



Research Report

Different kinds of simulation during literary reading: Insights from a combined fMRI and eye-tracking study



Marloes Mak ^{a,*}, Myrthe Faber ^{b,c} and Roel M. Willems ^{a,b,d}

^a Centre for Language Studies, Radboud University Nijmegen, Erasmusplein 1, 6525 HT Nijmegen, the Netherlands

^b Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, Kapittelweg 29, 6525 EN Nijmegen, the Netherlands

^c Department of Communication and Cognition, Tilburg Center for Cognition and Communication, Tilburg University, Warandelaan 2, 5037 AB Tilburg, the Netherlands

^d Max Planck Institute for Psycholinguistics, Wundtlaan 1, 6525 XD Nijmegen, the Netherlands

ARTICLE INFO

Article history:

Received 22 April 2022

Reviewed 22 June 2022

Revised 2 November 2022

Accepted 22 January 2023

Action editor Swathi Kiran

Published online 15 March 2023

Keywords:

Mental simulation

Literary reading

fMRI

Eye-tracking

Individual differences

ABSTRACT

Mental simulation is an important aspect of narrative reading. In a previous study, we 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, we investigated the existence of a common neural locus for these different kinds of simulation. We additionally investigated whether individual differences during reading, as indexed by the eye movements, are reflected in domain-specific activations in the brain. We 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).

© 2023 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

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

* Corresponding author.

E-mail address: marloes.mak@ru.nl (M. Mak).

<https://doi.org/10.1016/j.cortex.2023.01.014>

0010-9452/© 2023 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

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 a previous eye-tracking study we found that the three different kinds of mental simulation as described by Barsalou indeed have qualitatively different effects on eye movements during reading ([Mak & Willems, 2019](#)). Here we ask whether these different kinds of simulation share an overlapping neural locus, and/or are domain-specific processes. We 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, we attempt to find out whether there are individual differences in brain activation levels in the brain regions implicated in simulation.

Before explaining the hypotheses for the current study, we will first discuss what is known about eye movements and brain activation during reading.

1.1. Eye movements and brain activation during reading

A lot of research into eye movements during reading has shown that eye tracking can measure attention (e.g., to words, passages), and processing speed during reading (e.g., [Rayner, 1998, 2009](#)). For example, increased attention to certain words or passages is associated with longer gaze duration. This can be seen in important text features such as the lexical frequency of words or word length. 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. Similarly, shorter words are easier to process and thus warrant less attention than longer words. Indeed, highly frequent words and shorter words both are associated with shorter gaze durations (e.g., 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 predictable 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](#)). The predictability of a word within its context will be referred to as *surprisal* for the remainder of this paper ([Lopopolo, Frank, van den Bosch, & Willems, 2017](#)).

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.

Apart from eye movements, brain activation has also been found to be informative in the study of 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](#)). Activity in these brain areas during reading are modulated by word characteristics, such as word length, frequency and surprisal ([Desai, Choi, & Henderson, 2020](#); [Schuster, Hawelka, Hutzler, Kronbichler, & Richlan, 2016](#); [Shain, Blank, van Schijndel, Schuler, & Fedorenko, 2020](#)).

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](#)).

The networks described so far are related to the mental processing of language and its integration into a wider context and a wider knowledge about the world. However, reading is also a behavior in a more physical sense. When people are reading, they are making eye movements, and these eye movements are accompanied by brain activation. Therefore, it is important to point out what areas of the brain are involved in the execution of eye movements, even in the absence of congruent language. When comparing word reading with pseudoword reading (under the assumption that both require eye movements, but only word reading requires language processing), [Richlan et al. \(2014\)](#) found that eye movements in general activate mainly bilateral occipito-parietal brain areas,

whereas word reading (or, language processing) also activates language processing networks. Similarly, Choi, Desai, and Henderson (2014) found that eye movements were associated with activity in an eye movement network including the frontal eye fields, supplementary eye fields and intraparietal sulci, as well as an attentional network including the cingulate and parietal cortex.

1.2. Differences between motor, perceptual and mental simulation: eye movements

In our previous eye tracking study (Mak & Willems, 2019) we found differences in reading behavior between perceptual simulation, motor simulation, and mentalizing. 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 paper, we 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, we aim to replicate and extend the findings of our previous paper (Mak & Willems, 2019) in the context of a combined eye-tracking and fMRI experiment.

1.3. 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).

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, Lai, Hagoort, & Willems, 2015); left superior temporal gyrus, bilateral superior temporal sulcus, but also perisylvian 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., anterior medial prefrontal cortex (aMPFC), dorsomedial prefrontal cortex (dMPFC), anterior or posterior cingulate cortex, temporoparietal junction (TPJ); Frith & Frith, 2003; Hsu, Conrad, & Jacobs, 2014; Lai, Willems, & Hagoort, 2015; Nijhof & Willems, 2015; Saxe & Kanwisher, 2003; Tamir, Bricker, Dodell-Feder, & Mitchell, 2016).

A final possibility would be that we see both a common “situation-model building” simulation area, and modality-specific brain areas that respond to one specific kind of simulation. Evidence for both options has been found before (as has been reported above), but to our knowledge no previous studies have investigated the possibility of both common simulation areas and modality specific brain areas for all three kinds of simulation.

1.4. Individual differences in the effect of simulation

A second, more exploratory, question that we 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, we 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).

1.5. 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 us to link eye-tracking data to neuroimaging data, within participants. We were interested in the responses of participants at the word level: we 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 (underlinings) of motor, perceptual, and mental event content was acquired in a separate pretest with different participants.

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 and (while still in the scanner) 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.

2. Methods

We report how we determined our sample size, all data exclusions, all inclusion/exclusion criteria, whether inclusion/exclusion criteria were established prior to data analysis, all manipulations, and all measures in the study.

2.1. Pretest

In a pretest 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 pretest: 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 behaviour”. 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 to 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. 1. There was no clear association between the ratings per word for motor, perceptual and mental event descriptions on the one hand, and other word characteristics on the other hand (i.e., lexical frequency, word length, surprisal; see Appendix A). The rationale for using these ratings instead of, for example, existing ratings of concreteness (which is highly correlated with imageability; Brysbaert, Stevens, De Deyne, Voorspoels, & Storms, 2014) or sensory modality (Speed & Brysbaert, 2022), is that the ratings obtained in our pretest take the context of the stories into

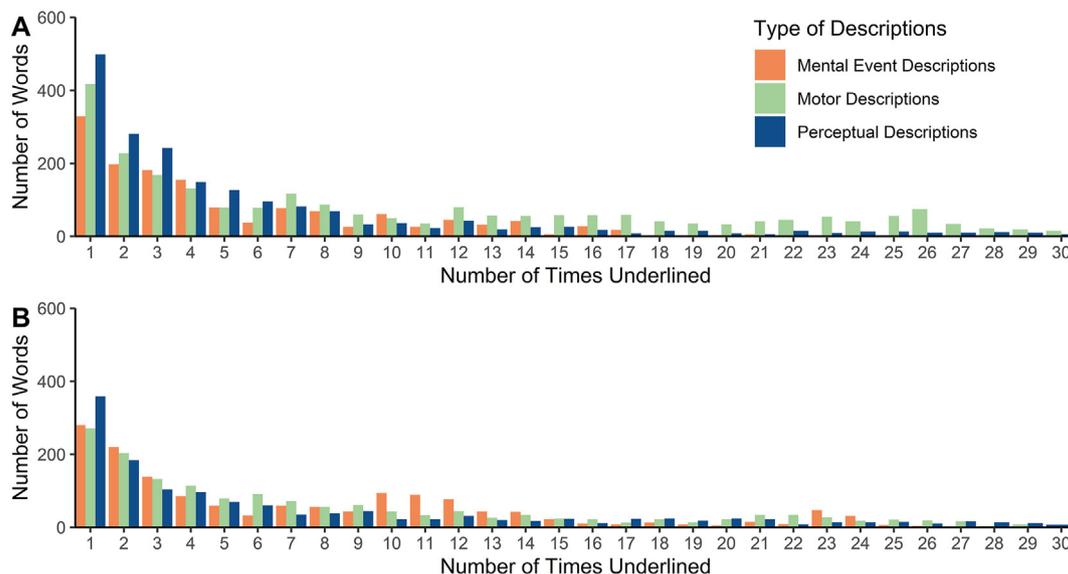


Fig. 1 – Number of times words were underlined for mental event, motor and perceptual descriptions in stories (A) *The people who had everything delivered* and (B) *Symbols and Signs*. Note: in story A, 1562 words, 659 words and 1070 words were underlined by none of the pre-test participants for mental 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.

account. Since we were interested in mental simulation during the reading of complete literary short stories (a contextualized process; see [Willems & Peelen, 2021](#)), we believed it fitting to use contextualized ratings.

