been associated with visual simulation (Chow et al., 2015), and were activated by perceptual descriptions in the current study as well. Additionally, I found activation related to perceptual descriptions in the inferior temporal gyrus, inferior frontal gyrus, amygdalae, temporal pole, frontal pole, precentral gyrus, superior lateral occipital cortex, frontal orbital cortex, and angular gyrus. The inferior temporal gyrus is considered to be an important part of the ventral visual pathway, a pathway implicated in object, face and scene perception (Grill-Spector & Weiner, 2014; Kanwisher, 2010). Other areas of this pathway include the parahippocampal gyrus, fusiform cortex and temporal pole, and this pathway is functionally connected to the amygdalae (for the processing of emotionally salient stimuli) and to the precentral gyrus (see for a review Kravitz, Saleem, Baker, Ungerleider, & Mishkin, 2013). Taken together, these findings provide evidence for the idea that a network of visual processing areas, and in particular the ventral pathway, might support perceptual simulation, suggesting a domain-specific mechanism.

Reading mental event descriptions was associated with activations in the temporal pole, parietal operculum, anterior cingulate (situated closely to the medial prefrontal cortex) and angular gyrus (which is closely connected with the temporoparietal junction; Igelström & Graziano, 2017), areas that have been previously associated with mentalizing outside of reading research (U. Frith & Frith, 2003; Igelström & Graziano, 2017; Laurita, Hazan, & Spreng, 2017; Paulus, Müller-Pinzler, Jansen, Gazzola, & Krach, 2015; Saxe & Kanwisher, 2003). Additionally, I found activation in the lingual gyrus, cuneal cortex, superior lateral occipital cortex, anterior and posterior middle temporal gyrus, superior and inferior frontal gyrus, frontal pole, posterior superior temporal gyrus, planum temporale, and the cerebellum. Interestingly, this list contains a number of language processing areas (e.g., middle temporal gyrus, inferior frontal gyrus, superior temporal gyrus, planum temporale, cerebellum; Hertrich, Dietrich, & Ackermann, 2020). This is in line with findings by Wallentin et al. (2011) who found that the inferior frontal cortex was involved in listening to emotionally intense parts of stories. Furthermore, Lai et al. (2015) and Hauptman, Blank, and Fedorenko (2022) found a similar combination of mentalizing and language processing areas activated in participants reading emotional sentences or non-literal language (see for an overview of how mentalizing and language processes are intertwined Hertrich et al., 2020).

4.4.2. Overlap Between the Three Kinds of Simulation

Apart from these modality-specific brain areas, all three kinds of simulation were associated with activity in an area in the supramarginal gyrus which is situated in the inferior parietal lobe. Although the supramarginal gyrus has been associated with phonological processing (Hartwigsen et al., 2010; Stoeckel, Gough, Watkins, & Devlin, 2009), it has also been associated with literary reading (Hartung et al., 2021) and with referential indexing: the integration of references into a context (Hagoort, 2019; Matchin, Hammerly, & Lau, 2017). Together with the temporoparietal junction (TPJ), the inferior parietal lobe (IPL) is a hub where multiple brain networks come together (i.e., the frontoparietal control network, default mode network, mentalizing network, cingulo-opercular network, ventral attention network; Igelström & Graziano, 2017). Functionally, the TPJ and IPL have been associated with a variety of tasks, from bottom-up perception tasks (such as automatic, stimulus-driven attention) to higher-order tasks (such as self-perception, mind wandering and social cognition; Igelström & Graziano, 2017). The fact that this area is involved in such a variety of brain networks and functions, implies that it is crucial for the integration and regulation of a multitude of neural processes (Igelström & Graziano, 2017). Therefore, it is likely that this area plays a role in situation model updating, which consists of the integration of new information from a variety of sources into an existing situation model (Zwaan et al., 1995; Zwaan & Radvansky, 1998). Situation model updating has been suggested to play a role in mental simulation in general (Kurby & Zacks, 2013; Zwaan, 2009), which is why it is not surprising that this area has been found to be associated with all three kinds of simulation. Taken together, the neuroimaging findings from the current study suggest that, although there are different kinds of simulation distinguishable at the neural level, these different kinds of simulation are all supported by the same, overarching neural mechanism.

4.4.3. Links with Individual Differences in Appreciation and Absorption

For all three kinds of descriptions, I found areas in which individual differences in the percent signal change as a response to these descriptions was associated with measures of story appreciation (story-specific effect) as well as personal characteristics (trait-based measures). Percent signal change was not associated with story world absorption (a state-based measure). Apparently, state-based individual differences do not explain the effects of reading on the neural level, whereas trait-based individual differences do. This is in line with results found in multiple previous studies, where trait-based individual differences were found to be more strongly associated with simulation than state-based individual differences (Faber, Mak, & Willems, 2020; Hartung, Hagoort, & Willems, 2017; Hartung et al., 2021; Mak, De Vries, & Willems, 2020; Van den Hoven et al., 2016). Seemingly, simulation is more strongly associated with stable characteristics than with reading experiences such as absorption and appreciation.

4.4.4. Conclusion

In the current experiment participants read literary short stories in the MRI scanner while their eye movements were being tracked. With this study, I aimed to determine if reading motor descriptions, perceptual descriptions and mental event descriptions are associated with differential neural activation. It was found that reading motor descriptions was associated with activation in modality specific motor areas and in areas involved in event segmentation and situation model building. Reading perceptual descriptions was associated with activation in modality specific perceptual processing areas. Reading mental event descriptions was associated with activation in mentalizing and language processing areas. All three kinds of descriptions were associated with activation in the supramarginal gyrus, which is part of a brain area involved in the integration and regulation of a multitude of neural processes. I propose that it is likely that this area is involved in referential indexing and situation model building, and is therefore also an important area for the integration of different kinds of mental simulation. Furthermore, within the activated areas I found some individual differences in percent signal change that were mainly associated with trait-level individual differences (i.e., perspective taking, transportability). This shows that trait-based personal characteristics can influence how strongly readers respond on the neural level to mental simulation as elicited by stories. Taken together, these findings suggest that mental simulation is supported by both domain-specific processes grounded in previous experiences, and by the neural mechanisms that underlie higher-order language processing (e.g., situation model building, event indexing, integration).

5 | Different Routes to Liking: How Readers Arrive at Narrative Evaluations

When two people read the same story, they might both end up liking it very much. However, this does not necessarily mean that their reasons for liking it were identical. I therefore ask what factors contribute to "liking" a story, and - most importantly - how people vary in this respect. I found that readers like stories because they find them interesting, amusing, suspenseful and/or beautiful. However, the degree to which these components of appreciation were related to how much readers liked stories differed between individuals. Interestingly, the individual slopes of the relationships between many of the components and liking were (positively or negatively) correlated. This indicated, for instance, that individuals displaying a relatively strong relationship between interest and liking, generally display a relatively weak relationship between sadness and liking. The individual differences in the strengths of the relationships between the components and liking were not related to individual differences in expertise, a characteristic strongly associated with aesthetic appreciation of visual art. The work presented in the current chapter illustrates that it is important to take into consideration the fact that individuals differ in how they arrive at their evaluation of literary stories, and that it is possible to quantify these differences in empirical experiments. This work suggests that future research should be careful about "overfitting" theories of aesthetic appreciation to an "idealized reader", but rather take into consideration variations across individuals in the reason for liking a particular story.

This Chapter Is Based on

Mak, Marloes, Faber, Myrthe, & Willems, Roel M. (2022). Different routes to liking: How readers arrive at narrative evaluations. *Cognitive Research: Principles and Implications, 7,* 72. DOI: 10.1186/s41235-022-00419-0

In Chapters 2-4, I described three experiments in which I tried to study mental simulation during literary reading by means of reading instructions, eye tracking, and fMRI. I also focused on individual differences in mental simulation, and asked whether these individual differences in mental simulation were related to individual differences in subjective reading experiences or personality characteristics.

One of the subjective reading experiences I have looked at in Chapters 2-4 is aesthetic appreciation. This was measured by means of a questionnaire containing questions of general story liking (or reading enjoyment), and adjectives on which readers could rate the stories they read. The current chapter focuses on these two ways of measuring aesthetic appreciation, and asks whether the adjective ratings can explain individual differences in reading enjoyment (and if so, if this happens the same way or differently in all readers).

5.1. Introduction

People often do not have to think long about whether they like something (e.g., architecture, art; see Jacobs, Hofmann, & Kinder, 2016). Indeed, it seems easy for readers to decide whether they "like" a story or not. Although such ratings of liking can give an impression of someone's aesthetic preferences, they do not offer any insight into what drives these evaluations. People might arrive at the same judgment in different ways: it is possible that someone for instance likes a story because of its emotional content, whereas another person likes it because they are interested in the topic. Here, I aim to explore how people differ in what determines whether they "like" a story or not in the context of literary reading. I particularly investigate whether and how the contribution of different cognitive and emotional processes varies across readers.

Models of aesthetic appreciation propose that both cognitive and affective processes play a role in aesthetic evaluation (Chatterjee & Vartanian, 2016; Jacobs, 2015a; Leder, Belke, Oeberst, & Augustin, 2004; Leder & Nadal, 2014), and that both of these processes can be either conscious or subconscious (i.e., automatic; see also Graf & Landwehr, 2015). In addition, sensory-motor processes, such as sensation and perception, might play a role, in particular in the context of engaging with aesthetic objects such as artworks (Chatterjee & Vartanian, 2016). How these processes interact with each other likely varies across individuals. For instance, expertise, taste, personality, and pre-existing mood are likely to affect how cognitive and affective processes influence evaluative decisions made by observers (Chatterjee & Vartanian, 2016; Leder et al., 2004). An art connoisseur for instance will experience a painting differently than a layman (and arrive at their evaluative decision differently): the connoisseur may rely more heavily on cognitive processes (e.g., trying to understand the meaning of the painting) whereas the layman may rely more on the positive or negative affect elicited by the painting (see Leder, Gerger, Brieber, & Schwarz, 2014, for evidence of reduced affective responses to artworks in art history students). This means that their aesthetic experience of the painting might differ, even if they both reach the same conclusion about the painting ("I like this painting").

Cognitive and affective processes are also thought to play a role in how people arrive at aesthetic evaluations in narrative reading (Jacobs, 2015a). According to the Neurocognitive Poetics Model (NCPM; Jacobs, 2015b), there are two routes of processing literary narratives. The fast, affective processing route results in "fiction feelings" (e.g., empathy, vicarious emotions, narrative absorption) via emotional contexts in narratives. Cognitive processing is thought to be a slow route that results in so-called "aesthetic feelings" (i.e., feelings induced by the aesthetic experience) via foregrounded elements in narratives (i.e., stylistic devices, defamiliarization). Previous work has suggested that motivational-emotional processes such as interest, meta-emotions, and taste might influence whether people are likely to prefer reading narratives that align with either route (Bartsch, Vorderer, Mangold, & Viehoff, 2008; Zillmann, 1988), but empirical evidence is markedly lacking in the field (Jacobs, 2015b).

Recent work has approached aesthetic emotions as a multidimensional construct, resulting in the development of the Aesthetic Emotions Scale (AESTHEMOS; Schindler et al., 2017). This scale captures emotions related to aesthetics (e.g., positive emotions such as beauty, fascination, and negative emotions such as ugliness), epistemics (e.g., interest), amusement (e.g., humor), and qualitative aspects of experience such as whether the reader feels activated or relaxed by the text (Schindler et al., 2017). Importantly, experiencing one emotion does not preclude the possibility of experiencing another (seemingly opposite) emotion (Schindler et al., 2017). Applications in the context of various aesthetic experiences (e.g., concerts, theatrical performances, exhibitions) highlight how people can experience many different emotions at the same time, and that the specific combination of experienced emotions can differ between people and between (types of) stimuli, together constituting a person's "signature" of affective aesthetic processing (Schindler et al., 2017).

Here, I aim to explore how people differ in what determines whether they "like" a story or not. I build on previous work that measured "aesthetics from below" (Knoop et al., 2016; cf. Fechner, 1876). Knoop et al. (2016) selected adjectives that could be used to describe readers' aesthetic experiences while engaging with literature (i.e., poetry, plays, comedies, novels, short stories). Ratings were gathered from 1544 participants, resulting in a list of 22 adjectives that were brought up by a minimum percentage (> 10%) of participants (Knoop et al., 2016; for a similar approach to capture the aesthetic appreciation of objects, see Jacobsen, Buchta, Köhler, & Schröger, 2004). From these lists, I took all adjectives (N = 13) that could be used for rating literary short stories (thus leaving out musical/poetry specific terms such as *melodious or poetic*[1]) and presented them, together with a question regarding general story liking, to 270 readers who read Dutch literary short stories (nine different stories in total) across three experimental studies. Since it is unclear how readers differ in their reliance on one or more aesthetic features to come to an overall 'liking' of a story, the main goal of this chapter will be to get better insight into such individual differences.

In this chapter I aim to answer five consecutive questions, to uncover what aspects of stories lead to story liking, and, importantly, whether and how this differs between readers. I ask (1) whether the adjectives derived from Knoop et al. (2016) tap into distinguishable components of literature appreciation. I obtain these components using principal components analysis, which results in clusters of adjectives and participant-level scores on each component. I ask (2) how these components are related to "story liking", and (3) whether there is variation between readers in how the components relate to story liking. Subsequently, I ask whether (4) the direction of the relationship between the components and liking is consistent across participants and (5) whether the variation in slopes between participants is systematically associated with reader characteristics (i.e., reading habits, print exposure, story world absorption). This last question sheds light on whether literary expertise matters with regard to how different components of appreciation contribute to the aesthetic evaluation of stories.

[1]The adjectives that I did not select, were either specific to different genres (theatre, poetry), or were not applicable to the stories that were used in this research (e.g., romantic, short, succinct).

5.2. Methods

Data sets collected in three previous studies were combined for this investigation. In all the previous studies participants read Dutch literary short stories, and completed an appreciation questionnaire (Mak & Willems, 2019) as well as questionnaires regarding story world absorption (Kuijpers et al., 2014), reading habits in daily life (Hartung et al., 2016; Mak & Willems, 2019), and print exposure (Author Recognition Test; Stanovich & West, 1989). I will describe each questionnaire in more detail below.

The first study (Mak & Willems, 2019) investigated different kinds of mental simulation during narrative reading, the second study (Eekhof, Eerland, & Willems, 2018) tested the influence of verb tense on mental simulation during literary reading, and the third study (Mak et al., 2020) investigated the influence of pre-reading instructions on reported mental imagery and other subjective reading experiences.

5.2.1. Participants

In total, 270 native speakers of Dutch were tested across three experimental studies (see Table 5.1 for sample characteristics). The majority of the participants were university or college students. Depending on the study, participants read two, three or four Dutch literary short stories (resulting in a total of nine different stories overall, for distribution across studies, see below), which resulted in 716 individual data points (i.e., completed questionnaires; one per participant/story combination). Of these 716 questionnaires, there were 13 questionnaires where at least one question was skipped by the participant. As a result, 703 data points were complete and could be entered into data analysis. Participants were recruited from the Radboud University participant pool, and received appropriate compensation (monetary or course credits) for their participation. All studies were approved by the local ethics committees (approval code 8976), and were conducted in accordance with the Declaration of Helsinki.

5.2.2. Materials

5.2.2.1. Stories

Characteristics of the stories read by the participants in the three studies are shown in Table 5.2. A short synopsis of all stories can be found in Appendix G. The common structure of all stories is that they describe an event or person,

Table 5.1.: *Sample Information for the Three Studies*

Study	N			*M*age (Range)
	Female	Male	Other	
Study 1 (Mak & Willems, 2019)	81	21	0	23.27 (18 – 40)
Study 2 (Eekhof et al., 2018)	33	9	1	23.26 (18 – 46)
Study 3 (Mak et al., 2020)	103	22	0	23.80 (18 – 61)

followed by some plot twist or extraordinary event, and end with a very open ending that leaves the reader feeling a bit alienated. Stories differed across studies, as they had been selected separately for each study, from the entire collection of Dutch literary short stories. However, all studies used literary stories, written by critically acclaimed authors and published by literary publishing houses. All stories belonged to the genre of "literary short story", were available in Dutch, and were readable in 10 to 15 minutes. Except for *Symbols and Signs*, all stories were originally written in Dutch. *Symbols and Signs* was read in a published translation, which was translated from English to Dutch by a professional translator. In Study 1 and Study 3, the stories were presented in their original form. In Study 2, the original stories, alongside slightly altered versions in which the verb tense was changed from present to past tense or vice versa (for reasons not relevant to the current study, and with no reported difference in readability between original and altered versions, see Eekhof et al., 2018).

5.2.2.2. Questionnaires

The Appreciation Questionnaire consisted of a general score of story liking (*How did you like the story*; 1 = It was very bad, 7 = It was very good) and 13 adjectives (e.g., *[did you find the story] Entertaining, ... Ominous, etc.*) that I adapted from Knoop et al. (2016). Studies 2 (Eekhof et al., 2018) and 3 (Mak et al., 2020) both omitted one adjective from the list (Study 2: Special; Study 3: Entertaining), resulting in 11 adjectives that were included in the lists in all three studies. The resulting 11 adjectives that were included in the analysis can be found in Table 5.3. Finally, six questions were asked regarding the enjoyment of the story (from Kuijpers et al., 2014; e.g., *I was constantly curious about how the story would end; I thought the story was written well, etc.*). These final six questions were omitted from the analyses in the current study, because they were highly

Table 5.2.: *Descriptive Information for the Stimulus Stories Used in the Three Previous Studies*

Study	Story	Author	Year of Publication	Word Count
Study 1 (Mak & Willems, 2019)	De Mensen Die Alles Lieten Bezorgen (The People That Had Everything Delivered)	Rob van Essen	2014	2988
	De Chinese Bruiloft (The Chinese Wedding)	Sanneke van Hassel	2012	2659
	Signalen en Symbolen (Symbols and Signs)	Vladimir Nabokov	1948 /2003	2143
Study 2 (Eekhof et al., 2018)	Het Is Muis (It Is Mouse)	Sanneke van Hassel	2012	2016
	Hoe de Wolven Dansen (How the Wolves Dance)	Jordi Lammers	2017	1176
	De Invaller (The Substitute)	René Appel	2003	743
	Ze Is Overal (She Is Everywhere)	Ed van Eeden	2015	1074
Study 3 (Mak et al., 2020)	Brommer op Zee (Moped on Sea)	Maarten Biesheuvel	1972	1827
	God en de Gekkenrechter (God and the Judge of the Insane)	Adriaan van Dis	1986	2026

correlated with the liking question, and were therefore not considered to be of added importance for the current investigation. Participants rated both the adjectives and the questions regarding enjoyment on a seven-point scale (1 = disagree, 7 = agree).

To compare the results on the appreciation questionnaire to other subjective reading experiences, I also measured story world absorption, which refers to an experiential state in which readers are focused on reading and the content of what is read (Kuijpers, 2014). In particular, if the reading process feels effortless, readers experience a narrative world and feel for or with characters, and mental imagery is rich and vivid (Kuijpers, 2014). Story world absorption was measured using the Story World Absorption Scale (SWAS; Kuijpers et al., 2014). The SWAS is a validated scale consisting of 18 items with high internal validity, which measure four aspects of story world absorption on the four subscales Attention, Transportation, Emotional Engagement and Mental Imagery (e.g., *When I finished the story I was surprised to see that time had gone by so fast; I could imagine what the world in which the story took place looked like*). Participants rated each question on a seven-point scale (1 = disagree, 7 = agree).

Additionally, I was interested in whether habitual readers differed in their appreciation of stories from participants who do not read much in daily life. Reading habits were measured using five multiple choice questions about reading habits in everyday life, with four or five answer options (Hartung et al., 2016; Mak & Willems, 2019; e.g., *How often do you read fiction?; How many books do you read each year?*). Additionally, participants were asked for their genre preference in an open-ended question, where they could list up to three genres they enjoyed reading (this question was added for purposes irrelevant to the current study, and will not be used in the analyses in this chapter). As an implicit measure of print exposure, participants completed the well-established Author Recognition Test (ART; Acheson et al., 2008; Stanovich & West, 1989; Dutch adaptation reported in Koopman, 2015), consisting of 42 names (30 real authors and 12 foils), where they had to indicate who they thought were genuine authors.

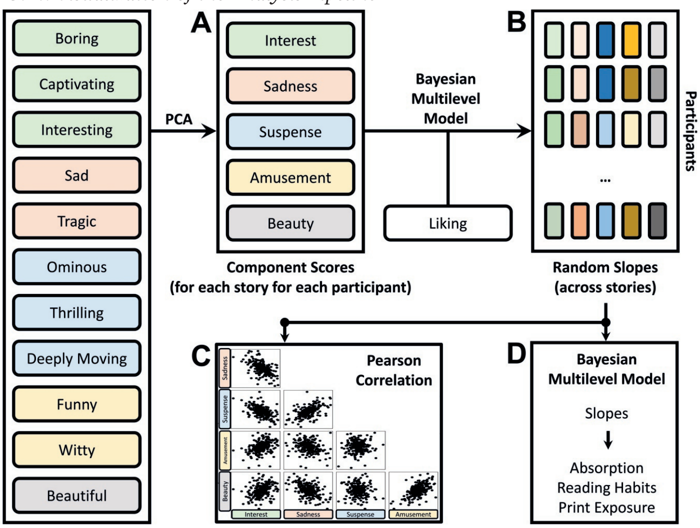
5.2.3. Procedure

In all studies, informed consent was obtained before the experiment, after which participants were instructed to read as naturally as possible. The stories (i.e., three stories in Study 1, four stories in Study 2 and two stories in Study 3; see Table 5.2) were read in a counterbalanced order. After reading the first story, participants completed the SWAS and Appreciation Questionnaire. These steps were repeated for the other stories in the experiment. After participants had read the last story and completed the corresponding questionnaires, they filled out the reading habits questionnaire and the Author Recognition Test.

5.2.4. Data Analysis

In Fig. 5.1 I give a schematic overview of the analysis pipeline. Each analysis step is described in detail below, the following description serves to give a rough overview. In the first step of the analysis, the 11 adjectives from the appreciation questionnaire (see Fig. 5.1, left column) were entered into a principal components analysis resulting in five components (Fig. 5.1A). Then, participants' scores (per story) on these components were linked to liking scores per story, while allowing for random slopes for the components over participants and over stories (Fig. 5.1B). With this analysis, I first focused on the population-level effects of the components, to find out whether the different components of appreciation each play a role in the eventual evaluation of stories. The by-participant variation in the random slopes across components was compared in a correlation analysis (Fig. 5.1C). Finally, variation in the random slopes was linked to absorption, reading habits and print exposure (Fig. 5.1D). With this final analysis, I zoom in on the participant level to acknowledge the individual differences in story liking and to try to explain some of these individual differences by linking them to concepts that are theorized to be related to aesthetic processes and may explain individual differences therein.

Figure 5.1.: *Visualization of the Analysis Pipeline*



5.3. Results

All data and analysis scripts are available on the Open Science Framework, `https://osf.io/h3ct6/`.

5.3.1. Question 1: Do Adjectives Tap Into Distinguishable Components of Literature Appreciation?

The first step of the analysis pipeline (see Fig. 5.1) was to reduce the 11 adjectives to a smaller number of components consisting of highly similar adjectives (Fig. 5.1A). Using the package *psych* (Revelle, 2014) in *R* version 4.0.3 (R Core Team, 2021), I conducted a principal component analysis (PCA) with oblique rotation (direct oblimin) on the 11 appreciation adjectives used in all three studies. The resulting components tap into distinct aspects of literature appreciation.

The Kaiser-Meyer-Olkin measure (KMO) was .83 (all KMO values for individual items $> .61$), indicating good sampling adequacy for this analysis. Bartlett's test of sphericity showed sufficient correlation between items, χ^2 (55) = 490.56, $p < .001$. The primary rationale for determining the number of components was maximization of explained variance (at least 80% of variance explained), along with interpretability of the component (i.e., reducing the number of dimensions while making sure that these components still represented the original data reasonably well). A 5-component solution explained 81% of the variance, and therefore represents the original data closely. For the 5-component solution, the mean communality was $> .7$, and the fit (fit based upon off diagonal values) was 97.2%.

The first component that I found corresponded to Interest (consisting of items Boring (-), Captivating, and Interesting); the second component to Sadness (Sad, Tragic); the third component to Suspense (Ominous, Suspenseful, Gripping); the fourth component to Amusement (Funny, Witty); and the final component to Beauty (Beautiful). The structure and pattern matrices for the factor loadings after rotation can be found in Table 5.3. All correlations between the components were below $r = .43$, confirming that the extracted components were indeed measuring separate constructs, and that such lists of adjectives can be used to measure distinct aspects of literature appreciation. Component scores per participant per story were used in the subsequent analyses.

5.3.2. Question 2 & 3: How Do Adjective Components Relate to "Story Liking"? Is There Variation Between Readers in the Way These Components Relate to "Story Liking"?

The components resulting from the PCA were used to assess how the adjectives related to "story liking". This relationship was analyzed (see Fig. 5.1B) with

Table 5.3.: *Pattern Matrix for the PCA of the 11 Adjectives on the Appreciation Questionnaire (N = 703). Factor loadings over .40 appear in bold*

	Pattern Matrix				
	Interest	Sadness	Suspense	Amusement	Beauty
Beautiful	0.11	-0.01	-0.12	0.02	**0.92**
Boring	**-0.90**	-0.01	-0.01	0.01	-0.06
Gripping	0.20	0.35	0.41	-0.03	0.32
Funny	0.25	-0.01	-0.25	**0.84**	-0.17
Interesting	**0.51**	0.05	0.18	0.13	0.32
Ominous	-0.05	0.12	**0.88**	-0.05	-0.10
Sad	-0.03	**0.93**	-0.09	-0.03	0.04
Suspenseful	0.29	-0.12	**0.75**	0.08	-0.02
Tragic	0.04	**0.91**	0.08	0.04	-0.07
Witty	-0.18	0.00	0.24	**0.83**	0.25
Captivating	**0.59**	0.00	0.25	0.16	0.24

a Bayesian Multilevel[2] Model using the package *brms* (Bürkner, 2017, 2018) and *Stan* (Stan Development Team, 2020) in *R* version 4.0.3 (R Core Team, 2021). The rationale for calculating a Bayesian multilevel model as opposed to a "classical" frequentist model, was that Bayesian models are more flexible and more capable of fitting complex models (e.g., Bürkner, 2018; Nalborczyk et al., 2019). Rather intuitively, Bayesian multilevel models calculate the range of the most probable values of each parameter, a 95% Credible Interval. If this Credible Interval does not cross zero for a given parameter, this indicates a 95% certainty that the true value of this parameter is distinguishable from zero.

I constructed a partially crossed model that predicted the answer on the general liking question (*How did you like the story?*) by the individual scores on the five components found in step 1, allowing random intercepts and slopes for all five predictors per participant and per story[3]. This random effect structure made sure that the model took the between subject and between story variation into account. As a result, the data were analyzed in such a way that the observations that belonged together (because they belonged to the same participant) were grouped together. Therefore, these random intercepts allowed me to

[2]Multilevel models are also known as mixed models or hierarchical models. The principle behind these types of models is that they are distinguishing a population level and a group level (in this case accounting for inter-participant and inter-story random effects).

[3]All components were incorporated in the same model, to account for any shared variance between the components. Note, however, that this also means that all results of the components are after controlling for the other components, possibly reducing the effect sizes for the associations between all components and liking.

control for the fact that all participants and all stories occurred more than once in the dataset. For the population-level intercept I used a weakly informative, normally-distributed prior with a mean of 0 and a standard deviation of 10. A weakly informative, normally-distributed prior with a mean of 0 and a standard deviation of 1 was set for the fixed effects. These priors are considered relatively conservative (McElreath, 2016). As variance can only be positive, weakly regularizing, half-cauchy priors with a mean of 0 and a standard deviation of 1 were used for the variance of the random effects as well as the overall variance (as suggested by Gelman, 2006; McElreath, 2016). The model was trained during 4000 iterations, using 4 chains, and using an MCMC sampler (for a complete model specification, see the analysis scripts on the Open Science Framework, `https://osf.io/h3ct6/`). The Gelman-Rubin diagnostic (Rhat) was 1.0 for all parameters, indicating that the model had converged.

