

## Mental simulation during literary reading: Individual differences revealed with eye-tracking

Marloes Mak & Roel M. Willems

To cite this article: Marloes Mak & Roel M. Willems (2019) Mental simulation during literary reading: Individual differences revealed with eye-tracking, *Language, Cognition and Neuroscience*, 34:4, 511-535, DOI: [10.1080/23273798.2018.1552007](https://doi.org/10.1080/23273798.2018.1552007)

To link to this article: <https://doi.org/10.1080/23273798.2018.1552007>



© 2018 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



[View supplementary material](#)



Published online: 01 Dec 2018.



[Submit your article to this journal](#)



Article views: 796



[View Crossmark data](#)



Citing articles: 1 [View citing articles](#)

## Mental simulation during literary reading: Individual differences revealed with eye-tracking

Marloes Mak <sup>a</sup> and Roel M. Willems<sup>a,b,c</sup>

<sup>a</sup>Centre for Language Studies, Radboud University Nijmegen, Nijmegen, The Netherlands; <sup>b</sup>Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, Nijmegen, The Netherlands; <sup>c</sup>Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

### ABSTRACT

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

### ARTICLE HISTORY

Received 15 February 2018  
Accepted 14 November 2018

### KEYWORDS

Motor simulation; perceptual simulation; mentalising; literary reading; eye-tracking; individual differences; absorption; appreciation

### Introduction

When people read stories, they sometimes vividly imagine the events occurring in the stories and in the story world in which these events are happening. The process underlying this vivid imagination has been called mental simulation. One result of mental simulation is that readers get the feeling that they are part of the story they are reading. Consequently, literary stories can take a strong grip on readers, although the strength of this “grip” can vary widely from one story to the next, and from one reader to the next. This sense of grip has been described in the literature as absorption<sup>1</sup> (Gerrig, 1993; Green & Brock, 2000; Jacobs & Willems, 2017; Kuijpers, Hakemulder, Tan, & Doicaru, 2014; Kuzmičová, 2012). In this study we will focus on mental simulation as an important driver of absorption. We distinguish between three kinds of mental simulation, and we have a particular focus on identifying individual differences in simulation.

Mental simulation has been defined as “... the re-enactment of perceptual, motor, and introspective states acquired during experience with the world, body, and mind.” (Barsalou, 2008, p. 618). Importantly, the definition of Barsalou suggests that mental simulation is not one of a kind. Indeed, from theoretical (Barsalou, 2008; and see

Shanton & Goldman, 2010, for a review of Simulation Theory) as well as empirical work (discussed below), it is known that mental simulation should be subdivided in different kinds of simulation. Language users are capable of simulating perceptual and motor events on the one hand, and mental processes of others on the other hand (also referred to as mentalising; e.g. Goldman, 2012). The effects of these three kinds of simulation on language processing have so far been studied mostly separately from each other (see Jacobs & Willems, 2017). In this study, we investigate them in one experiment. This enables us to disentangle the effects of different kinds of simulation on literary reading behaviour. An additional advantage of this study is that we use narratives as stimuli. We choose to investigate mental simulation in the context of narratives because narratives allow readers to construct a much richer mental story world as compared to single sentences or paragraphs that are sometimes used in research.

### Empirical background

The aforementioned three kinds of simulation have been extensively studied in different subdisciplines of cognitive science. Perceptual and motor simulation

**CONTACT** Marloes Mak  [H.Mak@let.ru.nl](mailto:H.Mak@let.ru.nl)

 Supplemental data for this article can be accessed <http://dx.doi.org/10.1080/23273798.2018.1552007>.

© 2018 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

(sometimes called “sensorimotor simulation”) have been studied in the tradition of embodied cognition. According to Zwaan (2009), there is an important link between situation models, simulation and grounding in perception and action. When people form situation models (for example when they encounter narratives), the events and event nodes within these situation models are grounded in perception and action through (sensorimotor) simulation.

There is some evidence to suggest that readers indeed form perceptual mental images when reading language. It has, for example, been found that reading implicit descriptions of shape (e.g. “the ranger saw the eagle in the nest” vs. “the ranger saw the eagle in the sky”) or orientation (e.g. “John put the pencil in the drawer” vs. “John put the pencil in the cup”) primes subsequent visual perception of the described object in the implied orientation, both in adults and children (e.g. Engelen, Bouwmeester, de Bruin, & Zwaan, 2011; Stanfield & Zwaan, 2001; Zwaan, Stanfield, & Yaxley, 2002). Reading words that imply a certain location on a vertical axis similarly primes perception of pictures of semantically related objects appearing in the implied location (e.g. “sky” primes the detection of a picture of a cloud if this picture is presented in the top half of the screen; Ostarek & Vigliocco, 2017; see also Estes, Verges, & Barsalou, 2008). Additionally, reading descriptions of biological motion has been associated with activation in a motion processing area (i.e. middle temporal gyrus; Deen & McCarthy, 2010; Samur, Lai, Hagoort, & Willems, 2015). Similar associations between reading auditory descriptions and activation in multiple areas involved in auditory processing (Kurby & Zacks, 2013) and between reading vivid visual descriptions and connectivity between different areas in the visual processing system (Chow et al., 2015) have also been found.

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

elicit activation in the same way and in the same areas; for a more elaborate review of the task effects at hand and the precise role of the motor cortex in action language processing see for example Kemmerer (2015) or Willems and Casasanto (2011).

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

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

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

mentalising (Hsu, Conrad, & Jacobs, 2014; Nijhof & Willems, 2015; Tamir, Bricker, Dodell-Feder, & Mitchell, 2016).

### **Relationship between kinds of simulation during narrative reading**

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

Both on theoretical as well as on empirical grounds there is good reason to expect that perceptual, motor, and mental state simulation play a role in narrative understanding. In the current study we aim to investigate these kinds of simulation and how they are inter-related within one experiment.

### **Current study**

Most of the studies described above looked at the relationship between one or two kinds of simulation and language processing, but the tasks used are

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

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

Additionally, we linked individual differences in simulation to individual differences in personality traits or characteristics that have been found to be related to absorption, such as fiction reading experience (Mar, Oatley, Hirsh, de la Paz, & Peterson, 2006), the tendency to get transported (see Green & Donahue, 2009) and perspective taking (Mar & Oatley, 2008; see Mumper & Gerrig,

2017 for a meta-analysis). As simulation is related to absorption, we expected that simulation would also be related to these traits. Any association between individual differences in simulation and individual differences in one or more of these personality traits or characteristics might give an indication as to why people seem to differ so much in their experiences during reading.

## Materials & Methods

This study was pre-registered in the Open Science Framework ([osf.io/qgx26](https://osf.io/qgx26)).

### Participants

We recruited 109 participants (85 females) from the participant database of the Radboud University. All participants were native speakers of Dutch, and had normal or corrected to normal vision. Based on poor quality of the eye-tracking data or insufficient performance on a comprehension check, data for 7 participants were rejected. Of these participants, 4 were female. The mean age of the remaining participants ( $N = 102$ ) was 23 years (range 18–40).

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

### Materials

Three existing Dutch short stories (see Table 1) were presented to the participants. The selection of the stories was made based on the length of the stories, the presence of simulation-eliciting content, and the probability

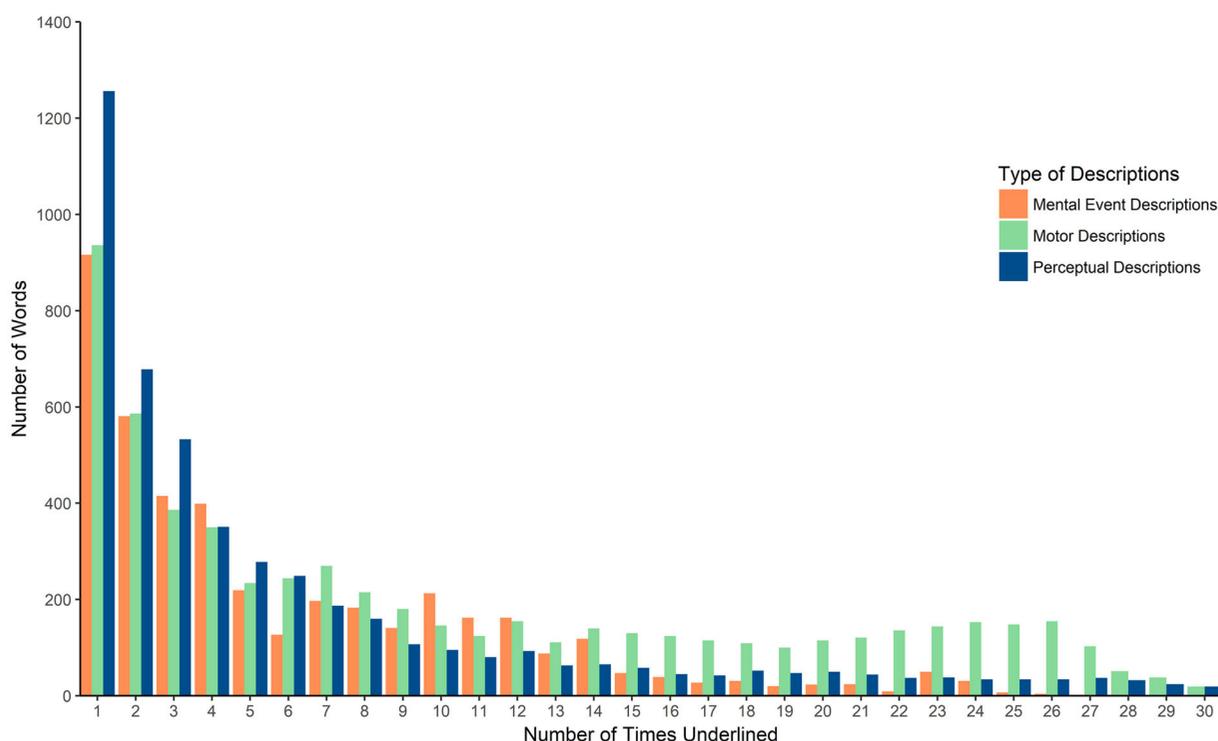
that the stories would be unknown to the target group (to ensure that all participants would read the stories for the first time). All stories were written by acclaimed writers (who all have received literary awards for their work) and have been published by literary publishers. Stories A (Van Essen, 2014) and B (Van Hassel, 2012) are written by contemporary Dutch writers, and story C (Nabokov, 2003) was translated from American English to Dutch. This story was taken from a professional and published translation. The stories were on average around 2600 words (2143, 2659, and 2988 words), and each story took around 10–15 min to read. A pre-test (see below) confirmed that all stories contained simulation-eliciting passages, indicating that all stories contained passages (or sentences or clusters of words) that were likely to elicit motor simulation, perceptual simulation or mentalising. All participants read all three stories (in counterbalanced order). None of the participants reported having read any of the three stories before.

### Simulation Scoring pre-test

For a pre-test, 90 participants were recruited from the same participant database. These participants did not take part in the main part of the study (i.e. the eye-tracking session). All participants read all three stories (in counterbalanced order), and were asked to focus on one of the three kinds of simulation. They were instructed to underline all the words, sentences, or passages that they considered to be part of either of three possible types of simulation-eliciting content: perceptual descriptions, motor descriptions, or descriptions of mental events (e.g. thoughts, feelings, opinions) that revealed what was going on in the mind of a character in the story. We defined perceptual descriptions as “things that are perceivable with the senses”, motor descriptions as “concrete acts or actions performed by a person or object”, and mental event descriptions as “explicit descriptions of the thoughts, feelings and opinions of a character” and/or “reflection by a character on his own or someone else’s thoughts, feelings or behaviour”. In addition to the definitions of the types of descriptions, participants were presented with a short explanation of what was meant by these definitions, including a couple of example sentences derived from different stories. We hence used subjective ratings for when simulation-eliciting content occurred in a story. Several previous studies have similarly used subjective ratings as a proxy for simulation (Nijhof & Willems, 2015), imagery (Kurby & Zacks, 2013), but also effects of other variables affecting the reading process, such as foregrounding (Van den Hoven, Hartung, Burke, & Willems, 2016).