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 based on [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 [Jobe, 2009](#)). We 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. All inclusion and exclusion criteria were established prior to data collection. 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).

2.3. Materials

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 & 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 [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 min 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. Legal copyright restrictions prevent public archiving of the stories used in this experiment which can be obtained from the copyright holders in the cited references.

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).

2.3.2.1. READING EXPERIENCES. Reading experiences (i.e., story world absorption, story appreciation) were measured using questionnaires. Story world absorption was measured with the Story World Absorption Scale (SWAS; [Kuijpers, Hakemulder, Tan, & Doicaru, 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 depict 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 rate each question on a 7-point scale (1 = disagree, 7 = agree). Legal copyright restrictions prevent public archiving of the SWAS which can be obtained from the copyright holders in the cited references. 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, Wagner, Jacobsen, & Menninghaus, 2016](#)). These adjectives are 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, under review](#)). 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 rate both the adjectives and the questions regarding enjoyment on a 7-point scale (1 = disagree, 7 = agree). The appreciation questionnaire can be found on <https://osf.io/9rwqn/>. Both the SWAS and the appreciation questionnaire were also used in the previous eye-tracking experiment.

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 have read any of the stories before. The comprehension check can be found on <https://osf.io/9rwqn/>.

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, Burke, Hagoort, & Willems, 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 reading habits questionnaire can be found on <https://osf.io/9rwqn/>.

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. Legal copyright restrictions prevent public archiving of the ART which can be obtained from the copyright holders in the cited references.

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. Legal copyright restrictions prevent public archiving of the IRI which can be obtained from the copyright holders in the cited references. We do not have legal permission to publicly archive all materials used in this study. Readers seeking access to the materials should contact the first author.

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 (see Appendix B). Prior to this experiment, it was unknown whether individual differences in reading behavior would be detectable at the whole brain level. We therefore included the localizer tasks (now described in Appendix B) to have the opportunity to obtain functional regions of interest (ROIs). One important downside of these localizer tasks, however, was that they contain decontextualized stimuli, whereas our experiment and research question were specifically aimed at the neural processes that underlie naturalistic, contextualized reading (see Willems & Peelen, 2021 for a review about the differences in processing in response to contextualized versus decontextualized stimuli). Given that our whole brain analysis yielded functional ROIs that could be interrogated further, data obtained using

the localizer tasks were not used. 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.

No part of the study procedures was pre-registered prior to the research being conducted.

2.5. Stimulus presentation

Stimuli were presented page by page on a projection screen (<http://www.macada-innovision.nl>) at the end of the bore, using a EIKI LC - XL100 beamer with a native resolution of 1024 × 768, 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 (110 × 100 mm) and the projection screen (369 × 277 mm) was 855 mm, and the distance between the mirror and the eye about 100 mm (depending on how high a participant's head lies in the head coil).

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 we 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, we did not discard the fMRI data for these participants. After preprocessing, we were 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, we needed to impute the eye fixations for these data. To stay as close as possible to the participant's natural reading behavior, we modelled the onset and duration of the fixations, but imputed the mean value of the word characteristics we 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 our analyses.

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 3 T MAGNETOM PrismaFit MR scanner (Siemens AG, Healthcare sector, Erlangen, Germany) with a 64-channel head-coil. Functional (TR = 1000 ms, TE = 34 ms, flip angle = 60°, Field of View = 210 mm, voxel size = 2.0 × 2.0 × 2.0 mm, number of slices = 66, Multi-band acceleration factor = 6, multi-slice mode = interleaved, echo spacing = .62 ms) and anatomical (Magnetization Prepared Rapid Acquisition Gradient Echo, voxel size = 1.0 × 1.0 × 1.0 mm) images were acquired in one session lasting about 60–90 min, 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 11 s) 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, 2 × 2x2mm) space using FNIRT (Andersson, Jenkinson, & Smith, 2007a; 2007b) nonlinear registration (≥ 12 DOF). Spatial smoothing was performed using SUSAN noise reduction (Smith & Brady, 1997) with a 5 mm FWHM Gaussian kernel. Grand-mean intensity normalisation 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 $\sigma = 45.0$ s.

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.

2.8. Data analysis

No part of the analyses was pre-registered prior to the research being conducted.

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. We analyzed how motor description, perceptual description 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. 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 (i.e., an indication of the accuracy of a language model, in this case the perplexity is an indication of the accuracy with which a word is predicted by the previous words, cf. Lopopolo et al., 2017), calculated using a 3-g model trained by SRILM (Stolcke, 2002) on 1 million sentences from the NLCOW2012 corpus (Schäfer & Bildhauer, 2012). Perplexity was equal to 10 to the power of negative surprisal. As in the analyses in Mak and Willems (2019), we used the values for the descriptions and word characteristics of the previous word, as this allowed us to look in the spillover regions. We 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), see <https://osf.io/8eqya/>. 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, Lævenbruck, 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 our model, we 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 we used an informative, normally-distributed prior with a mean of 250 and a standard deviation of 50, since gaze durations are generally between 200 and 300 ms 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.

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, we 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, we ran a GLM per participant (per run, one analysis for each story) where we modelled the onset and duration of each fixation, weighted by the scores for motor descriptions, perceptual descriptions, mental event descriptions, and the first principal component¹ of lexical frequency, word length and surprisal value of the word that was fixated (to control for these word characteristics). This way we 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 we 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 -0.5 -0.5] for motor > perceptual and mental event, perceptual > motor and mental event, mental event > motor and perceptual). Additionally, we 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, we 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, we 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. We 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. We selected regions of interest that were (1) significantly activated by one or all of the three types of descriptions on the group level, that were (2) part of large clusters of activation in response to the descriptions, and that were (3) good candidates for finding individual differences in simulation based on previous literature (e.g., Chow et al., 2015; Grill-Spector & Weiner, 2014; Igelström & Graziano, 2017; Kurby & Zacks, 2013; Moody & Gennari, 2010; Nijhof & Willems, 2015). We derived the regions of interest from the results of our group analysis: we extracted the by-participant by-story percent signal change from areas that were found to be commonly activated by these descriptions. We 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). We built separate models for Story World Absorption and for appreciation, to make sure that any conceptual overlap between the two would not skew our results. We 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), see <https://osf.io/8eqya/>. In our models, we 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 we 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.

¹ To avoid multicollinearity issues due to the high correlation between lexical frequency, word length and surprisal value, we entered the first principal component of these variables into the model instead of entering each of the variables separately into the model. For this principal component analysis (PCA), the Kaiser-Meyer-Olkin measure (KMO) was .71 (all KMO values for individual items >.65), indicating good sampling adequacy for this analysis. Bartlett's test of sphericity showed sufficient correlation between items, $\chi^2(3) = 597.40, p < .001$. 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) confirmed that it was appropriate to summarize the three word characteristics into one principal component. This principal component explained 84% of the variance.

3. Results

All anonymized data can be found on <https://doi.org/10.34973/hrax-yn39>. Data and scripts for the behavioral results

and individual difference analysis can be found on <https://osf.io/8eqya/>.

3.1. Behavioral results

3.1.1. Questionnaires

3.1.1.1. SWAS. The Story World Absorption Scale showed excellent reliability, Cronbach's $\alpha = .94$ (Story A: $M = 4.81$, $SD = .79$; Story B: $M = 3.55$, $SD = .83$).

3.1.1.2. APPRECIATION. In order to reduce the 13 adjectives to a smaller number of components consisting of highly similar adjectives, we 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), see <https://osf.io/8eqya/>.

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, $\chi^2(55) = 1879.28$, $p < .001$. 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 we 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 1.

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, we 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.

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, Cronbach's $\alpha = .80$.

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.

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 ($M = 5.05$, $SD = .92$) showed good reliability, Cronbach's $\alpha = .80$. The Perspective Taking subscale ($M = 4.95$, $SD = .75$) showed sufficient reliability, Cronbach's $\alpha = .76$.