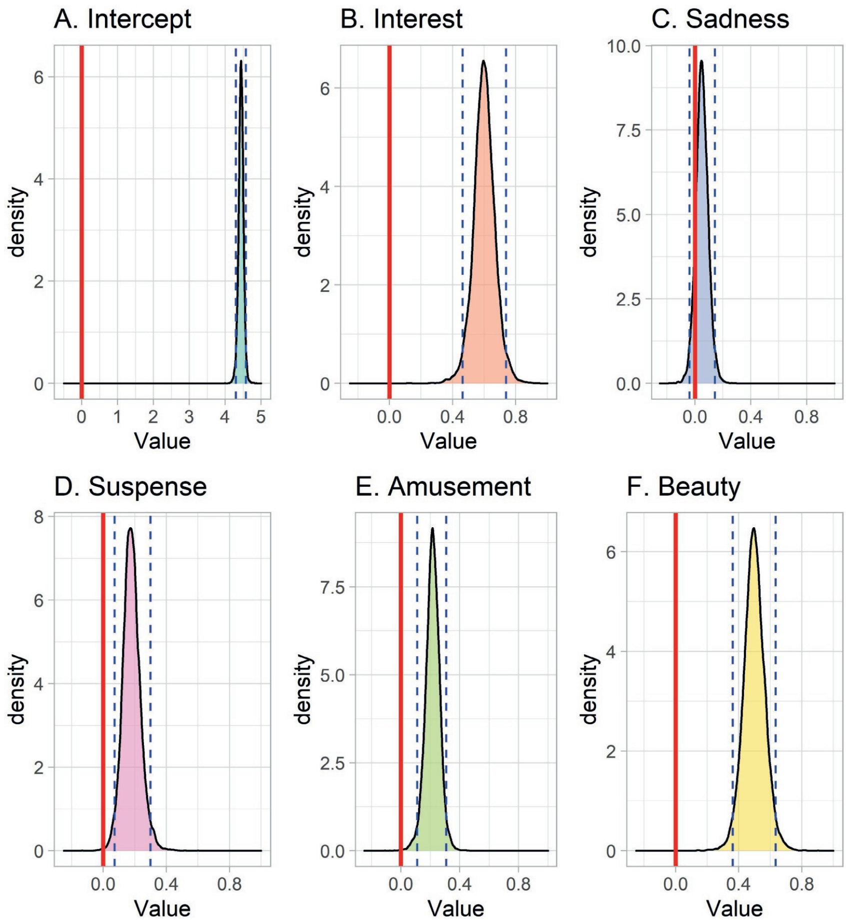
I found that the Interest component was positively associated with story liking, showing that stories that were considered more interesting were generally liked more (Table 5.4; Fig. 5.2B; mass > 0: 99.9%). Additionally, the relationship between Interest and liking varied between participants and between stories (the standard deviation of the slope of the Interest component = 0.16 [CI: 0.02 - 0.28] across participants; and 0.15 [CI: 0.03 - 0.32] across stories) . Posterior distributions of the individual slopes for the association between Interest and liking (per participant) showed that this association was positive for all participants (all participants showed a positive association between Interest and general story liking; the complete by-participant posterior distributions can be seen on `https://osf.io/cr7nh/`).

Table 5.4.: *Posterior Distributions (Median, MAD, 95% CI) of the Population-Level Associations between the Components and Liking. MAD = Median Absolute Deviation; CI = Credible Interval*

	Estimate (Median)	Estimate (MAD)	Lower Bound (95%CI)	Upper Bound (95%CI)
(Intercept)	4.44	0.06	4.30	4.58
Interest	0.60	0.06	0.47	0.73
Sadness	0.05	0.04	-0.03	0.14
Suspense	0.17	0.05	0.08	0.30
Amusement	0.22	0.05	0.11	0.31
Beauty	0.50	0.06	0.36	0.63

I found no conclusive evidence for an association between the Sadness component and story liking (Table 5.4; Fig. 5.2C): as the credible interval crossed

Figure 5.2.: *Posterior Distributions of the Population-Level Fixed Effects of the Relationships Between the Components and Liking. The Intercept (A) represents the average liking score. The blue dashed lines indicate the limits of the 95% credible interval. If the credible interval of a parameter does not cross zero, this means that it is likely that the true value for that parameter is different from zero. Code for this figure is adapted from* `https://www.rensvandeschoot.com/tutorials/brms-started/`.



zero, I cannot reasonably assume a positive relationship between Sadness and liking (mass > 0: 87.9%). However, I did find variation between participants (the standard deviation of the slope of the Sadness component = 0.23 [CI: 0.11 - 0.32] across participants). The posterior distributions of the individual slopes for the association between Sadness and liking (per participant) showed that some participants showed a positive association between Sadness and Liking, although there were also participants who showed no association or a negative association between Sadness and Liking (the complete by-participant posterior distributions can be seen on `https://osf.io/zvau8/`). Ultimately, the data suggest that some readers like a story more when they consider it to be sadder, whereas others are indifferent to the sadness of a story or actually dislike sad stories. There was no clear variation in the relationship between Sadness and liking across stories (the standard deviation of the slope of the Sadness component = 0.05 [CI: 0.00 - 0.17] across stories).

The Suspense component was positively associated with story liking (see Table 5.4; Fig. 5.2D; mass > 0: 99.7%). The relationship between Suspense and liking varied between participants and between stories (the standard deviation of the slope of the Suspense component = 0.18 [CI: 0.04 - 0.28] across participants; and 0.09 [CI: 0.01 - 0.27] across stories). The posterior distributions of the individual slopes for the association between Suspense and liking (per participant) suggested that a large part of the participants showed a positive association between Suspense and Liking, but there were also participants who showed no association or a negative association between Suspense and Liking (the complete by-participant posterior distributions can be seen on `https://osf.io/3vxsp/`). This suggests that many readers like a story more when they consider it to be more suspenseful, but some are indifferent to suspense, or dislike suspenseful stories.

The Amusement component showed a very similar pattern. Amusement was positively associated with story liking (see Table 5.4; Fig. 5.2E; mass > 0: 99.8%). Again, the relationship between Amusement and liking varied between participants (the standard deviation of the slope of the Amusement component = 0.20 [CI: 0.08 - 0.29] across participants). The posterior distributions of the individual slopes for the association between Amusement and liking suggested that a large part of the participants showed a positive association between Amusement and Liking, whereas some participants showed no association or a negative association between Amusement and Liking, indicating that many readers like a story more when they consider it to be more amusing, but some are indifferent, or dislike amusing stories (the complete by-participant posterior distributions can be seen on `https://osf.io/a93yk/`). There was no clear variation in the relationship between Amusement and liking across stories (the standard deviation of the slope of the Amusement component = 0.07 [CI: 0.00 - 0.22] across stories).

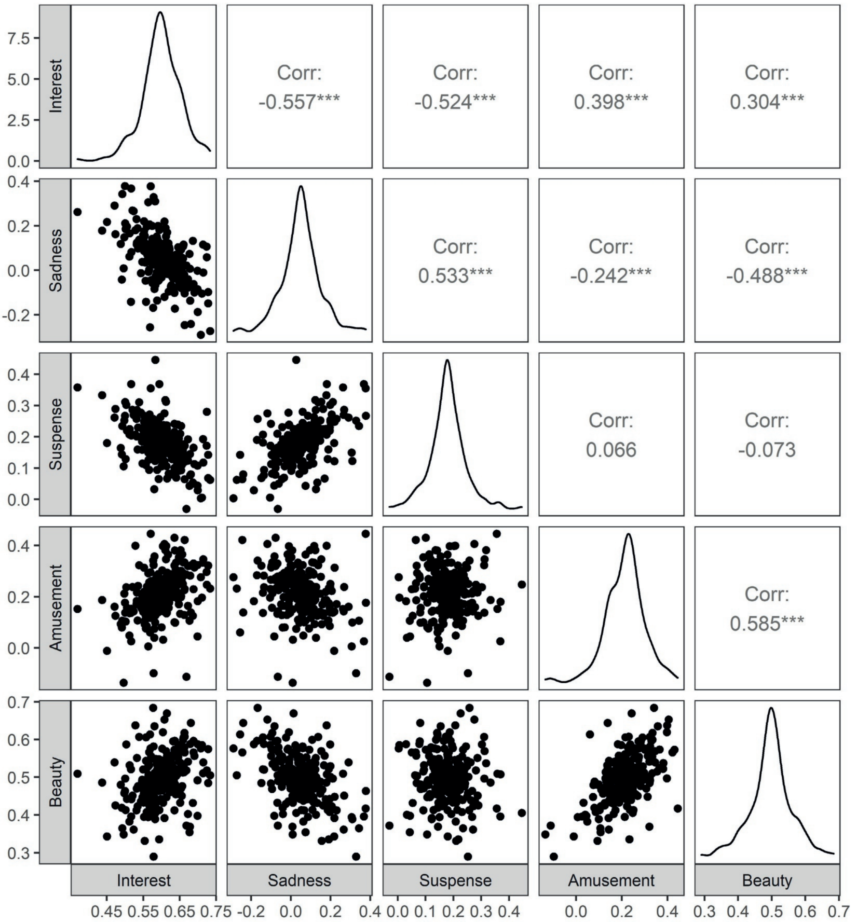
Finally, the Beauty component was positively associated with story liking, showing that stories that were considered more beautiful were generally liked more (see Table 5.4; Fig. 5.2F; mass > 0: 99.98%). The relationship between Beauty and liking varied between participants and between stories (the standard deviation of the slope of the Beauty component = 0.16 [CI: 0.03 - 0.27] across participants; and 0.14 [CI: 0.06 - 0.30] across stories). The posterior distributions of the individual slopes for the association between Beauty and liking (per participant) showed that this association was positive for all participants (all participants showed a positive association between Beauty and general story liking; the complete by-participant posterior distributions can be seen on `https://osf.io/n47hr/`).

5.3.3. Question 4: Is the Direction of the Relationship Between the Components and Liking Consistent Across Participants?

As the relationships between all components and liking reliably varied between participants, it would be interesting to know whether these relationships correlated with each other on the individual level (within participants). For instance, if a given participant displays a relatively strong association between Interest and liking, does this same participant also display a relatively strong association between Amusement and liking? To address this question, I first extracted the estimated slopes (median per participant, collapsed across the individual story-readings within each participant) for the associations between the components and general story liking (i.e., 270 coefficients for each of the five components) from the model reported above. All slopes were entered into a pair-wise correlation analysis (see Fig. 5.1C; Fig. 5.3), with Bonferroni correction for multiple comparisons.

The slopes for the relationship between Interest and liking were moderately negatively correlated to the slopes for the relationships between Sadness and liking (r = -.601; $p < .001$) and Suspense and liking, (r = -.542; $p < .001$), suggesting that participants displaying a relatively strong association between Interest and liking, displayed relatively weak associations between Sadness / Suspense and liking. Oppositely, the slopes for the relationship between Interest and liking were weakly positively correlated to the slopes for the relationships between Amusement and liking (r = .387; $p < .001$) and Beauty and liking (r = .338; $p < .001$), suggesting that participants displaying a relatively strong association between Interest and liking, also displayed relatively strong associations between Amusement / Beauty and liking.

Figure 5.3.: *Plot of the Correlations Between the Slopes for the Associations of the Components with Liking. Below the diagonal, scatterplots of the individual slopes are displayed. The diagonal represents density plots of the distributions of the slopes. Pearson correlation coefficients are given above the diagonal. *** indicates p < .001. Bonferroni correction for multiple comparisons was applied.*



The slopes for the relationship between Sadness and liking were moderately positively correlated to the slopes for the relationship between Suspense and liking (r = .549; p < .001), suggesting that relatively high associations between Sadness and liking co-occurred with relatively high associations between Suspense and liking. Oppositely, the slopes for the relationship between Sadness and liking were weakly negatively correlated to the slopes for the relationship between Amusement and liking (r = -.245; p < .001) and moderately negative correlated to the slopes for the relationship between Beauty and liking (r = -.504; $p < .001$), suggesting that participants displaying a relatively strong association between Sadness and liking, displayed relatively weak associations between Amusement / Beauty and liking.

The slopes for the relationship between Suspense and liking were not correlated to the slopes for the relationships between Amusement and liking ($r = .063$; $p = .30$) and between Beauty and liking ($r = -.093$; $p = .13$). This indicates that participants displaying a relatively strong association between Suspense and liking, do not reliably display weaker or stronger associations between Amusement / Beauty and liking.

Finally, the slopes for the relationship between Amusement and liking were moderately positively correlated to the slopes for the relationship between Beauty and liking ($r = .610$; $p < .001$). Relatively high associations between Amusement and liking co-occurred with relatively high associations between Beauty and liking.

5.3.4. Question 5: Is Variation in Slopes Between Participants Systematically Associated with Reader Characteristics?

To assess whether the variation in the slopes (between participants) was systematically associated with reader characteristics, I linked the median estimated slopes per participant (see Question 4) to the scores per participant for story world absorption (M = 4.25, SD = 1.07, Range 1.22 – 6.72), print exposure (Author Recognition Test; M = 7.40, SD = 4.42, Range 0 – 23), and reading habits (the scores on the Reading Habits questionnaire were z-transformed, as they were measured on slightly varying scales across experiments. Descriptive statistics for all questions on the reading habits questionnaire can be found in Appendix H).

I compared the individual slopes to the reader characteristics with Bayesian Multilevel Models (see Fig. 5.1D) using the package *brms* (Bürkner, 2017, 2018) and *Stan* (R Core Team, 2021). I constructed multilevel models that predicted average scores for story world absorption, reading habits and print exposure by the median estimated slopes per participant for the associations between all five components and liking. For story world absorption, there was more than one observation per participant and per story. Therefore, random intercepts for Participant and Story were included in the model for story world absorption. I used a weak, normally-distributed prior with a mean of 0 and a standard deviation of 10 for the population-level intercept. A normal prior with a mean of 0 and a standard deviation of 1 was set for the fixed effects. These priors are considered relatively conservative (McElreath, 2016). As variance can only be positive, half-cauchy priors with a mean of 0 and a standard deviation of 1 were used for the overall variance (as suggested by Gelman, 2006; McElreath, 2016), as well as the variance of the random effects (in the model for story world absorption). The model was trained during 4000 iterations, using 4 chains, and using an MCMC sampler (for a complete model specification, see the analysis scripts on the Open Science Framework, `https://osf.io/h3ct6/`). The Gelman-Rubin diagnostic (Rhat) was 1.0 for all parameters, indicating that the model had converged.

The variation in the slopes for the relationships between the components and liking were not reliably associated with story world absorption (see Table 5.5)[4], print exposure (see Table 5.6), or reading habits (see Table 5.7), all credible intervals crossed zero (see Table 5.5-5.7 for the mass > 0 for all posterior distributions). This means that the by-participant variability in slopes for the relationships between the components and liking cannot be explained by the variability in the measured reader characteristics.

Table 5.5.: *Posterior Distributions of the Associations between the Slopes and Absorption. The Median, Median Absolute Difference, 95%CI and Mass > 0 of the posterior distribution are given*

	Estimate (Median)	Estimate (MAD)	Lower Bound (95%CI)	Upper Bound (95%CI)	Mass > 0 (%)
(Intercept)	4.54	0.64	3.28	5.81	99.9
Interest Slope	0.16	0.82	-1.41	1.75	58.0
Sadness Slope	-0.45	0.52	-1.47	0.56	18.4
Suspense Slope	0.55	0.71	-0.84	1.90	78.5
Amusement Slope	0.02	0.54	-1.09	1.09	51.4
Beauty Slope	-0.99	0.70	-2.44	0.51	9.2

5.4. Discussion

In this study, I aimed to determine what makes readers consider a story to be good or bad, and how people differ in this respect. I found that adjectives used in previous studies (e.g., Knoop et al., 2016) tapped into distinguishable components of literature appreciation, that I labeled Interest, Sadness, Suspense,

[4]There were no effects on any of the subscales of the SWAS either.

Table 5.6.: *Posterior Distributions of the Associations between the Slopes and Print Exposure. The Median, Median Absolute Difference, 95%CI and Mass > 0 of the posterior distribution are given*

	Estimate (Median)	Estimate (MAD)	Lower Bound (95%CI)	Upper Bound (95%CI)	Mass > 0 (%)
(Intercept)	7.11	0.83	5.44	8.81	99.9
Interest Slope	0.03	0.96	-1.88	1.94	51.3
Sadness Slope	0.15	0.94	-1.68	2.02	56.5
Suspense Slope	0.00	0.98	-1.92	1.98	50.3
Amusement Slope	0.53	0.96	-1.29	2.47	71.2
Beauty Slope	0.26	0.97	-1.70	2.18	61.3

Table 5.7.: *Posterior Distributions of the Associations between the Slopes and Reading Habits. The Median, Median Absolute Difference, 95%CI and Mass > 0 of the posterior distribution are given*

	Estimate (Median)	Estimate (MAD)	Lower Bound (95%CI)	Upper Bound (95%CI)	Mass > 0 (%)
(Intercept)	-0.26	0.60	-1.46	0.93	33.4
Interest Slope	0.20	0.79	-1.35	1.73	59.0
Sadness Slope	-0.21	0.49	-1.19	0.78	33.4
Suspense Slope	-0.15	0.68	-1.20	1.48	57.9
Amusement Slope	0.32	0.54	-0.79	1.42	71.9
Beauty Slope	0.13	0.75	-1.29	1.59	58.0

Amusement and Beauty. Four out of five of these components (i.e., Interest, Suspense, Amusement, Beauty) were significantly positively associated with the general question regarding how much participants liked the story. However, Interest and Beauty were more strongly associated with story liking than the other components (i.e., Suspense and Amusement). When looking at individual slopes per participant, I discovered substantial variation in the associations between the five components and story liking on the individual level, suggesting that there might be distinct patterns of relative associations between these components and story liking. Additionally, although Sadness was on average not associated with liking, here I found individual variation as well, with some participants showing a positive association between Sadness and liking, and some participants a negative association.

5.4.1. Individual Differences in the Routes to Appreciation

I found that the individual slopes between the components on the one hand, and liking on the other hand, were weakly to moderately correlated. For some sets of components these slopes were positively related to each other, whereas for other sets of components these slopes were negatively related to each other. These different contributions showed patterns across participants. For example, in readers for whom Interest plays a relatively large role in how much they like a story, Sadness will generally play a relatively weak role. This suggests that readers differ in what drives them to positively evaluate stories.

Looking at the individual variation in the associations between specific components and liking, this association can be strong in some readers, but weak or even negative in other readers. This raises the question whether the assessed components of appreciation capture all reasons people like stories, or that there are other elements that also play into evaluations of stories. One likely possibility is that more cognitive (rather than affective) routes of aesthetic processing, such as foregrounding or stylistic elements in stories, contribute to the evaluation of literary story as well, and perhaps even more strongly in readers who respond weakly or negatively to the components assessed here.

Looking at the individual variation in the association between Sadness and liking specifically, readers differed in how negative emotions were related to their evaluations of stories. In some readers, negative emotions (Sadness) in response to stories lead them to like those stories more, whereas for others negative emotions in response to stories lead to a decrease in liking. The association between negative emotions and liking is reminiscent of the phenomenon of *mixed emotions* in literary reading: It is possible to feel sadness (often experienced as a negative, unappreciated emotion), but perceive this as an enjoyable experience, for example in “bittersweet” situations (e.g., Larsen & McGraw, 2011; Oceja & Carrera, 2009; Schimmack, 2001). An example of mixed emotions in response to fiction can be found in the work by Hanich, Wagner, Shah, Jacobsen, and Menninghaus (2014), which showed that in the context of film, experienced sadness (considered to be a negative emotion) is strongly positively correlated to enjoyment (a positive evaluation). The authors subsequently hypothesized that the correlation between sadness and enjoyment may not be a direct relationship, but may rather be mediated by the feeling of “being moved”. To elaborate, stories may elicit a feeling of sadness, which in turn contributes to the feeling of being moved, which is evaluated as a positive feeling. This way sadness can positively contribute to enjoyment, but only if this sadness results in or is interpreted as a feeling of being moved.

The paradoxical relationship between negative emotions and enjoyment is elaborated on by Menninghaus et al. (2017) in the *Distancing-Embracing model*. They state that the exceptional quality of art in being capable of leading to enjoyment through negative emotion lies in the processes of *distancing* and *embracing*. In this model, distancing refers to the sense of control art viewers feel when interacting with negatively valenced art: Viewers are aware that they can step away and stop looking as soon as they experience too many negative emotions due to the art work. This way they are confident they can distance themselves from these negative feelings before getting overwhelmed. Because of this process of distancing, art viewers can ultimately embrace an art work and the negative emotions associated with it. This might be through a feeling of being moved, or due to a process of cognitive dissonance resolution. A viewer may implicitly reason: This piece of art is eliciting negative emotions, and yet I am choosing to look at it, therefore I must like it. This way, in the aesthetic appreciation of art and literature, negative and positive emotions can both contribute to a positive evaluation of the object in question.

Indeed, as mentioned above, I found readers who displayed positive associations between (negative) emotions and liking, indicating that the processes of distancing and embracing when dealing with mixed emotions might influence "story liking" in some readers. However, there are also quite some readers who show a negative association between negative emotion in response to stories and liking, or are indifferent to negative emotion. The processes of distancing and embracing, and the phenomenon of mixed emotions therefore do not seem to manifest themselves equally in all readers.

Interestingly, individual variation in the relationships between the appreciation components and liking was not related to the experiential process of story world absorption (which conceptually differs from aesthetic experiences[5]) or to

[5]Aesthetic appreciation and absorption are separated in all leading models of empirical aesthetics. One example that is particularly relevant for the present study is the Neurocognitive Poetics model by Jacobs (2015b). According to the Neurocognitive Poetics Model (NCPM; Jacobs, 2015b) aesthetic appreciation and narrative absorption are related to two different modes of literature processing. The point made by Jacobs, is that literary texts can contain elements that evoke either "fast" processing, via the so-called affective route, or "slow" processing, via a cognitive processing route. Fast processing can be evoked by elements such as suspenseful, emotional, and empathy-inducing passages in stories, and will lead to experiences such as narrative absorption in readers. Slow processing can be evoked by elements of foregrounding in stories, such as stylistic devices and defamiliarization. This processing mode will lead to aesthetic experiences, such as aesthetic appreciation of stories. Here, I have defined absorption as "an experiential state in which readers are focused on reading and the content of what is read.

measures of daily life reading habits and print exposure. Although readers varied with respect to the relationship between aesthetic experiences and story liking, this did not translate to other measures often used in reading research (i.e., story world absorption, reading habits, print exposure). Apparently, aesthetic experiences are not directly associated with absorption, reading habits and print exposure, and they should not be used to make predictions about one another.

5.4.2. Cognitive and Affective Routes to Aesthetic Appreciation

As elaborated on in the introduction, there are several theories and models of aesthetic appreciation that highlight the different routes to appreciation (Chatterjee & Vartanian, 2016; Jacobs, 2015a; Leder et al., 2004). Both affective (e.g., emotions elicited by a narrative) and cognitive (e.g., being intellectually challenged by a narrative) processing can contribute to the evaluation of a narrative. These different processing styles can interact in readers (or readers may prefer one style over the other), leading to different evaluations of the same narratives by different readers. In the current study I find that, indeed, interaction between styles occurs in at least some readers, as both affective (e.g., sadness, amusement) and cognitive (e.g., interest) processes can be associated with general liking scores in one reader.

5.4.3. Limitations and Directions for Future Research

It is important to note that the five components of appreciation measured in this chapter, although a good start when it comes to measuring appreciation more comprehensively, will not be the *only* contributors to a reader's eventual evaluation. Especially the cognitive elements of aesthetic appreciation (Chatterjee & Vartanian, 2016; Jacobs, 2015b; Leder et al., 2004) were not sufficiently captured in the adjectives derived from the study by Knoop et al. (2016) and may contribute to liking just as much as the components studied here (or perhaps even more strongly in readers who display low associations between the five components as measured in the current study and liking).

[...] if the reading process feels effortless, readers experience a narrative world and feel for or with characters, and mental imagery is rich and vivid." In this sense, absorption coincides with the fast route of the NCPM. Aesthetic appreciation, however, is not limited to the fast route of the NCPM, but contains both emotional (fast route) and cognitive or aesthetic (slow route) processes.

To address these limitations, it thus seems important that the cognitive processes involved in appreciation are investigated more thoroughly in future research. For example, the degree to which a story is experienced as intellectually challenging or stylistically striking is not captured in the adjectives used in the current study. In this context, the judgment of beauty should receive special attention. As I simply asked participants to rate the stories for being beautiful, without defining beauty as either stylistic or emotional beauty, I cannot tell what participants' spontaneous criteria were when deciding on a rating for beauty (and thus whether this rating reflected a cognitive or emotional aesthetic process).

When studying individual variation in routes to appreciation, it is possible to distinguish two sub-questions. In the current study, I have investigated how participants vary in their routes to liking. I have shown that aesthetic processes can be positively associated with liking in some participants, and negatively associated with liking in other participants. An open question with regard to the individual variation between readers as found in my analyses is why readers vary in their routes to liking. Leder et al. (2014) state that level of expertise is an important factor determining whether someone will prefer a cognitive processing style over an affective processing style. Therefore, I hypothesized that reading habits or print exposure would be associated with individual variation between readers. However, in my results there is no indication that the differences between readers are due to their expertise, despite substantial variation in my sample. Both reading habits and print exposure could not sufficiently explain the differences between readers in the relationships between the components and liking.

Further exploration of the variation between readers could perhaps shine a light on different types of readers that may react differently to aesthetic experiences. For example, it would be interesting to answer the question whether there are mainly cognitively driven (i.e., distanced) or mainly affectively driven (i.e., identifying) readers, as well as readers who are somewhere in between (Riddell & Van Dalen-Oskam, 2018). In a future experiment studying why participants differ in their routes to liking, it would be interesting to let participants read and rate a larger number of texts (perhaps also including texts of different genres). This would also address an important limitation of the current study: As the data in this study were not sampled across genres (and participants each read only 2-4 stories), I cannot generalize across genres. Therefore, no conclusions about genre differences could be made based on these data. A future study in which the stories are thoroughly sampled for genre differences would help shed light on any story or genre differences.

Pinpointing how and why readers vary in their routes to liking could in the future perhaps also help direct individuals to books or stories that they will like, through the use of recommender systems: for example, readers that enjoy sad stories (or who have characteristics that are associated with enjoying sad stories) could be recommended to read books liked by similar readers, whereas readers who prefer amusing stories would receive different recommendations (e.g., Faridani, Jalali, & Jahan, 2017). This could result in more enjoyable reading experiences, which has been associated with a higher inclination to read again (Mol & Jolles, 2014), which in turn has been positively associated with school success (Chiu & McBride-Chang, 2006; Mol & Jolles, 2014; Retelsdorf et al., 2011) second language learning (Lao & Krashen, 2000; Lee et al., 2015; Yamashita, 2008), and social cognition and empathy (Fong et al., 2013; Johnson et al., 2013; Mar & Oatley, 2008; Oatley, 2016).

5.4.4. Conclusion

Looking at the findings from this study, it is important to note that, while I do not contest the merit of any theoretical model of appreciation, there is a danger of "overfitting" these models to an "idealized reader". I showed that readers can have strikingly different reasons for indicating that they like a story or not. As a consequence, a simple question about the "liking" of a particular story will not be informative about the variation in the reading experiences that readers have. My findings have illustrated how these individual differences contribute to evaluations, and have provided an example of how these differences could be quantitatively and empirically tested. This work might therefore motivate future empirical approaches to establishing individual differences in appreciation to get to a deeper understanding of what it means to "like" a story.

6 | General Discussion

At the beginning of this dissertation, I asked the questions "Do people use mental imagery during story reading, how does this relate to how readers experience stories, and how do people differ in these respects?" Throughout this dissertation, I focused on mental simulation during literary reading, starting from the definition given by Barsalou (2008, p. 618):

> *"Simulation is the re-enactment of perceptual, motor, and introspective states acquired during experience with the world, body, and mind."*

To answer the questions asked in the introduction, I performed three studies. These studies are described in Chapters 2, 3 and 4. Chapter 5 describes a re-analysis of questionnaire data from three previous studies, in which individual differences in the relationships between the subjective experiences of general liking, aesthetic appreciation, and narrative absorption are explored further.

In the current Chapter, I will discuss the findings from these studies to answer the research questions asked in the introduction. In sections 6.1. – 6.3. I will describe what the studies presented in this dissertation tell us about the research questions. In section 6.4. I will place the findings from these studies in light of the three theories on mental simulation during literary reading that were discussed in the introduction, and I will discuss the implications of the findings. Section 6.5 will summarize the conclusions that can be drawn from the work presented in this dissertation.

6.1. Do People Use Mental Imagery During Story Reading?

I will answer this research question in three parts. First (section 6.1.1.), I will discuss what the findings reported in this dissertation indicate when it comes to the roles of automatic or subconscious mental simulation and deliberate, conscious mental imagery during literary reading. Next, I will discuss evidence for the distinction between motor simulation, perceptual simulation and mentalizing (cf. Barsalou, 2008), both at the level of eye movements (section 6.1.2.) and the level of brain activity (section 6.1.3.). In section 6.1.4., I will combine these three parts to evaluate what has been learned about whether people use mental imagery during story reading.