**Table 1.** Title, author, year of publication and word count of the experimental stories.

	Title	Author	Year of publication	Word count
Story A	De mensen die alles lieten bezorgen <i>(The people that had everything delivered)</i>	Rob van Essen	2013	2988
Story B	De Chinese bruiloft <i>(The Chinese wedding)</i>	Sanneke van Hassel	2012	2659
Story C	Signalen en symbolen <i>(Symbols and Signs)</i>	Vladimir Nabokov		1948



**Figure 1.** Distribution of the number of times words were underlined for mental event descriptions, motor descriptions and perceptual descriptions. 4235 out of 7790 words were underlined at least once for mental event descriptions, 5838 words were underlined at least once for motor descriptions and 4822 words were underlined at least once for perceptual descriptions.

We asked participants to underline all words or passages in each of the three stories that they considered to be part of one of the three types of simulation-eliciting content, resulting in scores between 0 and 30 for every word in each of the three stories, for all three types of simulation-eliciting content: a score of 0 if none of the participants had underlined it and a score of 30 if every participant had underlined it (with higher scores theoretically resulting in a higher probability this word would be mentally simulated). The underlining of perceptual descriptions was performed by 24 females and 6 males, the underlining of motor descriptions was performed by 22 females and 8 males, and mental event descriptions were underlined by 19 females, 10 males and 1 unspecified.

The distribution of scores per kind of simulation is shown in Figure 1, and the average number of times the words were underlined per kind of simulation and per story can be seen in Table 2. Importantly, 2968,

1952 and 3555 words were underlined by none of the pre-test participants for perceptual, motor and mental events descriptions, respectively. This indicated considerable consensus between participants about whether words were part of a description. The number of times a word was underlined was similar for all percentiles of sentence length (with the last word set at 100% and the middle word at 50%) for all types of underlining (see Supplemental Material A1–3). This means that underlinings did not occur systematically more at the beginning or end of the sentence.

### Apparatus

For eye-movement data collection, a monocular desktop-mounted EyeLink1000plus eye-tracking system was used. Data were recorded with a sampling rate of 500 Hz. Head movements were minimised using a

**Table 2.** Descriptive statistics of the underlined words: Mean, standard deviation and maximal observed value of the number of times words were underlined for each type of description in each story (the maximal possible value of the number of times words were underlined is, in all instances, 30).

	Perceptual Descriptions		Motor Descriptions		Mental Event Descriptions	
	<i>M</i> ( <i>SD</i> )	max. observed value	<i>M</i> ( <i>SD</i> )	max. observed value	<i>M</i> ( <i>SD</i> )	max. observed value
Story A	5.69 (6.31)	30	9.88 (8.62)	30	5.21 (4.57)	25
Story B	6.08 (6.56)	30	11.95 (8.98)	30	5.54 (4.68)	23
Story C	7.27 (7.75)	30	8.00 (7.47)	29	7.32 (6.51)	27

head stabiliser. This allowed us to ensure all participants were seated at 108 cm from the screen (i.e. distance from the eye to the bottom of the screen).

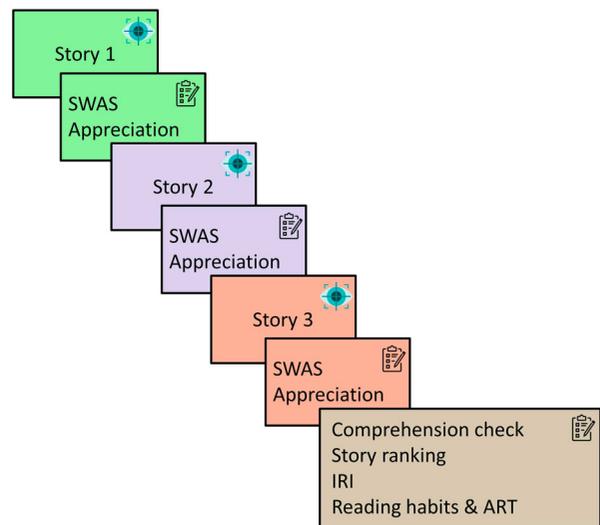
### Stimulus presentation

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

### Procedure

At the beginning of the experiment, participants were instructed to move as little as possible, but to read as naturally as possible, the way they would read a story outside of the laboratory. Since the eye-tracking data were collected monocularly, the dominant eye was tracked. To identify the dominant eye, participants performed an eye dominance test. In 7 participants, it was not possible to track the dominant eye, so in these participants the non-dominant eye was tracked. The stories were presented in a sound proof booth. There was no time restriction, participants could proceed to the next section of a story by pressing the space bar as soon as they had finished reading the current section. In addition to the eye-tracking part of the study, questionnaires (see below) were presented as paper and pencil tests outside of the booth, to enable participants to take a break from the computer screen.

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



**Figure 2.** Graphic depiction of the time line of the procedure of the experiment. The stories were presented in counterbalanced order.

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

### Additional measures

All used questionnaires can be found in Supplemental Material B. After reading each story, participants filled out a simulation and an appreciation questionnaire. The simulation questionnaire consisted of the story world absorption scale (SWAS; Kuijpers et al., 2014; e.g. *When I finished the story I was surprised to see that time had gone by so fast; I could imagine what the world in which the story took place looked like*), complemented with six additional questions (partly based on items originally designed by Kuijpers et al., 2014) regarding the experience of different kinds of simulation (mainly perceptual and motor simulation, e.g. *I could see the events in the story happening as if I could see through the eyes of the main character; I could easily depict the characters in the story*; See Supplemental Material B1 for a list of the questions added to the SWAS, all other used

questionnaires can be found in Supplemental Material B2–8). The SWAS is a validated scale consisting of 18 items with high internal validity (Kuijpers et al., 2014), which measures 4 dimensions of story world absorption via the subscales Attention, Transportation, Emotional Engagement and Mental Imagery. Participants rated each question on a 7-point scale (1 = disagree, 7 = agree).

The appreciation questionnaire consisted of a general score of story liking (*How did you like the story*; 1 = It was very bad, 7 = It was very good) and thirteen adjectives (e.g. [*did you find the story*] *Entertaining, ... Ominous*) that could be used to describe the stories (adapted from Knoop, Wagner, Jacobsen, & Menninghaus, 2016). These adjectives were taken from a list of adjectives that were most often used by people to describe their opinion of poetry, and which are also used to describe aesthetic appeal in the domain of literature (Knoop et al., 2016). Finally, 6 questions were asked regarding the enjoyment of the story (from Kuijpers et al., 2014; e.g. *I was constantly curious about how the story would end; I thought the story was written well*). Participants rated both the adjectives and the questions regarding enjoyment on a 7-point scale (1 = disagree, 7 = agree).

At the end of the experiment, participants completed additional, more general, questionnaires. First, participants were presented with 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 had read the stories with normal attention (example question, *Why did Jeffrey and Rita leave the flat?*). Subsequently, they were asked to rank the stories from most appreciated to least appreciated, and they were asked to indicate whether they had read the story before. Next, they answered six questions about their reading habits in everyday life, choosing from 4 or 5 optional answers (adapted from Hartung et al., 2016; e.g. *How often do you read fiction; How often do you read non-fiction; How many books do you read each year*), and filled out the Fantasy and Perspective Taking subscales of the Interpersonal Reactivity Index (IRI; Davis, 1980; Dutch translation adapted from De Corte et al., 2007) on a 7-point scale (e.g. *Becoming extremely involved in a good book or movie is somewhat rare for me; When I'm upset at someone, I usually try to "put myself in his shoes" for a while*). The Fantasy subscale measures the extent to which someone gets mentally very involved in the stories they encounter, to the point at which they imagine themselves being part of the story. The Perspective Taking subscale measures the extent to which someone is able to take someone else's perspective in daily life. Finally, as an implicit measure of reading experience, participants completed the Author

Recognition Test (ART; Stanovich & West, 1989; Dutch adaptation reported in Koopman, 2015), consisting of 42 names (30 real authors and 12 foils), where they had to indicate who they thought were genuine authors.

### **Eye-movement data pre-processing**

Before data analysis, all fixations were checked for all sections of all stories. This was done to make sure that they did not drift off so much that they entered a different interest area, thus corrupting the data. If necessary, they were manually aligned using SR Research's EyeLink Data Viewer. If this was impossible, because fixations did not lie on clear lines (corresponding to the lines on the pages presented on the screen), individual sections were excluded. If more than 6 sections of one story had to be excluded (more than 20% of the data for that story), data for entire story-readings were excluded in order to reduce noise in the data. If the entire story-reading had to be excluded for more than one story in the same participant, all data for this participant was excluded.

For 62 participants, no sections of any of the stories had to be removed. For 40 participants at least one section had to be removed. For Story A, at least 1 but no more than 6 sections had to be removed for 9 participants (on average 1.56 sections). For Story B, at least 1 but no more than 6 sections had to be removed from the analysis for 14 participants (on average 2.14 sections). For Story C, at least 1 but no more than 6 sections had to be removed for 21 participants (on average 2.05 sections). For 4 participants, the number of excluded sections exceeded 6, resulting in the exclusion of one story-reading for this participant. This happened twice for Story A, and twice for Story C, resulting in the exclusion of the data for one story-reading for 4 participants. In total, eye-movement data pre-processing resulted in the loss of 2.26% of the total amount of data.

### **Comprehension check**

Seventy-four participants answered all multiple choice questions (4 possible answers) in the comprehension check correctly. Participants were allowed to answer 1 question per story incorrectly. If participants answered more than 1 question incorrectly for a given story, it was concluded that they had not paid sufficient attention to this story, and data for this story-reading was excluded for these participants. If the entire story-reading had to be excluded for more than one story, all data for this participant was excluded. Eight participants answered more than one question incorrectly for one of the three stories. This was the case four times

for Story 2 and four times for Story 3, resulting in the exclusion of the data for one story-reading for 8 participants (an additional loss of 2.42% of the total amount of data).

### **Data-analysis: step 1**

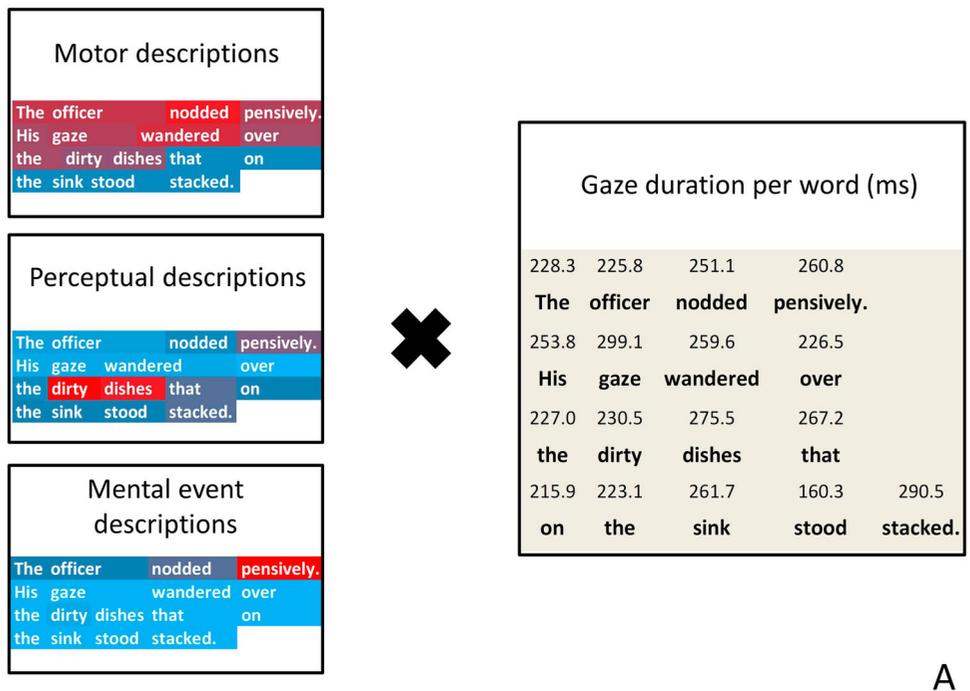
In the first part of our analysis, we used the “*lme4*” package in R (Bates, Mächler, Bolker, & Walker, 2015; R version 3.5.1) to analyse our data with a linear mixed effects regression model that predicted gaze duration for each individual word (i.e. the total duration of all fixations on a word the first time that word is read) by simulation-eliciting content (as the effect of interest), with lexical frequency, word length and surprisal value as covariates (see Figure 3A; values for all predictors were at the word-level). *P*-values were estimated using the “*lmerTest*” package (Kuznetsova, Brockhoff, & Christensen, 2017). We controlled for lexical frequency, word length and surprisal value, because previous studies have shown that high frequency words are associated with shorter gaze durations than low frequency words (see Rayner, 1998 for a review), longer words with longer gaze durations than shorter words (e.g. Rayner & Fischer, 1996; Rayner, Sereno, & Raney, 1996), and words that are more likely to occur given their context (low surprisal value) with shorter gaze durations than unlikely words (Goodkind & Bicknell, 2018; Hale, 2001; Levy, 2008). The effects of the different types of descriptive words (motor descriptions, perceptual descriptions and descriptions of mental events) were allowed to vary per story per participant (i.e. different intercepts and slopes were allowed for stories 1, 2, and 3 for each participant<sup>2</sup>). This resulted in a total of 294 different coefficients<sup>3</sup> for each predictor (102 participants times 3 stories, minus single story-readings of 4 participants based on insufficient quality of the eye-tracking data and of 8 participants based on poor performance on the comprehension check). Data for the first word of each slide were excluded, as previous research has shown that fixations on these words are disproportionately long, due to the after effect of the fixation cross (Van den Hoven et al., 2016).