3.1.2. Eye tracking data

Our 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 2, Fig. 2B). Perceptual and mental event descriptions were associated with longer gaze durations on the next word (i.e., slower reading; see Table 2, Fig. 2C and 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

Table 1 – Pattern matrix for the PCA of the 11 adjectives on the appreciation questionnaire (N = 703).

	Pattern Matrix				
	Interest	Sadness	Special	Positive Affect	Ominous
Beautiful	.31		.33	.45	.40
Boring	-.54	-.11	-.19	-.27	.25
Gripping	.59	.28	.17		.20
Entertaining	.38	-.11		.34	-.49
Funny	.17	-.16		.36	-.59
Interesting	.76	.17	.11		.14
Ominous					.89
Sad	-.17	.89			
Suspenseful	.91	-.13	-.11	-.23	
Tragic	.16	.91		-.11	
Witty			-.10	.93	
Captivating	.73			.28	-.17
Special			.99	-.10	

Note: Factor loadings over .40 appear in bold.

Table 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	.73	-4.05	-1.16
Perceptual Descriptions	9.60	.88	7.87	11.29
Mental Event Descriptions	2.69	.78	1.11	4.19
Lexical Frequency	1.45	1.14	-.76	3.70
Word Length	-.37	.87	-2.09	1.31
Surprisal Value	1.64	.96	-.23	3.51

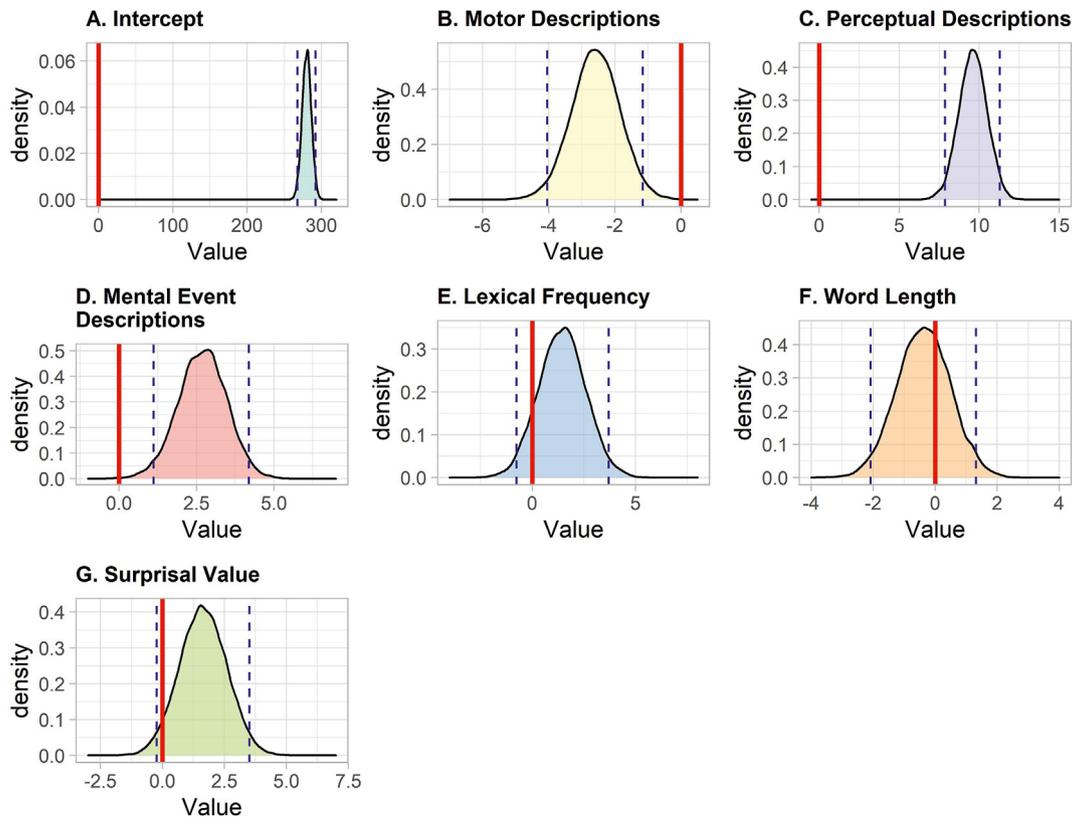


Fig. 2 – Posterior distributions of the fixed effects of the relationships between the three types of descriptions and word characteristics and gaze durations.

characteristics (lexical frequency, word length and surprisal value) did not reliably influence gaze durations in the spillover region (see Table 2, Fig. 2E and F).

Looking at the individual variation in the relationship between the three types of descriptions and gaze durations, we see 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. 3).

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 MN152 space after normalization with FNIRT*.

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 3, for each of the types of description.

3.2.1.1. MOTOR DESCRIPTIONS. Fig. 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 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.

3.2.1.2. PERCEPTUAL DESCRIPTIONS. Fig. 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, superior 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.

3.2.1.3. MENTAL EVENT DESCRIPTIONS. Fig. 6 visualizes the main results for the contrast for mental event descriptions versus

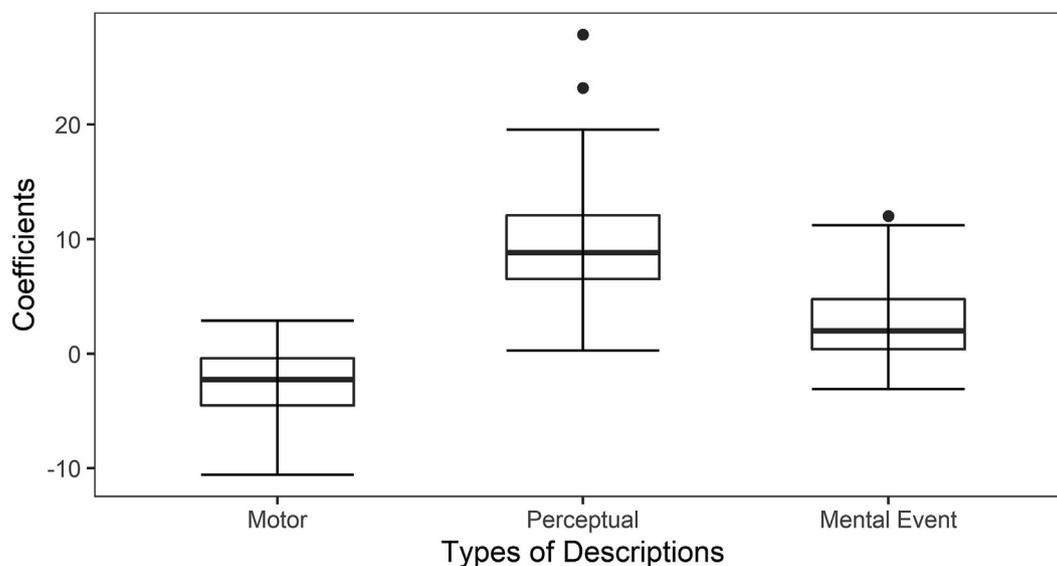


Fig. 3 – Estimated Means from participants' Individual Posterior Distributions for the Relationships Between the Types of Descriptions and Gaze Durations.

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).

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 baseline²; conjunction analysis) was found in the left anterior supramarginal gyrus (see Fig. 7; see Table 4 for the peak coordinates). No other reliable clusters of activation were visible in all three contrasts.

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, we calculated spherical regions of interest (10 mm diagonally around a peak coordinate) in regions that had been significantly activated by the three types of descriptions on the group level, and that we believed would be good candidates for finding individual differences in simulation. For the motor contrast, there regions included the left supramarginal gyrus, the left frontal orbital gyrus, the left and right precuneus, and the cingulate gyrus. For the perceptual contrast, our 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, we 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, we 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. We decided to analyze the 90th percentile of the percent signal change, meaning that within our ROIs we 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. We chose to look at the 90th percentile of the percent signal change, to make our findings more robust to outliers than if we would have looked in the 95th percentile.

We 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). We will report all statistically reliable results here (i.e., the instances where questionnaire scores were reliably associated with the percent signal change; see Fig. 8), and provide tables with all results for all ROIs in Appendix C.