6.1.1. Automaticity Versus Deliberateness of Mental Simulation

Regarding the first question from this dissertation “Do people use mental imagery during story reading?”, the findings from the experiment reported in Chapter 2 show that mental imagery during story reading does probably not occur in an explicit, intentional form in the naturalistic context of story reading. In Chapter 2, I described an experiment in which readers were instructed to either (1) read stories while vividly imagining everything in the stories, (2) focus on the grammatical structure of the stories, or to (3) read as if they were reading for leisure. I found small differences between the first two groups (vivid imagery vs grammatical structure) on subjective reports of mental imagery, but no differences between any of the groups in other reading experiences. From this study, it can be concluded that people can intentionally use mental imagery to a limited degree only. That is, readers who were asked to intentionally increase their mental imagery, did not report significantly more mental imagery than readers who read for leisure (which was considered the baseline condition in the experiment). However, readers who were asked to intentionally increase their mental imagery did report more mental imagery than readers who had been told to focus on the grammatical structure of the narrative (which was meant to distract from using mental imagery). Additionally, any increases or differences in reported mental imagery were not related to the experienced absorption and appreciation (constructs that have been found to be strongly associated with mental imagery; Bilandzic & Busselle, 2011; Green & Brock, 2002; Kuijpers et al., 2014). Together, these findings suggest that the mental imagery experienced during story reading is not easily voluntarily influenced.

However, when looking at natural reading processes, I found in Chapters 3 and 4 that there are individual differences in mental simulation (see also section 6.3.). In Chapters 3 and 4, participants were told to read stories as if they were reading for leisure (comparable to the baseline condition used in Chapter 2). During reading, participants had their eye movements and their brain activity tracked, to determine how participants responded to encountering simulation-eliciting descriptions in the stories they read. This way, it was possible to observe natural reading processes, as reading was not being influenced by additional tasks. Any differences between participants could therefore be attributed to individual differences in mental simulation instead of differences between experimental tasks. As I did find substantial individual differences in the strength of the relationship between simulation-eliciting descriptions on the one hand

and reading behavior and brain activity on the other hand, two conclusions can be drawn. First, it can be concluded that readers are susceptible to simulation-eliciting descriptions in stories, suggesting that they are employing some kind of mental imagery or mental simulation. Secondly, the finding that individual differences do occur naturally, but are not easily voluntarily influenced, suggests that it is indeed more likely that people use automatic or subconscious mental simulation during reading instead of deliberate, conscious mental imagery.

6.1.2. Three Kinds of Mental Simulation – Evidence from Eye Movements

As outlined in the introduction, mental simulation has been defined as "... the re-enactment of perceptual, motor, and introspective states acquired during experience with the world, body, and mind" (Barsalou, 2008; p. 618). In line with this definition, the findings in Chapter 3 and Chapter 4 show that it is indeed possible to distinguish these three distinct kinds of mental simulation (i.e., perceptual simulation, motor simulation, and mental event simulation or mentalizing) at the level of reading behavior. In two experiments in which readers' eye movements were being tracked while they read literary short stories, it was found that reading speed increases when people encounter motoric language in stories, whereas reading speed decreases when people encounter perceptual language. On average, reading speed also decreases when people encounter language describing mental events (i.e., thoughts, emotions), although this is less pronounced than with perceptual language.

6.1.3. Three Kinds of Mental Simulation – Evidence from Neuroimaging

The findings described in section 6.1.2. suggest that motoric language, perceptual language, and language describing mental events are processed in different ways. This could indicate that there are also different brain regions involved in the processing of these three kinds of language. This hypothesis was supported by the neuroimaging findings from Chapter 4. Apart from the abovementioned difference between motor simulation, perceptual simulation and mentalizing in reading behavior (i.e., gaze duration), in the study reported in Chapter 4 I found evidence for the distinction between these three kinds of simulation at the neural level. Different sets of modality-specific brain areas or networks were associated with encountering motoric language, perceptual language, or language describing mental events. Motor descriptions in stories were associated with increased activation in multiple areas that previously have been associated with motor simulation (e.g., Kurby & Zacks, 2013; Moody & Gennari, 2010; Nijhof & Willems, 2015), perceptual descriptions were associated with increased activation in areas associated with perceptual simulation (e.g., Chow et al., 2015; Kanwisher, 2010; Kravitz et al., 2013; Shi et al., 2022), and mental event descriptions were associated with brain areas that have been previously associated with mentalizing (e.g., U. Frith & Frith, 2003; Igelström & Graziano, 2017; Laurita et al., 2017; Paulus et al., 2015; Saxe & Kanwisher, 2003).

Apart from these modality-specific brain areas, all three kinds of simulation were associated with activity in an area in the supramarginal gyrus which is situated in the inferior parietal lobe. This area has been associated with referential indexing: the integration of references into a context (Matchin et al., 2017). Together with the temporoparietal junction (TPJ), the inferior parietal lobe (IPL) is a hub where multiple brain networks come together (i.e., the frontoparietal control network, default mode network, cingulo-opercular network, ventral attention network; see Igelström & Graziano, 2017 for a review). Functionally, the TPJ and IPL have been associated with a variety of tasks, from bottom-up perception tasks (such as automatic, stimulus-driven attention) to higher-order tasks (such as self-perception, mind wandering and social cognition; Igelström & Graziano, 2017). The fact that this area is involved in such a variety of brain networks and functions, implies that it is crucial for the integration and regulation of a multitude of neural processes (Igelström & Graziano, 2017). Therefore, this area may play a role in situation model updating, which consists of the integration of new information from a variety of sources into an existing situation model (Zwaan et al., 1995; Zwaan & Radvansky, 1998). Situation model updating has been suggested to play a role in mental simulation in general (Kurby & Zacks, 2013; Zwaan, 2009), which is another reason why is not be surprising that this area has been found to be associated with all three kinds of simulation in Chapter 4.

6.1.4. Interim Summary

In sum, regarding the initial research question "Do people use mental imagery during story reading?", it can be concluded that readers most probably do not intentionally form a very vivid mental image of the events in the story (cf. the instructions given in Chapter 2: *During reading, try to vividly imagine the events* *happening in the story. Vividly imagine what you see, hear, feel or smell. For example, envision the characters and places described in the story, imagine what the conversations and environmental sounds sound like, what the odors smell like, how the physical experiences of the characters feel*). More probably, what people imagine during story reading is akin to mental simulation, which has been described as an experience that is an outcome of automatic associations between certain types of words and certain neural processes or as a re-experience of perceptual, motor, and introspective events, states or processes (Barsalou, 2008; Shanton & Goldman, 2010). These associations (or re-experiences) are different for different kinds of mental simulation. The difference between the kinds of mental simulation can be measured both at the neural level and on the level of reading behavior. Furthermore, all three kinds of simulation are associated with activation in modality-specific brain areas, indicating that mental simulation is indeed (at least partially) grounded in previous experiences. Differences between the speed of processing of the kinds of simulation (as seen in the eye tracking data) may be explained by differences in the speed of processing in one brain area/network versus the other, or by differences in automaticity (perhaps motor simulation is a much more automatic process than perceptual simulation, which could be the case if motor simulation happened much more often in many more tasks in daily life than perceptual simulation).

An interesting remaining question is how the findings at the neural level and at the level of reading behavior relate to each other. Different kinds of mental simulation were associated with either faster or slower reading, and it would be interesting to find out exactly how these changes in reading behavior relate to (changes in) brain activity. Similarly, an open question remains how individual differences in the strength of the association between simulation and reading behavior relate to brain activity. These are questions that the experiments and analyses reported in this dissertation have not been able to answer, but that deserve further exploration.

Having discussed the first research question, I will now turn to the second research question "How does mental simulation relate to the way readers experience stories?".

6.2. How Does Mental Simulation Relate to the Way Readers Experience Stories?

Chapter 1 discussed that existing literature showed how mental imagery is theoretically related to absorption and appreciation (e.g., Bilandzic & Busselle, 2011; Kuijpers et al., 2014); the remaining question was whether this is similarly the case for mental simulation. To be able to answer this question, I administered absorption and appreciation questionnaires in all of the studies described in this dissertation. From Chapter 2, I can conclude that *inducing* individual differences in mental imagery does not result in individual differences in other reading experiences, such as story world absorption or narrative appreciation.

In Chapter 3, however, I found that *naturally occurring* individual differences in the relationship between simulation and reading behavior are related to absorption and appreciation. An important difference between the studies reported in Chapter 2 and Chapter 3 is that while in Chapter 2 mental imagery was studied using a task in which readers were explicitly asked to use mental imagery during reading, Chapter 3 studied mental simulation without any specific task (i.e., people were told to read as they would normally, and no attempt was made to influence this process or to induce individual differences). This implies that, first of all, mental simulation is not easily influenced from the outside, but rather that differences in the degree of experienced simulation stem from individual differences. These differences are possibly rooted in personal experiences with the world, body and mind (cf. Barsalou, 2008; see section 6.3. for elaboration). Secondly, the finding that naturally occurring individual differences in simulation were associated with absorption and appreciation suggests that there is indeed a relationship between mental simulation and absorption/appreciation.

The strength of the association between simulation and absorption and appreciation is still unclear. In Chapter 4, I tested whether individual differences in simulation at the neural level would be associated with absorption and appreciation as well. Individual differences in simulation at the neural level were operationalized as individual differences in percent signal change in reaction to motor, perceptual and event descriptions in brain areas that were found to be associated with one or all of these kinds of descriptions. For all three kinds of descriptions, there were areas in which the percent signal change as a response to these descriptions was associated with measures of story appreciation (story-specific effects) as well as personality characteristics (trait-based measures). Percent signal change was not associated with story world absorption (state-based measures). Apparently, state-based individual differences do not explain the effects of reading at the neural level, whereas trait-based individual differences do. This is in line with results found in multiple previous studies, where trait-based individual differences (as opposed to state-based individual differences) were more strongly associated with simulation as well (Faber et al., 2020; Hartung, Hagoort, & Willems, 2017; Hartung et al., 2021; Mak et al., 2020; Van den Hoven et al., 2016). Seemingly, simulation is more strongly associated with stable characteristics than with reading experiences such as absorption and appreciation. This does not rule out any association between mental simulation and experiences such as story world absorption, but at least this association seems to be overshadowed by stable characteristics.

Importantly, this question was studied using a selection of literary short stories. An advantage of this approach is that I have gathered a large amount of data from multiple experiments concerning mental simulation and reading experiences using this selection of stories. In my analyses, I have looked at simulation-eliciting descriptions in the stories, and linked those to eye tracking and neuroimaging data. Although this was beyond the scope of this dissertation, future research can link the eye tracking and neuroimaging data to other aspects of the contents of the stories that were read. To this effect, the data collected in the experiments reported in this dissertation have been or will be made openly available (see Appendix I).

Having discussed the second research question, I will now turn to my final research question "How do people differ in mental simulation and other reading experiences?"

6.3. How Do People Differ in Mental Simulation?

From the aforementioned results, an answer to the third question posed in the introduction (How do people differ in mental simulation and in the relationship between mental simulation and reading experiences?) can be distilled. As a starting point, mental simulation was described as an experience that is an outcome of automatic associations between certain types of words and certain neural processes or as a re-experience of perceptual, motor, and introspective events, states or processes (Barsalou, 2008; Shanton & Goldman, 2010). As this would mean that mental simulation is dependent on acquired associations, grounded in personal experiences, I argued it was to be expected that this would also vary between individuals. This is indeed what I found when looking at the relationship between mental simulation and reading behavior in Chapter 3 and Chapter 4. The relationship between mental simulation and reading behavior (motor simulation is associated with faster reading, perceptual and mental event simulation with slower reading), varied between readers. In some readers, this relationship was much more pronounced than in other readers. As mentioned above, these individual differences were related to individual differences in absorption and appreciation. For example, the Attention subcomponent of absorption attenuated the relationship between mental simulation and reading behavior, whereas the Emotional Response component of appreciation strengthened this relationship.

An interesting open question remaining here, is what the meaning is of individual differences in simulation is, in light of theories about the role of simulation in literary reading. There has been debate about whether simulation is *necessary* for language processing (see for an overview of this discussion Meteyard, Cuadrado, Bahrami, & Vigliocco, 2012; Michel, 2021; Ostarek & Huettig, 2019), and in the same vein it could be asked whether simulation is necessary for reading experiences such as absorption or appreciation. In the current dissertation, I have shown that there are individual differences in mental simulation, which suggest that simulation may not be (equally) necessary for reading experiences in all readers. Moreover, it is very well possible that there are readers who do not use mental simulation altogether. The remaining question is therefore not whether simulation is necessary for reading experiences such as absorption and appreciation *overall*, but rather *under what circumstances* or *in which readers* it is necessary. Further investigation of individual differences in mental simulation, to be discovered with eye tracking and fMRI, might help shed light on this question.

Another unanswered question is what these individual differences in simulation imply at the individual level. For example, different groups of simulators have been called “enactors”, “observers”, “verbalizers”, or “mentalizers” (Hartung, Hagoort, & Willems, 2017; Kozhevnikov et al., 2005), but this does not explain what sets these groups of people apart and what causes different simulation styles. As mentioned above, individual variation could be due to differences in personality characteristics or life experiences (cf. Barsalou, 2008; Jacobs, 2015b), but it might also be related to the individual’s tendency to get transported (see Green & Donahue, 2009) and to individual perspective taking capacities (Mar & Oatley, 2008; see Mumper & Gerrig, 2017 for a meta-analysis).

Having discussed the three research questions central to this dissertation, I will now briefly reflect on the implications of my findings for the three theories on mental simulation during literary reading as discussed in Chapter 1 (section 1.1.3.).

6.4. Implications for Theory and Practice

6.4.1. Varieties of Mental Imagery During Literary Reading

The first theory presented in section 1.1.3., is the theory that multiple varieties of mental simulation exist during literary reading (Kuzmičová, 2014). Kuzmičová speaks of varieties of mental simulation in the verbal domain and in the referential domain. The verbal domain contains the two varieties rehearsal-imagery (perceiving the words in stories as if reading them aloud), and speech-imagery (hearing the voices of the characters in your mind, as if witnessing their conversations). The referential domain consists of description-imagery (forming (visual) mental pictures of the story from an observer's perspective), and enactment-imagery (forming mental pictures of the story from the character's perspective, as if acting out the events occurring in the story). The findings from the studies presented in the current dissertation would fit nicely into the categories from the referential domain, where motor simulation would overlap with enactment-imagery, and perceptual simulation would overlap with description imagery. Kuzmičová (2014) further hypothesized that readers constantly switch between simulation varieties, which fits with the findings that in general readers who exhibit large behavioral effects of motor simulation also exhibit large behavioral effects of perceptual simulation (and mentalizing). Finally, the different varieties of simulation are suggested to vary depending on an interplay between text characteristics and internal reader characteristics (Kuzmičová, 2014), which fits nicely with the individual differences results from the studies presented here.

6.4.2. Neurocognitive Poetics Model

The Neurocognitive Poetics Model (NCPM; Jacobs, 2015b) states that there are two routes of literary reading. The first route, the fast route, is provoked by backgrounded elements in stories, such as familiar words and phrases, high frequency words, and highly imageable words. This route evokes fluent reading (an automatic and subconscious reading style) through implicit processing and fiction feelings and is hypothesized to be related to immersive processes during reading. The slow route is provoked by foregrounded elements in stories: for example, metaphors, abstract and defamiliarizing language, rhyme and rhetorical devices. Foregrounded elements are hypothesized to evoke dysfluent reading through explicit processing and aesthetic feeling (Jacobs calls this the aesthetic trajectory). The outcome of dysfluent reading is the aesthetic appreciation of literature and poetry. As hypothesized in Chapter 1, the results for the studies presented in the current dissertation point towards mental simulation being an automatic, subconscious process, thus fitting into the fast route of the NCPM. Interestingly, the NPCM suggests that the fast route should result in immersive processes, such as story world absorption. The evidence for such a connection between mental simulation and absorption is thin, based on the current results. More research is needed to determine if mental simulation is indeed an important part of the fast route of the NCPM, and if so, why this is not as strongly associated with immersive reading experiences as expected.

6.4.3. Simulating Feelings

The third and final theory on mental simulation during literary reading presented in section 1.1.3., is the theory that readers not only simulate actions or perceptions, but also character's feelings (Miall & Kuiken, 2002). Miall and Kuiken (2002) suggest that the simulation of feelings can happen on different levels, depending on the "depth" of these feelings. From shallow to deep, these levels are called evaluative feelings (general enjoyment or reading pleasure), narrative feelings (empathy for and sympathy with characters or responses to specific story events), aesthetic feelings (responses to stylistic elements in stories), and self-modifying feelings (a deep identification with the story and story characters, as a result of both perspective taking and stylistic elements). In terms of the findings from the current studies, it could be argued that narrative feelings, and perhaps self-modifying feelings, are triggered by mental event descriptions, as mental event description are associated with activity in mentalizing areas, and thus involved in social cognitive processes such as Theory of Mind, empathizing and sympathizing (C. Frith & Frith, 1999; U. Frith & Frith, 2003).

A possible association between the levels of evaluative feelings and aesthetic feeling is provided by the findings from Chapter 5. In this chapter, a reanalysis is presented of data from three previous studies in which an appreciation questionnaire was used. In this appreciation questionnaire, readers were asked about their enjoyment of the stories they read, *and* they were asked to rate the stories on a number of adjectives concerning their aesthetic appraisal of the stories (adjectives were derived from Knoop et al., 2016). In the analysis of these data presented in Chapter 5, I found that it was possible to predict enjoyment based on the aesthetic appraisals, but there were also marked individual differences: the specific adjectives (or types of appraisals) that predicted enjoyment, differed from one person to the next. These findings indeed suggest an association between Miall and Kuiken's (2002) levels of evaluative feelings (i.e., enjoyment) and aesthetic feelings, although this association will not look the same in all readers.

6.4.4. Implications for Practice: Education

It has been suggested that mental simulation is a process through which subjective reading experiences (e.g., transportation, enjoyment) could be influenced (e.g., De Koning & Van der Schoot, 2013; Green, 2004). In turn, these reading experiences have been found to influence a range of important reading outcomes. For example, reading enjoyment makes students more inclined to read again (Mol & Jolles, 2014), thus encouraging students to read more. Reading frequency has been found to be positively associated with school success (Chiu & McBride-Chang, 2006; Mol & Jolles, 2014; Retelsdorf et al., 2011), second language learning (Lao & Krashen, 2000; Lee et al., 2015; Yamashita, 2008), and social cognition and empathy (Fong et al., 2013; Johnson et al., 2013; Mar & Oatley, 2008; Oatley, 2016). This makes educators highly motivated to increase reading enjoyment to promote reading in students. To this effect, mental simulation has been specifically promoted in literature education (e.g., De Koning & Van der Schoot, 2013). However, the results from the experiments reported in this dissertation suggest that mental simulation is not easily influenced from the outside, and therefore does not seem to be an optimal candidate for influencing subjective reading experiences or reading outcomes.

6.5. Conclusions

The experiments reported in the current dissertation shed light on the role of mental simulation during literary reading. It was found that when people read literary stories, they can mentally simulate the events and objects in these stories. However, such simulation is usually not in the form of explicit and voluntary mental imagery. Instead, mental simulation during literary reading is an implicit, involuntary process. Different kinds of simulation can be distinguished on the level of eye movements and on the level of brain activity during reading. The three kinds of mental simulation that were distinguished in the current dissertation are motor simulation, perceptual simulation, and mentalizing (in line with the theoretical distinction made by Barsalou, 2008).

Apart from differences between kinds of simulation, the studies in this dissertation also point in the direction of individual differences. Some readers seem to be more prone to mental simulation than others. As the studies reported in this dissertation were all correlational studies, it is impossible to make any causal claims based on these data. However, the results seem to point in a certain direction: in the current dissertation, the aim was to find out whether state-based subjective reading experiences (such as story world absorption) could explain individual differences in mental simulation, but it turned out that throughout the reported studies there are certain trait-based personality characteristics (i.e., perspective taking, transportability) that appear to be more strongly associated with individual differences in simulation than state-based individual differences.

References

Abrams, M., & Harpham, G. (2009). A glossary of literary terms. In *College composition and communication* (9th editio ed., Vol. 8). Boston: Wadsworth Cengage Learning.

Acheson, D., Wells, J., & MacDonald, M. (2008). New and updated tests of print exposure and reading abilities in college students. *Behavior Research Methods*, *40*(1), 278–289. doi: 10.3758/BRM.40.1.278

Altmann, U., Bohrn, I., Lubrich, O., Menninghaus, W., & Jacobs, A. (2014). Fact vs fiction-how paratextual information shapes our reading processes. *Social Cognitive and Affective Neuroscience*, *9*, 22–29. doi: 10.1093/scan/nss098

Andersson, J., Jenkinson, M., & Smith, S. (2007a). *Non-linear optimisation: FMRIB Technial Report TR07JA1* (Tech. Rep.). Oxford, UK: University of Oxford FMRIB Centre.

Andersson, J., Jenkinson, M., & Smith, S. (2007b). *Non-linear registration, aka spatial normalisation: FMRIB Technial Report TR07JA2* (Tech. Rep.). Oxford, UK: University of Oxford FMRIB Centre.

Andrews, K. (2008). It's in your nature: a pluralistic folk psychology. *Synthese*, *165*, 13–29. doi: 10.1007/s11229-007-9230-5

Appel, M., & Richter, T. (2010). Transportation and need for affect in narrative persuasion: A mediated moderation model. *Media Psychology*, *13*(2), 101–135. doi: 10.1080/15213261003799847

Appel, R. (2003). *De invaller.* Retrieved from https://www.nrc.nl/%0Anieuws/2003/06/16/de-invaller-7642950-a1353672

Apperly, I., & Butterfill, S. (2009). Do humans have two systems to track beliefs and belief-like states? *Psychological Review*, *116*(4), 953–970. doi: 10.1037/a0016923

Barsalou, L. (2008). Grounded cognition. *Annual Review of Psychology*, *59*(1), 617–645. doi: 10.1146/annurev.psych.59.103006.093639

Bartsch, A., Vorderer, P., Mangold, R., & Viehoff, R. (2008). Appraisal of emotions in media use: Toward a process model of meta-emotion and emotion regulation. *Media Psychology*, *11*(1), 7–27. doi: 10.1080/

15213260701813447

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. doi: 10.18637/jss.v067.i01

Beckmann, C., Jenkinson, M., & Smith, S. (2003). General multilevel linear modeling for group analysis in fmri. *NeuroImage*, *20*, 1052–1063. doi: 10.1016/S1053-8119(03)00435-X

Bergen, B., & Wheeler, K. (2010). Grammatical aspect and mental simulation. *Brain and Language*, *112*, 150–158. doi: 10.1016/j.bandl.2009.07.002

Biesheuvel, J. (1972). *In de bovenkooi*. Amsterdam, The Netherlands: Meulenhoff.

Bilandzic, H., & Busselle, R. (2011). Enjoyment of films as a function of narrative experience, perceived realism and transportability. *Communications*, *36*, 29–50. doi: 10.1515/COMM.2011.002

Binder, J., Conant, L., Humphries, C., Fernandino, L., Simons, S., Aguilar, M., & Desai, R. (2016). Toward a brain-based componential semantic representation. *Cognitive Neuropsychology*, *33*(3–4), 130–174. doi: 10.1080/02643294.2016.1147426

Bruner, J. (1986). *Actual minds, possible worlds*. Cambridge, Massachusetts: Harvard University Press.

Brysbaert, M., Sui, L., Dirix, N., & Hintz, F. (2020). Dutch author recognition test. *Journal of Cognition*, *3*(1), 1–14. doi: 10.5334/joc.95

Burke, M. (2011). *Literary reading, cognition and emotion: An exploration of the oceanic mind. literary reading, cognition and emotion: An exploration of the oceanic mind*. Oxford, UK: Routledge. doi: 10.4324/9780203840306

Busselle, R., & Bilandzic, H. (2008). Fictionality and perceived realism in experiencing stories: A model of narrative comprehension and engagement. *Communication Theory*, *18*, 255–280. doi: 10.1111/j.1468-2885.2008.00322.x

Busselle, R., & Bilandzic, H. (2009). Measuring narrative engagement. *Media Psychology*, *12*(4), 321–347. doi: 10.1080/15213260903287259

Bürkner, P.-C. (2017). brms : An r package for bayesian multilevel models using stan. *Journal of Statistical Software*, *80*(1). doi: 10.18637/jss.v080.i01

Bürkner, P.-C. (2018). Advanced bayesian multilevel modeling with the r package brms. *The R Journal*, *10*(1), 395. doi: 10.32614/RJ-2018-017

Calvo, M., & Meseguer, E. (2002). Eye movements and processing stages in reading: relative contribution of visual, lexical, and contextual factors. *The Spanish Journal of Psychology*, *5*(1), 66–77. doi: 10.1017/S1138741600005849

Chatterjee, A., & Vartanian, O. (2016). Neuroscience of aesthetics. *Annals of the New York Academy of Sciences*, *1369*, 172–194. doi: 10.1111/nyas.13035

Chiu, M., & McBride-Chang, C. (2006). Gender, context, and reading: A comparison of students in 43 countries. *Scientific Studies of Reading*, *10*(4), 331–362. doi: 10.1207/s1532799xssr1004_1

Chow, H., Mar, R., Xu, Y., Liu, S., Wagage, S., & Braun, A. (2015). Personal experience with narrated events modulates functional connectivity within visual and motor systems during story comprehension. *Human Brain Mapping*, *36*, 1494–1505. doi: 10.1002/hbm.22718

Connell, L., & Lynott, D. (2016). Do we know what we're simulating? information loss on transferring unconscious perceptual simulation to conscious imagery. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(8), 1218–1232. doi: 10.1037/xlm0000245

Csikszentmihalyi, M. (1990). Flow: The psychology of optimal experience. *Journal of Leisure Research*, *24*(1), 93–94. doi: 10.1017/CBO9781107415324.004

Dal Cin, S., Zanna, M., & Fong, G. (2004). Narrative persuasion and overcoming resistance. In E. Knowles & J. Linn (Eds.), *Resistance and persuasion* (p. 175–191). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Davis, M. (1980). A multidimensional approach to individual differences in empathy. *JSAS Catalog of Selected Documents in Psychology*, *10*, 85.

Davis, M. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology*, *44*(1), 113–126. doi: 10.1037/0022-3514.44.1.113

De Corte, K., Buysse, A., Verhofstadt, L., Roeyers, H., Ponnet, K., & Davis, M. (2007). Measuring empathic tendencies: Reliability and validity of the dutch version of the interpersonal reactivity index. *Psychologica Belgica*, *47*(4), 235–260. doi: 10.5334/pb-47-4-235

Deen, B., & McCarthy, G. (2010). Reading about the actions of others: Biological motion imagery and action congruency influence brain activity. *Neuropsychologia*, *48*, 1607–1615. doi: 10.1016/j.neuropsychologia.2010.01.028

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

De Lange, F., Roelofs, K., & Toni, I. (2008). Motor imagery: A window into the mechanisms and alterations of the motor system. *Cortex*, *44*(5), 494–506. doi: 10.1016/j.cortex.2007.09.002

Dijkstra, N., Bosch, S., & Gerven, M. (2017). Vividness of visual imagery depends on the neural overlap with perception in visual areas. *The Journal of Neuroscience*, *37*(5), 1367–1373. doi: 10.1523/JNEUROSCI.3022-16.2016

Dunn, T., Baguley, T., & Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology*, *105*, 399–412. doi: 10.1111/bjop.12046

Eekhof, L., Eerland, A., & Willems, R. (2018). Readers' insensitivity to tense revealed: No differences in mental simulation during reading of present and past tense stories. *Collabra: Psychology*, *4*(1), 16. doi: 10.1525/collabra.121

Eekhof, L., Kuijpers, M., Faber, M., Gao, X., Mak, M., Van den Hoven, E., & Willems, R. (2021). Lost in a story, detached from the words. *Discourse Processes*, *58*(7), 595-616. doi: 10.1080/0163853X.2020.1857619

Emmott, C. (1997). *Narrative comprehension: A discourse perspective*. Oxford University Press.

Engelen, J., Bouwmeester, S., Bruin, A., & Zwaan, R. (2011). Perceptual simulation in developing language comprehension. *Journal of Experimental Child Psychology*, *110*, 659–675. doi: 10.1016/j.jecp.2011.06.009

Estes, Z., Verges, M., & Barsalou, L. (2008). Head up, foot down: Object words orient attention to the objects' typical location. *Psychological Science*, *19*(2), 93–97. doi: 10.1111/j.1467-9280.2008.02051.x

Faber, M., Mak, M., & Willems, R. (2020). Word skipping as an indicator of individual reading style during literary reading. *Journal of Eye Movement Research*, *13*(3), 1–9. doi: 10.16910/jemr.13.3.2

Faridani, V., Jalali, M., & Jahan, M. (2017). Collaborative filtering-based recommender systems by effective trust. *International Journal of Data Science and Analytics*, *3*(4), 297–307. doi: 10.1007/s41060-017-0049-y

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using g*power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*, 1149–1160. doi: 10.3758/BRM.41.4.1149

Fayn, K., Tiliopoulos, N., & MacCann, C. (2015). Interest in truth versus beauty: Intellect and openness reflect different pathways towards interest. *Personality and Individual Differences*, *81*, 47–52. doi: 10.1016/j.paid.2014.12.031

Fechner, G. (1876). *Vorschule der Ästhetik*. Leipzig: Breitkopf & Härtel.