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

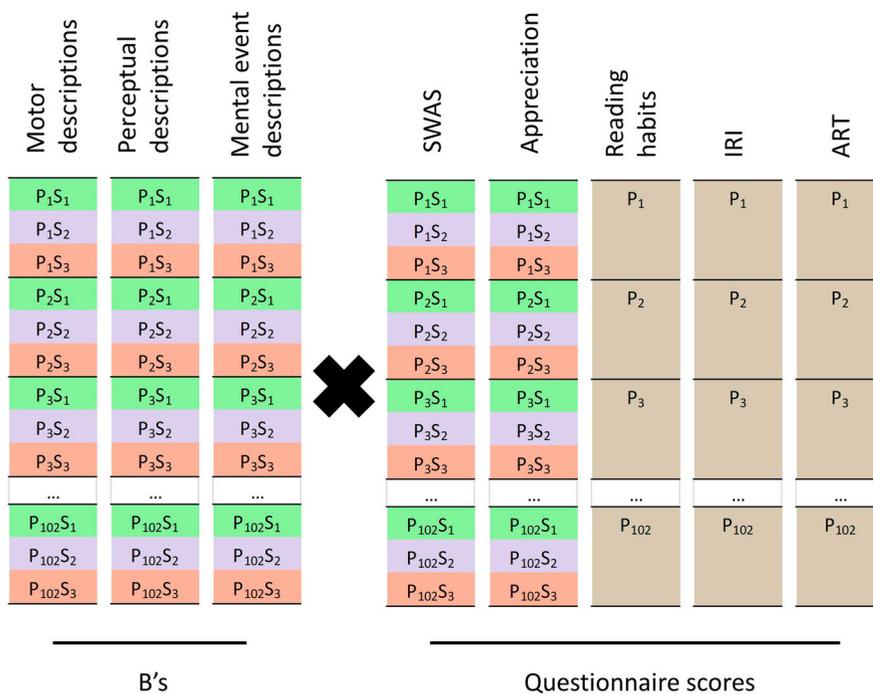
were excluded from the analysis (resulting in the loss of another 3.20% of the total amount of data).

A model of eye movements during reading, the E-Z Reader model (Reichle, Pollatsek, Fisher, & Rayner, 1998; see Reichle, Warren, & McConnell, 2009 for version 10 of this model), predicts spill-over effects in reading behaviour, when considering reading behaviour on the word level (these spill-over effects have been previously described by Mitchell, 1984; Rayner, 1998; see Reichle & Drieghe, 2015 for an account on how E-Z reader can be used to explain spill-over effects). The E-Z Reader model assumes that words are processed serially, meaning that the processing of one word has to be completed before processing of the next word can be started. However, after a first stage of initial processing (which is based on the “familiarity” of the word), a saccade can already be made towards the next word. As a result, deeper processing (based on the meaning of the word) and integration of the word into a sentence representation can actually take place while the gaze has already shifted towards the next word, resulting in spill-over effects (the effect of a variable on the processing speed at word  $n$  is reflected in the gaze duration towards word  $n + 1$ ). The effect of simulation-eliciting content on gaze duration that we would like to unveil would not primarily lie in the processing of the familiarity of the word being processed, but rather in the processing of the meaning of the word, and would therefore be expected to be reflected by the gaze duration in the spill-over region (as later processing of words is reflected by spill-over effects; Rayner, Ashby, Pollatsek, & Reichle, 2004). To account for these spill-over effects, the scores for all predictors included in the model (i.e. lexical frequency, word length, surprisal value, motor descriptions, perceptual descriptions and mental event descriptions) were taken from the previous word (all predictors thus taken to predict gaze duration at the next word in the story instead of the word they were derived from; comparable to approaches reported by, among others, Calvo & Meseguer, 2002; Frisson, Koole, Hughes, Olson, & Wheeldon, 2014; Kliegl, Nuthmann, & Engbert, 2006; Rayner et al., 2004; Schroyens, Vitu, Brysbaert, & D’Ydewalle, 1999). However, please note that after analysing the data at the level of the target word (i.e. if predictors derived from the current word were included instead of predictors derived from the previous word), results remained highly similar: Statistically significant effects were found, with the same direction for the three types of descriptions as in the spill-over analysis.

We constructed a generalised linear mixed effects regression model to predict the probability of regressing into an interest area (i.e. word). Regressions are right-to-left eye-movements, indicating a difficulty in the



A



B

**Figure 3.** A. In the first step of the analysis, scores from the underlining pre-test for motor, perceptual and mental event descriptions per word (with red words having a high score in the underlining pre-test, that is they score highly on that type of description) were linked to gaze duration scores (or to scores indicating whether or not a regression back to a word had been made) per word. B. In the second step of the analysis, the coefficients for the relationships between motor, perceptual and mental event descriptions and gaze duration/regression probability (per participant per story) were linked to the questionnaire data (per participant; for the SWAS and Appreciation questionnaires per participant per story). The rationale of this second analysis was to see if individual differences (as discovered in our questionnaire data) were related to individual differences in the way reading behaviour was affected by the different kinds of simulation, as established in step 1 of the analysis.

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

We decided to look at gaze duration, as this has been found to be a good measure of difficulty of processing of a word (Just & Carpenter, 1980; Rayner, 1998). Another possibility would have been to look at first fixation duration. However, first fixation duration and gaze duration are not independent of one another: first fixation durations are always part of, and often equal to, gaze durations, making it statistically undesirable to perform analyses on both variables (Kliegl & Laubrock, 2017). Therefore, we choose to use gaze durations in our analyses, as this is “considered [to be] the upper bound of early processing” in reading research (Kliegl & Laubrock, 2017, p. 77). Apart from gaze duration, we looked at the probability of regressing back to a word, as a measure of difficulty in the processing of a previous portion of the text or of incorporating a word into a mental representation of a sentence (Rayner et al., 2004). Readers have been found to be quite accurate in making a saccade back to the word with which they have trouble integrating, indicating that the word on which the eye lands after a regression is usually the word they found difficult to process (Rayner, 1998). Because of the strong co-dependency of different eye-tracking measures, we decided to choose our measures of interest prior to our study to avoid the pitfall of looking at too many different measures and subsequently reporting “spuriously significant results” (Kliegl & Laubrock, 2017, p. 78).

### Data-analysis: step 2

In the second part of our analysis, we wanted to link individual differences in the relationships between the

different kinds of simulation and gaze duration to individual differences in absorption, appreciation, perspective taking-ability, and reading experience. The purpose of this analysis was to see whether self-report measures of reading experiences such as transportation, mental imagery, appreciation, etc. that have been used in previous studies would be associated with simulation as measured using eye-tracking. In order to test this, we derived the slopes per participant per story for the relationships between the different kinds of simulation and gaze duration/regression probability (i.e. 294 different coefficients for motor, perceptual and mental event descriptions) from the predictions of both models from the first part of the analysis, and investigated how these were related to absorption and appreciation. In this part of the analysis, we constructed three models, each predicting the coefficients of one of the three types of descriptions (per participant and per story), by the questionnaire scores per participant (and per story for the simulation and appreciation questionnaires), allowing for random intercepts per participant (see Figure 3B).

## Results

### Questionnaires

#### SWAS

The four subscales of the Story World Absorption Scale all showed good or excellent reliability; Attention (5 items), Cronbach’s  $\alpha = .90$ ; Transportation (5 items),  $\alpha = .87$ ; Emotional Engagement (6 items),  $\alpha = .90$ ; Mental Imagery (8 items),  $\alpha = .91$ . Descriptives per subscale and per story are given in Table 3.

#### Appreciation questionnaire

The Appreciation Questionnaire was divided into two parts for the analysis. The first part, consisting of thirteen adjectives that could be used to describe the stories, was analysed using a principal components analysis (PCA)

**Table 3.** Mean, standard deviation, median and interquartile range for the scores on the SWAS subscales, per story.

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

**Table 4.** Summary of the Principal Components Analysis results for the 13 adjectives on the appreciation questionnaire ( $N = 294$ ). Factor loadings over .40 appear in bold

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

with oblique rotation (direct oblimin). Using the Kaiser-Meyer-Olkin measure, it was determined that the sampling adequacy for this analysis was good,  $KMO = .87$  (all  $KMO$  values for individual items  $> .75$ ). There was sufficient correlation between items, as indicated by Bartlett's test of sphericity,  $\chi^2(78) = 2061.961$ ,  $p < .001$ . An initial analysis showed that two components had eigenvalues over 1 (Kaiser's criterion). However, a model with 2 components did not fit the data well enough (fit based upon off diagonal values). Therefore, in the final model three components were retained. This model explained 68% of the variance. The first component contained items measuring the evoked interest in the story (beautiful, boring (-), captivating, interesting), the second component contained items measuring the emotional response to the story (sad, tragic, ominous, deeply moving, suspenseful), and the third component contained items measuring the positive affect elicited by the story (witty, funny, entertaining, special). The structure and pattern matrices for the factor loadings after rotation can be seen in Table 4. Factor scores per participant and story were used in the subsequent analyses.

A second part of the questionnaire consisted of a general score of story liking, and 6 questions regarding the enjoyment of the story,  $\alpha = .93$  (7 items). The answers on these questions were collapsed into a mean score for General Appreciation. These General Appreciation scores turned out to be highly correlated with the Evoked Interest factor scores,  $r_s = .846$ ,  $p < .001$ . To prevent too high levels of multicollinearity, it was decided to use only the Evoked Interest factor score in further analyses as an indicator of evoked interest/general appreciation (EI/GA).

### Top 3 questionnaire

After reading all three stories, participants were asked to rank the stories from most appreciated to least

appreciated. Most participants preferred story A, ranked story B as second best and story C as least appreciated (see Table 5). Note that participants read all stories in counterbalanced order, this preference was not an order effect.

### Reading habits and author recognition test

Reading experience was measured both directly using a reading habits questionnaire, and indirectly using the Author Recognition Test (ART). Because answers on the reading habits questionnaire were measured on a scale ranging from 1 to 5 on four of the five questions, but from 1 to 4 on the final question,  $z$  scores were calculated for all questions on this questionnaire (higher values indicating more reading experience). Overall reliability was sufficient if the question about non-fiction reading was excluded,  $\alpha = .71$ . The scores on the ART were positively skewed ( $M = 7.324$ ,  $SD = 4.695$ , median = 6.000, IQR = 4.000–9.000) with higher values indicating more (literary) reading experience.