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 = .10; 95% CI = .01; .18), showing that experiencing a story as special is associated with higher percent signal change in this brain

² The implicit baseline consisted of all parts of the signal where there was no fixation on words (e.g., saccades, blinks).

Table 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.

Region	L/R	Peak MNI coordinates			Z-max	Volume (voxels)
		x	y	z		
motor descriptions						
Precuneous cortex	R	6	−60	34	7.1	9408
Precuneous cortex	L	−8	−58	18	7.03	
Cingulate gyrus	L	−2	−42	42	6.59	5094
Cingulate gyrus	R	14	−50	38	6.37	
Angular gyrus	R	54	−56	16	6.69	1424
Lateral occipital cortex (superior division)	R	48	−64	30	6.08	
Lateral occipital cortex (superior division)	L	−40	−72	38	6.96	3797
Angular gyrus	L	−40	−56	14	5.61	
Superior frontal gyrus	R	20	28	46	5.93	712
Middle frontal gyrus	R	26	32	44	5.58	
Paracingulate gyrus	R	6	52	10	5.49	310
Subcallosal cortex	L	−10	28	−10	5.2	
Frontal pole	R	24	44	48	5.18	220
Frontal medial cortex	R	4	54	−8	5.01	
Superior frontal gyrus	L	−20	24	44	5.86	352
Middle frontal gyrus	L	−32	4	66	4.08	
Middle frontal gyrus	R	44	14	36	4.4	190
Lingual gyrus	L	−26	−42	−8	4.86	
Parahippocampal gyrus (posterior division)	L	−22	−40	−12	4.82	595
Cerebellum: Left Crus II	L	−10	−80	−40	5.74	
Cerebellum: Left VI	L	−14	−68	−30	4.36	352
Cerebellum: Left Crus I	L	−26	−74	−36	4.09	
Cerebellum: Left IX	L	−8	−50	−46	6.22	190
Cerebellum: Right IX	R	8	−50	−52	4.48	
Cerebellum: Left Crus II	L	−50	−52	−46	3.96	3.43
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	1016
Amygdala	L	−18	0	−20	5.36	
Inferior temporal gyrus (temporooccipital part)	L	−50	−56	−12	6.42	235
Temporal pole	L	−30	2	−32	5.71	
Temporal pole	R	26	8	−40	5.67	481
Temporal fusiform cortex (anterior division)	R	36	−6	−34	5.55	
Inferior temporal gyrus (temporooccipital part)	R	48	−50	−18	5.45	468
Parahippocampal gyrus (anterior division)	R	30	0	−34	4.94	
Temporal fusiform cortex (posterior division)	R	36	−32	−24	4.92	147
Amygdala	R	18	−4	−18	5.57	
Frontal pole	R	50	38	12	5.49	345
Interior frontal gyrus (pars triangularis)	R	44	30	16	5.27	
Precentral gyrus	L	−44	4	24	6.27	239
Inferior frontal gyrus (pars opercularis)	L	−46	8	24	6.19	
Precentral Gyrus	R	40	4	26	4.81	236
Inferior frontal gyrus (pars triangularis)	L	−44	34	14	4.54	
Frontal pole	R	30	34	−10	4.73	450
Frontal orbital cortex	R	24	28	−12	4.7	
Frontal orbital cortex	L	−24	30	−12	4.61	214
Frontal pole	L	−28	36	−20	3.4	
Lateral occipital cortex (superior division)	L	−28	−74	34	5.76	4.2
Lateral occipital cortex (superior division)	R	36	−58	48	5.24	
Angular gyrus	R	34	−54	42	4.2	
mental event descriptions						
Cerebellum: Right Crus I	R	30	−82	−32	7.63	11,753
Lingual gyrus	L	−6	−72	−4	6.74	
Lingual gyrus	R	2	−72	−2	6.48	4.06
Cuneal cortex	R	18	−74	28	4.95	
Cuneal cortex	L	−8	−82	34	4.06	5.11
Lateral occipital cortex (superior division)	R	26	−86	26	5.11	
Lateral occipital cortex (superior division)	L	−20	−85	22	4.21	6.26
Cerebellum: Right Crus II	R	30	−76	−42	6.26	

Table 3 – (continued)

Region	L/R	Peak MNI coordinates			Z-max	Volume (voxels)
		x	y	z		
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 (part triangularis)	L	−52	22	−6	6.34	2051
Temporal Pole	L	−52	16	−10	6.25	
Middle temporal gyrus (posterior division)	L	−50	−38	−4	5.02	4.95
Superior temporal gyrus (posterior division)	L	−66	−24	0	4.95	
Angular gyrus	L	−48	−50	30	4.76	4.45
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	188
Temporal pole	R	50	8	−42	4.6	
Middle temporal gyrus (posterior division)	R	60	−10	−28	3.63	3.6
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 VIIa	R	4	−68	−38	4.22	

region. In the left frontal orbital gyrus, we found a positive association between percent signal change and the Positive Affect component of appreciation (Estimate = .07, 95% CI = .03; .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, we found a positive association between percent signal change and the Fantasy subscale of the IRI (Estimate = .09, 95% CI = .03; .14), indicating that people who are prone to getting mentally involved in stories showed higher percent signal change in the cingulate gyrus.

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 = .08, 95% CI = .02; .13) and right (Estimate = .08, 95% CI = .01; .15) inferior temporal gyrus, and in the right temporal fusiform cortex (Estimate = .07, 95% CI = .00; .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 = .08, 95% CI = .02; .14) and in the left temporal fusiform cortex (Estimate = .06, 95% CI = .01; .10), indicating that people who more often take other people's perspectives in daily life show higher percent signal change in these areas.

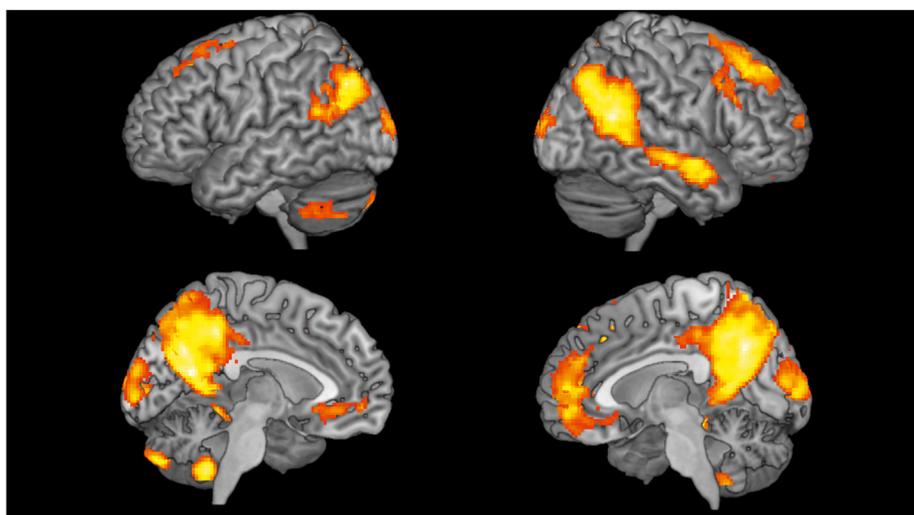


Fig. 4 – Visualization of activations specific to motor descriptions (contrast Motor versus Perceptual & Mentalizing [1 –.5 –.5]). The left panel represents the left hemisphere (lateral and medial), the right panel represents the right hemisphere (lateral and medial).