Fernandino, L., Humphries, C., Seidenberg, M., Gross, W., Conant, L., & Binder, J. (2015). Predicting brain activation patterns associated with individual lexical concepts based on five sensory-motor attributes. *Neuropsychologia*, *76*, 17–26. doi: 10.1016/j.neuropsychologia.2015.04.009

Fialho, O. (n.d.). *Transformative reading scale.* (Manuscript in preparation)

Filik, R., & Leuthold, H. (2013). The role of character-based knowledge in online narrative comprehension: Evidence from eye movements and erps. *Brain Research*, *1506*, 94–104. doi: 10.1016/j.brainres.2013.02.017

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

Fong, K., Mullin, J., & Mar, R. (2013). What you read matters: The role of fiction genre in predicting interpersonal sensitivity. *Psychology of Aesthetics, Creativity, and the Arts*, *7*(4), 370–376. doi: 10.1037/a0034084

Foroni, F., & Semin, G. (2013). Comprehension of action negation involves inhibitory simulation. *Frontiers in Human Neuroscience*, *7*, 1–7. doi: 10.3389/fnhum.2013.00209

Frisson, S., Koole, H., Hughes, L., Olson, A., & Wheeldon, L. (2014). Competition between orthographically and phonologically similar words during sentence reading: Evidence from eye movements. *Journal of Memory and Language*, *73*, 148–173. doi: 10.1016/j.jml.2014.03.004

Frith, C., & Frith, U. (1999). Interacting minds - a biological basis. *Science*, *286*(5445), 1692–1695. doi: 10.1126/science.286.5445.1692

Frith, U., & Frith, C. (2003). Development and neurophysiology of mentalizing. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *358*, 459–473. doi: 10.1098/rstb.2002.1218

Gallagher, S. (2015). The new hybrids: Continuing debates on social perception. *Consciousness and Cognition*, *36*, 452–465. doi: 10.1016/j.concog.2015.04.002

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by browne and draper. *Bayesian Analysis*, *1*(3), 515–534. doi: 10.1214/06-BA117A

Gerrig, R. (1993). *Experiencing narrative worlds: On the psychological activities of reading*. Boston: Yale University Press.

Glenberg, A., & Kaschak, M. (2002). Grounding language in action. *Psychonomic Bulletin & Review*, *9*(3), 558–565. doi: 10.3758/BF03196313

Goldman, A. (2006). *Simulating minds: The philosophy, psychology, and neuroscience of mindreading*. Oxford, UK: Oxford University Press. doi: 10.1093/0195138929.001.0001

Goldman, A. (2012). Theory of mind. In E. Margolis, R. Samuels, & S. Stich (Eds.), *The oxford handbook of philosophy of cognitive science.* Oxford University Press. doi: 10.1093/oxfordhb/9780195309799.013.0017

Goodkind, A., & Bicknell, K. (2018). Predictive power of word surprisal for reading times is a linear function of language model quality. In A. Sayeed, C. Jacobs, T. Linzen, & M. Schijndel (Eds.), *Proceedings of the 8th workshop on cognitive modeling and computational linguistics, cmcl 2018* (p. 10–18). Stroudsburg, PA, USA: Association for Computational Linguistics.

Graf, L., & Landwehr, J. (2015). A dual-process perspective on fluency-based aesthetics. *Personality and Social Psychology Review*, *19*(4), 395–410. doi: 10.1177/1088868315574978

Green, M. (2004). Transportation into narrative worlds: The role of prior knowledge and perceived realism. *Discourse Processes*, *38*(2), 247–266.

Green, M., & Brock, T. (2000). The role of transportation in the persuasiveness of public narratives. *Journal of Personality and Social Psychology*, *79*(5), 701–721. doi: 10.1037/0022-3514.79.5.701

Green, M., & Brock, T. (2002). In the mind's eye: Transportation-imagery model of narrative persuasion. In M. Green, J. Strange, & T. Brock (Eds.), *Narrative impact: Social and cognitive foundations* (p. 315–341). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers. doi: 10.1533/9780857095657.2.362

Green, M., Brock, T., & Kaufman, G. (2004). Understanding media enjoyment: The role of transportation into narrative worlds. *Communication Theory*, *14*(4), 311–327. doi: 10.1111/j.1468-2885.2004.tb00317.x

Green, M., & Donahue, J. (2009). Simulated worlds: Transportation into narratives. In K. Markman, W. Klein, & J. Suhr (Eds.), *Handbook of imagination and mental simulation* (p. 241–256). New York: Psychology Press.

Green, M., Kass, S., Carrey, J., Herzig, B., Feeney, R., & Sabini, J. (2008). Transportation across media: Repeated exposure to print and film. *Media Psychology*, *11*(4), 512–539. doi: 10.1080/15213260802492000

Greve, D., & Fischl, B. (2009). Accurate and robust brain image alignment using boundary-based registration. *NeuroImage*, *48*(1), 63–72. doi: 10.1016/j.neuroimage.2009.06.060

Grill-Spector, K., & Weiner, K. (2014). The functional architecture of the ventral temporal cortex and its role in categorization. *Nature Reviews Neuroscience*, *15*(8), 536–548. doi: 10.1038/nrn3747

Hagoort, P. (2019). The neurobiology of language beyond single-word processing. *Science*, *366*, 55–58. doi: 10.1126/science.aax0289

Hale, J. (2001). A probabilistic earley parser as a psycholinguistic model. In *Proceedings of the second meeting of the north american chapter of the association for computational linguistics on language technologies, naacl 2001* (p. 1–8). Stroudsburg, PA, USA: Association for Computational Linguistics. doi: 10.3115/1073336.1073357

Hall, A., & Zwarun, L. (2012). Challenging entertainment: Enjoyment, transportation, and need for cognition in relation to fictional films viewed online. *Mass Communication and Society*, *15*(3), 384–406. doi: 10.1080/15205436.2011.583544

Hanich, J., Wagner, V., Shah, M., Jacobsen, T., & Menninghaus, W. (2014). Why we like to watch sad films. the pleasure of being moved in aesthetic experiences. *Psychology of Aesthetics, Creativity, and the Arts*, *8*(2), 130–143. doi: 10.1037/a0035690

Hartung, F., Burke, M., Hagoort, P., & Willems, R. (2016). Taking perspective: Personal pronouns affect experiential aspects of literary reading. *PLoS ONE*, *11*(5), 1–18. doi: 10.1371/journal.pone.0154732

Hartung, F., Hagoort, P., & Willems, R. (2017). Readers select a comprehension mode independent of pronoun: Evidence from fmri during narrative comprehension. *Brain and Language*, *170*, 29–38. doi: 10.1016/j.bandl.2017.03.007

Hartung, F., Wang, Y., Mak, M., Willems, R., & Chatterjee, A. (2021). Aesthetic appraisals of literary style and emotional intensity in narrative engagement are neurally dissociable. *Communications Biology*, *4*(1), 1401. doi: 10.1038/s42003-021-02926-0

Hartung, F., Withers, P., Hagoort, P., & Willems, R. (2017). When fiction is just as real as fact: No differences in reading behavior between stories believed to be based on true or fictional events. *Frontiers in Psychology*, *8*, 1–14. doi: 10.3389/fpsyg.2017.01618

Hartwigsen, G., Baumgaertner, A., Price, C., Koehnke, M., Ulmer, S., & Siebner, H. (2010). Phonological decisions require both the left and right supramarginal gyri. *Proceedings of the National Academy of Sciences*, *107*(38), 16494–16499. doi: 10.1073/pnas.1008121107

Hasson, U., Egidi, G., Marelli, M., & Willems, R. (2018). Grounding the neurobiology of language in first principles: The necessity of non-language-centric explanations for language comprehension. *Cognition*, *180*, 135–157. doi: 10.1016/j.cognition.2018.06.018

Hauk, O., Johnsrude, I., & Pulvermüller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron*, *41*, 301–307. doi: 10.1016/S0896-6273(03)00838-9

Hauptman, M., Blank, I., & Fedorenko, E. (2022). Non-literal language processing is jointly supported by the language and theory of mind networks: Evidence from a novel meta-analytic fmri approach. *BioRxiv*, *preprint*, 1–65. doi: 10.1101/2022.03.08.481056

Hertrich, I., Dietrich, S., & Ackermann, H. (2020). The margins of the language network in the brain. *Frontiers in Communication*, *5*, 1–26. doi: 10.3389/fcomm.2020.519955

Hsu, C.-T., Conrad, M., & Jacobs, A. (2014). Fiction feelings in harry potter: Haemodynamic response in the mid-cingulate cortex correlates with immersive reading experience. *Neuroreport*, *25*, 1356–1361. doi: 10.1097/WNR.0000000000000272

Igelström, K., & Graziano, M. (2017). The inferior parietal lobule and temporoparietal junction: A network perspective. *Neuropsychologia*, *105*, 70–83. doi: 10.1016/j.neuropsychologia.2017.01.001

Jacobs, A. (2015a). Neurocognitive poetics: Methods and models for investigating the neuronal and cognitive-affective bases of literature reception. *Frontiers in Human Neuroscience*, *9*, 1–22. doi: 10.3389/fnhum.2015.00186

Jacobs, A. (2015b). Towards a neurocognitive poetics model of literary reading. In R. Willems (Ed.), *Cognitive neuroscience of natural language use* (p. 135–159). Cambridge: Cambridge University Press. doi: 10.1017/CBO9781107323667.007

Jacobs, A., Hofmann, M., & Kinder, A. (2016). On elementary affective decisions: To like or not to like, that is the question. *Frontiers in Psychology*, *7*, 1–17. doi: 10.3389/fpsyg.2016.01836

Jacobs, A., & Willems, R. (2018). The fictive brain: Neurocognitive correlates of engagement in literature. *Review of General Psychology*, *22*(2), 147–160. doi: 10.1037/gpr0000106

Jacobsen, T., Buchta, K., Köhler, M., & Schröger, E. (2004). The primacy of beauty in judging the aesthetics of objects. *Psychological Reports*, *94*, 1253–1260. doi: 10.2466/pr0.94.3c.1253-1260

Janssen, T., Braaksma, M., & Couzijn, M. (2009). Self-questioning in the literature classroom: Effects on students' interpretation and appreciation of short stories. *L1 Educational Studies in Language and Literature*, *9*(1), 91–116. doi: 10.17239/L1ESLL-2009.09.01.05

Jeannerod, M. (1994). The representing brain: Neural correlates of motor intention and imagery. *Behavioral and Brain Sciences*, *17*(2), 187–245.

Jeannerod, M. (2001). Neural simulation of action: A unifying mechanism for motor cognition. *NeuroImage*, *14*, 103– 109. doi: 10.1006/nimg.2001.0832

Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *NeuroImage*, *17*(2), 825–841. doi: 10.1016/S1053-8119(02)91132-8

Jenkinson, M., & Smith, S. (2001). A global optimisation method for robust affine registration of brain images. *Medical Image Analysis*, *5*(2), 143–156. doi: 10.1016/S1361-8415(01)00036-6

Jobe, T. (2009). *Power analysis for mixed-effect models in r.* Retrieved from `https://toddjobe.blogspot.com/2009/09/power-analysis-for-mixed-effect-models.html`

Johnson, D., Cushman, G., Borden, L., & McCune, M. (2013). Potentiating empathic growth: Generating imagery while reading fiction increases empathy and prosocial behavior. *Psychology of Aesthetics, Creativity, and the Arts*, *7*(3), 306–312. doi: 10.1037/a0033261

Just, M., & Carpenter, P. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, *87*(4), 329–354. doi: 10.1037/0033-295X.87.4.329

Kanwisher, N. (2010). Functional specificity in the human brain: A window into the functional architecture of the mind. *Proceedings of the National Academy of Sciences*, *107*(25), 11163–11170. doi: 10.1073/pnas.1005062107

Kemmerer, D. (2015). Are the motor features of verb meanings represented in the precentral motor cortices? Yes, but within the context of a flexible, multilevel architecture for conceptual knowledge. *Psychonomic Bulletin & Review*, *22*, 1068–1075. doi: 10.3758/s13423-014-0784-1

Keuleers, E., Brysbaert, M., & New, B. (2010). Subtlex-nl: A new measure for dutch word frequency based on film subtitles. *Behavior Research Methods*, *42*(3), 643–650. doi: 10.3758/BRM.42.3.643

Kliegl, R., Grabner, E., Rolfs, M., & Engbert, R. (2004). Length, frequency, and predictability effects of words on eye movements in reading. *European Journal of Cognitive Psychology*, *16*(1–2), 262–284. doi: 10.1080/09541440340000213

Kliegl, R., & Laubrock, J. (2017). Eye-movement tracking during reading. In A. Groot & P. Hagoort (Eds.), *Research methods in psycholinguistics and the neurobiology of language: A practical guide* (p. 68–88). Hoboken, NJ: Wiley.

Kliegl, R., Nuthmann, A., & Engbert, R. (2006). Tracking the mind during reading: The influence of past, present, and future words on fixation durations. *Journal of Experimental Psychology: General*, *135*(1), 12–35. doi: 10.1037/0096-3445.135.1.12

Knoop, C., Wagner, V., Jacobsen, T., & Menninghaus, W. (2016). Mapping the aesthetic space of literature "from below". *Poetics*, *56*, 35–49. doi: 10.1016/j.poetic.2016.02.001

Koopman, E. (2015). Empathic reactions after reading: The role of genre, personal factors and affective responses. *Poetics*, *50*, 62–79. doi: 10.1016/j.poetic.2015.02.008

Kosslyn, S. (1994). *Image and brain: The resolution of the imagery debate*. Cambridge, Massachusetts: MIT Press.

Kosslyn, S., Ganis, G., & Thompson, W. (2001). Neural foundations of imagery. *Nature Reviews Neuroscience*, *2*, 635–642. doi: 10.1038/35090055

Kovács, A., Kühn, S., Gergely, G., Csibra, G., & Brass, M. (2014). Are all beliefs equal? implicit belief attributions recruiting core brain regions of theory of mind. *PLoS ONE*, *9*(9), 106558. doi: 10.1371/journal.pone.0106558

Kozhevnikov, M., Kosslyn, S., & Shephard, J. (2005). Spatial versus object visualizers: A new characterization of visual cognitive style. *Memory & Cognition*, *33*(4), 710–726. doi: 10.3758/BF03195337

Kraaykamp, G., & Van Eijck, K. (2005). Personality, media preferences, and cultural participation. *Personality and Individual Differences*, *38*, 1675–1688. doi: 10.1016/j.paid.2004.11.002

Kravitz, D., Saleem, K., Baker, C., Ungerleider, L., & Mishkin, M. (2013). The ventral visual pathway: an expanded neural framework for the processing of object quality. *Trends in Cognitive Sciences*, *17*(1), 26–49. doi: 10.1016/j.tics.2012.10.011

Kuijpers, M. (2014). *Absorbing stories: The effects of textual devices on absorption and evaluative responses* (Dissertation). Utrecht University.

Kuijpers, M., Douglas, S., & Kuiken, D. (2018). Personality traits and reading habits that predict absorbed narrative fiction reading. *Psychology of Aesthetics, Creativity, and the Arts*, 1–15. doi: 10.1037/aca0000168

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

Kuiken, D., & Douglas, S. (2017). Forms of absorption that facilitate the aesthetic and explanatory effects of literary reading. In F. Hakemulder, M. Kuijpers, E. Tan, K. Bálint, & M. Doicaru (Eds.), *Narrative absorption* (p. 217–249). Amsterdam/Philadelphia: John Benjamins Publishing Company. doi: 10.1075/lal.27.12kui

Kuiken, D., & Douglas, S. (2018). Living metaphor as the site of bidirectional literary engagement. *Scientific Study of Literature*, *8*(1), 47–76. doi: 10.1075/ssol.18004.kui

Kurby, C., & Zacks, J. (2008). Segmentation in the perception and memory of events. *Trends in Cognitive Sciences*, *12*(2), 72–79. doi: 10.1016/j.tics.2007.11.004

Kurby, C., & Zacks, J. (2013). The activation of modality-specific representations during discourse processing. *Brain and Language*, *126*, 338–349. doi: 10.1016/j.bandl.2013.07.003

Kuzmičová, A. (2012). Presence in the reading of literary narrative: A case for motor enactment. *Semiotica*, *189*(1), 23–48. doi: 10.1515/semi.2011.071

Kuzmičová, A. (2014). Literary narrative and mental imagery: a view from embodied cognition. *Style*, *48*(3), 275–293. doi: 10.5325/style.48.3.275

Kuznetsova, A., Brockhoff, P., & Christensen, R. (2017). lmertest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. doi: 10.18637/jss.v082.i13

Labov, W. (2010). Narratives of personal experience. In P. Hogan (Ed.), *Cambridge encyclopedia of the language sciences* (p. 546–548). Cambridge: Cambridge University Press.

Labov, W., & Waletzky, J. (1967). Narrative analysis: oral versions of personal experience. In J. Helm (Ed.), *Essays on the verbal and visual arts* (p. 12–44). Seattle: University of Washington Press.

Lai, V., Willems, R., & Hagoort, P. (2015). Feel between the lines: Implied emotion in sentence comprehension. *Journal of Cognitive Neuroscience*, *27*(8), 1528–1541. doi: 10.1162/jocn_a_00798

Lammers, J. (2017). *500 à 1000: Hoe de wolven dansen.* Retrieved from http://www.revisor.nl/%0Aentry/2363/hoe-de-wolven-dansen (January 10, 2020)

Lao, C., & Krashen, S. (2000). The impact of popular literature study on literacy development in efl: more evidence for the power of reading. *System*, *28*(2), 261–270. doi: 10.1016/S0346-251X(00)00011-7

Larsen, J., & McGraw, A. (2011). Further evidence for mixed emotions. *Journal of Personality and Social Psychology*, *100*(6), 1095–1110. doi: 10.1037/a0021846

Laurita, A., Hazan, C., & Spreng, R. (2017). Dissociable patterns of brain activity for mentalizing about known others: a role for attachment. *Social Cognitive and Affective Neuroscience*, *12*(7), 1072–1082. doi: 10.1093/scan/nsx040

Leder, H., Belke, B., Oeberst, A., & Augustin, D. (2004). A model of aesthetic appreciation and aesthetic judgments. *British Journal of Psychology*, *95*, 489–508. doi: 10.1348/0007126042369811

Leder, H., Gerger, G., Brieber, D., & Schwarz, N. (2014). What makes an art expert? emotion and evaluation in art appreciation. *Cognition and Emotion*, *28*(6), 1137–1147. doi: 10.1080/02699931.2013.870132

Leder, H., & Nadal, M. (2014). Ten years of a model of aesthetic appreciation and aesthetic judgments : The aesthetic episode - developments and challenges in empirical aesthetics. *British Journal of Psychology*, *105*, 443–464. doi: 10.1111/bjop.12084

Lee, J., Schallert, D., & Kim, E. (2015). Effects of extensive reading and translation activities on grammar knowledge and attitudes for efl adolescents. *System*, *52*, 38–50. doi: 10.1016/j.system.2015.04.016

Lenth, R. V. (2020). emmeans: Estimated marginal means, aka least-squares means [Computer software manual]. Retrieved from https://CRAN.R-project.org/package=emmeans (R package version 1.5.3)

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*, 1126–1177. doi: 10.1016/j.cognition.2007.05.006

Lotze, M., & Halsband, U. (2006). Motor imagery. *Journal of Physiology-Paris*, *99*(4–6), 386–395. doi: 10.1016/j.jphysparis.2006.03.012

Mak, M., De Vries, C., & Willems, R. (2020). The influence of mental imagery instructions and personality characteristics on reading experiences. *Collabra: Psychology*, *6*(1), 43. doi: 10.1525/collabra.281

Mak, M., Faber, M., & Willems, R. (2022). Different routes to liking: How readers arrive at narrative evaluations. *Cognitive Research: Principles and*

Implications, *7*, 72. doi: 10.1186/s41235-022-00419-0

Mak, M., & Willems, R. (2019). Mental simulation during literary reading: Individual differences revealed with eye-tracking. *Language, Cognition and Neuroscience*, *34*(4), 511–535. doi: 10.1080/23273798.2018.1552007

Malanchini, M., Wang, Z., Voronin, I., Schenker, V., Plomin, R., Petrill, S., & Kovas, Y. (2017). Reading self-perceived ability, enjoyment and achievement: A genetically informative study of their reciprocal links over time. *Developmental Psychology*, *53*(4), 698–712. doi: 10.1037/dev0000209

Mar, R., & Oatley, K. (2008). The function of fiction is the abstraction and simulation of social experience. *Perspectives on Psychological Science*, *3*(3), 173–192. doi: 10.1111/j.1745-6924.2008.00073.x

Mar, R., Oatley, K., Hirsh, J., Paz, J., & Peterson, J. (2006). Bookworms versus nerds: Exposure to fiction versus non-fiction, divergent associations with social ability, and the simulation of fictional social worlds. *Journal of Research in Personality*, *40*, 694–712. doi: 10.1016/j.jrp.2005.08.002

Marino, B., Borghi, A., Buccino, G., & Riggio, L. (2017). Chained activation of the motor system during language understanding. *Frontiers in Psychology*, *8*, 1–11. doi: 10.3389/fpsyg.2017.00199

Martín-Loeches, M., Casado, P., Hernández-Tamames, J., & Álvarez Linera, J. (2008). Brain activation in discourse comprehension: A 3t fmri study. *NeuroImage*, *41*, 614–622. doi: 10.1016/j.neuroimage.2008.02.047

Matchin, W., Hammerly, C., & Lau, E. (2017). The role of the ifg and psts in syntactic prediction: Evidence from a parametric study of hierarchical structure in fmri. *Cortex*, *88*, 106–123. doi: 10.1016/j.cortex.2016.12.010

McElreath, R. (2016). *Statistical rethinking*. Boca Raton, Florida: Chapman and Hall/CRC. doi: 10.1201/9781315372495

Menninghaus, W., Wagner, V., Hanich, J., Wassiliwizky, E., Jacobsen, T., & Koelsch, S. (2017). The distancing-embracing model of the enjoyment of negative emotions in art reception. *Behavioral and Brain Sciences*, *40*, 347. doi: 10.1017/S0140525X17000309

Meteyard, L., Cuadrado, S., Bahrami, B., & Vigliocco, G. (2012). Coming of age: A review of embodiment and the neuroscience of semantics. *Cortex*, *48*(7), 788–804. doi: 10.1016/j.cortex.2010.11.002

Miall, D., & Kuiken, D. (1995). Aspects of literary response: A new questionnaire. *Research in the Teaching of English*, *29*(1), 37–58. doi: 10.2307/40171422

Miall, D., & Kuiken, D. (2002). A feeling for fiction: becoming what we behold. *Poetics*, *30*, 221–241. doi: 10.1016/S0304-422X(02)00011-6

Michel, C. (2021). Overcoming the modal/amodal dichotomy of concepts. *Phenomenology and the Cognitive Sciences*, *20*(4), 655–677. doi: 10.1007/s11097-020-09678-y

Mitchell, D. (1984). An evaluation of subject-paced reading tasks and other methods for investigating immediate processes in reading. In D. Kieras & M. Just (Eds.), *New methods in reading comprehension research* (p. 69–89). Hillsdale, N.J: Erlbaum.

Mol, S., & Jolles, J. (2014). Reading enjoyment amongst non-leisure readers can affect achievement in secondary school. *Frontiers in Psychology*, *5*, 1–10. doi: 10.3389/fpsyg.2014.01214

Moody, C., & Gennari, S. (2010). Effects of implied physical effort in sensory-motor and pre-frontal cortex during language comprehension. *NeuroImage*, *49*(1), 782–793. doi: 10.1016/j.neuroimage.2009.07.065

Mumper, M., & Gerrig, R. (2017). Leisure reading and social cognition: A meta-analysis. *Psychology of Aesthetics, Creativity, and the Arts*, *11*(1), 109–120. doi: 10.1037/aca0000089

Nabokov, V. (2003). Signalen en symbolen. In *Een russische schoonheid 1.* Amsterdam: De Bezige Bij.

Nalborczyk, L., Batailler, C., Lœvenbruck, H., Vilain, A., & Bürkner, P.-C. (2019). An introduction to bayesian multilevel models using brms: A case study of gender effects on vowel variability in standard indonesian. *Journal of Speech, Language, and Hearing Research*, *62*(5), 1225–1242. doi: 10.1044/2018_JSLHR-S-18-0006

Nijhof, A., Brass, M., Bardi, L., & Wiersema, J. (2016). Measuring mentalizing ability: A within-subject comparison between an explicit and implicit version of a ball detection task. *PLoS ONE*, *11*(10), 1–15. doi: 10.1371/journal.pone.0164373

Nijhof, A., & Willems, R. (2015). Simulating fiction: Individual differences in literature comprehension revealed with fmri. *PLoS ONE*, *10*(2), 1–17. doi: 10.1371/journal.pone.0116492

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

Oatley, K. (2012). The cognitive science of fiction. *Wiley Interdisciplinary Reviews: Cognitive Science*, *3*, 425–430. doi: 10.1002/wcs.1185

Oatley, K. (2016). Fiction: Simulation of social worlds. *Trends in Cognitive Sciences*, *20*(8), 618–628. doi: 10.1016/j.tics.2016.06.002

Oceja, L., & Carrera, P. (2009). Beyond a single pattern of mixed emotional experience: Sequential, prevalence, inverse, and simultaneous. *European Journal of Psychological Assessment*, *25*(1), 58–67. doi: 10.1027/1015-5759.25.1.58

Ostarek, M., & Huettig, F. (2019). Six challenges for embodiment research. *Current Directions in Psychological Science*, *28*(6), 593–599. doi: 10.1177/0963721419866441

Ostarek, M., & Vigliocco, G. (2017). Reading sky and seeing a cloud: On the relevance of events for perceptual simulation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(4), 579–590. doi: 10.1037/xlm0000318

Panero, M., Weisberg, D., Black, J., Goldstein, T., Barnes, J., Brownell, H., & Winner, E. (2016). Does reading a single passage of literary fiction really improve theory of mind? An attempt at replication. *Journal of Personality and Social Psychology*, *111*(5), 46– 54. doi: 10.1037/pspa0000064

Parsons, L., Fox, P., Downs, J., Glass, T., Hirsch, T., Martin, C., & Lancaster, J. (1995). Use of implicit motor imagery for visual shape discrimination as revealed by pet. *Nature*, *375*, 54–58.

Paulus, F., Müller-Pinzler, L., Jansen, A., Gazzola, V., & Krach, S. (2015). Mentalizing and the role of the posterior superior temporal sulcus in sharing others' embarrassment. *Cerebral Cortex*, *25*(8), 2065–2075. doi: 10.1093/cercor/bhu011

Peters, G. (2014). The alpha and the omega of scale reliability and validity: Why and how to abandon cronbach's alpha and the route towards more comprehensive assessment of scale quality. *European Health Psychologist*, *16*(2), 56–69.

Pollatsek, A., Juhasz, B., Reichle, E., Machacek, D., & Rayner, K. (2008). Immediate and delayed effects of word frequency and word length on eye movements in reading: A reversed delayed effect of word length. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(3), 726–750. doi: 10.1037/0096-1523.34.3.726

Power, J., Barnes, K., Snyder, A., Schlaggar, B., & Petersen, S. (2012). Spurious but systematic correlations in functional connectivity mri networks arise from subject motion. *NeuroImage*, *59*(3), 2142–2154. doi: 10.1016/j.neuroimage.2011.10.018

Pylyshyn, Z. (2003). *Seeing and visualizing: It's not what you think*. Cambridge, Massachusetts: MIT Press.

R Core Team. (2021). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from https://www.R-project.org/

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*(3), 372–422. doi: 10.1037/0033-2909.124.3.372

Rayner, K., Ashby, J., Pollatsek, A., & Reichle, E. (2004). The effects of frequency and predictability on eye fixations in reading: Implications for the e-z reader model. *Journal of Experimental Psychology: Human Perception and Performance*, *30*(4), 720–732. doi: 10.1037/0096-1523.30.4.720

Rayner, K., & Fischer, M. (1996). Mindless reading revisited: eye movements during reading and scanning are different. *Perception & Psychophysics*, *58*(5), 734–747. doi: 10.3758/BF03213106

Rayner, K., Sereno, S., & Raney, G. (1996). Eye movement control in reading: A comparison of two types of models. *Journal of Experimental Psychology: Human Perception and Performance*, *22*(5), 1188–1200. doi: 10.1037/0096-1523.22.5.1188

Reichle, E., & Drieghe, D. (2015). Using e-z reader to examine the consequences of fixation-location measurement error. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(1), 262–270. doi: 10.1037/a0037090

Reichle, E., Pollatsek, A., Fisher, D., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review*, *105*(1), 125–157. doi: 10.1037/0033-295X.105.1.125

Reichle, E., Reineberg, A., & Schooler, J. (2010). Eye movements during mindless reading. *Psychological Science*, *21*(9), 1300–1310. doi: 10.1177/0956797610378686

Reichle, E., Warren, T., & McConnell, K. (2009). Using e-z reader to model the effects of higher level language processing on eye movements during reading. *Psychonomic Bulletin and Review*, *16*(1), 1–21. doi: 10.3758/PBR.16.1.1

Retelsdorf, J., Köller, O., & Möller, J. (2011). On the effects of motivation on reading performance growth in secondary school. *Learning and Instruction*, *21*, 550–559. doi: 10.1016/j.learninstruc.2010.11.001

Revelle, W. (2014). Classical test theory and the measurement of reliability. In

An introduction to psychometric theory with applications in r (p. 205–239). Retrieved from http://www.personality-project.org/r/book/

Revelle, W. (2020). psych: Procedures for psychological, psychometric, and personality research [Computer software manual]. Evanston, Illinois. Retrieved from https://CRAN.R-project.org/package=psych (R package version 2.0.12)

Richlan, F., Gagl, B., Hawelka, S., Braun, M., Schurz, M., Kronbichler, M., & Hutzler, F. (2014). Fixation-related fmri analysis in the domain of reading research: Using self-paced eye movements as markers for hemodynamic brain responses during visual letter string processing. *Cerebral Cortex*, *24*, 2647–2656. doi: 10.1093/cercor/bht117

Riddell, A., & Van Dalen-Oskam, K. (2018). Readers and their roles: Evidence from readers of contemporary fiction in the netherlands. *PLOS ONE*, *13*(7), 0201157. doi: 10.1371/journal.pone.0201157

Ryan, M. (2001). *Narrative as virtual reality: Immersion and interactivity in literature and electronic media*. Baltimore/London: Johns Hopkins University Press.