### Interpersonal reactivity index

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

**Table 5.** Percentage of times each story was ranked as most appreciated, intermediate or least appreciated ( $n = 101$ ).

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

regularity, about things that might happen to me) was dropped,  $\alpha = .71$ .

## Eye-tracking data

### Gaze duration

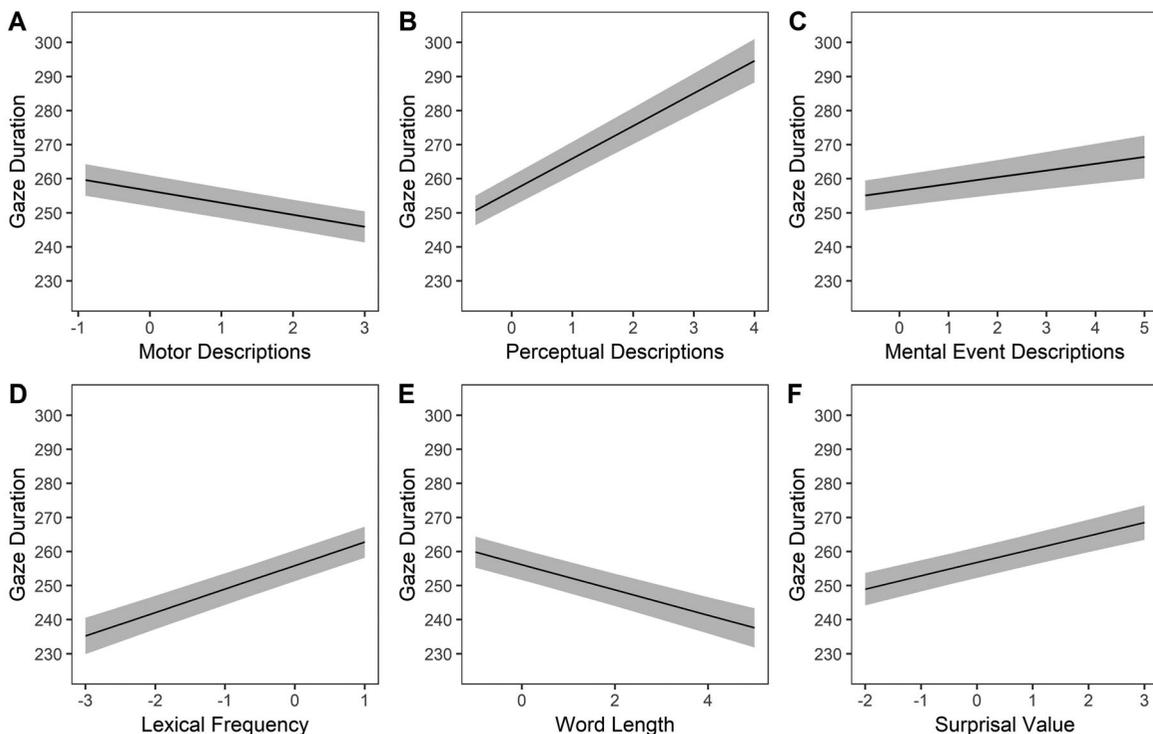
For the full model summaries, see Supplemental Material C. A Linear Mixed Effects Regression model was created, that predicted gaze duration by the number of times the previous word was underlined for motor descriptions, perceptual descriptions and mental event descriptions, controlling for lexical frequency, number of characters and surprisal value, and allowing random slopes per story per participant for the three different types of descriptions. All predictors were centred and scaled, to improve model fit. Variance Inflation Factors (VIFs) were calculated for this model, to check for multicollinearity (VIFs were calculated using the function reported on <https://github.com/aufrank/R-hacks/blob/master/mer-utils.R>). All VIFs were below 5, the VIFs for the underlining-scores were all around 1. This indicates that multicollinearity was not problematic in our models and all planned predictors were entered into the models.

As can be seen in Table 6 and Figure 4, motor descriptions were associated with a decrease in gaze duration (increased reading speed), whereas perceptual descriptions and descriptions of mental events were associated with an increase in gaze duration (decreased reading

**Table 6.** Coefficients of the model predicting gaze duration by motor descriptions, perceptual descriptions and mental event descriptions, taking into account the influence of lexical frequency, number of characters and surprisal value (all predictors taken from the previous word).

	<i>B</i>	<i>SE</i>	<i>df</i>	<i>t</i> -value	<i>p</i> -value
(Intercept)	255.500	2.295	294.1	111.317	<.001
Motor descriptions	-3.525	0.286	274.5	-12.316	<.001
Perceptual descriptions	9.555	0.354	318.3	26.977	<.001
Mental event descriptions	1.980	0.321	326.2	6.165	<.001
Lexical frequency	6.884	0.462	449200	14.908	<.001
Number of characters	-3.707	0.354	464900	-10.463	<.001
Surprisal value	3.912	0.387	462500	10.097	<.001

speed). More frequent or more unexpected (as reflected by a high surprisal value) words were read slower than infrequent or more expected words, as reflected by an increase in gaze duration towards frequent words and unexpected words. Longer words (words consisting of more characters) were associated with a decrease in gaze duration, reflecting increased reading speed. The effects of lexical frequency and word length were reversed to the effects that would have been expected based on previous research (see Data Analysis: Step 1). It should be noted that these effects were as expected in the analysis for the target word (in that analysis more frequent words were associated with a decrease in gaze duration (faster reading



**Figure 4.** Effects plots for the predictors of gaze duration in the spillover area. Note that all predictors are centred and scaled. Gaze duration is given in milliseconds, the grey areas indicate the 95-percent confidence intervals.

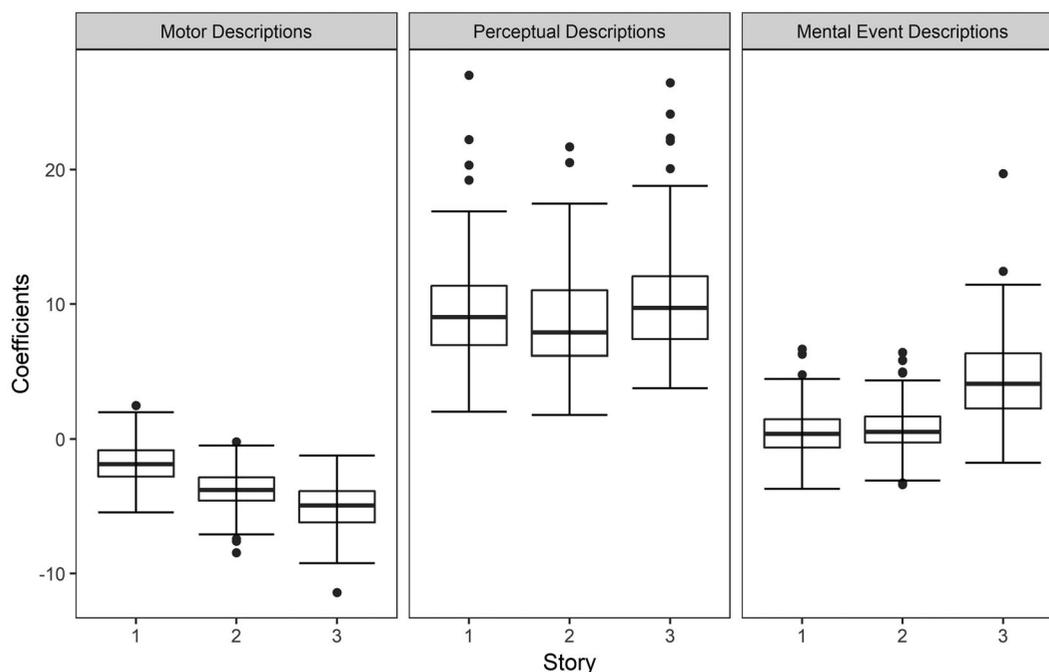
and longer words with an increase in gaze duration (slower reading); for an overview of the results of the analysis for the target word, see Supplemental Material D; for an elaborate discussion of these results, see the Discussion section). As can be seen in Figure 5, the associations between all types of descriptions and gaze duration varied between participants (and between different stories within participants; more detailed figures per participant can be found in Supplemental Material E1–3).

At a closer look, interesting individual differences in the association between simulation and gaze duration are visible (see Figure 5). For motor descriptions, the response of participants was rather homogeneous: nearly all participants showed decreased gaze duration after reading motor descriptions. Comparably, all participants showed an increase in gaze duration after reading perceptual descriptions. The association between mental event descriptions and gaze duration, however, seemed more variable between (and within, see Supplemental Material E3) participants, sometimes being associated with an increase and sometimes with a decrease in gaze duration (see Figure 5). Even though on average there was a decrease in reading speed for mental event descriptions, this was not always the case on the individual level. In fact, a considerable number of participants showed an increased instead of a decreased reading speed when reading mental event descriptions, in particular for stories 1 and 2.

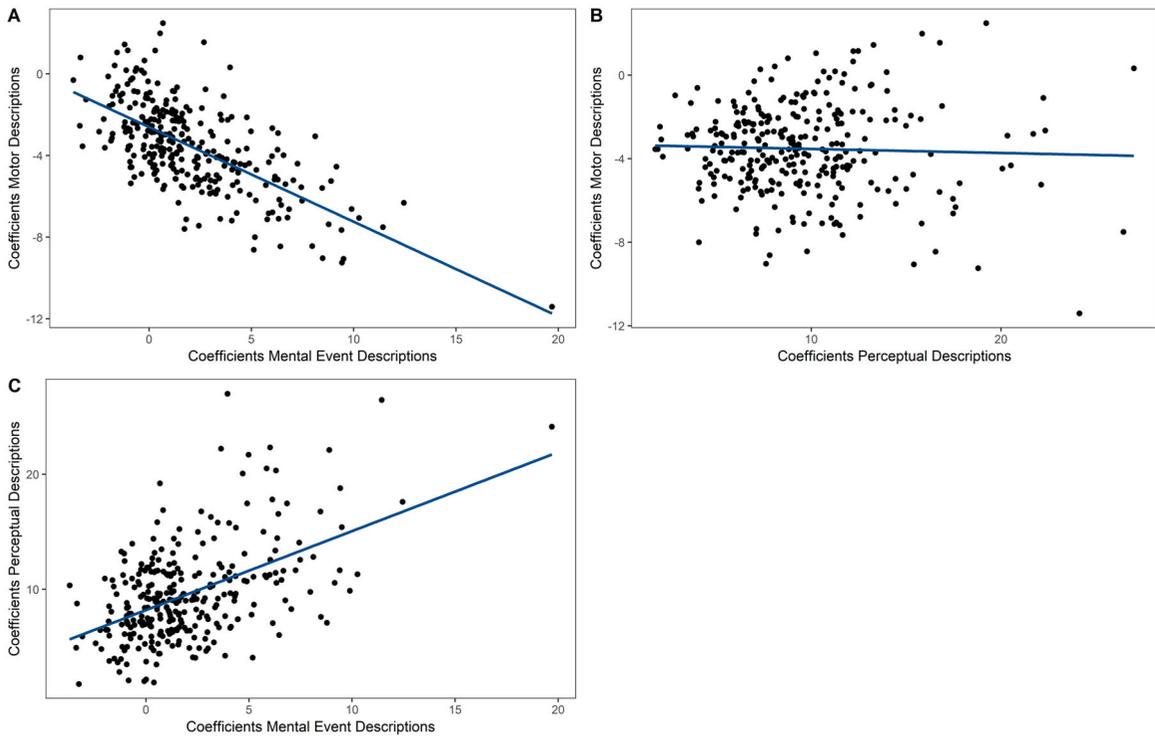
When comparing the coefficients for the relationships of the different types of descriptions with gaze duration, a significant negative correlation between the coefficients for the associations of motor descriptions and mental event descriptions with gaze duration appeared,  $r_s = -.622$ ,  $p < .001$ .<sup>4</sup> This indicated that participants showing a strong negative relationship between motor descriptions and gaze duration, showed a strong positive relationship between mental event descriptions and gaze duration, and vice versa (see Figure 6A). In contrast, no correlation was found between the coefficients for the associations of motor descriptions and perceptual descriptions with gaze duration,  $r_s = -.003$ ,  $p = .964$  (see Figure 6B). There was a significant positive correlation between the coefficients for the associations of perceptual descriptions and mental event descriptions with gaze duration,  $r_s = .403$ ,  $p < .001$ , indicating that participants showing a strong positive relationship between perceptual descriptions and gaze duration also showed a strong positive relationship between mental event descriptions and gaze duration (see Figure 6C).