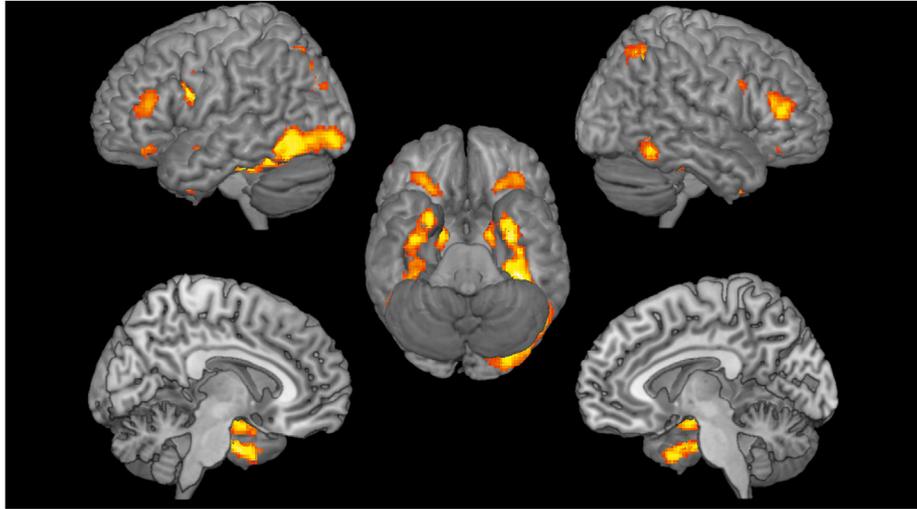


Fig. 5 – Visualization of activations specific to perceptual descriptions (contrast Perceptual versus Motor & Mentalizing [–.5 1 –.5]). The left panel represents the left hemisphere (lateral and medial), the right panel represents the right hemisphere (lateral and medial), the middle panel represents the inferior view (bilateral).

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 = .05, 95% CI = .00; .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 = –.05, 95% CI = –.09; .01), indicating that participants who experienced the stories as witty and beautiful showed a lower percent signal change in this area.

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, we 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, we attempted to find out whether individual differences in reading experiences can be linked to differences in activation levels in specific brain regions.

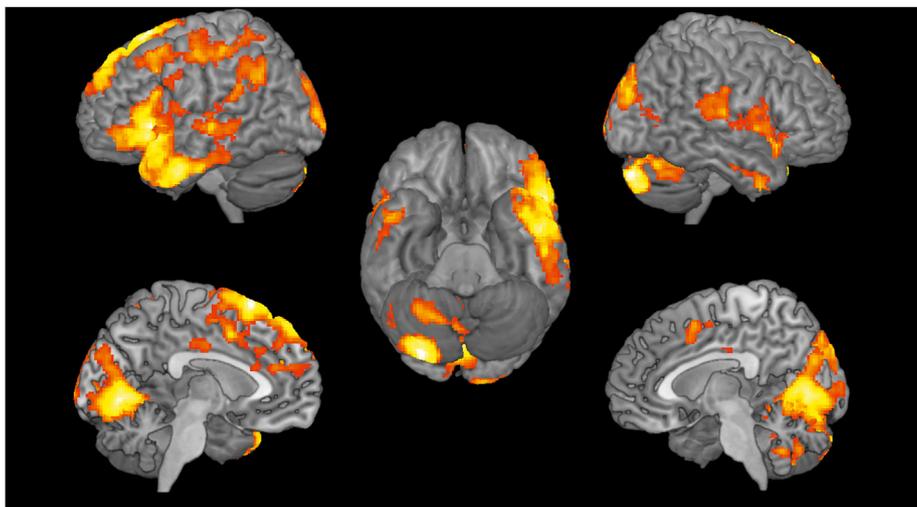


Fig. 6 – Visualization of activations specific to mental event descriptions (contrast Mentalizing versus Motor & Perceptual [–.5 –.5 1]). The left panel represents the left hemisphere (lateral and medial), the right panel represents the right hemisphere (lateral and medial), the middle panel represents the inferior view (bilateral).

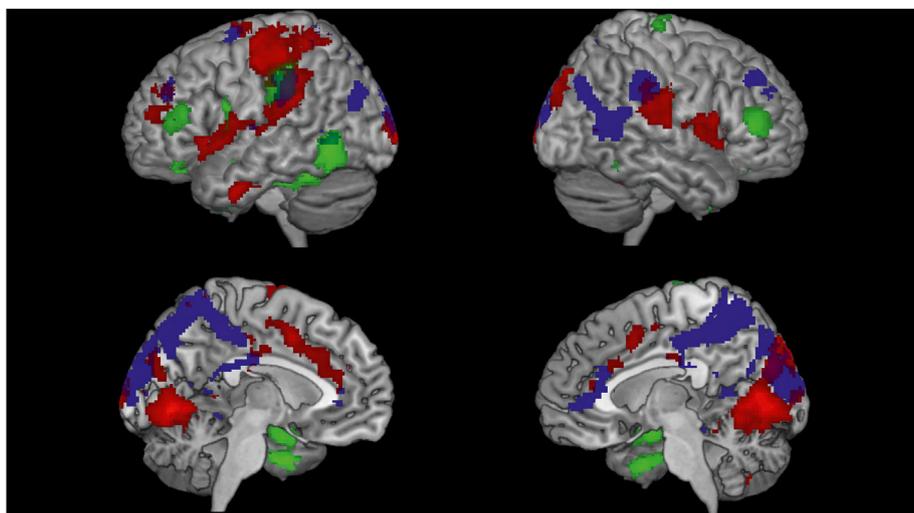


Fig. 7 – Visualization of the overlap of the activations to the three types of descriptions. Blue indicates activations to motor descriptions (contrast Motor versus Baseline [1 0 0]), green indicates activations to perceptual descriptions (contrast Perceptual versus Baseline [0 1 0]), and red indicates 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).

4.1. Differences in the effect of the three kinds of simulation on brain activation

The first question that we 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 our study, we found both modality-specific activations and a brain area that was activated in response to all kinds of simulation. We 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 our expectations, we 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). However, we did not find activation in important motor areas such as precentral and postcentral cortex, superior temporal sulcus, cingulate cortex, supplementary motor area, middle and superior temporal gyrus, inferior parietal lobule, intraparietal sulcus. These areas have been implicated in motor simulation in previous studies focusing on the processing of contextualized language (e.g., Kurby & Zacks, 2013; Nijhof & Willems, 2015). Although no activation in response to motor descriptions was found in these primary and premotor areas (as one may expect based on the grounded cognition account of mental simulation; but note that a lack of activation in the primary motor cortex in response to motor simulation has been found before; Willems, Toni, Hagoort, & Casasanto, 2010), activation was found in the precuneus, which is situated next to the primary sensory and

motor areas and is extensively connected with the premotor and supplementary motor areas (Cavanna & Trimble, 2006). Functionally, the precuneus has been implicated in motor imagery (Cavanna & Trimble, 2006), although it has been found to play a role in episodic memory, self-processing and mentalizing as well (Cavanna & Trimble, 2006; Schurz, Radua, Aichhorn, Richlan, & Perner, 2014).

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., 2007, 2009). 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).

Table 4 – Regions generally activated for all three types of descriptions (motor, perceptual and mental events) compared to the implicit baseline.

Region	L/R	Peak MNI coordinates			Z-max	Volume (voxels)
		x	y	z		
Supramarginal gyrus (anterior division)	L	−62	−34	36	5.08	152