Samur, D., Lai, V., Hagoort, P., & Willems, R. (2015). Emotional context modulates embodied metaphor comprehension. *Neuropsychologia*, *78*, 108–114. doi: 10.1016/j.neuropsychologia.2015.10.003

Samur, D., Tops, M., & Koole, S. (2018). Does a single session of reading literary fiction prime enhanced mentalising performance? Four replication experiments of Kidd and Castano (2013). *Cognition and Emotion*, *32*(1), 130–144. doi: 10.1080/02699931.2017.1279591

Sanders, J., & Van Krieken, K. (2018). Exploring narrative structure and hero enactment in brand stories. *Frontiers in Psychology*, *9*, 1–17. doi: 10.3389/fpsyg.2018.01645

Saxe, R., & Kanwisher, N. (2003). People thinking about thinking people: The role of the temporo-parietal junction in “theory of mind. *NeuroImage*, *19*(4), 1835–1842. doi: 10.1016/S1053-8119(03)00230-1

Schimmack, U. (2001). Pleasure, displeasure, and mixed feelings: Are semantic opposites mutually exclusive? *Cognition and Emotion*, *15*(1), 81–97. doi: 10.1080/0269993004200123

Schindler, I., Hosoya, G., Menninghaus, W., Beermann, U., Wagner, V., Eid, M., & Scherer, K. (2017). Measuring aesthetic emotions: A review of the literature and a new assessment tool. *PLOS ONE*, *12*(6), 0178899. doi: 10.1371/journal.pone.0178899

Schroyens, W., Vitu, F., Brysbaert, M., & D'Ydewalle, G. (1999). Eye movement control during reading: Foveal load and parafoveal processing. *Quarterly Journal of Experimental Psychology*, *52*(4), 1021–1046. doi: 10.1080/713755859

Schutte, N., & Malouff, J. (2004). University student reading preferences in relation to the big five personality dimensions. *Reading Psychology*, *25*, 273–295. doi: 10.1080/02702710490522630

Sereno, S., & Rayner, K. (2003). Measuring word recognition in reading: eye movements and event-related potentials. *Trends in Cognitive Sciences*, *7*(11), 489–493. doi: 10.1016/j.tics.2003.09.010

Shanton, K., & Goldman, A. (2010). Simulation theory. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*, 527–538. doi: 10.1002/wcs.33

Shi, J., Huang, H., Jiang, R., Mao, X., Huang, Q., & Li, A. (2022). The right inferior frontal gyrus plays an important role in unconscious information processing: Activation likelihood estimation analysis based on functional magnetic resonance imaging. *Frontiers in Neuroscience*, *16*, 1–11. doi: 10.3389/fnins.2022.781099

Smirnov, D., Glerean, E., Lahnakoski, J., Salmi, J., Jääskeläinen, I., Sams, M., & Nummenmaa, L. (2014). Fronto-parietal network supports context-dependent speech comprehension. *Neuropsychologia*, *63*, 293–303. doi: 10.1016/j.neuropsychologia.2014.09.007

Smith, S. (2002). Fast robust automated brain extraction. *Human Brain Mapping*, *17*(3), 143–155. doi: 10.1002/hbm.10062

Smith, S., & Brady, J. (1997). Susan—a new approach to low level image processing. *International Journal of Computer Vision*, *23*, 45–78. doi: 10.1023/A:1007963824710

Speed, L., & Vigliocco, G. (2014). Eye movements reveal the dynamic simulation of speed in language. *Cognitive Science*, *38*, 367–382. doi: 10.1111/cogs.12096

Speer, N., Reynolds, J., Swallow, K., & Zacks, J. (2009). Reading stories activates neural representations of visual and motor experiences. *Psychological Science*, *20*(8), 989–999. doi: 10.1111/j.1467-9280.2009.02397.x

Speer, N., Zacks, J., & Reynolds, J. (2007). Human brain activity time-locked to narrative event boundaries. *Psychological Science*, *18*(5), 449–455. doi: 10.1111/j.1467-9280.2007.01920.x

Stan Development Team. (2020). Rstan: the r interface to stan [Computer software manual]. Retrieved from `http://mc-stan.org`

Stanfield, R., & Zwaan, R. (2001). The effect of implied orientation derived from verbal context on picture recognition. *Psychological Science*, *12*(2), 153–156. doi: 10.1111/1467-9280.00326

Stanovich, K., & West, R. (1989). Exposure to print and orthographic processing. *Reading Research Quarterly*, *24*(4), 402–433. doi: 10.2307/747605

Stich, S., & Nichols, S. (1992). Folk psychology: Simulation or tacit theory? *Mind & Language*, *7*(1–2), 35–71. doi: 10.1111/j.1468-0017.1992.tb00196.x

Stoeckel, C., Gough, P., Watkins, K., & Devlin, J. (2009). Supramarginal gyrus involvement in visual word recognition. *Cortex*, *45*(9), 1091–1096. doi: 10.1016/j.cortex.2008.12.004

Tamir, D., Bricker, A., Dodell-Feder, D., & Mitchell, J. (2016). Reading fiction and reading minds: the role of simulation in the default network. *Social Cognitive and Affective Neuroscience*, *11*(2), 215–224. doi: 10.1093/scan/nsv114

Tettamanti, M., Buccino, G., Saccuman, M., Gallese, V., Danna, M., Scifo, P., & Perani, D. (2005). Listening to action-related sentences activates fronto-parietal motor circuits. *Journal of Cognitive Neuroscience*, *17*(2), 273–281. doi: 10.1162/0898929053124965

Toolan, M. (2016). *Making sense of narrative text: Situation, repetition, and picturing in the reading of short stories.* London/New York.

Troscianko, E. (2013). Reading imaginatively: The imagination in cognitive science and cognitive literary studies. *Journal of Literary Semantics*, *42*(2), 181–198. doi: 10.1515/jls-2013-0009

Tukachinsky, R. (2014). Experimental manipulation of psychological involvement with media. *Communication Methods and Measures*, *8*(1), 1–33. doi: 10.1080/19312458.2013.873777

Van den Hoven, E., Hartung, F., Burke, M., & Willems, R. (2016). Individual differences in sensitivity to style during literary reading: Insights from eye-tracking. *Collabra*, *2*(1), 25. doi: 10.1525/collabra.39

Van Dis, A. (1986). *Casablanca*. Amsterdam, The Netherlands: Meulenhoff.

Van Eeden, E. (2015). *Ze is overal.* Amsterdam. (Stichting Collectieve Propaganda van het Nederlandse Boek.)

Van Essen, R. (2014). De mensen die alles lieten bezorgen. In *Hier wonen ook mensen* (p. 113–123). Utrecht: Atlas Contact.

Van Hassel, S. (2012). De chinese bruiloft. In *Ezels.* Amsterdam: De Bezige Bij.

Van Krieken, K., Hoeken, H., & Sanders, J. (2017). Evoking and measuring identification with narrative characters - a linguistic cues framework. *Frontiers in Psychology*, *8*, 1–16. doi: 10.3389/fpsyg.2017.01190

Van Overwalle, F., & Vandekerckhove, M. (2013). Implicit and explicit social mentalizing: Dual processes driven by a shared neural network. *Frontiers in Human Neuroscience*, *7*, 1–6. doi: 10.3389/fnhum.2013.00560

Wallentin, M., Nielsen, A., Vuust, P., Dohn, A., Roepstorff, A., & Lund, T. (2011). Amygdala and heart rate variability responses from listening to emotionally intense parts of a story. *NeuroImage*, *58*, 963–973. doi: 10.1016/j.neuroimage.2011.06.077

Wallentin, M., Simonsen, A., & Nielsen, A. (2013). Action speaks louder than words: Empathy mainly modulates emotions from theory of mind-laden parts of a story. *Scientific Study of Literature*, *3*(1), 137–153. doi: 10.1075/ssol.3.1.11wal

Weibel, D., Wissmath, B., & Mast, F. (2011). Influence of mental imagery on spatial presence and enjoyment assessed in different types of media. *Cyberpsychology, Behavior, and Social Networking*, *14*(10), 607–612. doi: 10.1089/cyber.2010.0287

Willems, R., & Casasanto, D. (2011). Flexibility in embodied language understanding. *Frontiers in Psychology*, *2*, 1–11. doi: 10.3389/fpsyg.2011.00116

Willems, R., & Cristia, A. (2017). Hemodynamic methods : fmri and fnirs. In A. Groot & P. Hagoort (Eds.), *Research methods in psycholinguistics and the neurobiology of language : A practical guide* (p. 266–287). Oxford: Wiley-Blackwell.

Willems, R., & Francken, J. (2012). Embodied cognition: Taking the next step. *Frontiers in Psychology*, *3*, 1–3. doi: 10.3389/fpsyg.2012.00582

Willems, R., & Jacobs, A. (2016). Caring about dostoyevsky: The untapped potential of studying literature. *Trends in Cognitive Sciences*, *20*(4), 243–245. doi: 10.1016/j.tics.2015.12.009

Willems, R., Labruna, L., D'Esposito, M., Ivry, R., & Casasanto, D. (2011). A functional role for the motor system in language understanding: Evidence from theta-burst transcranial magnetic stimulation. *Psychological Science*, *22*(7), 849–854. doi: 10.1177/0956797611412387

Willems, R., Toni, I., Hagoort, P., & Casasanto, D. (2010). Neural dissociations between action verb understanding and motor imagery. *Journal of Cognitive Neuroscience*, *22*(10), 2387–2400. doi: 10.1162/jocn.2009.21386

Willems, R., & Van Gerven, M. (2018). New fmri methods for the study of language. In S.-A. Rueschemeyer & M. Gaskell (Eds.), *The oxford handbook of psycholinguistics* (p. 974–991). Oxford, New York: Oxford University Press. doi: 10.1093/oxfordhb/9780198786825.013.42

Wiltshire, T., Lobato, E., McConnell, D., & Fiore, S. (2015). Prospects for direct social perception: a multi-theoretical integration to further the science of social cognition. *Frontiers in Human Neuroscience*, *8*, 1–22. doi: 10.3389/fnhum.2014.01007

Woolrich, M. (2008). Robust group analysis using outlier inference. *NeuroImage*, *41*(2), 286–301. doi: 10.1016/j.neuroimage.2008.02.042

Woolrich, M., Behrens, T., Beckmann, C., Jenkinson, M., & Smith, S. (2004). Multilevel linear modelling for fmri group analysis using bayesian inference. *NeuroImage*, *21*(4), 1732–1747. doi: 10.1016/j.neuroimage.2003.12.023

Woolrich, M., Jbabdi, S., Patenaude, B., Chappell, M., Makni, S., Behrens, T., & Smith, S. (2009). Bayesian analysis of neuroimaging data in fsl. *NeuroImage*, *45*, 173– 186. doi: 10.1016/j.neuroimage.2008.10.055

Woolrich, M., Ripley, B., Brady, M., & Smith, S. (2001). Temporal autocorrelation in univariate linear modeling of fmri data. *NeuroImage*, *14*(6), 1370–1386. doi: 10.1006/nimg.2001.0931

Worsley, K. (2001). Statistical analysis of activation images. In P. Jezzard, P. Matthews, & S. Smith (Eds.), *Functional magnetic resonance imaging* (p. 251–270). Oxford University Press. doi: 10.1093/acprof:oso/9780192630711.003.0014

Yamashita, J. (2008). Extensive reading and development of different aspects of l2 proficiency. *System*, *36*(4), 661–672. doi: 10.1016/j.system.2008.04.003

Zillmann, D. (1988). Mood management through communication choices. *American Behavioral Scientist*, *31*(3), 327–340. doi: 10.1177/000276488031003005

Zwaan, R. (2009). Mental simulation in language comprehension and social cognition. *European Journal of Social Psychology*, *39*, 1142–1150. doi: 10.1002/ejsp.661

Zwaan, R., Langston, M., & Graesser, A. (1995). The construciton of situation models in narrative comprehension: An event-indexing model. *Psychological Science*, *6*(5), 292–297.

Zwaan, R., & Radvansky, G. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, *123*(2), 162–185. doi:

10.1037/0033-2909.123.2.162

Zwaan, R., Stanfield, R., & Yaxley, R. (2002). Language comprehenders mentally represent the shapes of objects. *Psychological Science*, *13*(2), 168–171. doi: 10.1111/1467-9280.00430

A Number of Times Words Were Underlined, Depending on Sentence Position (Chapter 3)

Figure A.1.: *Number of Times Words Were Underlined for Being Part of a Motor Description, Depending on Sentence Position*

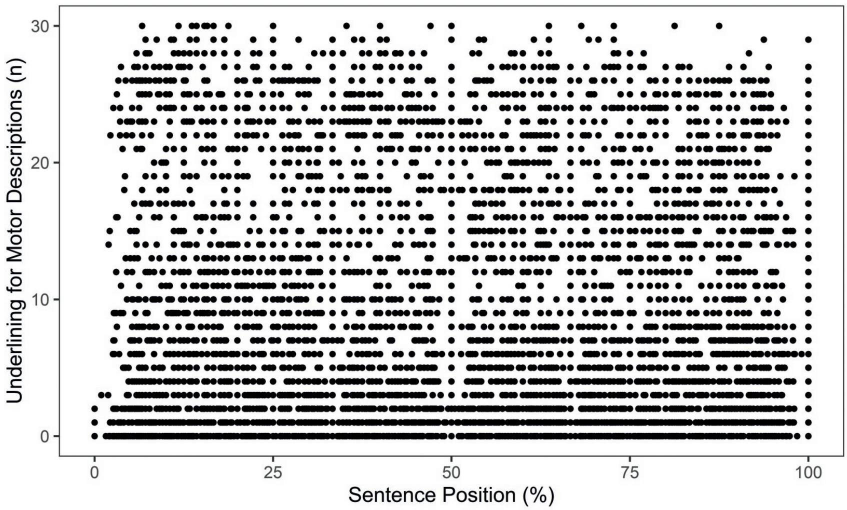


Figure A.2.: *Number of Times Words Were Underlined for Being Part of a Perceptual Description, Depending on Sentence Position*

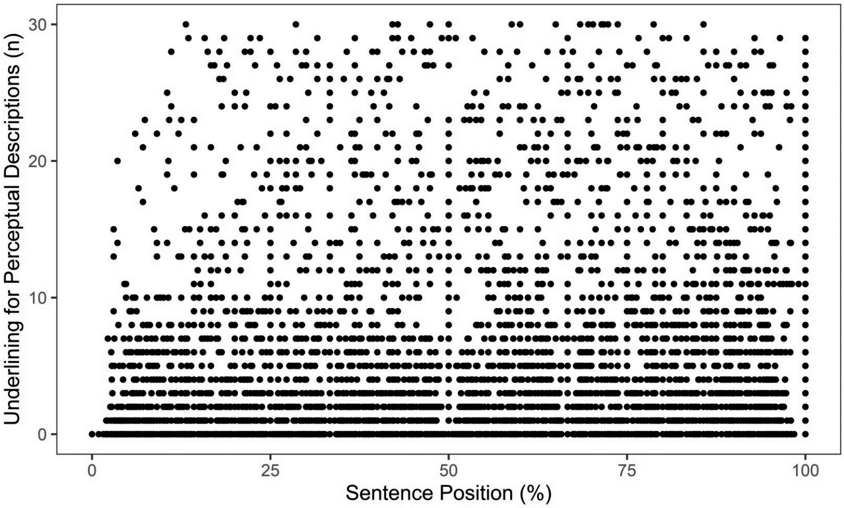
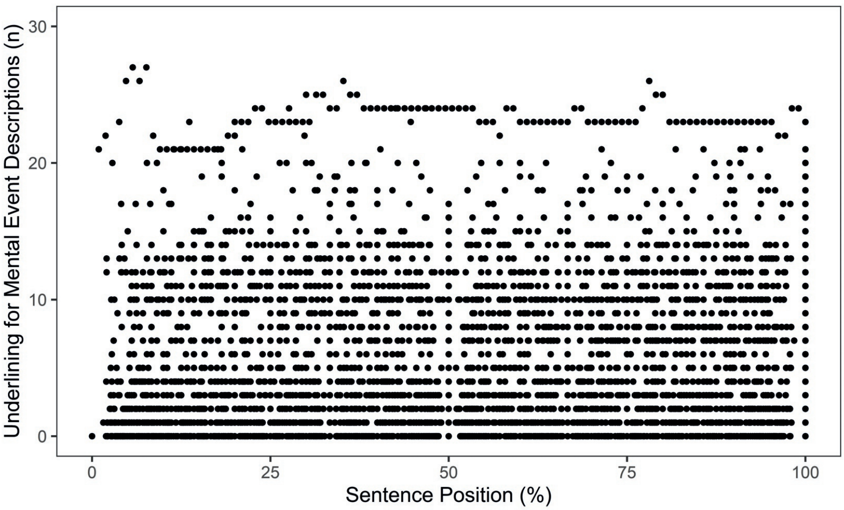


Figure A.3.: *Number of Times Words Were Underlined for Being Part of a Mental Event Description, Depending on Sentence Position*



B | Questionnaires (Chapter 3)

B.1. Items Added to the Story World Absorption Scale, and the Subscale They Were Added to

Item	SWAS subscale
I feel like I really got to know the characters in the story	Emotional Engagement
I could see the events in the story happening as if I could see through the eyes of the main character	Mental Imagery
While reading the story, I could see the events that happened in the story before my mind's eye as if I saw a movie	Mental Imagery
I thought the events in the story were described vividly	Mental Imagery
While reading the story, I could see the actions of the characters before my eye	Mental Imagery
I could easily depicture the characters in the story	Mental Imagery

B.2. Story World Absorption Scale (Including Additional Items)

Indicate to what degree you agree with the following questions

1. I felt sympathy for the main character
2. When I was reading the story it sometimes seemed as if I were in the story world too
3. I saw what happened in the story through the eyes of the main character
4. I felt for what happened in the story

5. When I was reading the story I was focused on what happened in the story
6. When I was reading the story I had an image of the main character in mind
7. The world of the story sometimes felt closer to me than the world around me
8. When I was reading the story I could see the situations happening in the story as a movie being played inside my mind
9. I felt absorbed in the story
10. When I was reading the story I could see the situations happening in the story being played out before my eyes
11. I thought the situations in the story were described picturesque
12. The story gripped me in such a way that I could close myself off for things that were happening around me
13. When reading the story there were moments in which I felt that the story world overlapped with my own world
14. When I was reading the story I could see the actions of the characters before my eyes
15. When I finished the story I was surprised to see that time had gone by so fast
16. Because all of my attention went into the story, I sometimes felt as if I could not exist separate from the story
17. I could imagine the characters in the story very well
18. I felt how the main character was feeling
19. I could imagine what the world in which the story took place looked like
20. When I was finished with reading the story it felt like I had taken a trip to the world of the story
21. I felt connected to the main character in the story
22. I feel like I really got to know the characters in the story

23. I was reading in such a concentrated manner that I had forgotten the world around me
24. When I read the story I could imagine what it must be like to be in the shoes of the main character

B.3. Appreciation Questionnaire

1. How did you like the story?
2. Did you find the story
    a) Beautiful
    b) Boring
    c) Gripping
    d) Entertaining
    e) Funny
    f) Interesting
    g) Ominous
    h) Sad
    i) Suspenseful
    j) Tragic
    k) Witty
    l) Captivating
    m) Special
3. Indicate to what degree you agree with the following questions
    a) I was constantly curious about how the story would end
    b) I thought it was fun to read this story
    c) I want to read the story again sometimes
    d) I thought the story was written well
    e) I rather did not want the story to end
    f) I would recommend this story to somebody else

B.4. Comprehension Check

Story 1: De mensen die alles lieten bezorgen (The people who had everything delivered)

What was striking about the flat where Jeffrey and Rita stayed?

1. There always sounded loud music
2. The kitchen window was always open
3. They never closed the curtains
4. There was a scooter in front of the door

How could Jeffrey and Rita stay inside for their entire vacation, without having a shortage in food or groceries?

1. They brought a lot of groceries in their large suitcases
2. The main character bought groceries for them when he went to the Albert Heijn (supermarket)
3. The neighbours left a lot of supplies
4. Their food and groceries were delivered to their apartment

Why did Jeffrey and Rita leave the flat?

1. Their holiday was over
2. There was a fire in the flat
3. Rita had turned ill
4. They were ejected by the police

Story 2: De Chinese bruiloft (The Chinese wedding)

What did Geert have to give Jing Jing before he could marry her?

1. A kitten
2. Money

3. Roses

4. Jewellery

Where did the marriage of Geert and Jing Jing take place?

1. At city hall

2. In a church

3. At the venue of the party

4. On the beach

Where did the main character take Jing Jing's parents?

1. To a mall

2. To the old port of Rotterdam

3. To the hairdresser's

4. To the airport

Story 3: Signalen en Symbolen (Signs and Symbols)

What kind of present had the parents bought for their son's birthday?

1. Fluffy new slippers for cold, rainy days

2. A coffeemaker to make 12 cups of coffee

3. A typewriter to write letters home

4. A basket with ten different fruit jellies in ten little jars

What was the weather like when the parents went to visit their son?

1. It was raining cats and dogs

2. It was freezing cold

3. It was a sunny spring day

4. It was very warm and humid

Why did the father stagger into the living room, when he came out of bed at night?

1. He woke up because he had a headache
2. He heard the telephone ring and wanted to answer it
3. He wanted to get his son out of the clinic, to live at home with his parents
4. He was looking for his dental plates but could not find them

B.5. Story Ranking Questionnaire

1. Rank the stories, with 1 being the story you liked best, and 3 being the story you liked the least.
   A. De mensen die alles lieten bezorgen (The people who had everything delivered)
   B. De Chinese bruiloft (The Chinese wedding)
   C. Signalen en symbolen (Signs and symbols)
2. Did you already know (one of) these stories, previous to the experiment?

B.6. Reading Habits Questionnaire (Direct Measure of Reading Experience)

1. How often do you read fiction (e.g., novels, fairy tales, poetry)?
   - Daily
   - More than two days a week, but not daily
   - One day per week
   - Hardly ever
   - Never
2. How often do you read non-fiction (e.g., newspapers, news websites, magazines)?
   - Daily
   - More than two days a week, but not daily

- One day per week
- Hardly ever
- Never

3. How often do you consume fiction in another way (for example in movies, television series, videogames)?
    - Daily
    - More than two days a week, but not daily
    - One day per week
    - Hardly ever
    - Never
4. How much do you like reading fiction?
    - Very much
    - Quite a lot
    - I do not like it, but I do not dislike it either
    - Not very much
    - Totally not
5. How many books do you read each year?
    - More than 15
    - About 1 each month
    - 1 or 2
    - I do not read books
6. What genre of fiction do you prefer? Choose the options (at least 1, at most 3) that best reflect your preference.
    - Science fiction
    - Fantasy
    - Thriller
    - Literary fiction
    - Historical fiction
    - Chicklit
    - Other

B.7. Interpersonal Reactivity Index (Fantasy and Perspective Taking Subscales)

Indicate to what degree you agree with the following questions

1. I daydream and fantasize, with some regularity, about things that might happen to me.
2. I sometimes find it difficult to see things from the “other guy’s” point of view.
3. I really get involved with the feelings of the characters in a novel.
4. I am usually objective when I watch a movie or play, and I don’t often get completely caught up in it.
5. I try to look at everybody’s side of a disagreement before I make a decision.
6. I sometimes try to understand my friends better by imagining how things look from their perspective.
7. Becoming extremely involved in a good book or movie is somewhat rare for me.
8. If I’m sure I’m right about something, I don’t waste much time listening to other people’s arguments.
9. After seeing a play or movie, I have felt as though I were one of the characters.
10. I believe that there are two sides to every question and try to look at them both.
11. When I watch a good movie, I can very easily put myself in the place of a leading character.
12. When I’m upset at someone, I usually try to “put myself in his shoes” for a while.
13. When I am reading an interesting story or novel, I imagine how I would feel if the events in the story were happening to me.
14. Before criticizing somebody, I try to imagine how I would feel if I were in their place.

B.8. Author Recognition Test (Instruction)

Below are a couple of names. Some are names (or pseudonyms) of fiction writers, some are fake names. **Indicate which names are familiar as names of writers, by underlining them**. If you underline a fake name, this will count as a negative score. So don't guess the answer, **but underline only the names you recognize with certainty**. You don't need to have read any books by the authors.

C | Model Summaries for the Full Models for Gaze Duration (Spill-Over) and Number of Regressions (Chapter 3)

Summaries for both the *lmer* and *lmerTest* models are given for each of the dependent variables.