#### Individual differences: gaze duration

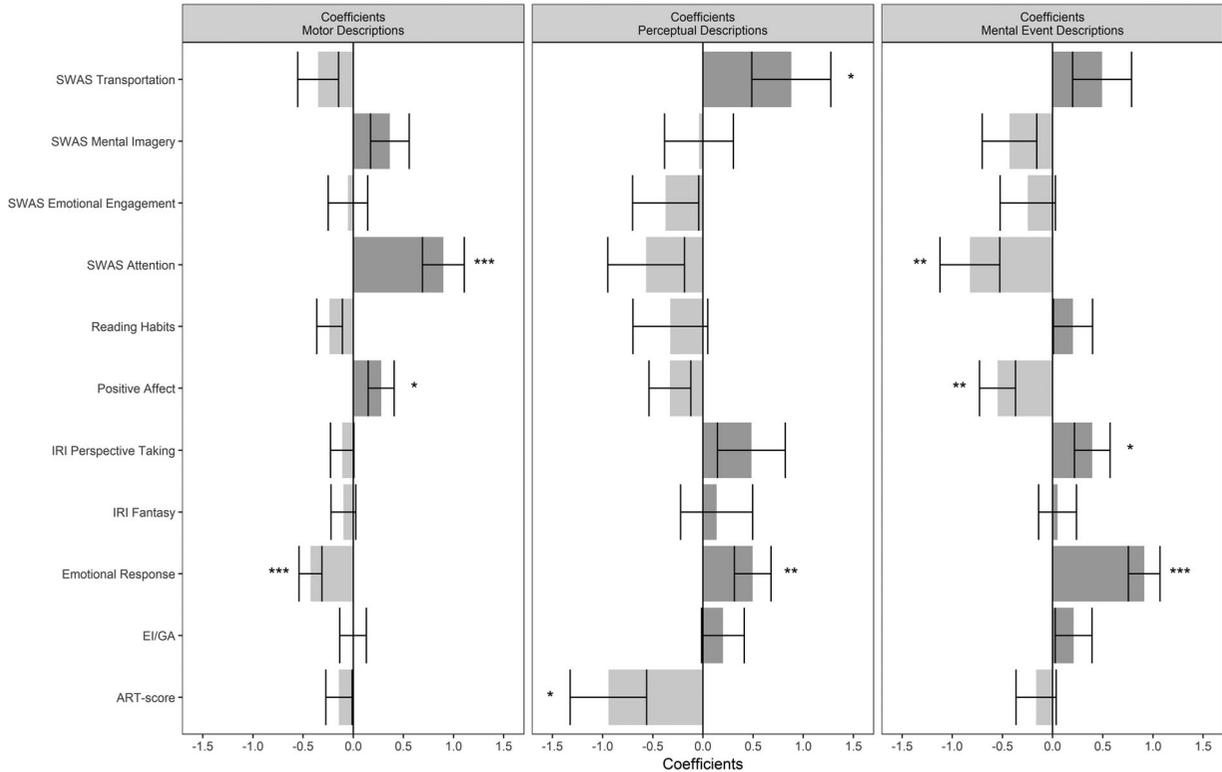
To test whether self-report measures of reading experiences were associated with simulation as measured using eye-tracking, coefficients for the associations of the three types of descriptions with gaze duration were derived per story per participant (total number of coefficients = 294). We created three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



**Figure 5.** Range of coefficients across participants for the relationships between different types of descriptions and gaze duration, depicted per story and per type of description.



**Figure 6.** Correlation between the coefficients of the relationships of (A) motor descriptions and mental event descriptions with gaze duration, (B) motor descriptions and perceptual descriptions with gaze duration, and (C) of perceptual and mental event descriptions with gaze duration.



**Figure 7.** Coefficients per predictor for each of the three models predicting the strength of the relationships between the three types of descriptions and gaze duration by individual differences measured with questionnaires. Negative coefficients appear in light grey, positive coefficients in dark grey. Errorbars indicate Standard Errors. EI/GA = Evoked Interest/General Appreciation. Significant predictors are marked (\*  $p < .05$ , \*\*  $p < .01$ , \*\*\*  $p < .001$ ).

**Table 7.** Coefficients of the model predicting the effect of motor descriptions on gaze duration by scores on the questionnaires.

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

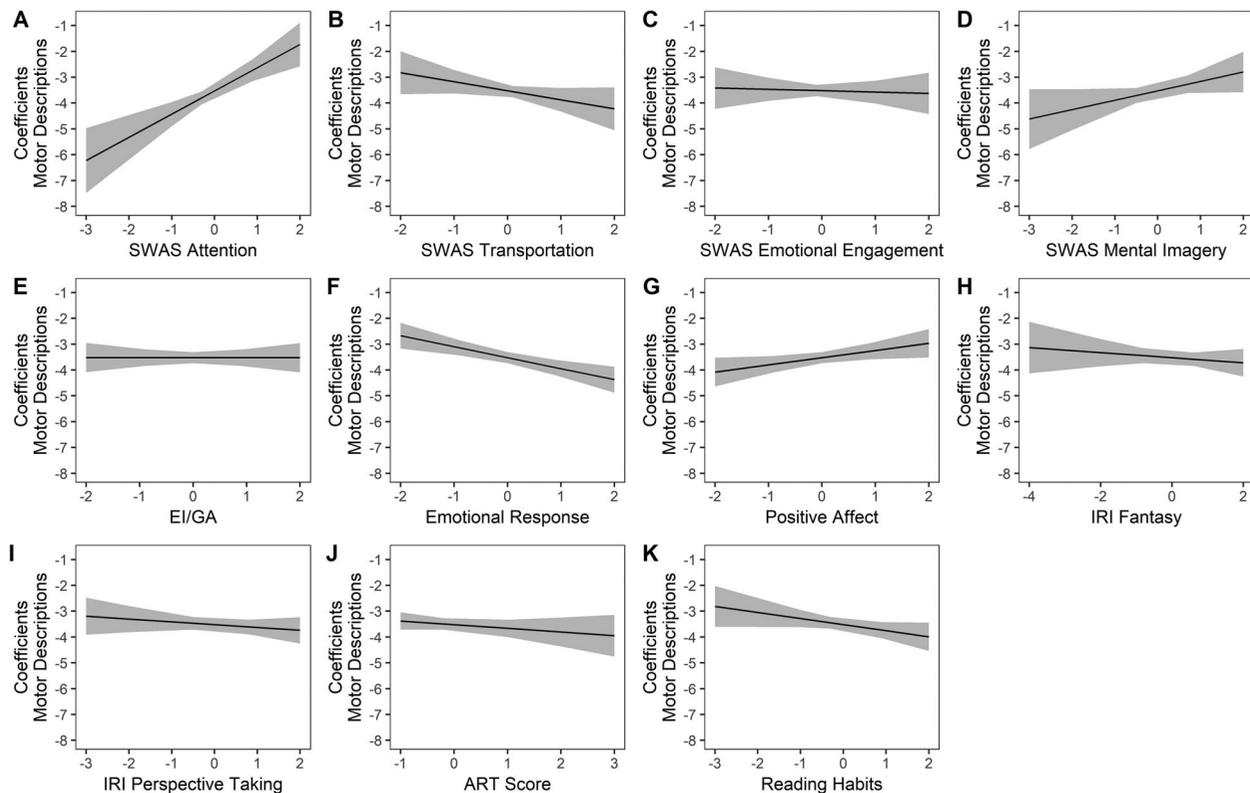
Significant predictors are marked (\*  $p < .05$ , \*\*  $p < .01$ , \*\*\*  $p < .001$ ).

these associations by scores on the four subscales of the SWAS, the three factor scores for appreciation, the Fantasy and Perspective Taking subscales of the Interpersonal Reactivity Index and the direct and indirect measures of reading experience (i.e. the reading habits questionnaire and Author Recognition Test). Random intercepts were allowed per participant. Again, all predictors were centred and scaled, to improve model fit. VIFs were below 5 for all predictors (indicating that multicollinearity was not problematic in our models), and all planned predictors were entered into the models. An

overview of the results of the three models can be seen in Figure 7.

**Motor simulation.** As can be seen in Table 7 and Figure 8, the Attention subscale of the SWAS, the evoked emotional response, and the positive affect elicited by the story were significantly associated with the strength of the relationship between motor descriptions and gaze duration. Mental imagery and positive affect were positively associated with the strength of the relationship between motor descriptions and gaze duration, implying that this relationship (faster reading of motor descriptions) was attenuated (motor descriptions were read relatively slower) in people who reported higher mental imagery or positive affect after reading a story. Evoked emotional response was negatively associated, implying that people who reported experiencing a high level of emotion while reading a story, read motor descriptions even faster (the strength of the relationship between motor descriptions and gaze duration was increased).

**Perceptual simulation.** Table 8 and Figure 9 show that the Transportation subscale of the SWAS was significantly associated with the strength of the relationship between perceptual descriptions and gaze duration, as well as the emotional response evoked by the story

**Figure 8.** Effects plots for the predictors of the strength of the relationship between motor descriptions and gaze duration. Note that all predictors are centred and scaled. The grey areas indicate the 95-percent confidence intervals.

**Table 8.** Coefficients of the model predicting the effect of perceptual descriptions on gaze duration by scores on the questionnaires.

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

Significant predictors are marked (\*  $p < .05$ , \*\*  $p < .01$ , \*\*\*  $p < .001$ ).

and the ART-score (indirect measure of reading experience). Reading experience was negatively associated with the strength of the relationship between perceptual descriptions and gaze duration, implying that in people reporting more reading experience, the relationship between perceptual descriptions and gaze duration (slower reading of perceptual descriptions) was attenuated (perceptual descriptions were read relatively faster). Transportation and the evoked emotional response, however, were positively associated: Individuals reporting a higher level of transportation or

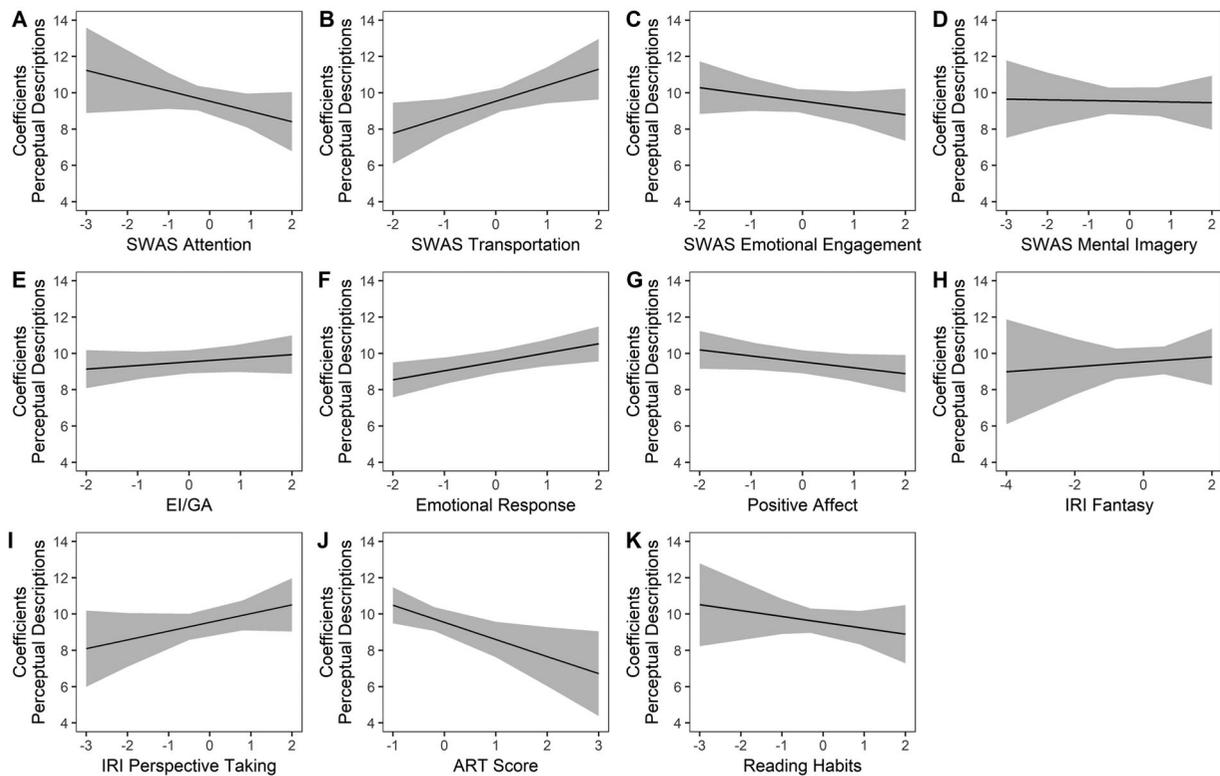
**Table 9.** Coefficients of the model predicting the effect of mental event descriptions on gaze duration by scores on the questionnaires.