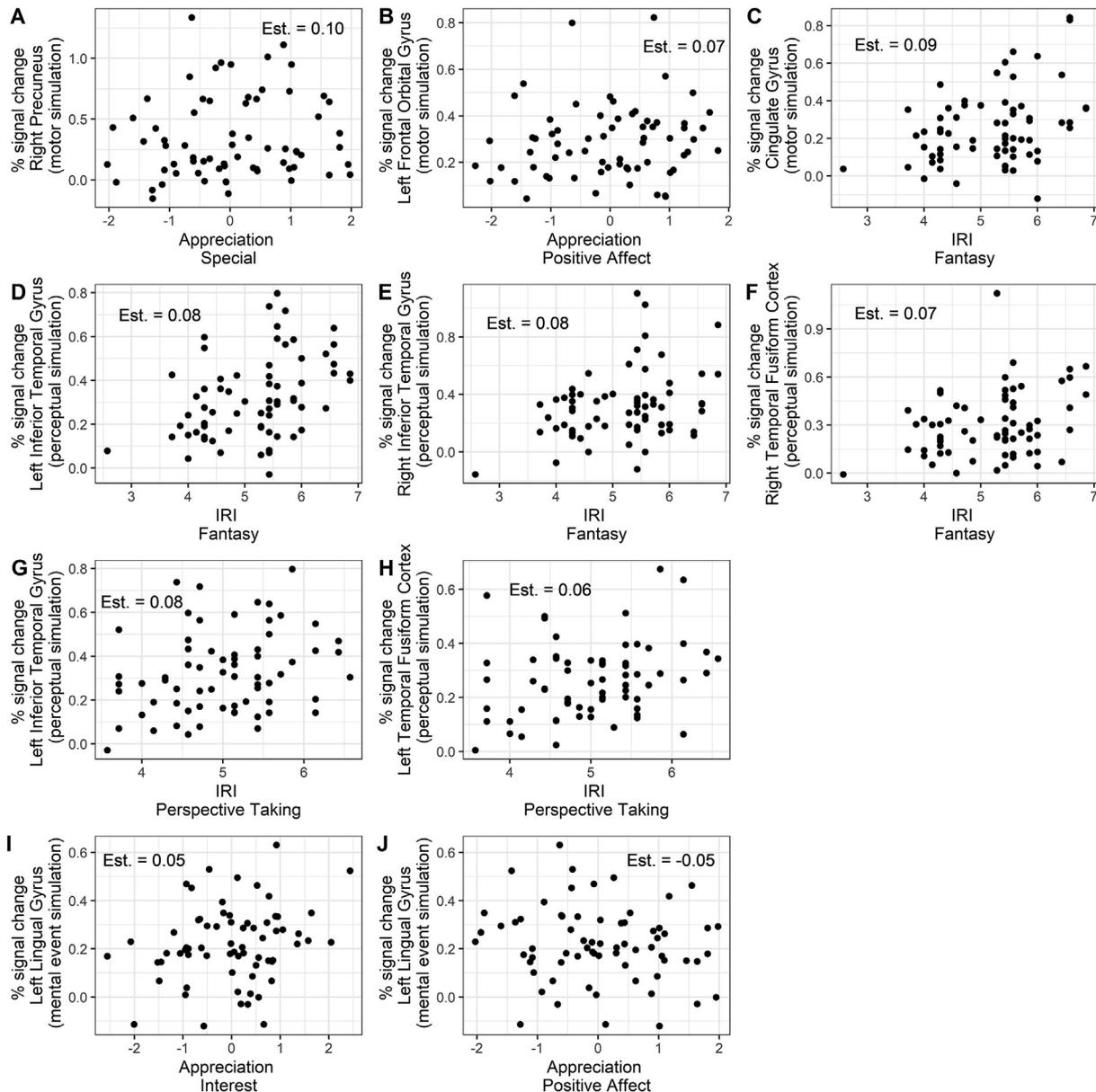


Fig. 8 – Scatter plots and estimated slopes for the relationships between percent signal change and components of appreciation or subscales of the Interpersonal Reactivity Index. Plots A–C are relationships found for motor simulation, plots D–H are relationships found for perceptual simulation, and plots I and J are relationships found for mental event simulation.

For perceptual simulation, we found several modality-specific areas that we 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, we also 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, our 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 (Frith & Frith, 2003; Igelström & Graziano, 2017; Laurita, Hazan, & Spreng, 2017; Paulus, Müller-Pinzler, Jansen, Gazzola, & Krach, 2015; Saxe & Kanwisher, 2003). Additionally, we 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).

4.2. 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 (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, Langston, & Graesser, 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.

4.3. Links with individual differences in appreciation and absorption

For all three kinds of descriptions, we 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 effects) as well as personal 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 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 (as opposed to state-based individual differences) were also more strongly associated with simulation (Faber, Mak, & Willems, 2020; Hartung et al., 2021; Hartung, Hagoort, & Willems, 2017; Mak, De Vries, & Willems, 2020; Van den Hoven, Hartung, Burke, & Willems, 2016). Seemingly, simulation is more strongly associated with stable characteristics than with reading experiences such as absorption and appreciation.

4.4. Recommendations for future research

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 our previous study (Mak & Willems, 2019). 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 (Mak & Willems, 2019). Given the importance of replication and reproducibility of results, we were happy to see that the eye tracking results are largely consistent with earlier results obtained in a non-fMRI setting. This indicates that the reading in the fMRI scanner appears to be consistent with reading in other settings, suggesting that our findings are likely to generalize to other reading contexts, which is encouraging for future studies attempting to find the relationship between eye movements and brain activation during reading.

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 paper have not been able to answer, but that deserve further exploration.

The dataset collected for this paper, can be used to answer many more questions concerning eye movements and brain activation during reading. For example, in the current study we used contextualized ratings of the words that were being read as regressors in our models. However, the effects of other aspects of the words in the stories that were read could be studied using existing ratings of, for example, concreteness (which is highly correlated with imageability; Brysbaert et al.,

2014), sensory modality (Speed & Brysbaert, 2022), or other word characteristics.

4.5. Conclusion

In the current experiment participants read literary short stories in the MRI scanner while their eye movements were being tracked. With this study, we 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. We propose that it is likely that this area is involved in referential indexing and situation model building. Furthermore, within the activated areas we 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 certain 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).

Author contributions

MM: Conceptualization, Data curation, Formal analysis, Methodology, Project administration, Visualization, Writing – original draft, Writing - review & editing.

MF: Conceptualization, Formal analysis, Writing – original draft, Writing - review & editing.

RW: Conceptualization, Funding acquisition, Methodology, Supervision, Writing – original draft, Writing - review & editing.

All authors read and approved the final manuscript.

Funding

This work was supported by the Netherlands Organization for Scientific Research (NWO) grants awarded to RW [Grant Number Vidi-276-89-007, awarded to RW; and Grant number VI. Veni.191G.001; awarded to MF].

Data statement

For this paper, a dataset was created that is available from <https://doi.org/10.34973/hrax-yn39>. It is stored on the repository of the Donders Center for Cognitive Neuroimaging.

The data have been archived in a Research Documentation Collection, were anonymized and archived as a Data Sharing Collection in the Donders Repository, according to the policies of the Donders Institute. The data and analysis code for the behavioral results and the individual difference analysis have been published on <https://osf.io/8eqya/>.

Declarations of interest

None.

Open Practices Section

The study in this article earned Open Data badge for transparent practices. The data for this study is available at: <https://osf.io/8eqya/3>. The section Data Availability project or the sentence the authors do not have permission to share data. should not appear anywhere in any version of this article.

Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.cortex.2023.01.014>.

REFERENCES

- Andersson, J. L. R., Jenkinson, M., & Smith, S. M. (2007a). *Non-linear optimisation*. Oxford: FMRIB technical report TR07JA1.
- Andersson, J. L. R., Jenkinson, M., & Smith, S. M. (2007b). *Non-linear registration, aka Spatial normalisation*. Oxford: FMRIB technical report TR07JA2.
- Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, 59(1), 617–645. <https://doi.org/10.1146/annurev.psych.59.103006.093639>
- Beckmann, C. F., Jenkinson, M., & Smith, S. M. (2003). General multilevel linear modeling for group analysis in FMRI. *NeuroImage*, 20, 1052–1063. [https://doi.org/10.1016/S1053-8119\(03\)00435-X](https://doi.org/10.1016/S1053-8119(03)00435-X)
- Brysbaert, M., Stevens, M., De Deyne, S., Voorspoels, W., & Storms, G. (2014). Norms of age of acquisition and concreteness for 30,000 Dutch words. *Acta Psychologica*, 150, 80–84. <https://doi.org/10.1016/j.actpsy.2014.04.010>
- Bürkner, P.-C. (2017). Brms : An R package for Bayesian multilevel models using stan. *Journal of Statistical Software*, 80(1). <https://doi.org/10.18637/jss.v080.i01>
- Bürkner, P.-C. (2018). Advanced Bayesian multilevel modeling with the R package brms. *The R Journal*, 10(1), 395. <https://doi.org/10.32614/RJ-2018-017>
- Cavanna, A. E., & Trimble, M. R. (2006). The precuneus: A review of its functional anatomy and behavioural correlates. *Brain: a Journal of Neurology*, 129, 564–583. <https://doi.org/10.1093/brain/awl004>
- Choi, W., Desai, R. H., & Henderson, J. M. (2014). The neural substrates of natural reading: A comparison of normal and nonword text using eyetracking and fMRI. *Frontiers in Human Neuroscience*, 8(DEC), 1–11. <https://doi.org/10.3389/fnhum.2014.01024>
- Chow, H. M., Mar, R. A., Xu, Y., Liu, S., Wagage, S., & Braun, A. R. (2015). Personal experience with narrated events modulates functional connectivity within visual and motor systems