Appendix C-1: Analysis Gaze Duration

lmer

```
Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: gazdur ~ corpfreq_p_c + nchar_p_c + surprisal_p_c +
mot_ul_total_p_c +
    sens_ul_total_p_c + emo_ul_total_p_c + (1 + mot_ul_total_p_c +
    sens_ul_total_p_c + emo_ul_total_p_c | subject/Story)
   Data: rtdata2[rtdata2$wrdno_slide > 1, ]
Control: lmerControl(optimizer = "nloptwrap", calc.derivs = FALSE)

     AIC      BIC   logLik deviance df.resid
 5938637  5938946 -2969290  5938581   466552

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.0857 -0.5674 -0.1995  0.3134 23.0000

Random effects:
 Groups        Name              Variance  Std.Dev.  Corr
 Story:subject (Intercept)       6.284e+04 250.67803
               mot_ul_total_p_c  9.118e-01   0.95487 -1.00
               sens_ul_total_p_c 8.250e+00   2.87223  0.99 -0.99
               emo_ul_total_p_c  5.755e+00   2.39904  0.95 -0.95  0.94
 subject       (Intercept)       4.626e+00   2.15082
               mot_ul_total_p_c  3.534e-02   0.18800  0.99
               sens_ul_total_p_c 2.016e-01   0.44904  0.15  0.23
               emo_ul_total_p_c  2.410e-03   0.04909 -0.08 -0.02  0.94
 Residual                        1.961e+04 140.02490
Number of obs: 466580, groups:  Story:subject, 294; subject, 102

Fixed effects:
                   Estimate Std. Error t value
(Intercept)       255.58675   14.62294  17.478
corpfreq_p_c        4.41013    0.29894  14.753
nchar_p_c          -1.39317    0.13326 -10.454
surprisal_p_c       2.10867    0.21192   9.951
mot_ul_total_p_c   -0.38727    0.06407  -6.044
sens_ul_total_p_c   1.53406    0.17671   8.681
emo_ul_total_p_c    0.38797    0.14725   2.635

Correlation of Fixed Effects:
            (Intr) crpf__ nchr__ srpr__ mt____ sn____
corpfrq_p_c -0.001
nchar_p_c    0.001  0.519
surprsl_p_c -0.001  0.672 -0.064
mt_l_ttl_p_ -0.865  0.001  0.011 -0.020
sns_l_ttl__  0.938  0.024  0.005 -0.004 -0.798
em_l_ttl_p_  0.903 -0.020 -0.032 -0.009 -0.749  0.852
```

Appendix C-1: Analysis Gaze Duration

lmerTest

```
Linear mixed model fit by maximum likelihood. t-tests use Satterthwaite's method
[lmerModLmerTest]
Formula: gazdur ~ corpfreq_p_c + nchar_p_c + surprisal_p_c + mot_ul_total_p_c +
    sens_ul_total_p_c + emo_ul_total_p_c + (1 + mot_ul_total_p_c +
    sens_ul_total_p_c + emo_ul_total_p_c | subject/Story)
   Data: rtdata2[rtdata2$wrdno_slide > 1, ]
Control: lmerControl(optimizer = "nloptwrap", calc.derivs = FALSE)

     AIC      BIC   logLik deviance df.resid
 5938637  5938946 -2969290  5938581   466552

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.0857 -0.5674 -0.1995  0.3134 23.0000

Random effects:
 Groups        Name              Variance  Std.Dev.  Corr
 Story:subject (Intercept)       6.284e+04 250.67803
               mot_ul_total_p_c  9.118e-01   0.95487 -1.00
               sens_ul_total_p_c 8.250e+00   2.87223  0.99 -0.99
               emo_ul_total_p_c  5.755e+00   2.39904  0.95 -0.95  0.94
 subject       (Intercept)       4.626e+00   2.15082
               mot_ul_total_p_c  3.534e-02   0.18800  0.99
               sens_ul_total_p_c 2.016e-01   0.44904  0.15  0.23
               emo_ul_total_p_c  2.410e-03   0.04909 -0.08 -0.02  0.94
 Residual                        1.961e+04 140.02490
Number of obs: 466580, groups:  Story:subject, 294; subject, 102

Fixed effects:
                    Estimate Std. Error         df t value Pr(>|t|)
(Intercept)        2.556e+02  1.462e+01  1.480e+04  17.478  < 2e-16 ***
corpfreq_p_c       4.410e+00  2.989e-01  4.402e+05  14.753  < 2e-16 ***
nchar_p_c         -1.393e+00  1.333e-01  4.649e+05 -10.454  < 2e-16 ***
surprisal_p_c      2.109e+00  2.119e-01  4.606e+05   9.951  < 2e-16 ***
mot_ul_total_p_c  -3.873e-01  6.407e-02  1.128e+01  -6.044 7.54e-05 ***
sens_ul_total_p_c  1.534e+00  1.767e-01  3.094e+01   8.681 8.52e-10 ***
emo_ul_total_p_c   3.880e-01  1.472e-01  2.114e+01   2.635   0.0154 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:
            (Intr) crpf__ nchr__ srpr__ mt____ sn____
corpfrq_p_c -0.001
nchar_p_c    0.001  0.519
surprsl_p_c -0.001  0.672 -0.064
mt_l_ttl_p_ -0.865  0.001  0.011 -0.020
sns_l_ttl__  0.938  0.024  0.005 -0.004 -0.798
em_l_ttl_p_  0.903 -0.020 -0.032 -0.009 -0.749  0.852
```

Appendix C-2: Analysis Number of Regressions

lmer

```
Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: regcnt ~ corpfreq_c + nchar_c + surprisal_c + mot_ul_total_c +
    sens_ul_total_c + emo_ul_total_c + (1 + mot_ul_total_c +
sens_ul_total_c + emo_ul_total_c | subject/Story)
   Data: rtdata2[rtdata2$wrdno_slide > 1, ]
Control: lmerControl(optimizer = "nloptwrap", calc.derivs = FALSE)

      AIC       BIC    logLik  deviance  df.resid
 643123.9  643433.1 -321533.9  643067.9    461949

Scaled residuals:
    Min      1Q  Median      3Q     Max
-1.6621 -0.5882 -0.4147 -0.1159 13.3682

Random effects:
 Groups        Name            Variance  Std.Dev.  Corr
 Story:subject (Intercept)     1.088e-03 0.0329922
               mot_ul_total_c  4.173e-07 0.0006460 -0.70
               sens_ul_total_c 3.287e-07 0.0005733 -0.27  0.88
               emo_ul_total_c  1.615e-06 0.0012707 -0.24  0.64  0.70
 subject       (Intercept)     4.020e-02 0.2004952
               mot_ul_total_c  1.107e-06 0.0010522 -0.86
               sens_ul_total_c 1.897e-06 0.0013773 -0.69  0.86
               emo_ul_total_c  2.256e-06 0.0015018 -0.89  0.87  0.93
 Residual                      2.349e-01 0.4846525
Number of obs: 461977, groups:  Story:subject, 294; subject, 102

Fixed effects (lmer):
                  Estimate Std. Error t value
(Intercept)      0.2674833  0.0199606   13.40
corpfreq_c       0.0231091  0.0010049   23.00
nchar_c         -0.0133288  0.0004092  -32.58
surprisal_c      0.0230039  0.0007100   32.40
mot_ul_total_c  -0.0017303  0.0001414  -12.24
sens_ul_total_c -0.0010362  0.0001814   -5.71
emo_ul_total_c  -0.0017441  0.0002253   -7.74

Correlation of Fixed Effects:
            (Intr) crpfr_ nchr_c srprs_ mt_l__ sns___
corpfreq_c   0.002
nchar_c     -0.002  0.499
surprisal_c  0.000  0.688 -0.056
mot_l_ttl_c -0.648  0.021  0.026 -0.029
sns_l_ttl_c -0.523  0.084  0.036 -0.025  0.528
emo_l_ttl_c -0.591 -0.038 -0.069 -0.021  0.603  0.495
```

Appendix C-2: Analysis Number of Regressions

lmerTest

```
Linear mixed model fit by maximum likelihood. t-tests use Satterthwaite's method
[lmerModLmerTest]
Formula: regcnt ~ corpfreq_c + nchar_c + surprisal_c + mot_ul_total_c +
    sens_ul_total_c + emo_ul_total_c + (1 + mot_ul_total_c +
    sens_ul_total_c + emo_ul_total_c | subject/Story)
   Data: rtdata2[rtdata2$wrdno_slide > 1, ]
Control: lmerControl(optimizer = "nloptwrap", calc.derivs = FALSE)

      AIC       BIC    logLik  deviance  df.resid
 643123.9  643433.1 -321533.9  643067.9    461949

Scaled residuals:
    Min      1Q  Median      3Q     Max
-1.6621 -0.5882 -0.4147 -0.1159 13.3682

Random effects:
 Groups        Name            Variance  Std.Dev.  Corr
 Story:subject (Intercept)     1.088e-03 0.0329922
               mot_ul_total_c  4.173e-07 0.0006460 -0.70
               sens_ul_total_c 3.287e-07 0.0005733 -0.27  0.88
               emo_ul_total_c  1.615e-06 0.0012707 -0.24  0.64  0.70
 subject       (Intercept)     4.020e-02 0.2004952
               mot_ul_total_c  1.107e-06 0.0010522 -0.86
               sens_ul_total_c 1.897e-06 0.0013773 -0.69  0.86
               emo_ul_total_c  2.256e-06 0.0015018 -0.89  0.87  0.93
 Residual                      2.349e-01 0.4846525
Number of obs: 461977, groups:  Story:subject, 294; subject, 102

Fixed effects:
                  Estimate Std. Error         df t value Pr(>|t|)
(Intercept)      2.675e-01  1.996e-02  3.705e+05  13.401  < 2e-16 ***
corpfreq_c       2.311e-02  1.005e-03  4.191e+05  22.997  < 2e-16 ***
nchar_c         -1.333e-02  4.092e-04  4.566e+05 -32.577  < 2e-16 ***
surprisal_c      2.300e-02  7.100e-04  4.536e+05  32.399  < 2e-16 ***
mot_ul_total_c  -1.730e-03  1.414e-04  3.995e+01 -12.236 4.30e-15 ***
sens_ul_total_c -1.036e-03  1.814e-04  4.596e+01  -5.712 7.84e-07 ***
emo_ul_total_c  -1.744e-03  2.253e-04  4.367e+01  -7.741 1.00e-09 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:
            (Intr) crpfr_ nchr_c srprs_ mt_l__ sns___
corpfreq_c   0.002
nchar_c     -0.002  0.499
surprisal_c  0.000  0.688 -0.056
mot_l_ttl_c -0.648  0.021  0.026 -0.029
sns_l_ttl_c -0.523  0.084  0.036 -0.025  0.528
emo_l_ttl_c -0.591 -0.038 -0.069 -0.021  0.603  0.495
```

D | Model Summary for the Full Model for Gaze Duration at the Target Word (Chapter 3)

Summaries for both the *lmer* and *lmerTest* models are given.

lmer

```
Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: gazdur ~ corpfreq_c + nchar_c + surprisal_c + mot_ul_total_c +
    sens_ul_total_c + emo_ul_total_c + (1 + mot_ul_total_c +
sens_ul_total_c + emo_ul_total_c | subject/Story)
   Data: rtdata2[rtdata2$wrdno_slide > 1, ]
Control: lmerControl(optimizer = "nloptwrap", calc.derivs = FALSE)

     AIC      BIC   logLik deviance df.resid
 5774075  5774385 -2887010  5774019   459912

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.7888 -0.5788 -0.1745  0.3441 24.7186

Random effects:
 Groups        Name            Variance  Std.Dev.  Corr
 Story:subject (Intercept)     1.611e+02  12.69220
               mot_ul_total_c  1.863e-02   0.13648 -0.16
               sens_ul_total_c 1.841e-01   0.42902  0.23  0.05
               emo_ul_total_c  5.480e-01   0.74027  0.69 -0.65  0.29
 subject       (Intercept)     1.153e+03  33.94894
               mot_ul_total_c  1.769e-03   0.04206 1.00
               sens_ul_total_c 3.269e-01   0.57175 0.69  0.69
               emo_ul_total_c  6.308e-02   0.25115 0.97  0.97  0.83
 Residual                      1.652e+04 128.54387
Number of obs: 459940, groups:  Story:subject, 294; subject, 102

Fixed effects:
                 Estimate Std. Error t value
(Intercept)     246.08993    3.44885   71.35
corpfreq_c       -2.93429    0.26680  -11.00
nchar_c           7.71545    0.10868   70.99
surprisal_c       3.10905    0.18873   16.47
mot_ul_total_c   -0.24364    0.02516   -9.68
sens_ul_total_c   0.32753    0.06911    4.74
emo_ul_total_c    0.19892    0.06474    3.07

Correlation of Fixed Effects:
            (Intr) crpfr_ nchr_c srprs_ mt_l__ sns___
corpfreq_c   0.004
nchar_c     -0.003  0.498
surprisal_c -0.001  0.687 -0.057
mot_l_ttl_c  0.152  0.030  0.038 -0.043
sns_l_ttl_c  0.565  0.058  0.024 -0.017  0.102
emo_l_ttl_c  0.463 -0.034 -0.060 -0.019  0.104  0.322
```

lmerTest

```
Linear mixed model fit by maximum likelihood. t-tests use Satterthwaite's method
[lmerModLmerTest]
Formula: gazdur ~ corpfreq_c + nchar_c + surprisal_c + mot_ul_total_c +
    sens_ul_total_c + emo_ul_total_c + (1 + mot_ul_total_c +
    sens_ul_total_c + emo_ul_total_c | subject/Story)
   Data: rtdata2[rtdata2$wrdno_slide > 1, ]
Control: lmerControl(optimizer = "nloptwrap", calc.derivs = FALSE)

     AIC      BIC   logLik deviance df.resid
 5774075  5774385 -2887010  5774019   459912

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.7888 -0.5788 -0.1745  0.3441 24.7186

Random effects:
 Groups        Name            Variance  Std.Dev.  Corr
 Story:subject (Intercept)     1.611e+02  12.69220
               mot_ul_total_c  1.863e-02   0.13648 -0.16
               sens_ul_total_c 1.841e-01   0.42902  0.23  0.05
               emo_ul_total_c  5.480e-01   0.74027  0.69 -0.65  0.29
 subject       (Intercept)     1.153e+03  33.94894
               mot_ul_total_c  1.769e-03   0.04206 1.00
               sens_ul_total_c 3.269e-01   0.57175 0.69  0.69
               emo_ul_total_c  6.308e-02   0.25115 0.97  0.97  0.83
 Residual                      1.652e+04 128.54387
Number of obs: 459940, groups:  Story:subject, 294; subject, 102

Fixed effects:
                  Estimate Std. Error         df t value Pr(>|t|)
(Intercept)      2.461e+02  3.449e+00  1.127e+02  71.354  < 2e-16 ***
corpfreq_c      -2.934e+00  2.668e-01  3.963e+05 -10.998  < 2e-16 ***
nchar_c          7.715e+00  1.087e-01  4.513e+05  70.991  < 2e-16 ***
surprisal_c      3.109e+00  1.887e-01  4.509e+05  16.474  < 2e-16 ***
mot_ul_total_c  -2.436e-01  2.516e-02  2.419e+02  -9.685  < 2e-16 ***
sens_ul_total_c  3.275e-01  6.911e-02  1.112e+02   4.739  6.4e-06 ***
emo_ul_total_c   1.989e-01  6.474e-02  2.212e+02   3.072  0.00239 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
            (Intr) crpfr_ nchr_c srprs_ mt_l__ sns___
corpfreq_c   0.004
nchar_c     -0.003  0.498
surprisal_c -0.001  0.687 -0.057
mot_l_ttl_c  0.152  0.030  0.038 -0.043
sns_l_ttl_c  0.565  0.058  0.024 -0.017  0.102
emo_l_ttl_c  0.463 -0.034 -0.060 -0.019  0.104  0.322
```

E | Beta Coefficients per Participant and per Story for the Effects of Descriptions on Gaze Duration (Chapter 3)

Figure E.1.: *Beta Coefficients per Participant and per Story for the Effect of Motor Descriptions on Gaze Duration*

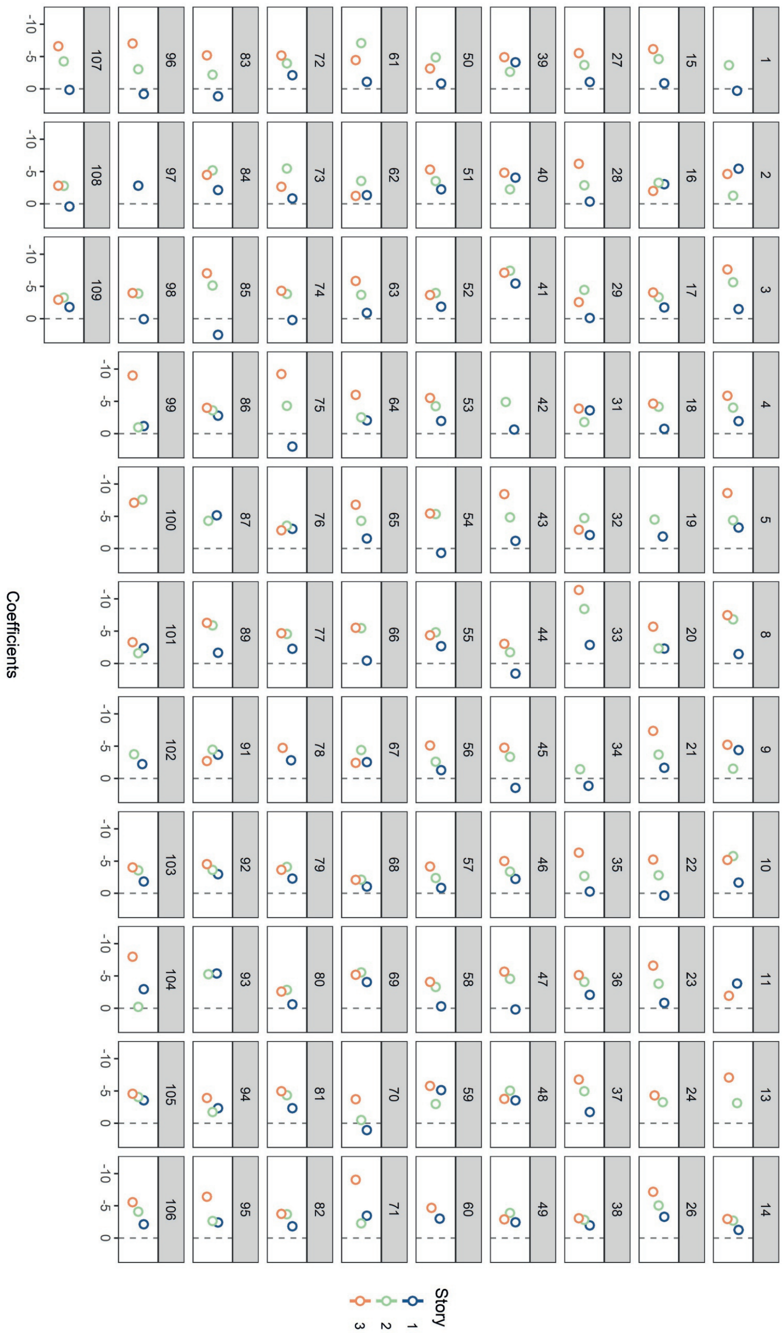


Figure E.2.: *Beta Coefficients per Participant and per Story for the Effect of Perceptual Descriptions on Gaze Duration*

Figure E.3.: *Beta Coefficients per Participant and per Story for the Effect of Mental Event Descriptions on Gaze Duration*

F Tables with All Results for All ROIs for the Individual Differences Analyses (Chapter 4)

F.1. Motor Descriptions

F.1.1. Absorption

Table F.1.: *Posterior Distributions (Mean, SE, 95% Credible Interval) of the Associations between Absorption and Percent Signal Change in the Supramarginal Gyrus as a Response to Motor Descriptions*

			95% Credible Interval	
	Estimate	Est. Error	Lower Bound	Upper Bound
Intercept	0.50	0.18	0.16	0.85
Mean SWAS	-0.03	0.03	-0.10	0.04
Story (B vs A)	-0.20	0.07	-0.34	-0.07

Table F.2.: *Posterior Distributions (Mean, SE, 95% Credible Interval) of the Associations between Absorption and Percent Signal Change in the Left Frontal Orbital Gyrus as a Response to Motor Descriptions*

			95% Credible Interval	
	Estimate	Est. Error	Lower Bound	Upper Bound
Intercept	0.20	0.12	-0.04	0.44
Mean SWAS	0.03	0.02	-0.02	0.08
Story (B vs A)	-0.07	0.05	-0.16	0.02

Table F.3.: *Posterior Distributions (Mean, SE, 95% Credible Interval) of the Associations between Absorption and Percent Signal Change in the Left Precuneus as a Response to Motor Descriptions*

			95% Credible Interval	
	Estimate	Est. Error	Lower Bound	Upper Bound
Intercept	0.87	0.36	0.17	1.58
Mean SWAS	-0.06	0.07	-0.20	0.08
Story (B vs A)	-0.35	0.14	-0.62	-0.09

Table F.4.: *Posterior Distributions (Mean, SE, 95% Credible Interval) of the Associations between Absorption and Percent Signal Change in the Right Precuneus as a Response to Motor Descriptions*

			95% Credible Interval	
	Estimate	Est. Error	Lower Bound	Upper Bound
Intercept	0.71	0.27	0.18	1.25
Mean SWAS	-0.05	0.05	-0.16	0.06
Story (B vs A)	-0.30	0.11	-0.52	-0.10

Table F.5.: *Posterior Distributions (Mean, SE, 95% Credible Interval) of the Associations between Absorption and Percent Signal Change in the Cingulate Gyrus as a Response to Motor Descriptions*

			95% Credible Interval	
	Estimate	Est. Error	Lower Bound	Upper Bound
Intercept	0.06	0.16	-0.23	0.37
Mean SWAS	0.05	0.03	-0.01	0.11
Story (B vs A)	-0.05	0.06	-0.17	0.07

F.1.2. Appreciation

Table F.6.: *Posterior Distributions (Mean, SE, 95% Credible Interval) of the Associations between Appreciation and Percent Signal Change in the Supramarginal Gyrus as a Response to Motor Descriptions*

			95% Credible Interval	
	Estimate	Est. Error	Lower Bound	Upper Bound
Intercept	0.36	0.05	0.26	0.46
Interest	-0.04	0.04	-0.11	0.03
Negative Affect	0.01	0.03	-0.05	0.07
Ominous	0.02	0.04	-0.05	0.09
Positive Affect	0.03	0.03	-0.03	0.08
Special	-0.01	0.03	-0.07	0.04
Story (B vs A)	-0.20	0.09	-0.37	-0.04

Table F.7.: *Posterior Distributions (Mean, SE, 95% Credible Interval) of the Associations between Appreciation and Percent Signal Change in the Left Frontal Orbital Gyrus as a Response to Motor Descriptions*

			95% Credible Interval	
	Estimate	Est. Error	Lower Bound	Upper Bound
Intercept	0.31	0.03	0.25	0.37
Interest	0.00	0.02	-0.04	0.05
Negative Affect	-0.02	0.02	-0.05	0.02
Ominous	0.01	0.02	-0.03	0.05
Positive Affect	0.07	0.02	0.03	0.11
Special	-0.00	0.02	-0.04	0.03
Story (B vs A)	-0.05	0.05	-0.16	0.05

Table F.8.: *Posterior Distributions (Mean, SE, 95% Credible Interval) of the Associations between Appreciation and Percent Signal Change in the Left Precuneus as a Response to Motor Descriptions*

			95% Credible Interval	
	Estimate	Est. Error	Lower Bound	Upper Bound
Intercept	0.67	0.10	0.48	0.86
Interest	-0.05	0.07	-0.19	0.08
Negative Affect	0.03	0.06	-0.09	0.15
Ominous	0.06	0.07	-0.08	0.19
Positive Affect	-0.06	0.06	-0.19	0.06
Special	0.10	0.06	-0.01	0.22
Story (B vs A)	-0.45	0.16	-0.77	-0.12

Table F.9.: *Posterior Distributions (Mean, SE, 95% Credible Interval) of the Associations between Appreciation and Percent Signal Change in the Right Precuneus as a Response to Motor Descriptions*

			95% Credible Interval	
	Estimate	Est. Error	Lower Bound	Upper Bound
Intercept	0.53	0.07	0.38	0.67
Interest	-0.04	0.05	-0.14	0.06
Negative Affect	0.01	0.05	-0.08	0.10
Ominous	0.05	0.05	-0.05	0.15
Positive Affect	-0.04	0.04	-0.13	0.04
Special	0.10	0.04	0.01	0.18
Story (B vs A)	-0.36	0.12	-0.59	-0.12

Table F.10.: *Posterior Distributions (Mean, SE, 95% Credible Interval) of the Associations between Appreciation and Percent Signal Change in the Cingulate Gyrus as a Response to Motor Descriptions*

			95% Credible Interval	
	Estimate	Est. Error	Lower Bound	Upper Bound
Intercept	0.34	0.04	0.25	0.42
Interest	-0.02	0.03	-0.08	0.04
Negative Affect	0.02	0.03	-0.03	0.07
Ominous	0.05	0.03	-0.01	0.11
Positive Affect	0.03	0.03	-0.02	0.08
Special	0.04	0.02	-0.01	0.09
Story (B vs A)	-0.18	0.07	-0.32	-0.04

F.1.3. IRI/ART/Reading Habits

Table F.11.: *Posterior Distributions (Mean, SE, 95% Credible Interval) of the Associations between IRI/ART/Reading Habits and Percent Signal Change in the Supramarginal Gyrus as a Response to Motor Descriptions*

			95% Credible Interval	
	Estimate	Est. Error	Lower Bound	Upper Bound
Intercept	0.10	0.27	-0.42	0.63
IRI Fantasy	0.02	0.03	-0.04	0.09
IRI Perspective Taking	0.02	0.04	-0.05	0.10
ART	-0.00	0.01	-0.02	0.02
Reading Habits	0.00	0.04	-0.08	0.09
Story (B vs A)	-0.16	0.05	-0.26	-0.05

Table F.12.: *Posterior Distributions (Mean, SE, 95% Credible Interval) of the Associations between IRI/ART/Reading Habits and Percent Signal Change in the Left Frontal Orbital Gyrus as a Response to Motor Descriptions*

			95% Credible Interval	
	Estimate	Est. Error	Lower Bound	Upper Bound
Intercept	0.16	0.19	-0.22	0.54
IRI Fantasy	0.02	0.02	-0.02	0.07
IRI Perspective Taking	0.01	0.03	-0.05	0.06
ART	0.00	0.01	-0.01	0.02
Reading Habits	-0.01	0.03	-0.06	0.06
Story (B vs A)	-0.11	0.03	-0.18	-0.05

Table F.13.: *Posterior Distributions (Mean, SE, 95% Credible Interval) of the Associations between IRI/ART/Reading Habits and Percent Signal Change in the Left Precuneus as a Response to Motor Descriptions*

			95% Credible Interval	
	Estimate	Est. Error	Lower Bound	Upper Bound
Intercept	0.44	0.58	-0.71	1.56
IRI Fantasy	0.02	0.07	-0.12	0.16
IRI Perspective Taking	0.03	0.08	-0.13	0.20
ART	-0.02	0.02	-0.07	0.02
Reading Habits	0.04	0.09	-0.14	0.21
Story (B vs A)	-0.27	0.10	-0.46	-0.08

Table F.14.: *Posterior Distributions (Mean, SE, 95% Credible Interval) of the Associations between IRI/ART/Reading Habits and Percent Signal Change in the Right Precuneus as a Response to Motor Descriptions*

			95% Credible Interval	
	Estimate	Est. Error	Lower Bound	Upper Bound
Intercept	0.57	0.42	-0.26	1.39
IRI Fantasy	0.00	0.05	-0.10	0.11
IRI Perspective Taking	-0.00	0.06	-0.12	0.12
ART	-0.02	0.02	-0.05	0.01
Reading Habits	0.06	0.06	-0.06	0.19
Story (B vs A)	-0.23	0.08	-0.39	-0.08

Table F.15.: *Posterior Distributions (Mean, SE, 95% Credible Interval) of the Associations between IRI/ART/Reading Habits and Percent Signal Change in the Cingulate Gyrus as a Response to Motor Descriptions*

			95% Credible Interval	
	Estimate	Est. Error	Lower Bound	Upper Bound
Intercept	-0.09	0.22	-0.54	0.35
IRI Fantasy	0.09	0.03	0.03	0.14
IRI Perspective Taking	-0.01	0.03	-0.07	0.06
ART	-0.00	0.01	-0.02	0.01
Reading Habits	-0.02	0.03	-0.09	0.05
Story (B vs A)	-0.12	0.04	-0.20	-0.04

F.2. Perceptual Descriptions

F.2.1. Absorption

Table F.16.: *Posterior Distributions (Mean, SE, 95% Credible Interval) of the Associations between Absorption and Percent Signal Change in the Supramarginal Gyrus as a Response to Perceptual Descriptions*

			95% Credible Interval	
	Estimate	Est. Error	Lower Bound	Upper Bound
Intercept	0.09	0.27	-0.45	0.61
Mean SWAS	0.05	0.05	-0.06	0.16
Story (B vs A)	0.12	0.10	-0.09	0.32

Table F.17.: *Posterior Distributions (Mean, SE, 95% Credible Interval) of the Associations between Absorption and Percent Signal Change in the Left Inferior Temporal Gyrus as a Response to Perceptual Descriptions*

			95% Credible Interval	
	Estimate	Est. Error	Lower Bound	Upper Bound
Intercept	0.08	0.16	-0.22	0.40
Mean SWAS	0.05	0.03	-0.01	0.11
Story (B vs A)	0.06	0.06	-0.06	0.18

Table F.18.: *Posterior Distributions (Mean, SE, 95% Credible Interval) of the Associations between Absorption and Percent Signal Change in the Right Inferior Temporal Gyrus as a Response to Perceptual Descriptions*

			95% Credible Interval	
	Estimate	Est. Error	Lower Bound	Upper Bound
Intercept	0.24	0.20	-0.15	0.64
Mean SWAS	0.02	0.04	-0.06	0.10
Story (B vs A)	-0.01	0.08	-0.17	0.14

Table F.19.: *Posterior Distributions (Mean, SE, 95% Credible Interval) of the Associations between Absorption and Percent Signal Change in the Left Temporal Fusiform Cortex as a Response to Perceptual Descriptions*

			95% Credible Interval	
	Estimate	Est. Error	Lower Bound	Upper Bound
Intercept	0.13	0.12	-0.10	0.36
Mean SWAS	0.03	0.02	-0.02	0.07
Story (B vs A)	0.02	0.05	-0.07	0.11

Table F.20.: *Posterior Distributions (Mean, SE, 95% Credible Interval) of the Associations between Absorption and Percent Signal Change in the Right Temporal Fusiform Cortex as a Response to Perceptual Descriptions*

			95% Credible Interval	
	Estimate	Est. Error	Lower Bound	Upper Bound
Intercept	0.18	0.17	-0.16	0.51
Mean SWAS	0.03	0.03	-0.03	0.10
Story (B vs A)	-0.03	0.07	-0.16	0.10

Table F.21.: *Posterior Distributions (Mean, SE, 95% Credible Interval) of the Associations between Absorption and Percent Signal Change in the Right Parahippocampal Gyrus as a Response to Perceptual Descriptions*

			95% Credible Interval	
	Estimate	Est. Error	Lower Bound	Upper Bound
Intercept	0.50	0.19	0.13	0.89
Mean SWAS	-0.02	0.04	-0.10	0.05
Story (B vs A)	0.00	0.07	-0.14	0.14

F.2.2. Appreciation

Table F.22.: *Posterior Distributions (Mean, SE, 95% Credible Interval) of the Associations between Appreciation and Percent Signal Change in the Supramarginal Gyrus as a Response to Perceptual Descriptions*

			95% Credible Interval	
	Estimate	Est. Error	Lower Bound	Upper Bound
Intercept	0.30	0.08	0.16	0.45
Interest	-0.03	0.06	-0.14	0.08
Negative Affect	0.02	0.05	-0.07	0.11
Ominous	-0.05	0.05	-0.15	0.06
Positive Affect	0.01	0.05	-0.08	0.10
Special	0.00	0.04	-0.08	0.09
Story (B vs A)	0.09	0.13	-0.16	0.34