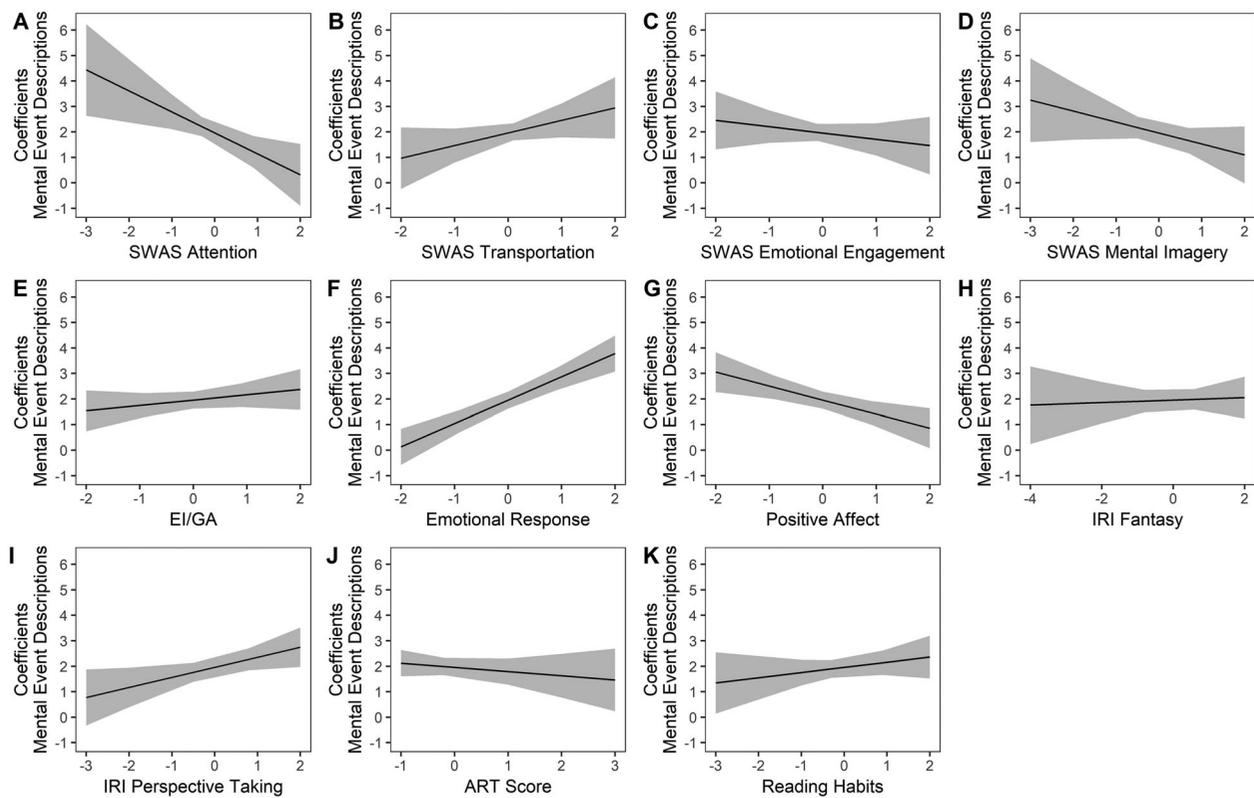
	<i>B</i>	<i>SE</i>	<i>df</i>	<i>t</i> -value	<i>p</i> -value
(Intercept)	1.969	0.168	100.68	11.734	<.001***
SWAS Attention	-0.824	0.298	260.89	-2.766	.006**
SWAS Transportation	0.494	0.294	235.54	1.679	.095
SWAS Emotion	-0.248	0.275	289.63	-0.902	.368
SWAS Mental Imagery	-0.429	0.272	272.57	-1.576	.116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

Significant predictors are marked (\*  $p < .05$ , \*\*  $p < .01$ , \*\*\*  $p < .001$ ).

experienced emotion while reading a story, read perceptual content even slower (the strength of the relationship between perceptual descriptions and gaze duration was increased).

**Mentalising.** The strength of the relationship between mental event descriptions and gaze duration was associated with scores on the Attention subscale of the SWAS, as well as the evoked emotional response and the positive affect elicited by the story, and the perspective taking-ability of the participants (see Table 9 and

**Figure 9.** Effects plots for the predictors of the strength of the relationship between perceptual descriptions and gaze duration. Note that all predictors are centred and scaled. The grey areas indicate the 95-percent confidence intervals.



**Figure 10.** Effects plots for the predictors of the strength of the relationship between mental event descriptions and gaze duration. Note that all predictors are centred and scaled. The grey areas indicate the 95-percent confidence intervals.

Figure 10). Attention and elicited positive affect both had a negative association with the strength of the relationship between mental event descriptions and gaze duration: in readers who reported higher attention or higher positive affect, the strength of this relationship was attenuated (resulting in some participants even showing an increased instead of decreased reading speed when reading mental event descriptions). In contrast, evoked emotional response and participants' perspective taking abilities had a positive association with the strength of the relationship between mental event descriptions and gaze duration. The strength of this relationship was increased in participants who reported experiencing a high level of emotion while reading a story or who reported often considering other people's perspectives.

### Regression probability

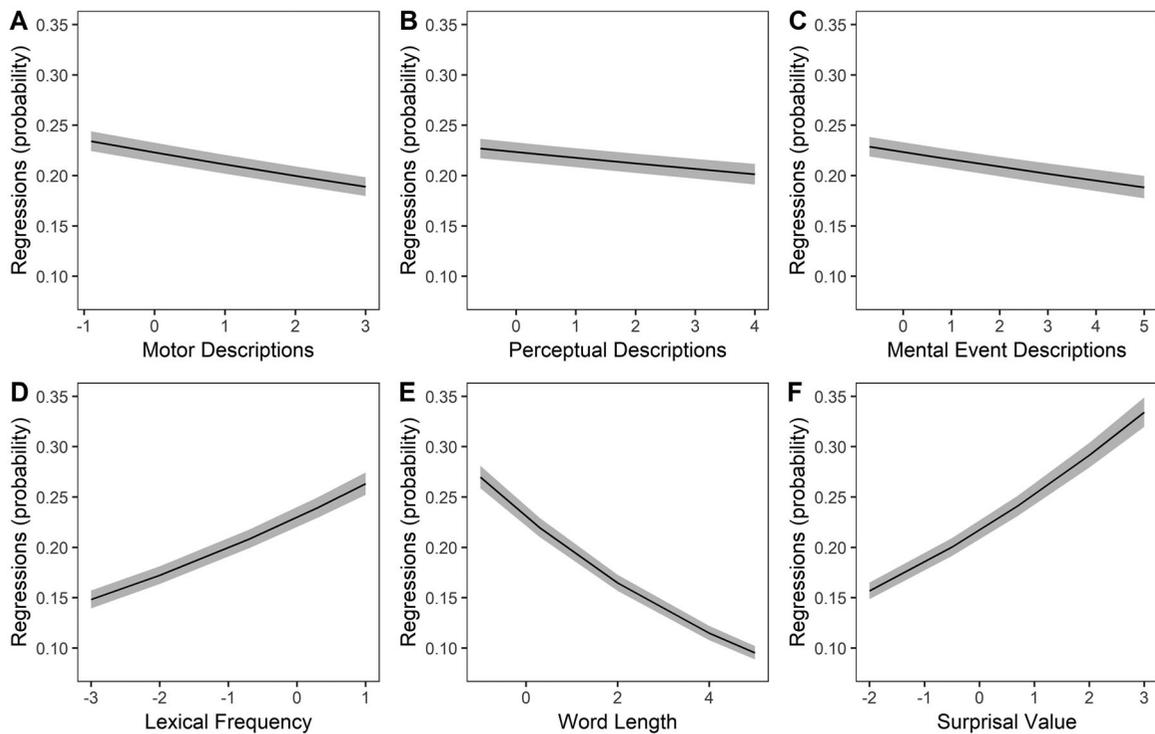
For the full model summaries, see Supplemental Material C. A Generalised Linear Mixed Effects model was created, that predicted the probability of regressing back to a word by the number of times this word was underlined for motor descriptions, perceptual descriptions and descriptions of mental events, controlling for lexical frequency, number of

characters and surprisal value, and allowing random intercepts and slopes for underlining-scores per story per participant. Again, all predictors were centred and scaled, to improve model fit. All VIFs for this model were below 5, the VIFs for the underlining-scores were all close to 1.

As can be seen in Table 10 and Figure 11, motor descriptions, perceptual descriptions and descriptions of mental events were all associated with a decrease in the probability of regressing back to a word. More frequent or more unexpected (as reflected by a high surprisal value) words were more likely to be looked back to than infrequent or more expected words, as reflected

**Table 10.** Coefficients of the model predicting the probability of regressing back to a word by motor descriptions, perceptual descriptions and mental event descriptions, taking into account the influence of lexical frequency, number of characters and surprisal value

	<i>B</i>	<i>SE</i>	<i>z</i> -value	<i>p</i> -value
(Intercept)	-1.203	0.028	-43.075	<.001
Motor descriptions	-0.070	0.004	-16.959	<.001
Perceptual descriptions	-0.033	0.004	-7.869	<.001
Mental event descriptions	-0.043	0.004	-9.767	<.001
Lexical frequency	0.180	0.008	22.894	<.001
Number of characters	-0.209	0.006	-36.109	<.001
Surprisal value	0.199	0.007	30.139	<.001

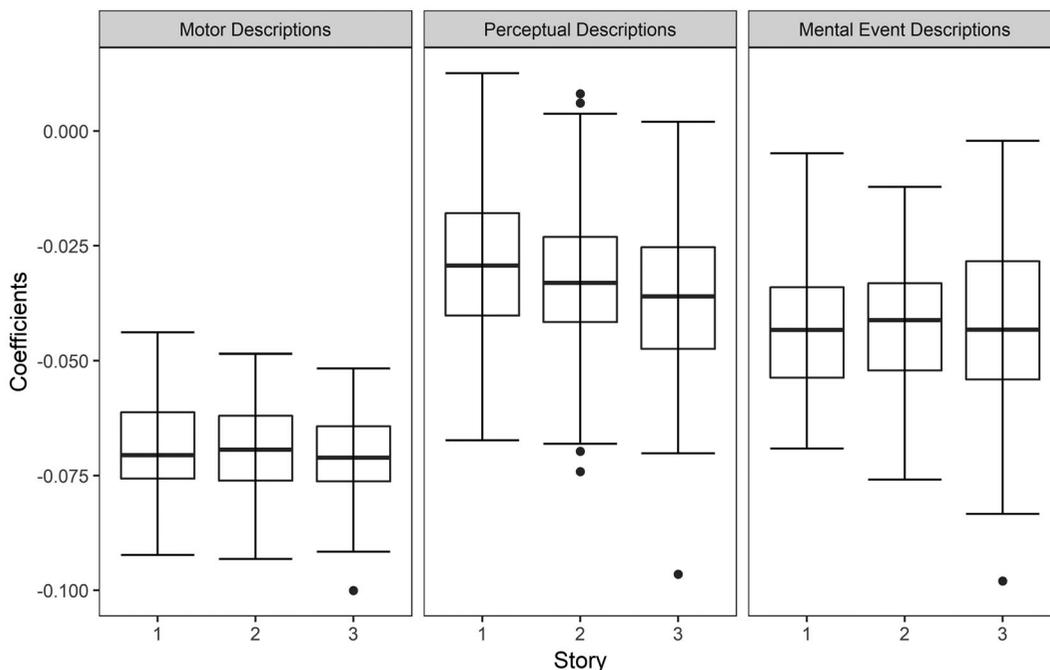


**Figure 11.** Effects plots for the predictors of the probability of regressing back to a word. Note that all predictors are centred and scaled. The grey areas indicate the 95-percent confidence intervals.

by an increase in the probability of regressing back to frequent words and unexpected words. Longer words (words consisting of more characters) were associated with a decrease in the probability of regressing back to that word.