- during story comprehension. *Human Brain Mapping*, 36, 1494–1505. <https://doi.org/10.1002/hbm.22718>
- Davis, M. H. (1980). A multidimensional approach to individual differences in empathy. *JSAS Catalog of Selected Documents in Psychology*, 10, 85.
- De Corte, K., Buysse, A., Verhofstadt, L. L., Roeyers, H., Ponnet, K., & Davis, M. H. (2007). Measuring empathic tendencies: Reliability and validity of the Dutch version of the interpersonal reactivity index. *Psychologica Belgica*, 47(4), 235–260. <https://doi.org/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. <https://doi.org/10.1016/j.neuropsychologia.2010.01.028>
- Desai, R. H., Choi, W., & Henderson, J. M. (2020). Word frequency effects in naturalistic reading. *Language, Cognition and Neuroscience*, 35(5), 583–594. <https://doi.org/10.1080/23273798.2018.1527376>
- 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. <https://doi.org/10.16910/jemr.13.3.2>
- Frith, U., & Frith, C. D. (2003). Development and neurophysiology of mentalizing. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 358, 459–473. <https://doi.org/10.1098/rstb.2002.1218>
- Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Analysis*, 1(3), 515–534. <https://doi.org/10.1214/06-BA117A>
- 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. Van Schijndel (Eds.), *Proceedings of the 8th workshop on cognitive modeling and computational linguistics, CMCL 2018* (pp. 10–18). Stroudsburg, PA, USA: Association for Computational Linguistics.
- Greve, D. N., & Fischl, B. (2009). Accurate and robust brain image alignment using boundary-based registration. *NeuroImage*, 48(1), 63–72. <https://doi.org/10.1016/j.neuroimage.2009.06.060>
- Grill-Spector, K., & Weiner, K. S. (2014). The functional architecture of the ventral temporal cortex and its role in categorization. *Nature Reviews Neuroscience*, 15(8), 536–548. <https://doi.org/10.1038/nrn3747>
- Hagoort, P. (2019). The neurobiology of language beyond single-word processing. *Science*, 366, 55–58. <https://doi.org/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* (pp. 1–8). Stroudsburg, PA, USA: Association for Computational Linguistics. <https://doi.org/10.3115/1073336.1073357>
- Hartung, F., Burke, M., Hagoort, P., & Willems, R. M. (2016). Taking perspective: Personal pronouns affect experiential aspects of literary reading. *Plos One*, 11(5), 1–18. <https://doi.org/10.1371/journal.pone.0154732>
- Hartung, F., Hagoort, P., & Willems, R. M. (2017). Readers select a comprehension mode independent of pronoun: Evidence from fMRI during narrative comprehension. *Brain and Language*, 170, 29–38. <https://doi.org/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. <https://doi.org/10.1038/s42003-021-02926-0>
- Hartwigsen, G., Baumgaertner, A., Price, C. J., Koehnke, M., Ulmer, S., & Siebner, H. R. (2010). Phonological decisions require both the left and right supramarginal gyri. *Proceedings of the National Academy of Sciences*, 107(38), 16494–16499. <https://doi.org/10.1073/pnas.1008121107>
- Hasson, U., Egidi, G., Marelli, M., & Willems, R. M. (2018). Grounding the neurobiology of language in first principles: The necessity of non-language-centric explanations for language comprehension. *Cognition*, 180, 135–157. <https://doi.org/10.1016/j.cognition.2018.06.018>
- 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 (preprint) *BioRxiv*, 1–65. <https://doi.org/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. <https://doi.org/10.3389/fcomm.2020.519955>
- Hsu, C.-T., Conrad, M., & Jacobs, A. M. (2014). Fiction feelings in Harry Potter: Haemodynamic response in the mid-cingulate cortex correlates with immersive reading experience. *Neuroreport*, 25, 1356–1361. <https://doi.org/10.1097/WNR.0000000000000272>
- Igelström, K. M., & Graziano, M. S. A. (2017). The inferior parietal lobule and temporoparietal junction: A network perspective. *Neuropsychologia*, 105, 70–83. <https://doi.org/10.1016/j.neuropsychologia.2017.01.001>
- 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. [https://doi.org/10.1016/S1053-8119\(02\)91132-8](https://doi.org/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. [https://doi.org/10.1016/S1361-8415\(01\)00036-6](https://doi.org/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>.
- 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. <https://doi.org/10.1073/pnas.1005062107>
- 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. <https://doi.org/10.3758/BRM.42.3.643>
- Knoop, C. A., Wagner, V., Jacobsen, T., & Menninghaus, W. (2016). Mapping the aesthetic space of literature “from below. *Poetics*, 56, 35–49. <https://doi.org/10.1016/j.poetic.2016.02.001>
- Koopman, E. M. (Emy) (2015). Empathic reactions after reading: The role of genre, personal factors and affective responses. *Poetics*, 50, 62–79. <https://doi.org/10.1016/j.poetic.2015.02.008>
- Kravitz, D. J., Saleem, K. S., Baker, C. I., Ungerleider, L. G., & 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. <https://doi.org/10.1016/j.tics.2012.10.011>
- Kuijpers, M. M. (2014). *Absorbing stories: The effects of textual devices on absorption and evaluative responses*. Doctoral dissertation. Utrecht University.
- Kuijpers, M. M., Hakemulder, F., Tan, E. S., & Doicaru, M. 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. <https://doi.org/10.1075/sslol.4.1.05kui>
- Kurby, C. A., & Zacks, J. M. (2008). Segmentation in the perception and memory of events. *Trends in Cognitive Sciences*, 12(2), 72–79. <https://doi.org/10.1016/j.tics.2007.11.004>