Table F.23.: *Posterior Distributions (Mean, SE, 95% Credible Interval) of the Associations between Appreciation and Percent Signal Change in the Left Inferior Temporal Gyrus as a Response to Perceptual Descriptions*

			95% Credible Interval	
	Estimate	Est. Error	Lower Bound	Upper Bound
Intercept	0.30	0.04	0.21	0.39
Interest	-0.01	0.03	-0.07	0.05
Negative Affect	0.01	0.03	-0.04	0.07
Ominous	-0.05	0.03	-0.11	0.01
Positive Affect	0.01	0.03	-0.04	0.06
Special	0.01	0.03	-0.04	0.06
Story (B vs A)	0.05	0.07	-0.10	0.19

Table F.24.: *Posterior Distributions (Mean, SE, 95% Credible Interval) of the Associations between Appreciation and Percent Signal Change in the Right Inferior Temporal Gyrus as a Response to Perceptual Descriptions*

			95% Credible Interval	
	Estimate	Est. Error	Lower Bound	Upper Bound
Intercept	0.33	0.06	0.22	0.44
Interest	0.00	0.04	-0.08	0.08
Negative Affect	-0.03	0.04	-0.10	0.05
Ominous	0.01	0.04	-0.07	0.09
Positive Affect	0.00	0.04	-0.06	0.07
Special	0.00	0.03	-0.06	0.06
Story (B vs A)	-0.03	0.10	-0.22	0.16

Table F.25.: *Posterior Distributions (Mean, SE, 95% Credible Interval) of the Associations between Appreciation and Percent Signal Change in the Left Temporal Fusiform Cortex as a Response to Perceptual Descriptions*

			95% Credible Interval	
	Estimate	Est. Error	Lower Bound	Upper Bound
Intercept	0.26	0.03	0.19	0.32
Interest	-0.02	0.02	-0.06	0.03
Negative Affect	0.02	0.02	-0.02	0.06
Ominous	-0.03	0.02	-0.08	0.02
Positive Affect	0.01	0.02	-0.03	0.05
Special	-0.00	0.02	-0.04	0.04
Story (B vs A)	0.00	0.06	-0.11	0.11

Table F.26.: *Posterior Distributions (Mean, SE, 95% Credible Interval) of the Associations between Appreciation and Percent Signal Change in the Right Temporal Fusiform Cortex as a Response to Perceptual Descriptions*

			95% Credible Interval	
	Estimate	Est. Error	Lower Bound	Upper Bound
Intercept	0.37	0.05	0.27	0.47
Interest	-0.01	0.03	-0.08	0.05
Negative Affect	0.03	0.03	-0.03	0.09
Ominous	0.04	0.04	-0.03	0.11
Positive Affect	-0.00	0.03	-0.06	0.06
Special	-0.02	0.03	-0.08	0.03
Story (B vs A)	-0.16	0.08	-0.32	0.01

Table F.27.: *Posterior Distributions (Mean, SE, 95% Credible Interval) of the Associations between Appreciation and Percent Signal Change in the Right Parahippocampal Gyrus as a Response to Perceptual Descriptions*

			95% Credible Interval	
	Estimate	Est. Error	Lower Bound	Upper Bound
Intercept	0.35	0.06	0.24	0.46
Interest	-0.00	0.04	-0.09	0.08
Negative Affect	0.01	0.04	-0.06	0.08
Ominous	-0.04	0.04	-0.12	0.03
Positive Affect	0.03	0.04	-0.04	0.10
Special	-0.05	0.03	-0.11	0.01
Story (B vs A)	0.10	0.09	-0.07	0.27

F.2.3. IRI/ART/Reading Habits

Table F.28.: *Posterior Distributions (Mean, SE, 95% Credible Interval) of the Associations between IRI/ART/Reading Habits and Percent Signal Change in the Supramarginal Gyrus as a Response to Perceptual Descriptions*

			95% Credible Interval	
	Estimate	Est. Error	Lower Bound	Upper Bound
Intercept	-0.41	0.38	-1.16	0.33
IRI Fantasy	0.06	0.05	-0.04	0.15
IRI Perspective Taking	0.07	0.05	-0.03	0.18
ART	0.01	0.01	-0.02	0.04
Reading Habits	0.09	0.06	-0.02	0.20
Story (B vs A)	0.06	0.07	-0.08	0.21

Table F.29.: *Posterior Distributions (Mean, SE, 95% Credible Interval) of the Associations between IRI/ART/Reading Habits and Percent Signal Change in the Left Inferior Temporal Gyrus as a Response to Perceptual Descriptions*

			95% Credible Interval	
	Estimate	Est. Error	Lower Bound	Upper Bound
Intercept	-0.43	0.22	-0.85	-0.00
IRI Fantasy	0.08	0.03	0.02	0.13
IRI Perspective Taking	0.08	0.03	0.02	0.14
ART	-0.00	0.01	-0.02	0.01
Reading Habits	0.03	0.03	-0.03	0.10
Story (B vs A)	-0.00	0.04	-0.08	0.08

Table F.30.: *Posterior Distributions (Mean, SE, 95% Credible Interval) of the Associations between IRI/ART/Reading Habits and Percent Signal Change in the Right Inferior Temporal Gyrus as a Response to Perceptual Descriptions*

			95% Credible Interval	
	Estimate	Est. Error	Lower Bound	Upper Bound
Intercept	-0.38	0.29	-0.94	0.19
IRI Fantasy	0.08	0.04	0.01	0.15
IRI Perspective Taking	0.07	0.04	-0.02	0.15
ART	-0.01	0.01	-0.03	0.01
Reading Habits	0.03	0.04	-0.06	0.11
Story (B vs A)	-0.03	0.06	-0.14	0.08

Table F.31.: *Posterior Distributions (Mean, SE, 95% Credible Interval) of the Associations between IRI/ART/Reading Habits and Percent Signal Change in the Left Temporal Fusiform Cortex as a Response to Perceptual Descriptions*

			95% Credible Interval	
	Estimate	Est. Error	Lower Bound	Upper Bound
Intercept	-0.19	0.16	-0.51	0.13
IRI Fantasy	0.04	0.02	-0.00	0.08
IRI Perspective Taking	0.06	0.02	0.01	0.10
ART	-0.00	0.01	-0.01	0.01
Reading Habits	0.04	0.02	-0.01	0.09
Story (B vs A)	-0.01	0.03	-0.07	0.05

Table F.32.: *Posterior Distributions (Mean, SE, 95% Credible Interval) of the Associations between IRI/ART/Reading Habits and Percent Signal Change in the Right Temporal Fusiform Cortex as a Response to Perceptual Descriptions*

			95% Credible Interval	
	Estimate	Est. Error	Lower Bound	Upper Bound
Intercept	-0.21	0.25	-0.70	0.27
IRI Fantasy	0.07	0.03	0.00	0.13
IRI Perspective Taking	0.05	0.04	-0.02	0.12
ART	-0.01	0.01	-0.03	0.01
Reading Habits	0.04	0.04	-0.04	0.11
Story (B vs A)	-0.06	0.05	-0.16	0.03

Table F.33.: *Posterior Distributions (Mean, SE, 95% Credible Interval) of the Associations between IRI/ART/Reading Habits and Percent Signal Change in the Right Parahippocampal Gyrus as a Response to Perceptual Descriptions*

			95% Credible Interval	
	Estimate	Est. Error	Lower Bound	Upper Bound
Intercept	0.12	0.34	-0.55	0.76
IRI Fantasy	0.03	0.04	-0.05	0.11
IRI Perspective Taking	0.03	0.05	-0.06	0.13
ART	-0.00	0.01	-0.03	0.02
Reading Habits	0.06	0.05	-0.04	0.17
Story (B vs A)	0.04	0.05	-0.06	0.13

F.3. Mental Event Descriptions

F.3.1. Absorption

Table F.34.: *Posterior Distributions (Mean, SE, 95% Credible Interval) of the Associations between Absorption and Percent Signal Change in the Supramarginal Gyrus as a Response to Mental Event Descriptions*

			95% Credible Interval	
	Estimate	Est. Error	Lower Bound	Upper Bound
Intercept	0.25	0.13	0.01	0.51
Mean SWAS	-0.02	0.03	-0.07	0.03
Story (B vs A)	-0.01	0.05	-0.11	0.09

Table F.35.: *Posterior Distributions (Mean, SE, 95% Credible Interval) of the Associations between Absorption and Percent Signal Change in the Left Insula as a Response to Mental Event Descriptions*

			95% Credible Interval	
	Estimate	Est. Error	Lower Bound	Upper Bound
Intercept	0.36	0.11	0.16	0.58
Mean SWAS	-0.04	0.02	-0.08	0.00
Story (B vs A)	-0.04	0.04	-0.12	0.05

Table F.36.: *Posterior Distributions (Mean, SE, 95% Credible Interval) of the Associations between Absorption and Percent Signal Change in the Right Insula as a Response to Mental Event Descriptions*

			95% Credible Interval	
	Estimate	Est. Error	Lower Bound	Upper Bound
Intercept	0.30	0.11	0.09	0.51
Mean SWAS	-0.03	0.02	-0.07	0.01
Story (B vs A)	-0.02	0.04	-0.10	0.06

Table F.37.: *Posterior Distributions (Mean, SE, 95% Credible Interval) of the Associations between Absorption and Percent Signal Change in the Left Lingual Gyrus as a Response to Mental Event Descriptions*

			95% Credible Interval	
	Estimate	Est. Error	Lower Bound	Upper Bound
Intercept	0.37	0.13	0.12	0.63
Mean SWAS	-0.02	0.03	-0.07	0.03
Story (B vs A)	-0.10	0.05	-0.20	-0.00

Table F.38.: *Posterior Distributions (Mean, SE, 95% Credible Interval) of the Associations between Absorption and Percent Signal Change in the Right Lingual Gyrus as a Response to Mental Event Descriptions*

			95% Credible Interval	
	Estimate	Est. Error	Lower Bound	Upper Bound
Intercept	0.37	0.13	0.10	0.63
Mean SWAS	-0.02	0.03	-0.07	0.03
Story (B vs A)	-0.12	0.05	-0.22	-0.02

Table F.39.: *Posterior Distributions (Mean, SE, 95% Credible Interval) of the Associations between Absorption and Percent Signal Change in the Left Temporal Pole as a Response to Mental Event Descriptions*

			95% Credible Interval	
	Estimate	Est. Error	Lower Bound	Upper Bound
Intercept	0.31	0.16	-0.00	0.63
Mean SWAS	-0.01	0.03	-0.07	0.05
Story (B vs A)	-0.00	0.06	-0.12	0.12

F.3.2. Appreciation

Table F.40.: *Posterior Distributions (Mean, SE, 95% Credible Interval) of the Associations between Appreciation and Percent Signal Change in the Supramarginal Gyrus as a Response to Mental Event Descriptions*

			95% Credible Interval	
	Estimate	Est. Error	Lower Bound	Upper Bound
Intercept	0.15	0.03	0.08	0.21
Interest	-0.00	0.02	-0.05	0.05
Negative Affect	-0.02	0.02	-0.06	0.02
Ominous	-0.04	0.02	-0.09	0.01
Positive Affect	-0.01	0.02	-0.06	0.03
Special	-0.02	0.02	-0.06	0.02
Story (B vs A)	0.06	0.06	-0.06	0.17

Table F.41.: *Posterior Distributions (Mean, SE, 95% Credible Interval) of the Associations between Appreciation and Percent Signal Change in the Left Insula as a Response to Mental Event Descriptions*

			95% Credible Interval	
	Estimate	Est. Error	Lower Bound	Upper Bound
Intercept	0.17	0.03	0.11	0.23
Interest	-0.03	0.02	-0.07	0.02
Negative Affect	0.01	0.02	-0.03	0.05
Ominous	-0.01	0.02	-0.05	0.03
Positive Affect	-0.01	0.02	-0.05	0.03
Special	-0.02	0.02	-0.05	0.02
Story (B vs A)	-0.01	0.05	-0.11	0.09

Table F.42.: *Posterior Distributions (Mean, SE, 95% Credible Interval) of the Associations between Appreciation and Percent Signal Change in the Right Insula as a Response to Mental Event Descriptions*

			95% Credible Interval	
	Estimate	Est. Error	Lower Bound	Upper Bound
Intercept	0.18	0.03	0.12	0.23
Interest	-0.00	0.02	-0.05	0.04
Negative Affect	0.01	0.02	-0.03	0.05
Ominous	0.00	0.02	-0.04	0.04
Positive Affect	-0.02	0.02	-0.05	0.02
Special	-0.02	0.02	-0.06	0.01
Story (B vs A)	-0.01	0.05	-0.11	0.08

Table F.43.: *Posterior Distributions (Mean, SE, 95% Credible Interval) of the Associations between Appreciation and Percent Signal Change in the Left Lingual Gyrus as a Response to Mental Event Descriptions*

			95% Credible Interval	
	Estimate	Est. Error	Lower Bound	Upper Bound
Intercept	0.24	0.03	0.18	0.31
Interest	0.05	0.02	0.00	0.09
Negative Affect	-0.04	0.02	-0.08	0.00
Ominous	-0.01	0.02	-0.06	0.04
Positive Affect	-0.05	0.02	-0.09	-0.01
Special	0.01	0.02	-0.03	0.04
Story (B vs A)	-0.03	0.06	-0.15	0.08

Table F.44.: *Posterior Distributions (Mean, SE, 95% Credible Interval) of the Associations between Appreciation and Percent Signal Change in the Right Lingual Gyrus as a Response to Mental Event Descriptions*

			95% Credible Interval	
	Estimate	Est. Error	Lower Bound	Upper Bound
Intercept	0.29	0.04	0.21	0.36
Interest	0.01	0.03	-0.04	0.05
Negative Affect	-0.00	0.02	-0.05	0.04
Ominous	-0.01	0.03	-0.06	0.04
Positive Affect	-0.04	0.02	-0.09	0.00
Special	0.00	0.02	-0.04	0.04
Story (B vs A)	-0.11	0.06	-0.23	0.01

Table F.45.: *Posterior Distributions (Mean, SE, 95% Credible Interval) of the Associations between Appreciation and Percent Signal Change in the Left Temporal Pole as a Response to Mental Event Descriptions*

			95% Credible Interval	
	Estimate	Est. Error	Lower Bound	Upper Bound
Intercept	0.22	0.04	0.14	0.31
Interest	0.02	0.03	-0.04	0.07
Negative Affect	-0.03	0.03	-0.08	0.02
Ominous	-0.05	0.03	-0.11	0.01
Positive Affect	-0.01	0.03	-0.07	0.04
Special	0.01	0.02	-0.04	0.06
Story (B vs A)	0.10	0.07	-0.04	0.25

F.3.3. IRI/ART/Reading Habits

Table F.46.: *Posterior Distributions (Mean, SE, 95% Credible Interval) of the Associations between IRI/ART/Reading Habits and Percent Signal Change in the Supramarginal Gyrus as a Response to Mental Event Descriptions*

			95% Credible Interval	
	Estimate	Est. Error	Lower Bound	Upper Bound
Intercept	-0.02	0.18	-0.38	0.34
IRI Fantasy	-0.01	0.02	-0.06	0.04
IRI Perspective Taking	0.05	0.03	-0.00	0.10
ART	0.00	0.01	-0.01	0.02
Reading Habits	-0.05	0.03	-0.11	0.00
Story (B vs A)	0.02	0.04	-0.05	0.09

Table F.47.: *Posterior Distributions (Mean, SE, 95% Credible Interval) of the Associations between IRI/ART/Reading Habits and Percent Signal Change in the Left Insula as a Response to Mental Event Descriptions*

			95% Credible Interval	
	Estimate	Est. Error	Lower Bound	Upper Bound
Intercept	0.19	0.17	-0.13	0.52
IRI Fantasy	-0.00	0.02	-0.04	0.04
IRI Perspective Taking	0.00	0.02	-0.04	0.05
ART	-0.00	0.01	-0.02	0.01
Reading Habits	-0.02	0.02	-0.06	0.03
Story (B vs A)	0.02	0.03	-0.05	0.08

Table F.48.: *Posterior Distributions (Mean, SE, 95% Credible Interval) of the Associations between IRI/ART/Reading Habits and Percent Signal Change in the Right Insula as a Response to Mental Event Descriptions*

			95% Credible Interval	
	Estimate	Est. Error	Lower Bound	Upper Bound
Intercept	0.11	0.17	-0.21	0.45
IRI Fantasy	0.00	0.02	-0.04	0.04
IRI Perspective Taking	0.01	0.02	-0.04	0.06
ART	-0.00	0.01	-0.02	0.01
Reading Habits	-0.02	0.02	-0.06	0.03
Story (B vs A)	0.02	0.03	-0.04	0.08

Table F.49.: *Posterior Distributions (Mean, SE, 95% Credible Interval) of the Associations between IRI/ART/Reading Habits and Percent Signal Change in the Left Lingual Gyrus as a Response to Mental Event Descriptions*

			95% Credible Interval	
	Estimate	Est. Error	Lower Bound	Upper Bound
Intercept	0.30	0.20	-0.11	0.70
IRI Fantasy	-0.00	0.03	-0.05	0.05
IRI Perspective Taking	-0.01	0.03	-0.07	0.05
ART	0.00	0.01	-0.01	0.02
Reading Habits	-0.01	0.03	-0.07	0.05
Story (B vs A)	-0.07	0.04	-0.14	0.01

Table F.50.: *Posterior Distributions (Mean, SE, 95% Credible Interval) of the Associations between IRI/ART/Reading Habits and Percent Signal Change in the Right Lingual Gyrus as a Response to Mental Event Descriptions*

			95% Credible Interval	
	Estimate	Est. Error	Lower Bound	Upper Bound
Intercept	0.26	0.20	-0.13	0.65
IRI Fantasy	0.01	0.03	-0.03	0.07
IRI Perspective Taking	-0.00	0.03	-0.06	0.05
ART	-0.00	0.01	-0.02	0.01
Reading Habits	-0.02	0.03	-0.08	0.04
Story (B vs A)	-0.09	0.04	-0.17	-0.02

Table F.51.: *Posterior Distributions (Mean, SE, 95% Credible Interval) of the Associations between IRI/ART/Reading Habits and Percent Signal Change in the Left Temporal Pole as a Response to Mental Event Descriptions*

			95% Credible Interval	
	Estimate	Est. Error	Lower Bound	Upper Bound
Intercept	0.16	0.23	-0.29	0.60
IRI Fantasy	-0.03	0.03	-0.09	0.03
IRI Perspective Taking	0.04	0.03	-0.02	0.11
ART	0.00	0.01	-0.01	0.02
Reading Habits	0.06	0.04	-0.01	0.13
Story (B vs A)	0.02	0.05	-0.07	0.11

G | Synopsis of All Stories (Chapter 5)

The People That Had Everything Delivered (used in Chapter 2, 3, 4, 5)
The story recounts the experiences of a man who lives in an apartment building in Amsterdam. His neighbors rent out their apartment while they are on holiday for the Christmas days, and a morbidly obese British couple stays there. When the wife has a heart attack she has to be lifted out of the apartment by a firetruck, as there is no elevator in the building.

The Chinese Wedding (used in Chapter 3, 5)
In this story, a Dutch male student is best man at his roommate's wedding to a Chinese exchange student. He ends up spending the day with the parents of the bride who don't speak a word of English, which leads to some very awkward moments.

Symbols and Signs (used in Chapter 3, 4, 5)
This story tells about an older Russian Couple, that lives in the USA. Their son is hospitalized, and on his birthday, they can't visit him as he just attempted suicide. The rest of the day, the mother is reminiscing about her son's life, and later that night the father decides that he wants to bring him home to take care of him by themselves.

It Is Mouse (used in Chapter 5)
In this story, a preschool teacher is very annoyed by one of her students which results in her trying to ignore the child. This results in the child dying in an unhappy accident at the preschool, due to the teacher's negligence.

How the Wolves Dance (used in Chapter 5)
This story is about someone who wakes up during the night to discover dancing wolves in his living room.

The Substitute (used in Chapter 5)
In this story, a girl is sitting in a cafe, and she thinks that she is being stalked by a man who eventually turns out to be an acquaintance – as he used to be her substitute teacher. The story ends on the cliffhanger that perhaps, after all, he is not exactly who he claims he is.

She Is Everywhere (used in Chapter 5)
This story is about a man in the library in Utrecht, who has the feeling that he is being followed around by his ex (the subtext is that he has a psychosis, since he is not really being followed around).

Moped on Sea (used in Chapter 2, 5)
This is a surrealistic story about a boy on a boat and his encounter with a man riding a moped at sea in the middle of the night.

God and the Judge of the Insane (used in Chapter 2, 5)
In this story, the author narrates the story of a mentally instable man who is convinced that he is God, and believes that therefore all his excrements are holy and should not be thrown away. Apart from that, he terrorizes the neighborhood, leading to his institutionalization later on in the story, after which he finally seems to realize that he was mistaken in thinking that he was God.

H | Descriptive Statistics Reading Habits Questionnaire (Chapter 5)

Table H.1.: *Numbers of Responses for Each Answer Option per Question on the Reading Habits Questionnaire, for Study 1 (N=102)*

Question 1: How often do you read fiction (e.g., novels, fairy tales, poetry)?				
Never	Hardly ever	One day per week	More than two days a week, but not daily	Daily
5	55	23	18	8
Question 2: How often do you read non-fiction (e.g., newspapers, news websites, magazines)? (NA=1)				
Never	Hardly ever	One day per week	More than two days a week, but not daily	Daily
3	3	12	23	67
Question 3: How often do you consume fiction (for example in movies, television series, video games)?				
Never	Hardly ever	One day per week	More than two days a week, but not daily	Daily
0	4	12	39	54
Question 4: How much do you like reading fiction?				
Totally not	Not very much	I don't like it, but I don't dislike it either	Quite a lot	Very much
0	5	14	58	32
Question 5: How many books do you read each year? (NA=1)				
I don't read books	1 or 2	About 1 each month	More than 15	
5	46	41	16	

Table H.2.: *Numbers of Responses for Each Answer Option per Question on the Reading Habits Questionnaire, for Study 2 (N=43)*

Question 1: How often do you read fiction (e.g., novels, fairy tales, poetry)?				
Never	Hardly ever	One day per week	More than two days a week, but not daily	Daily
3	15	14	8	3
Question 2: How often do you read non-fiction (e.g., newspapers, news websites, magazines)?				
Never	Hardly ever	One day per week	More than two days a week, but not daily	Daily
2	0	3	11	27
Question 3: How often do you consume fiction (for example in movies, television series, video games)?				
Never	Hardly ever	One day per week	More than two days a week, but not daily	Daily
0	4	5	15	18
Question 4: How much do you like reading fiction?				
Totally not	Not very much	I don't like it, but I don't dislike it either	Quite a lot	Very much
0	4	4	18	17
Question 5: How many books do you read each year?				
I don't read books	1 or 2	About 1 each month	More than 15	
3	14	22	4	

Table H.3.: *Numbers of Responses for Each Answer Option per Question on the Reading Habits Questionnaire, for Study 3 (N=125)*

Question 1: How often do you read fiction (e.g., novels, fairy tales, poetry)?					
Never	Hardly ever	Once a month	Once a week	A couple of times per week	Daily
10	32	24	16	32	11
Question 2: How often do you read non-fiction (e.g., newspapers, news websites, magazines)?					
Never	Hardly ever	Once a month	Once a week	A couple of times per week	Daily
16	16	13	7	37	35
Question 3: How often do you consume fiction (for example in movies, television series, video games)?					
Never	Hardly ever	Once a month	Once a week	A couple of times per week	Daily
18	17	10	9	19	51
Question 4: How much do you like reading fiction? (NA=2)					
Totally not	Not very much	I don't like it, but I don't dislike it either	Quite a lot	Very much	
18	18	11	38	37	
Question 5: How many books do you read each year?					
I don't read books	1 or 2	Between 2 and 5	Between 5 and 10	Between 10 and 15	More than 15
12	24	21	28	20	19

I | Data Management

Datasets Used in this Disseration

For this dissertation, three datasets were created, and one existing dataset was reused. The dataset that was reused was a dataset collected by Eekhof, Eerland and Willems (2018), and can be found on `https://osf.io/qynhu/`.

For Chapter 2, a dataset was created that can be found on `https://osf.io/98ntg/`. On this page, analysis scripts are also available.

For Chapter 3, a dataset was created that can be found on `https://osf.io/qgx26/`. On this page, analysis scripts are also available. This entire data collection has also been published open access (Mak & Willems, 2021) and can be downloaded from DANS-EASY (doi: 10.17026/dans-zqk-zmqs).

For Chapter 4, a dataset was created that is not available yet. It is currently stored on in the project folder assigned by the Donders Institute (project 3011086.01), on their servers. Access is possible via a SSH connection, but only for registered project members. The processed and analyzed data is stored similarly. Eventually, these data will be archived in a Research Documentation Collection, according to the policies of the Donders Institute. Next, the data will be anonymized and archived as a Data Sharing Collection in the Donders Repository, according to the policies of the Donders Institute. This will be done upon publication of the data.

For Chapter 5, the data from Chapters 2 and 3 were reused, together by the aforementioned dataset collected by Eekhof et al. (2018). The combined dataset, including analysis scripts, is available on `https://osf.io/h3ct6/`.

Personal Data and Privacy

Only anonimyzed data has been shared. Eye tracking and questionnaire data were anonimyzed during data collection, and cannot be traced back to individual participants.

As MRI data can more easily be traced back to individuals, these data have not been shared yet (and the raw data will not be shared). Before sharing the fMRI data collected in Chapter 4, these data first need to be defaced. Until then, the data will remain for at least 10 years on the servers of the Donders Institute for the purpose of scientific integrity.

Subject payment forms and consent forms have been collected and stored separately from the research data.

Data Sharing

To promote Open Science, an effort has been made to always share all research data and analysis scripts. As can be seen in the list of used datasets, all data and scripts have been shared open access for Chapters 2, 3, and 5. The data used in these chapters can be reused by others, if they refer to the original datasets.

The data collected for Chapter 4 has not been shared yet, but will be shared as soon as possible, and will then also be available for reuse by other researchers.

Nederlandse samenvatting

De vraag die ik in dit proefschrift opwerp is of mensen zich zaken voorstellen tijdens het lezen, hoe dit voorstellingsproces er dan ongeveer uit ziet, of iedereen dit op dezelfde manier doet, en of het proces van voorstellen ook samenhangt met de manier waarop mensen gelezen verhalen waarderen.

Ik heb dit vraagstuk onderzocht aan de hand van drie experimenten. In het eerste experiment kregen mensen verhalen te lezen en kregen daarbij een leesinstructie. Sommigen kregen de opdracht om zich de gebeurtenissen in de verhalen zo levendig mogelijk voor te stellen, anderen moesten simpelweg lezen zoals ze normaal gesproken ook zouden lezen, en weer anderen moesten juist niet op de inhoud van het verhaal letten maar op de schrijfstijl. De verwachting was dat mensen die de opdracht kregen om zich de gebeurtenissen levendig voor te stellen, na het lezen op een vragenlijst ook in sterkere mate zouden aangeven dat ze een voorstelling hadden van het verhaal dan de beide andere groepen. Dit bleek echter niet overtuigend het geval te zijn. De vraag was toen: is dat omdat mensen zich niet zoveel voorstellen tijdens het lezen, of omdat ze deze voorstelling gewoon niet *op bevel* aan of uit kunnen zetten?

In de andere twee experimenten heb ik het daarom anders aangepakt: geen leesinstructies meer, maar iedereen laten lezen zoals ze dat normaal ook zouden doen, en dan observeren wat er gebeurt. In het tweede experiment heb ik mensen laten lezen achter een computerscherm terwijl ik hun oogbewegingen volgde met een zogenoemde *eye-tracker*, en in het derde experiment heb ik mensen laten lezen terwijl ze in een MRI scanner lagen, én heb ik tegelijkertijd ook hun oogbewegingen gevolgd met een *eye-tracker*. Ik heb vervolgens gekeken naar drie soorten beschrijvingen uit de verhalen die je je zou kunnen voorstellen: beschrijvingen van acties, beschrijvingen van hoe zaken eruit zien (bijvoorbeeld personages, objecten, of de omgeving), en beschrijvingen van gedachten en emoties.

Wat bleek was dat mensen gemiddeld genomen sneller gaan lezen als ze beschrijvingen van acties lezen, en langzamer als ze beschrijvingen lezen van hoe zaken eruit zien, of van gedachten en emoties. Daarnaast zag ik aan de hersenactiviteit dat bij het lezen van beschrijvingen van acties vooral hersengebieden actief worden die betrokken zijn bij het waarnemen en plannen van acties, terwijl bij het lezen van beschrijvingen van hoe zaken eruit zien vooral hersengebieden actief worden die betrokken zijn bij het waarnemen van objecten, en bij het lezen van beschrijvingen van gedachten en emoties vooral hersengebieden actief worden die betrokken zijn bij het zogenoemde *mentalizing*: het begrijpen van wat er omgaat in andere mensen.