#### *Individual differences: regression probability*

The relationships between the three types of descriptions and the regression probability did not vary much between participants (and between different stories within participants; see also Figure 12). As a result, no



**Figure 12.** Range of coefficients across participants for the relationships between different types of descriptions and regression probability, depicted per story and per type of description.

notable associations were found between any of the measures of individual differences (i.e. SWAS, appreciation, IRI, and measures of reading experience) and the strength of the relationship between any type of description and regression probability.

## Discussion

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

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

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

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

In our analysis, we controlled for a number of factors known to influence reading behaviour. Higher surprisal value (indicating lower expectancy of a word given its context) was associated with prolonged gaze duration (slower reading) in the spill-over region (i.e. gaze durations on words occurring after unexpected words were relatively long). Interestingly, lexical frequency was associated with prolonged gaze duration (i.e. gaze durations on words occurring after frequent words

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

As we controlled for lexical frequency, word length and surprisal value in our analyses, it is unlikely that differences in these characteristics between motor descriptions and other parts of the stories caused the negative association between motor simulation and gaze duration. Moreover, when looking at the distribution of the data from the pre-test (see Supplemental Material A1–3), we see that motor descriptions occurred comparably often in all parts of the sentences. The same was true for perceptual descriptions and mental event descriptions. As a result, it is also unlikely that the relationship between motor descriptions and gaze duration merely reflects an effect of position in the sentence. Apparently, motor descriptions are processed differently from the rest of the text (and differently from other types of descriptions), suggesting that motor descriptions might actually be easier to process than other parts of

the text (as they are associated with faster reading). How this exactly relates to motor simulation, and whether motor simulation is indeed “easier” than other types of simulation, still remains to be seen.

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

The emotional response evoked by the stories, on the other hand, was positively related to the strength of the relationships between all kinds of simulation and gaze duration. Participants who found the stories more sad, tragic, ominous, deeply moving and suspenseful showed relatively strong relationships between simulation and gaze duration. This can be interpreted as evidence that participants who were moved by the stories they read were more prone to mentally simulate the events happening in the story (or the other way around: participants who are more prone to simulate the events happening in a story are to a larger extent moved by the stories). This is an extension of what Oatley (1995) suggested about simulation and emotion. With simulation, he meant identification with characters and the simulation of the emotions of the characters in a story, which is reminiscent of the concept of mentalising (simulation of mental events,

such as thoughts and emotions) we use here. He suggested that readers simulate the emotions of a character by tapping into their own emotional experience. This implies that when readers report more emotions elicited by a story, this results from a more vivid simulation of the emotions described in this story. This explains why in this study the strength of the relationship of mentalising (and to a lesser extent motor and perceptual simulation) with gaze duration was larger in participants reporting a higher emotional response to the stories.

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

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

Reading experience was negatively associated with the strength of the relationship between perceptual simulation and gaze duration (meaning that more experienced readers read perceptual descriptions relatively faster). Interestingly, a comparable result was found in a combined analysis of several eye-tracking

datasets including the present dataset (Eekhof et al., n.d.). It was observed that participants showing the weakest relationships between word characteristics (such as lexical frequency) and gaze durations, reported relatively high reading experience. This suggests that more experienced readers are more “detached” from low-level word characteristics. The results from the current study suggest that, to some extent, the same can be said about simulation-eliciting content: more experienced readers seem to be less influenced by this kind of content.

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

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

## Conclusion

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

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

## Notes

1. Many researchers have tried to capture the experience of becoming part of a story, resulting in constructs such as immersion (Ryan, 2001; see also Jacobs, 2015), absorption (e.g. Kuijpers et al., 2014), transportation (Gerrig, 1993; Green & Brock, 2000) or presence (Kuzmičová, 2012). For the sake of clarity, we will refer to this experience as absorption for the remainder of this article.
2. As a result of the nature of our random effect structure, random effects were calculated for (1 | story:subject) – which has 294 levels (i.e. 294 individual subject and story combinations). Although it can be argued that in our experiment story is not necessarily nested in subject, crossed random effects (where random effects are calculated for (1 | subject) and (1 | story), but not for the interaction between the two) would mean that random effects would be calculated for (1 | story) – which has only 3 levels (namely, three stories were used in this experiment). As it is advised to only calculate random effects for variables with more than 5 or 6 levels, this approach would, statistically speaking, not have been favourable.
3. A reviewer pointed out to us that these coefficients, derived from the random slopes in our model, are strictly speaking Best Linear Unbiased Predictions (or BLUPs), as opposed to the population-level coefficients derived from the model summary (Best Linear Unbiased Estimates, or BLUEs). However, using such terminology in

the text of this article, might make it unnecessarily complex, therefore the BLUPs will be referred to as “coefficients”.

4. Note that this analysis was exploratory, and was not part of our pre-registration.

## Acknowledgements

The authors thank two anonymous reviewers for their helpful comments on earlier versions of this manuscript, and Alessandro Lopopolo for the calculation of the surprisal values used in our analyses.

## Disclosure statement

No potential conflict of interest was reported by the authors.

## Funding

This work was supported by the Netherlands Organization for Scientific Research (NWO) [Grant Number Vidi-276-89-007].

## ORCID

Marloes Mak  <http://orcid.org/0000-0001-8183-4039>

## References

- Altmann, U., Bohrn, I. C., Lubrich, O., Menninghaus, W., & Jacobs, A. M. (2014). Fact vs fiction-how paratextual information shapes our reading processes. *Social Cognitive and Affective Neuroscience*, 9, 22–29. doi:10.1093/scan/nss098
- Apperly, I. A., & Butterfill, S. A. (2009). Do humans have two systems to track beliefs and belief-like states? *Psychological Review*, 116(4), 953–970. doi:10.1037/a0016923
- Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, 59(1), 617–645. doi:10.1146/annurev.psych.59.103006.093639
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. doi:10.18637/jss.v067.i01
- Bergen, B., & Wheeler, K. (2010). Grammatical aspect and mental simulation. *Brain and Language*, 112, 150–158. doi:10.1016/j.bandl.2009.07.002
- Binder, J. R., Conant, L. L., Humphries, C. J., Fernandino, L., Simons, S. B., Aguilar, M., & Desai, R. H. (2016). Toward a brain-based componential semantic representation. *Cognitive Neuropsychology*, 33(3–4), 130–174. doi:10.1080/02643294.2016.1147426
- Bruner, J. S. (1986). *Actual minds, possible worlds*. Cambridge, Massachusetts: Harvard University Press.
- Burke, M. (2011). *Literary reading, cognition and emotion: An exploration of the oceanic mind*. *Literary reading, cognition and emotion: An exploration of the oceanic mind*. Oxford, UK: Routledge. doi:10.4324/9780203840306
- Busselle, R., & Bilandzic, H. (2009). Measuring narrative engagement. *Media Psychology*, 12(4), 321–347. doi:10.1080/15213260903287259

- Calvo, M. G., & Meseguer, E. (2002). Eye movements and processing stages in reading: Relative contribution of visual, lexical, and contextual factors. *The Spanish Journal of Psychology*, 5(1), 66–77. doi:10.1017/S1138741600005849
- 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. doi: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.
- Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology*, 44(1), 113–126. doi:10.1037/0022-3514.44.1.113
- De Corte, K., Buysse, A., Verhofstadt, L. 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. doi:10.5334/pb-47-4-235
- Deen, B., & McCarthy, G. (2010). Reading about the actions of others: Biological motion imagery and action congruency influence brain activity. *Neuropsychologia*, 48, 1607–1615. doi:10.1016/j.neuropsychologia.2010.01.028
- Eekhof, L., Kuijpers, M. M., Faber, M., Gao, X., Mak, M., Van den Hoven, E., & Willems, R. M. (n.d.). Lost in a story, detached from the words.
- Engelen, J. A. A., Bouwmeester, S., de Bruin, A. B. H., & Zwaan, R. A. (2011). Perceptual simulation in developing language comprehension. *Journal of Experimental Child Psychology*, 110, 659–675. doi:10.1016/j.jecp.2011.06.009
- Estes, Z., Verges, M., & Barsalou, L. W. (2008). Head up, foot down: Object words orient attention to the objects' typical location. *Psychological Science*, 19(2), 93–97. doi:10.1111/j.1467-9280.2008.02051.x
- Fernandino, L., Humphries, C. J., Seidenberg, M. S., Gross, W. L., Conant, L. L., & Binder, J. R. (2015). Predicting brain activation patterns associated with individual lexical concepts based on five sensory-motor attributes. *Neuropsychologia*, 76, 17–26. doi:10.1016/j.neuropsychologia.2015.04.009
- Filik, R., & Leuthold, H. (2013). The role of character-based knowledge in online narrative comprehension: Evidence from eye movements and ERPs. *Brain Research*, 1506, 94–104. doi:10.1016/j.brainres.2013.02.017
- Fischer, M. H., & Zwaan, R. A. (2008). Embodied language: A review of the role of the motor system in language comprehension. *Quarterly Journal of Experimental Psychology*, 61(6), 825–850. doi:10.1080/17470210701623605
- Feroni, F., & Semin, G. R. (2013). Comprehension of action negation involves inhibitory simulation. *Frontiers in Human Neuroscience*, 7, 1–7. doi:10.3389/fnhum.2013.00209
- Frisson, S., Koole, H., Hughes, L., Olson, A., & Wheeldon, L. (2014). Competition between orthographically and phonologically similar words during sentence reading: Evidence from eye movements. *Journal of Memory and Language*, 73, 148–173. doi:10.1016/j.jml.2014.03.004
- Frith, C. D., & Frith, U. (1999). Interacting minds - a biological basis. *Science*, 286(5445), 1692–1695. doi:10.1126/science.286.5445.1692
- Frith, U., & Frith, C. D. (2003). Development and neurophysiology of mentalizing. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 358, 459–473. doi:10.1098/rstb.2002.1218
- Gerrig, R. J. (1993). *Experiencing narrative worlds: On the psychological activities of reading*. Boston: Yale University Press.
- Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin & Review*, 9(3), 558–565. doi:10.3758/BF03196313
- Goldman, A. I. (2012). Theory of mind. In E. Margolis, R. Samuels, & S. P. Stich (Eds.), *The Oxford Handbook of Philosophy of cognitive Science* (pp. 402–424). New York: Oxford University Press. doi:10.1093/oxfordhb/9780195309799.013.0017
- Goodkind, A., & Bicknell, K. (2018). Predictive power of word surprisal for reading times is a linear function of language model quality. In A. Sayeed, C. Jacobs, T. Linzen, & M. 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.
- Green, M. C., & Brock, T. C. (2000). The role of transportation in the persuasiveness of public narratives. *Journal of Personality and Social Psychology*, 79(5), 701–721. doi:10.1037/0022-3514.79.5.701
- Green, M. C., Brock, T. C., & Kaufman, G. F. (2004). Understanding media enjoyment: The role of transportation into narrative worlds. *Communication Theory*, 14(4), 311–327. doi:10.1111/j.1468-2885.2004.tb00317.x
- Green, M. C., & Donahue, J. K. (2009). Simulated worlds: Transportation into narratives. In K. D. Markman, W. M. P. Klein, & J. A. Suhr (Eds.), *Handbook of imagination and mental simulation* (pp. 241–256). New York: Psychology Press.
- Hale, J. (2001). A probabilistic early 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. doi:doi: 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), e0154732. doi: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. doi:10.1016/j.bandl.2017.03.007
- Hauk, O., Johnsrude, I., & Pulvermüller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron*, 41, 301–307. doi:10.1016/S0896-6273(03)00838-9
- 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. doi:10.1097/WNR.0000000000000272
- Jacobs, A. M. (2015). Neurocognitive poetics: Methods and models for investigating the neuronal and cognitive-affective bases of literature reception. *Frontiers in Human Neuroscience*, 9, 1–22. doi:10.3389/fnhum.2015.00186
- Jacobs, A. M., & Willems, R. M. (2017). The fictive brain: Neurocognitive correlates of engagement in literature. *Review of General Psychology*. doi:10.1037/gpr0000106

- Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, 87(4), 329–354. doi:10.1037/0033-295X.87.4.329
- Kemmerer, D. (2015). Are the motor features of verb meanings represented in the precentral motor cortices? Yes, but within the context of a flexible, multilevel architecture for conceptual knowledge. *Psychonomic Bulletin & Review*, 22, 1068–1075. doi:10.3758/s13423-014-0784-1
- Keuleers, E., Brysbaert, M., & New, B. (2010). SUBTLEX-NL: A new measure for Dutch word frequency based on film subtitles. *Behavior Research Methods*, 42(3), 643–650. doi:10.3758/BRM.42.3.643
- Kliegl, R., & Laubrock, J. (2017). Eye-movement tracking during reading. In A. M. B. De Groot & P. Hagoort (Eds.), *Research Methods in Psycholinguistics and the Neurobiology of language: A Practical Guide* (pp. 68–88). Hoboken, NJ: Wiley.
- Kliegl, R., Nuthmann, A., & Engbert, R. (2006). Tracking the mind during reading: The influence of past, present, and future words on fixation durations. *Journal of Experimental Psychology: General*, 135(1), 12–35. doi:10.1037/0096-3445.135.1.12
- Knoop, C. A., Wagner, V., Jacobsen, T., & Menninghaus, W. (2016). Mapping the aesthetic space of literature “from below.”. *Poetics*, 56, 35–49. doi:10.1016/j.poetic.2016.02.001
- Koopman, E. M. (Emy) (2015). Empathic reactions after reading: The role of genre, personal factors and affective responses. *Poetics*, 50, 62–79. doi:10.1016/j.poetic.2015.02.008
- Kovács, ÁM, Kühn, S., Gergely, G., Csibra, G., & Brass, M. (2014). Are all beliefs equal? Implicit belief attributions recruiting core brain regions of theory of mind. *PLoS ONE*, 9(9), e106558. doi:10.1371/journal.pone.0106558
- 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. doi:10.1075/ssol.4.1.05kui
- Kurby, C. A., & Zacks, J. M. (2013). The activation of modality-specific representations during discourse processing. *Brain and Language*, 126, 338–349. doi:10.1016/j.bandl.2013.07.003
- Kuzmičová, A. (2012). Presence in the reading of literary narrative: A case for motor enactment. *Semiotica*, 189(1), 23–48. doi:10.1515/semi.2011.071
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). Lmertest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13), 1–26. doi:10.18637/jss.v082.i13
- Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, 106, 1126–1177. doi:10.1016/j.cognition.2007.05.006
- Mar, R. A., & Oatley, K. (2008). The function of fiction is the abstraction and simulation of social experience. *Perspectives on Psychological Science*, 3(3), 173–192. doi:10.1111/j.1745-6924.2008.00073.x
- Mar, R. A., Oatley, K., Hirsh, J., dela Paz, J., & Peterson, J. B. (2006). Bookworms versus nerds: Exposure to fiction versus non-fiction, divergent associations with social ability, and the simulation of fictional social worlds. *Journal of Research in Personality*, 40, 694–712. doi:10.1016/j.jrp.2005.08.002
- Marino, B. F., Borghi, A. M., Buccino, G., & Riggio, L. (2017). Chained activation of the motor system during language understanding. *Frontiers in Psychology*, 8, 1–11. doi:10.3389/fpsyg.2017.00199
- Miall, D. S., & Kuiken, D. (2002). A feeling for fiction: Becoming what we behold. *Poetics*, 30, 221–241. doi:10.1016/S0304-422X(02)00011-6
- Mitchell, D. C. (1984). An evaluation of subject-paced reading tasks and other methods for investigating immediate processes in reading. In D. Kieras & M. A. Just (Eds.), *New methods in reading comprehension research* (pp. 69–89). Hillsdale, NJ: Erlbaum.
- Mumper, M. L., & Gerrig, R. J. (2017). Leisure reading and social cognition: A meta-analysis. *Psychology of Aesthetics, Creativity, and the Arts*, 11(1), 109–120. doi:10.1037/aca0000089
- Nabokov, V. (2003). *Signalen en symbolen. In Een Russische schoonheid 1*. De Bezige Bij: Amsterdam.
- Nijhof, A. D., Brass, M., Bardi, L., & Wiersema, J. R. (2016). Measuring mentalizing ability: A within-subject comparison between an explicit and implicit version of a ball detection task. *PLoS ONE*, 11(10), e0164373. doi:10.1371/journal.pone.0164373
- Nijhof, A. D., & Willems, R. M. (2015). Simulating fiction: Individual differences in literature comprehension revealed with fMRI. *PLoS ONE*, 10(2), e0116492. doi:10.1371/journal.pone.0116492
- Oatley, K. (1995). A taxonomy of the emotions of literary response and a theory of identification in fictional narrative. *Poetics*, 23, 53–74. doi:10.1016/0304-422X(94)P4296-S
- Oatley, K. (2012). The cognitive science of fiction. *Wiley Interdisciplinary Reviews: Cognitive Science*, 3, 425–430. doi:10.1002/wcs.1185
- Ostarek, M., & Vigliocco, G. (2017). Reading sky and seeing a cloud: On the relevance of events for perceptual simulation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 43(4), 579–590. doi:10.1037/xlm0000318
- Pollatsek, A., Juhasz, B. J., Reichle, E. D., Machacek, D., & Rayner, K. (2008). Immediate and delayed effects of word frequency and word length on eye movements in reading: A reversed delayed effect of word length. *Journal of Experimental Psychology: Human Perception and Performance*, 34(3), 726–750. doi:10.1037/0096-1523.34.3.726
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124(3), 372–422. doi:10.1037/0033-2909.124.3.372
- Rayner, K., Ashby, J., Pollatsek, A., & Reichle, E. D. (2004). The effects of frequency and predictability on eye fixations in reading: Implications for the E-Z Reader model. *Journal of Experimental Psychology: Human Perception and Performance*, 30(4), 720–732. doi:10.1037/0096-1523.30.4.720
- Rayner, K., & Fischer, M. H. (1996). Mindless reading revisited: Eye movements during reading and scanning are different. *Perception & Psychophysics*, 58(5), 734–747. doi:10.3758/BF03213106
- Rayner, K., Sereno, S. C., & Raney, G. E. (1996). Eye movement control in reading: A comparison of two types of models. *Journal of Experimental Psychology: Human Perception and Performance*, 22(5), 1188–1200. doi:10.1037/0096-1523.22.5.1188
- Reichle, E. D., & Drieghe, D. (2015). Using E-Z Reader to examine the consequences of fixation-location measurement error. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 41(1), 262–270. doi:10.1037/a0037090

- Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review*, 105(1), 125–157. doi:10.1037/0033-295X.105.1.125
- Reichle, E. D., Reineberg, A. E., & Schooler, J. W. (2010). Eye movements during mindless reading. *Psychological Science*, 21(9), 1300–1310. doi:10.1177/0956797610378686
- Reichle, E. D., Warren, T., & McConnell, K. (2009). Using E-Z reader to model the effects of higher level language processing on eye movements during reading. *Psychonomic Bulletin and Review*, 16(1), 1–21. doi:10.3758/PBR.16.1.1
- Ryan, M. L. (2001). *Narrative as Virtual Reality: Immersion and Interactivity in literature and Electronic Media*. Baltimore/London: Johns Hopkins University Press.
- Samur, D., Lai, V. T., Hagoort, P., & Willems, R. M. (2015). Emotional context modulates embodied metaphor comprehension. *Neuropsychologia*, 78, 108–114. doi:10.1016/j.neuropsychologia.2015.10.003
- Schroyens, W., Vitu, F., Brysbaert, M., & D'Ydewalle, G. (1999). Eye movement control during reading: Foveal load and parafoveal processing. *The Quarterly Journal of Experimental Psychology Section A*, 52(4), 1021–1046. doi:10.1080/713755859
- Sereno, S. C., & Rayner, K. (2003). Measuring word recognition in reading: Eye movements and event-related potentials. *Trends in Cognitive Sciences*, 7(11), 489–493. doi:10.1016/j.tics.2003.09.010
- Shanton, K., & Goldman, A. (2010). Simulation theory. *Wiley Interdisciplinary Reviews: Cognitive Science*, 1, 527–538. doi:10.1002/wcs.33
- Stanfield, R. A., & Zwaan, R. A. (2001). The effect of implied orientation derived from verbal context on picture recognition. *Psychological Science*, 12(2), 153–156. doi:10.1111/1467-9280.00326
- Stanovich, K. E., & West, R. F. (1989). Exposure to print and orthographic processing. *Reading Research Quarterly*, 24(4), 402–433. doi:10.2307/747605
- 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. doi:10.1093/scan/nsv114
- Tettamanti, M., Buccino, G., Saccuman, M. C., Gallese, V., Danna, M., Scifo, P., ... Perani, D. (2005). Listening to action-related sentences activates fronto-parietal motor circuits. *Journal of Cognitive Neuroscience*, 17(2), 273–281. doi:10.1162/0898929053124965
- 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. doi:10.1525/collabra.39
- Van Essen, R. (2014). De mensen die alles lieten bezorgen. In *Hier wonen ook mensen* (pp. 113–123). Utrecht: Atlas Contact.
- Van Hassel, S. (2012). De Chinese Bruiloft. In *Ezels*. Amsterdam: De Bezige Bij.
- van Krieken, K., Hoeken, H., & Sanders, J. (2017). Evoking and measuring identification with narrative characters - A linguistic cues framework. *Frontiers in Psychology*, 8, 1–16. doi:10.3389/fpsyg.2017.01190
- Van Overwalle, F., & Vandekerckhove, M. (2013). Implicit and explicit social mentalizing: Dual processes driven by a shared neural network. *Frontiers in Human Neuroscience*, 7, 1–6. doi:10.3389/fnhum.2013.00560
- Wallentin, M., Simonsen, A., & Nielsen, A. H. (2013). Action speaks louder than words: Empathy mainly modulates emotions from theory of mind-laden parts of a story. *Scientific Study of Literature*, 3(1), 137–153. doi:10.1075/ssol.3.1.11wal
- Willems, R. M., & Casasanto, D. (2011). Flexibility in embodied language understanding. *Frontiers in Psychology*, 2, 1–11. doi:10.3389/fpsyg.2011.00116
- Willems, R. M., & Francken, J. C. (2012). Embodied cognition: Taking the next step. *Frontiers in Psychology*, 3, 1–3. doi:10.3389/fpsyg.2012.00582
- Zwaan, R. A. (2009). Mental simulation in language comprehension and social cognition. *European Journal of Social Psychology*, 39, 1142–1150. doi:10.1002/ejsp.661
- Zwaan, R. A., Stanfield, R. A., & Yaxley, R. H. (2002). Language comprehenders mentally represent the shapes of objects. *Psychological Science*, 13(2), 168–171. doi:10.1111/1467-9280.00430