- Kurby, C. A., & Zacks, J. M. (2013). The activation of modality-specific representations during discourse processing. *Brain and Language*, 126, 338–349. <https://doi.org/10.1016/j.bandl.2013.07.003>
- Lai, V. T., Willems, R. M., & Hagoort, P. (2015). Feel between the lines: Implied emotion in sentence comprehension. *Journal of Cognitive Neuroscience*, 27(8), 1528–1541. https://doi.org/10.1162/jocn_a_00798
- Laurita, A. C., Hazan, C., & Spreng, R. N. (2017). Dissociable patterns of brain activity for mentalizing about known others: A role for attachment. *Social Cognitive and Affective Neuroscience*, 12(7), 1072–1082. <https://doi.org/10.1093/scan/nsx040>
- Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, 106, 1126–1177. <https://doi.org/10.1016/j.cognition.2007.05.006>
- Lopopolo, A., Frank, S. L., van den Bosch, A., & Willems, R. M. (2017). Using stochastic language models (SLM) to map lexical, syntactic, and phonological information processing in the brain. *Plos One*, 12(5), Article e0177794. <https://doi.org/10.1371/journal.pone.0177794>
- Mak, M., De Vries, C., & Willems, R. M. (2020). The influence of mental imagery instructions and personality characteristics on reading experiences. *Collabra: Psychology*, 6(1), 43. <https://doi.org/10.1525/collabra.281>
- Mak, M., Faber, M., & Willems, R. M. (n.d.). Different routes to liking: How readers arrive at narrative evaluations.
- 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. <https://doi.org/10.1080/23273798.2018.1552007>
- Martín-Loeches, M., Casado, P., Hernández-Tamames, J. A., & Álvarez-Linera, J. (2008). Brain activation in discourse comprehension: A 3t fMRI study. *NeuroImage*, 41, 614–622. <https://doi.org/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; a Journal Devoted To the Study of the Nervous System and Behavior*, 88, 106–123. <https://doi.org/10.1016/j.cortex.2016.12.010>
- McElreath, R. (2016). *Statistical rethinking*. Boca Raton, Florida: Chapman and Hall/CRC. <https://doi.org/10.1201/9781315372495>
- Moody, C. L., & Gennari, S. P. (2010). Effects of implied physical effort in sensory-motor and pre-frontal cortex during language comprehension. *NeuroImage*, 49(1), 782–793. <https://doi.org/10.1016/j.neuroimage.2009.07.065>
- Nabokov, V. (2003). *Signalen en Symbolen. In Een Russische schoonheid 1*. Amsterdam: De Bezige Bij.
- Nalborczyk, L., Batailler, C., Løvenbruck, 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: JSLHR*, 62(5), 1225–1242. <https://doi.org/10.1044/2018.JSLHR-S-18-0006>
- Nijhof, A. D., & Willems, R. M. (2015). Simulating fiction: Individual differences in literature comprehension revealed with fMRI. *Plos One*, 10(2), 1–17. <https://doi.org/10.1371/journal.pone.0116492>
- Paulus, F. M., 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. <https://doi.org/10.1093/cercor/bhu011>
- Power, J. D., Barnes, K. A., Snyder, A. Z., Schlaggar, B. L., & Petersen, S. E. (2012). Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *NeuroImage*, 59(3), 2142–2154. <https://doi.org/10.1016/j.neuroimage.2011.10.018>
- R Core Team. (2021). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. 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. <https://doi.org/10.1037/0033-2909.124.3.372>
- Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology*, 62(8), 1457–1506. <https://doi.org/10.1080/17470210902816461>
- Revelle, W. (2020). *psych: Procedures for psychological, psychometric, and personality research*. Evanston, Illinois, USA: Northwestern University. Retrieved from <https://cran.r-project.org/package=psych>.
- Richlan, F., Gagl, B., Hawelka, S., Braun, M., Schurz, M., Kronbichler, M., et al. (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. <https://doi.org/10.1093/cercor/bht117>
- Samur, D., Lai, V. T., Hagoort, P., & Willems, R. M. (2015). Emotional context modulates embodied metaphor comprehension. *Neuropsychologia*, 78, 108–114. <https://doi.org/10.1016/j.neuropsychologia.2015.10.003>
- 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. [https://doi.org/10.1016/S1053-8119\(03\)00230-1](https://doi.org/10.1016/S1053-8119(03)00230-1)
- Schäfer, R., & Bildhauer, F. (2012). Building large corpora from the web using a new efficient tool chain. In N. Calzolari, K. Choukri, T. Declerck, M. Ugur Dogan, B. Maegaard, J. Mariani, et al. (Eds.), *Proceedings of the 8th international conference on language Resources and evaluation, LREC 2012, Istanbul, Turkey, May 23–25, 2012* (pp. 486–493).
- Schurz, M., Radua, J., Aichhorn, M., Richlan, F., & Perner, J. (2014). Fractionating theory of mind: A meta-analysis of functional brain imaging studies. *Neuroscience and Biobehavioral Reviews*, 42, 9–34. <https://doi.org/10.1016/j.neubiorev.2014.01.009>
- Schuster, S., Hawelka, S., Hutzler, F., Kronbichler, M., & Richlan, F. (2016). Words in context: The effects of length, frequency, and predictability on brain responses during natural reading. *Cerebral Cortex*, 26, 3889–3904. <https://doi.org/10.1093/cercor/bhw184>
- Shain, C., Blank, I. A., van Schijndel, M., Schuler, W., & Fedorenko, E. (2020). fMRI reveals language-specific predictive coding during naturalistic sentence comprehension. *Neuropsychologia*, 138, Article 107307. <https://doi.org/10.1016/j.neuropsychologia.2019.107307>
- Smirnov, D., Glerean, E., Lahnakoski, J. M., Salmi, J., Jääskeläinen, I. P., Sams, M., et al. (2014). Fronto-parietal network supports context-dependent speech comprehension. *Neuropsychologia*, 63, 293–303. <https://doi.org/10.1016/j.neuropsychologia.2014.09.007>
- Smith, S. M. (2002). Fast robust automated brain extraction. *Human Brain Mapping*, 17(3), 143–155. <https://doi.org/10.1002/hbm.10062>
- Smith, S. M., & Brady, J. M. (1997). SUSAN—a new approach to low level image processing. *International Journal of Computer Vision*, 23, 45–78. <https://doi.org/10.1023/A:1007963824710>
- Speed, L. J., & Brysbaert, M. (2022). Dutch sensory modality norms. *Behavior Research Methods*, 54(3), 1306–1318. <https://doi.org/10.3758/s13428-021-01656-9>
- Speer, N. K., Reynolds, J. R., Swallow, K. M., & Zacks, J. M. (2009). Reading stories activates neural representations of visual and motor experiences. *Psychological Science*, 20(8), 989–999. <https://doi.org/10.1111/j.1467-9280.2009.02397.x>
- Speer, N. K., Zacks, J. M., & Reynolds, J. R. (2007). Human brain activity time-locked to narrative event boundaries.

- Psychological Science, 18(5), 449–455. <https://doi.org/10.1111/j.1467-9280.2007.01920.x>
- Stanovich, K. E., & West, R. F. (1989). Exposure to print and orthographic processing. *Reading Research Quarterly*, 24(4), 402–433. <https://doi.org/10.2307/747605>
- Stoeckel, C., Gough, P. M., Watkins, K. E., & Devlin, J. T. (2009). Supramarginal gyrus involvement in visual word recognition. *Cortex; a Journal Devoted To the Study of the Nervous System and Behavior*, 45(9), 1091–1096. <https://doi.org/10.1016/j.cortex.2008.12.004>
- Stolcke, A. (2002). SRILM - an extensible language modeling toolkit. In *7th international conference on spoken language processing, ICSLP 2002*.
- Tamir, D. I., Bricker, A. B., Dodell-Feder, D., & Mitchell, J. P. (2016). Reading fiction and reading minds: The role of simulation in the default network. *Social Cognitive and Affective Neuroscience*, 11(2), 215–224. <https://doi.org/10.1093/scan/nsv114>
- Team, S. D. (2020). RStan: the R interface to Stan. Retrieved from <http://mc-stan.org>.
- Van Essen, R. (2014). *De mensen die alles lieten bezorgen. In Hier wonen ook mensen* (pp. 113–123). Utrecht: Atlas Contact.
- Van den Hoven, E., Hartung, F., Burke, M., & Willems, R. M. (2016). Individual differences in sensitivity to style during literary reading: Insights from eye-tracking. *Collabra*, 2(1), 25. <https://doi.org/10.1525/collabra.39>
- Wallentin, M., Nielsen, A. H., Vuust, P., Dohn, A., Roepstorff, A., & Lund, T. E. (2011). Amygdala and heart rate variability responses from listening to emotionally intense parts of a story. *NeuroImage*, 58, 963–973. <https://doi.org/10.1016/j.neuroimage.2011.06.077>
- Willems, R. M., & Peelen, M. V. (2021). How context changes the neural basis of perception and language. *IScience*, 24, Article 102392. <https://doi.org/10.1016/j.isci.2021.102392>
- Willems, R. M., Toni, I., Hagoort, P., & Casasanto, D. (2010). Neural dissociations between action verb understanding and motor imagery. *Journal of Cognitive Neuroscience*, 22(10), 2387–2400. <https://doi.org/10.1162/jocn.2009.21386>
- Woolrich, M. W. (2008). Robust group analysis using outlier inference. *NeuroImage*, 41(2), 286–301. <https://doi.org/10.1016/j.neuroimage.2008.02.042>
- Woolrich, M. W., Behrens, T. E. J., Beckmann, C. F., Jenkinson, M., & Smith, S. M. (2004). Multilevel linear modelling for FMRI group analysis using Bayesian inference. *NeuroImage*, 21(4), 1732–1747. <https://doi.org/10.1016/j.neuroimage.2003.12.023>
- Woolrich, M. W., Jbabdi, S., Patenaude, B., Chappell, M., Makni, S., Behrens, T., et al. (2009). Bayesian analysis of neuroimaging data in FSL. *NeuroImage*, 45, S173–S186. <https://doi.org/10.1016/j.neuroimage.2008.10.055>
- Woolrich, M. W., Ripley, B. D., Brady, M., & Smith, S. M. (2001). Temporal autocorrelation in univariate linear modeling of FMRI data. *NeuroImage*, 14(6), 1370–1386. <https://doi.org/10.1006/nimg.2001.0931>
- Worsley, K. J. (2001). Statistical analysis of activation images. In P. Jezzard, P. M. Matthews, & S. M. Smith (Eds.), *Functional magnetic resonance imaging* (pp. 251–270). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780192630711.003.0014>
- Zwaan, R. A. (2009). Mental simulation in language comprehension and social cognition. *European Journal of Social Psychology*, 39, 1142–1150. <https://doi.org/10.1002/ejsp.661>
- Zwaan, R. A., Langston, M. C., & Graesser, A. C. (1995). The construction of situation models in narrative comprehension: An event-indexing model. *Psychological Science*, 6(5), 292–297.
- Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, 123(2), 162–185. <https://doi.org/10.1037/0033-2909.123.2.162>