Wat verder opviel was dat die gevonden effecten in de oogbewegingen en de hersenactiviteit varieerde *tussen* personen: sommige mensen reageren veel sterker op die beschrijvingen dan anderen. Dat verschil tussen mensen vind ik interessant: mogelijk zijn er nogal wat verschillen tussen mensen in hoeveel ze zich voorstellen tijdens het lezen. Het is in ieder geval *mogelijk* om je dingen voor te stellen, maar niet *noodzakelijk* - mensen kunnen ook verhalen lezen zonder zich dingen voor te stellen.

Die verschillen tussen mensen vormen de kwestie waar ik me in het laatste deel van mijn onderzoek over heb gebogen. Hierbij heb ik specifiek gekeken naar verschillen in waardering voor verhalen. Voor mijn eerdere experimenten had ik een vragenlijst voor waardering gemaakt die bestond uit twee onderdelen: een soort rapportcijfer voor het verhaal (hoe leuk vond je het verhaal, op een schaal van...), en een evaluatie van het verhaal aan de hand van een aantal bijvoeglijk naamwoorden (in hoeverre vond je het verhaal tragisch, grappig, saai et cetera). Wat ik me afvroeg was hoe mensen tot een oordeel kwamen over hoe leuk ze het verhaal vonden: is dat bijvoorbeeld omdat ze het verhaal tragisch vinden, of grappig, of saai? De bijvoeglijk naamwoorden op de lijst bleken in te delen in categorieën van waarderingen die konden voorspellen wat iemand van het verhaal vond. Mensen konden het verhaal verdrietig, interessant, spannend, vermakelijk of mooi vinden, en elk van die categorieën droegen in meer of mindere mate bij aan het leuk vinden van een verhaal. Wat opviel, was weer de aanzienlijke mate waarin mensen op dit punt verschillen: sommige mensen vonden verhalen leuk omdat ze die interessant, vermakelijk of mooi vonden, maar anderen juist omdat ze de verhalen verdrietig of spannend vonden.

Samenvattend komt het erop neer dat mensen zich inderdaad kunnen voorstellen wat er gebeurt in verhalen, maar dat dit proces vrij automatisch gaat (je kunt je niet op bevel dingen voorstellen tijdens het lezen). Dat dit voorstellen in elk geval mogelijk is, is te zien aan zowel oogbewegingen als aan hersenactiviteit tijdens het lezen. Verder is te zien dat mensen op deze punten verschillen: de één laat een veel groter effect zien van dit voorstellen dan de ander. Die individuele verschillen hangen deels samen met verschillen in bepaalde aanleg die mensen kunnen hebben, bijvoorbeeld voor fantasie of voor het innemen van het perspectief van een ander. De verschillen hangen ook deels samen worden met de waardering die mensen hebben voor een verhaal of de aandacht die ze hebben voor de verhaalwereld, maar deze samenhang lijkt een stuk minder sterk te zijn dan die met de aanleg die mensen al hebben.

Deze bevindingen vertellen ons iets over de manier waarop mensen taal verwerken. Er wordt in de wetenschap gediscussieerd of taal wel of niet gegrond is in onze ervaringen in het dagelijks leven. Een uiterste standpunt is dat alle taal uiteindelijk terug te voeren is op ons lichaam: als kind voelen, horen, ruiken en zien we dingen, en aan die ervaringen geven we op den duur woorden. De betekenis van die woorden is in ons brein opgeslagen bij de lichamelijke ervaringen. Een andere uiterste is dat het lichaam niet nodig is voor taal: taal bestaat juist uit abstracte symbolen die verder niets te maken hebben met ons lichaam of met onze ervaringen uit het dagelijks leven. De betekenis van woorden staat dan dus los van die lichamelijke ervaringen. In dit proefschrift heb ik gekeken naar het voorstellen van zaken uit verhalen – en dan specifiek die dingen die je kunt relateren aan (zintuiglijke) ervaringen uit je eigen leven. Als er bijvoorbeeld in een verhaal wordt gesproken over een *zachte, rode sjaal*, stel je je die sjaal dan tijdens het lezen voor? Als dat zo is, verwacht je dat tijdens het lezen hersenactiviteit te zien is die ook te zien is bij het voelen en zien van die zachte, rode sjaal. Veel mensen blijken inderdaad dit soort voorstellingen te maken als ze verhalen lezen, en op basis van de hersenactiviteit die ik zag is te zien dat inderdaad gebieden in de hersenen actief worden die ook actief zijn bij het verwerken van ervaringen in het dagelijks leven. Maar: er was verschil te zien tussen mensen, sommige mensen leken zich veel meer voorstellingen te maken tijdens het lezen dan anderen. De mensen die zich minder voorstellingen maken, verwerken taal mogelijk op een manier die veel minder gegrond is in hun ervaringen uit het dagelijks leven. De resultaten uit dit proefschrift laten dus zien dat taal deels wel terug te voeren is op ons lichaam of ervaringen uit het dagelijks leven, maar waarschijnlijk niet alle taal, en zeker niet even sterk in alle mensen.

Los van de wetenschappelijke relevantie van de resultaten uit dit proefschrift, zijn de resultaten van dit onderzoeken zijn relevant voor theorieën en strategieën rondom leesbevordering. Het aanwakkeren van het voorstellingsvermogen wordt wel eens gezien als een manier om lezen interessanter of aantrekkelijker te maken. Uit dit onderzoek bleek echter dat dit niet eenvoudig is. Simpele instructies om lezers zich op bevel dingen te laten voorstellen bleken de waardering voor verhalen niet te beïnvloeden. Als men via het voorstellingsvermogen lezen wil stimuleren, zal er een manier gevonden moeten worden om het voorstellingsproces te trainen.

Daarnaast zijn de individuele verschillen ook van belang: de één stelt zich uit zichzelf veel meer voor tijdens het lezen dan de ander, en hoe mensen ertoe komen om een verhaal leuk te vinden verschilt ook tussen personen. Aangezien motivatie voor lezen een gevolg is van positieve leeservaringen, is het dus belangrijk dat mensen van jongs af aan de boeken en verhalen lezen die zij zelf leuk vinden. In het literatuuronderwijs is het daarom ook belangrijk om in te spelen op die individuele verschillen: iedereen hetzelfde boek laten lezen zal voor veel leerlingen niet bevorderlijk werken. Het is juist belangrijk dat iedereen datgene leest wat aansluit bij zijn of haar persoonlijke voorkeur en manier van taalverwerking. Juist dan is de kans groot op positieve leeservaringen, wat gerelateerd is aan een hogere motivatie om meer te gaan lezen, met alle positieve gevolgen van dien.

English Summary

In this dissertation, I raise the question whether people envision elements of stories while reading, how they envision these elements, whether everyone does this the same way, and whether this process is also related to how people appreciate the stories they read.

I studied this question in three experiments. In the first experiment, people were asked to read stories after receiving a reading instruction. Some were instructed to envision the events in the stories as vividly as possible, others were simply told to read as they would normally read, and still others were told to pay attention to the writing style rather than the content of the story. I expected that people who were instructed to envision the events vividly, would indicate on a questionnaire that they had pictured the story more clearly than the other two groups. However, this did not prove to be the case convincingly. The question was: is this because people do not envision anything while reading, or because they simply cannot do this *on command*?

In the other two experiments I took a different approach: I used no more reading instructions, but let everyone read as they normally would, and observed what happened. In the second experiment I asked people to read stories from a computer screen while I followed their eye movements with an eye-tracker, and in the third experiment I asked people to read stories while lying in an MRI scanner, *and* at the same time I also followed their eye movements with an eye-tracker. I then looked at three types of descriptions from the stories that readers could envision: descriptions of actions, descriptions of what things look like (for example, characters, objects, or the environment), and descriptions of thoughts and emotions.

What I found was that, on average, people read faster when they read descriptions of actions, and slower when they read descriptions of what things look like, or of thoughts and emotions. I also saw from the brain activity that when reading descriptions of actions, brain areas involved in perceiving and planning actions become active, whereas when reading descriptions of what things look like, brain areas involved in perceiving objects become active, and when rea-

ding descriptions of thoughts and emotions, brain areas involved in *mentalizing* (understanding what is going on in other people) become active.

Strikingly, the effects found in eye movements and brain activity varied *between* individuals: some people respond much more strongly to descriptive language than others. I find that difference between people interesting: possibly there are quite a few differences between people in how much they envision during reading. In any case, it is *possible* to envision events in stories, but not *necessary* - people can also read stories without envisioning anything.

The difference between people is the issue I considered in the last part of my research. In doing so, I looked specifically at differences in appreciation for stories. For my earlier experiments, I had created a rating questionnaire that consisted of two parts: one part where readers were asked to rate the story (how much did you like the story, on a scale of...), and a part where readers could evaluate the story using a number of adjectives (to what extent did you find the story tragic, funny, boring et cetera). I wanted to find out how people come to a judgment about how much they liked the story: for example, is it because they found the story tragic, or funny, or boring? Based on the answers on the questonnaire, the adjectives on the list could be divided into categories of ratings that predicted what someone thought of the story. People could find the story sad, interesting, exciting, entertaining, or beautiful, and each of those categories contributed to liking a story to a greater or lesser extent. What was striking, again, was the significant degree to which people differed on this point: some people liked stories because they found them interesting, entertaining or beautiful, whereas others liked stories because they found the stories sad or exciting.

In summary, people can indeed envision the events happening in stories, and this process is quite automatic (one cannot envision things *on command* while reading). That this envisioning is at least possible can be seen in both eye movements and brain activity during reading. Furthermore, people differ on these points: some show a much greater effect of this envisioning than others. These individual differences are partly related to differences in certain predispositions that people may have, for example, for fantasy or perspective taking. The differences are also partly related to the appreciation people have for a story or the attention they have for the story world, but this correlation seems to be a lot weaker than that with the predispositions people already have.

These findings tell us something about the way people process language. Scientists debate whether or not language is grounded in our experiences in everyday life. On the one hand, there is the position that all language can ultimately be traced back to our bodies: as children, we feel, hear, smell and see things, and we eventually give words to those experiences. The meaning of those words is stored in our brain alongside the bodily experiences. On the other hand, there is the position that the body is not necessary for language: instead, language consists of abstract symbols that have nothing to do with our bodies or with our daily life experiences. In that view, the meaning of words is independent of bodily experiences. In this dissertation I looked at the envisioning of elements of stories - and then specifically those elements that can be related to (sensory) experiences from daily life. For example, if a story features a *soft, red scarf*, do people envision that scarf while reading? If so, one would expect to see brain activity during reading that can also be seen when feeling and seeing that soft, red scarf. Many people do indeed turn out to envision elements of stories when they read, and based on the brain activity I saw, it is indeed the case that areas of the brain become active that are also active when processing experiences in everyday life. But: a difference can be seen between people, some people seem to envision much more while reading than others. The people that envision less during reading, may process language in a way that is much less grounded in their everyday life experiences. Thus, the results from this dissertation show that while some language can be traced back to our bodies or daily life experiences, this is probably not the case for all language, and it is certainly not equally the case in all people.

Apart from the scientific relevance of the results from this dissertation, the results of this research are also relevant to theories and strategies surrounding reading promotion. Stimulating the process of envisioning elements of stories is sometimes seen as a way to make reading more interesting or appealing. However, the research reported in this dissertation showed that this is not easy. Simple instructions to get readers to envision elements of stories on command were not found to affect story appreciation. If reading is to be stimulated through envisioning, a way will have to be found to train this process.

In addition, the found individual differences are also important in the light of reading promotion: some people envision much more during reading than others, and how people come to like a story also differs between individuals.

Since reading motivation is a consequence of positive reading experiences, it is important that people read the books and stories they personally enjoy from an early age. In teaching literature, therefore, it is important to accommodate these individual differences: having everyone read the same book will be demotivating for many students. On the contrary, it is important that everyone reads books that fit their personal preferences and ways of language processing. The occurence of positive reading experiences will then be more likely, which often results in a higher motivation to read more, with all its positive consequences.

Acknowledgements

Completing a doctoral dissertation is often compared to giving birth. Having recently experienced both, I am not sure which is more intense, the delivery of a dissertation or the delivery of a human being. Either way, both are difficult to survive without other people. Here, I want to thank those people that have helped me with the first.

The very first person on this list is of course my co-promotor and daily supervisor, dr. Roel Willems. Thank you for trusting me to work on your project, and thank you for guiding me every step of the way. You have helped me discover who I am as a researcher and shown me that I could do this job without drowning in work. I also want to thank you for your patience and empathy when I got ill. It was very frustrating having to pause the project, but you encouraged me to take time to heal and stood up for me when it came to contract extensions. I am very happy to have been able to work on this dissertation with you as my supervisor.

Then, my promotor, prof. dr. José Sanders. You joined the project at a later stage, but your feedback and guidance when I was putting this dissertation together has been very helpful! Thank you for taking the time to read and comment on my work so tirelessly, even though you might be the busiest person on this planet. Thank you for helping me think about things I would otherwise have forgotten to consider.

I also want to thank my manuscript committee, prof. dr. Wilbert Spooren, prof. dr. Enny Das, prof. dr. Frank Hakemulder, dr. Laura Speed and dr. Moniek Kuijpers, for taking the time to read and comment on my dissertation. Thanks, also, to dr. Frederik van Dam, for joining the manuscript committee for my doctoral defense.

I am forever grateful for the amazing people who have closely worked with me on the studies reported in this dissertation, in particular Clarissa de Vries, Xin Gao, Lynn Eekhof and Myrthe Faber. I want to thank students Iris, Maarten, Cecile, Rosa and Katharina for their help with the data collection for Chapter

2. I also want to thank all Narrative Cognition & Communication members for their feedback on ongoing studies.

Writing this dissertation would have been much more difficult without the courses, workshops, but also the social events provided by my graduate schools. For putting all these things together, and for always being there to lend an ear, I want to thank Kevin Lam at the International Max Planck Research School for Language Sciences and Peter van der Heiden at the Graduate School for the Humanities

For their practical support, I also want to thank Margret van Beuningen and Bob Rosbag at the CLS lab, and Paul Gaalman at the MRI lab at the DCCN. A special thank you goes out to Karin Heidlmayr for showing me around in the fascinating world of combined fMRI and eye-tracking, and for sharing your code with me.

Thanks to the IGEL community, for your interest in and feedback on my work, in particular Moniek, Jana, Caroline, Victoria, but of course all the other members who have taken the time to discuss my work with me during training schools and conferences. Thank you for trusting me enough to let me join the board, it is an honour and a great experience.

Finally, it has been impossible to survive this era of my life without my social network, both in and outside of the workplace. I want to thank my office mates, co-PhD's and coffee & lunch buddies, in no particular order: Linde, Inge, Alessandro, Liza, Ilona, Lynn, Rianne, Beatrice, Alex, Mesian, Elena. Thank you for always being there for me, for joking around with me and helping me find a distraction when I needed to just not think for a minute, and for hearing me out whenever I got stuck. The past years have not always been easy for me, but it was always a relief to see your friendly faces!

And of course I should not forget my paranymphs and carpool-buddies Lynn Eekhof and Myrthe Faber. I love how we can discuss anything work-related *and* non-work-related. Lynn, you first came to our "group"(then consisting of Roel, you and myself, if I'm not mistaken), only a few months after the start of my PhD. It was a great pleasure to have you along for the ride. I greatly enjoyed our train journeys to Utrecht, our lunches, our walks, our brainstorming sessions, conference visits (or visit? Was it just once?) and our working together as a whole. Myrthe, I remember meeting you for the first time at the DCCN (was it after my Project Proposal Meeting?). I also remember being completely intimidated by how smart you are and how much you know about everything. I

don't know how I would have survived the fMRI project without you (and Koen, of course). Thank you for the practical help with data collection when I got ill and couldn't come to the lab myself, and for the endless Zoom calls when I was pre-processing and analyzing the data. And of course a big thank you for sharing all your recently obtained motherhood-wisdom with me, it has been incredibly special to be on this journey together.

Thanks to all my other friends for being there. Thanks especially to Loes and Sanne, you made travelling to Nijmegen much more fun when I could have dinner with you guys after work. Thanks Irene, and Geke, for the game nights, weekends in Belgium, evening walks, good talks.

My family. I don't think I would be where I am today without you. Papa and Lena, thanks for motivating me to pursue a PhD and this project in particular. Thanks for all the dinner conversations about my and your work (apologies to Emma, Martijn, Iwan, Liza and Huibert-Jan for having to sit through all this). It has been so helpful to have the both of you around, with all your knowledge and insight into the academic world and projects such as this. Emma and Martijn, thank you for being the best brother and sister. Let's stay this close for the rest of our lives.

I don't know how to thank you, Huibert-Jan. Without you, these past years would have been completely different. Without you, I don't know if I would have had the courage to apply to this project, or to visit conferences around the world. Thank you for being by my side through all the highs and lows. And thank you for being a life outside of academia for me. Without you, I think I would have lost myself in this project. Thank you for keeping me grounded.

And finally, Anne, the newest love of my life. Some day you will be able to read this. Lieve Anne, aan het begin had ik het over twee bevallingen, twee weken en vier dagen na elkaar. Op jou ben ik het trotst.

Curriculum Vitae

Marloes Mak MSc

Radboud University Nijmegen Department of Language and Communication

Centre for Language Studies

Erasmusplein 1 | 6525HT Nijmegen | The Netherlands | +31616684714 | marloes.mak@ru.nl

EDUCATION

PhD candidate **2016 – present**
Centre for Language Studies
Radboud University Nijmegen, The Netherlands

Effects of sensori-motor and emotional simulation on aesthetic appreciation of literary narratives

Supervisors: Dr. Roel Willems, Prof. Dr. José Sanders
Dissertation to be finalized within 2021.

International Max Planck Research School **2016 – present**
IMPRS for Language Sciences
Max Planck Institute for Psycholinguistics, Donders Institute for Brain Cognition and Behaviour, and Centre for Language studies
Cohort 2016

IGEL summer school **2018**
Summer school organized by the International Society for the Empirical Study of Literature (IGEL), Stavanger, Norway

Psychometrics: designing your own measuring instrument

Instructors: Dr. Moniek Kuijpers, Em. Prof. Dr. Don Kuiken, Prof. Dr. Arthur Jacobs, Dr. Jana Lüdtke

MSc in Neuropsychology **Graduation: March 2015**
Utrecht University, The Netherlands

Thesis: *Brain reserve and cognitive reserve: Effect modification in the relationship between hippocampal volume and memory functioning* (within the SMART-MR study of the Julius Centre at UMC Utrecht)

Von Humboldt College **Graduation: January 2014**
Honours Bachelor program of the Faculty of Social Sciences
Utrecht University, The Netherlands
Track: Neuropsychology & Cognitive and Neurobiological psychology

BSc in Psychology **Graduation: January 2014**
Utrecht University, The Netherlands
Track: Neuropsychology

ACADEMIC POSITIONS

Teacher **2021 – 2022**
Department of Language, Literature and Communication, Utrecht University, The Netherlands

Researcher **2021**
Centre for Language Studies & Department of Language and Communication, Radboud University Nijmegen, The Netherlands

Research Assistant **2016**
Department of Language, Literature and Communication, Utrecht University, The Netherlands

Junior Teacher **2016**
Department of Language, Literature and Communication, Utrecht University, The Netherlands

Student Assistant **2015**
Department of Language, Literature and Communication, Utrecht University, The Netherlands

Research Internship **2012 – 2013**
Department of Language, Literature and Communication, Utrecht University, The Netherlands

ACADEMIC SERVICE

Board member (Treasurer)
International Society for the Empirical Study of Literature (IGEL)

Founder & Organiser of EyeMeth
Group meetings (with the purpose of discussing methodological issues regarding eye-tracking) for researchers working with or planning to work with eye-tracking in the lab of the Centre for Language Studies at Radboud University Nijmegen

Ad-Hoc Reviewer
Poetics; Language, Cognition and Neuroscience; Journal of Eye Movement Research; Cognitive Processing; Behavior Research Methods

PUBLICATIONS

Mak, M., Faber, M., & Willems, R.M. (2022). Different routes to liking: How readers arrive at narrative evaluations. *Cognitive Research: Principles and Implications,* 7, 72. doi: 10.1186/s41235-022-00419-0

Hartung, F., Wang, Y., **Mak, M.**, Willems, R.M., & Chatterjee, A. (2021). Aesthetic appraisals of literary style and emotional intensity in

PUBLICATIONS (CONT)

narrative engagement are neurally dissociable. *Communications Biology, 4,* 1401. doi: 10.1038/s42003-021-02926-0

Mak, M., & Willems, R.M. (2021). Eyelit: Eye-movement and reader response data during literary reading. *Journal of Open Humanities Data, 7*, 1–6. doi: 10.5334/johd.49

Mak, M., & Willems, R.M. (2021). Mental Simulation during Literary Reading. In Kuiken, D., & Jacobs, A.M. (Eds.), *Handbook of Empirical Literary Studies* (pp. 63-84). Berlin, Boston: De Gruyter. Chapter doi: 10.1515/9783110645958-004

Mak, H.M.L., & Willems, R.M. (2021). *Eyelit: Eye-movement and reader response data during literary reading.* DANS-EASY [Dataset]. doi: 10.17026/dans-zqk-zmqs

Eekhof, L.S., Kuijpers, M.M., Faber, M., Gao, X., **Mak, M.**, Van den Hoven, E., & Willems, R.M. (2021). Lost in a Story, Detached from the Words. *Discourse Processes, 58*(7), 595-616. doi: 10.1080/0163853X.2020.1857619

Mak, M., Vries, C. de, & Willems, R.M. (2020). The Influence of Mental Imagery Instructions and Personality Characteristics on Reading Experiences. *Collabra: Psychology, 6*(1), 43. doi: 10.1525/collabra.281

Faber, M., **Mak, M.**, & Willems, R.M. (2020). Word skipping as an indicator of individual reading style during literary reading. *Journal of Eye Movement Research, 13*(3), 2. doi: 10.16910/jemr.13.3.2

Mak, M., & Willems, R.M. (2019). Mental Simulation during Literary Reading: Individual Differences Revealed with Eye-Tracking. *Language, Cognition and Neuroscience, 34* (4), 511-535. doi: 10.1080/23273798.2018.1552007

Tribushinina, E., **Mak, M.**, Dubinkina, E., & Mak, W.M. (2018). Adjective production by Russian-speaking children with Developmental Language Disorder and Dutch-Russian simultaneous bilinguals: Disentangling the profiles. *Applied Psycholinguistics, 39*(5), 1033-1064. doi: 10.1017/S0142716418000115

SUBMITTED

Mak, M., Faber, M., & Willems, R.M. (submitted). Different kinds of simulation during literary reading: Insights from a combined fMRI and eye-tracking study.

INVITED TALKS

Mental simulation and narratives. Evidence from eye tracking and fMRI (2021). Freie Universität Berlin, Germany.

The subjective experience of stories: Measuring mental simulation and aesthetic appreciation during literary reading (2019). Max Planck Institute for Empirical Aesthetics, Frankfurt, Germany.

Differential influences of sensory simulation, motor simulation and mentalizing in literary reading: An eye-tracking study (2017). Centre for Language Studies Colloquium Lunch, Radboud University Nijmegen, The Netherlands.

Using eye movement data to measure mental simulation (postponed). Workshop *Utilizing eye tracking in reading research*, Centre of Cognitive Science, Freiburg, Germany

CONFERENCE ABSTRACTS

Mak, M., Faber, M., & Willems, R.M. (2022). Different kinds of simulation during literary reading: Insights from a combined fMRI and eye tracking study. The 21st European Conference on Eye Movements, Leicester, UK (Talk).

Mak, M., Faber, M., & Willems, R.M. (2021). Aesthetic fingerprints: The case of narratives. The 17th IGEL Conference, Liverpool, UK. (Talk)

Mak, M., Faber, M., & Willems, R.M. (2019). Liking stories one way or the other: Clustering aesthetic appreciation for literary narratives. NVP Winter Conference, Egmond aan Zee, The Netherlands. (Poster)

Mak, M., & Willems, R.M. (2019). Eye-tracking narrative reading: Individual roles of perceptual simulation, motor simulation and mentalizing. The 21st meeting of the European Society for Cognitive Psychology, Tenerife, Spain. (Talk)

Mak, M., & Willems, R.M. (2019). Individual differences in reading experiences: The roles of Mental Imagery and Fantasy. The 41st Annual Meeting of the Cognitive Science Society, Montréal, Canada. (Talk)

Eekhof, L., **Mak, M.**, et al. (2019). Lost in a story, detached from the words. Grote Taaldag, Utrecht, The Netherlands. (Talk)

Mak, M., & Willems, R.M. (2018). Embodiment in narratives: An eye-tracking study of perceptual simulation, motor simulation and mentalizing during narrative reading. The 11th Embodied & Situated Language Processing Conference, Lancaster, UK. (Talk)

CONFERENCE ABSTRACTS (CONT)

Mak, M., & Willems, R.M. (2018). Individual differences in mental simulation during literary reading revealed with eye-tracking. The 16th IGEL Conference, Stavanger, Norway. (Talk)

Mak, M., & Willems, R.M. (2018). Sensory Simulation, Motor Simulation and Mentalizing during Narrative Reading: Insights from Eye-Tracking. IMPRS Conference on Interdisciplinary Approaches in the Language Sciences, Nijmegen, The Netherlands. (Poster)

Mak, M., & Willems, R.M. (2018). Sensory Simulation, Motor Simulation and Mentalizing during Narrative Reading: Insights from Eye-Tracking. CNS 2018 Annual Meeting, Boston, MA, USA. (Poster)

Mak, M., & Willems, R.M. (2018). Differential influences of sensory simulation, motor simulation and mentalizing in literary reading: An eye-tracking study. Grote Taaldag, Utrecht, the Netherlands. (Talk)

Mak, M., & Willems, R.M. (2017). Insights into mental simulation via narratives. NVP Winter Conference, Egmond aan Zee, The Netherlands. (Talk)

Mak, M., & Willems, R.M. (2017). The role of simulation during narrative reading: Insights from eyetracking. Event Representations in Brain and Language Development workshop, Nijmegen, The Netherlands. (Poster)

PUBLIC OUTREACH

Participation in De Kennis Van Nu (featuring an experimental paradigm I designed for my dissertation). Popular Science TV Show, episode “Gedachten de baas” (*Master of your mind*) (2019)

Interview with NEMO Kennislink (Popular Science website, linked to a Science Museum in Amsterdam; 2018). https://www.nemokennislink.nl/publicaties/boeken-als-oefenmateriaal-voor-empathie/ (in Dutch)

Scientific jury member at Manifestatie Wetenschaporiëntatie Nederland (2018; Annual finals of a Dutch country-wide science contest for secondary school (12 to 18 years old) students)

Invited Talk InScience Festival (Popular Science Festival; 2017).

Introductory talk and scientific jury member at Manifestatie Wetenschaporiëntatie Nederland (2017; Annual finals of a Dutch country-wide science contest for secondary school (12 to 18 years old) students).

PUBLIC OUTREACH (CONT)

Finalist in the Radboud Talks Science Communication Contest (3-minute TED-style talks; 2017). https://youtu.be/2KDwLo4kbC0 (in English)

Workshop "Experiment van de verbeelding" (*experiment of the imagination*) at the Wintertuin Festival (Popular Science Festival centered around Literature; 2016).

TEACHING AND SUPERVISION

GUEST LECTURES

Advanced empirical methods: Behavioral and neuroimaging methods and the case of literary reading.

Guest lecturer. MA Literary Studies and Linguistics, Institut für Anglistik, Amerikanistik und Romanistik, RWTH Aachen University, Germany. (2018)

Workshop Eye-tracking: Wat zeggen onze oogbewegingen over wat we ervaren tijdens het lezen?

Workshop for secondary school students, at the Manifestatie Wetenschaporiëntatie Nederland (2018; Annual finals of a Dutch country-wide science contest for secondary school (12 to 18 years old) students)

Bronnen zoeken.

Lecture and practical about finding scientific sources for secondary school students working on their research projects. Pre-University College Humanities. (2018)

STUDENT SUPERVISION

BA Supervisor **Nov 2021 – Jan 2022**
Supervising five BA students of Communication and Information Studies. Utrecht University, The Netherlands

MA Supervisor **Feb 2018 – June 2018**
Co-supervisor of a MA student Communication and Information Studies. Radboud University Nijmegen, The Netherlands

Research Internship Supervisor **Sep 2017 – Jan 2018**
Co-supervising a research intern (BA). Utrecht University & Radboud University Nijmegen, The Netherlands

TEACHING AND SUPERVISION (CONT)

BA Supervisor **Feb 2017 – June 2017**
Supervising six BA students of Communication and Information Studies. Radboud University Nijmegen, The Netherlands

COURSES

Methods and Statistics 1 **2016**
Teaching work groups (two groups). Department of Language, Literature and Communication, Utrecht University, The Netherlands

Cognitive Neuroscience for Humanities Scholars **2021**
Teaching work groups (three groups). Department of Language, Literature and Communication, Utrecht University, The Netherlands

Methods of Communication Research **2016**
Teaching work groups and practicals (two groups). Department of Language, Literature and Communication, Utrecht University, The Netherlands