ARTICLE

The influences of narrative perspective shift and scene detail on narrative semantic processing

Jian Jin[1,2] (ID) and Siyun Liu[2,3,4] (ID)

[1]School of Educational Science, Anhui Normal University, Wuhu, China
[2]School of Psychology, Central China Normal University, Wuhan, China
[3]Ministry of Education, Key Laboratory of Adolescent Cyberpsychology and Behavior, Central China Normal University, Wuhan, China
[4]Key Laboratory of Human Development and Mental Health of Hubei Province, Central China Normal University, Wuhan, China
Corresponding author: Siyun Liu; Email: liusy@ccnu.edu.cn

Abstract
The embodied view of semantic processing holds that readers achieve reading comprehension through mental simulation of the objects and events described in the narrative. However, it remains unclear whether and how the encoding of linguistic factors in narrative descriptions impacts narrative semantic processing. This study aims to explore this issue under the narrative context with and without perspective shift, which is an important and common linguistic factor in narratives. A sentence-picture verification paradigm combined with eye-tracking measures was used to explore the issue. The results showed that (1) the inter-role perspective shift made the participants' to evenly allocate their first fixation to different elements in the scene following the new perspective; (2) the internal–external perspective shift increased the participants' total fixation count when they read the sentence with the perspective shift; (3) the scene detail depicted in the picture did not influence the process of narrative semantic processing. These results suggest that perspective shift can disrupt the coherence of situation model and increase the cognitive load of readers during reading. Moreover, scene detail could not be constructed by readers in natural narrative reading.

Keywords: eye-tracking; narrative perspective shift; scene detail; semantic processing; sentence-picture verification task

1. Introduction

Semantic processing is crucial in reading comprehension. Many previous studies advocate the embodied view of semantic processing and believe that semantic processing involves readers' sensorimotor system (Barsalou, 2008; Dam et al.,

2010; García et al., 2019; Phillips et al., 2012; Zwaan, 2016). Along with this view, mental simulation theory further proposes that readers can automatically or subconsciously simulate the textual descriptions during semantic processing in natural reading (Mak, 2022; Mak & Willems, 2021; Zwaan, 2009). For example, narrative comprehension studies have found that readers obtain the meaning of narrative texts through automatically simulating the objects' physical features (e.g., orientation, shape and color) and the characters' visual perspectives, actions and emotions depicted in narratives (Brunyé et al., 2011; Ditman et al., 2010; Mak & Willems, 2019; Zwaan & Pecher, 2012). This strand of studies demonstrates that the mental simulation automatically elicited by textual descriptions contributes to readers' semantic processing of the text.

Textual descriptions are commonly composed of linguistic factors. Discourse researchers and linguists have long proposed that linguistic factors could influence readers' semantic processing and reading comprehension (Gernsbacher, 1997; Givón, 1992; McNamara & Magliano, 2009; Morrow, 1986). For example, the structure-building model (Gernsbacher, 1997) and event-indexing model (Zwaan et al., 1995) suggest that narrative linguistic information such as referents, time, location and causality affect readers' coherent mental representations or simulations of narrative content. McNamara and Magliano (2009) also believe that the linguistic features that denote referential cohesion (i.e., the amount of explicit overlap of referents, concepts and ideas between adjacent sentences or paragraphs in a text; e.g., two adjacent sentences with the same subject have high referential cohesion) and situational cohesion (i.e., the semantic connections of two events in terms of five dimensions about time, space, entity, causality and goals; e.g., two events with the same actor and happened in the same period have high situational cohesion) will affect the coherence of readers' mental representations about the protagonists, their goals and the objects they interact with, as well as the spatial locations and temporal information they interact (i.e., situation model construction) (Bailey et al., 2017; Zwaan & Radvansky, 1998).

Another series of studies on event segmentation provides insights into the relationship between linguistic factors and semantic processing. Event segmentation researchers found that individuals' understanding and memory of continuous information such as narrative content is achieved by spontaneously parsing it into discrete, meaningful events (Radvansky & Zacks, 2017; Shin & DuBrow, 2021). An effective process of event segmentation is beneficial to individuals' understanding and memory of continuous information (Flores et al., 2017; Zacks et al., 2009). Event segmentation is based on top-down conceptual factors (e.g., comprehenders' goal, perspective, context knowledge or attentional focus) and bottom-up sensory factors (e.g., perceptual change of the object in narrative situation) (Bestgen & Vonk, 2000; Mariola et al., 2022; Zacks & Swallow, 2007). For example, comprehends will segment continuous narrative information in a storybook or film into discrete units based on their orientated attention to the perceptual change occurred in the aspects of character, space or time in the narrative (Bailey et al., 2017; Zacks et al., 2009). However, readers' event segmentation should be more heavily driven by bottom-up factors than top-down factors (Newberry & Bailey, 2019). Therefore, narrative-based bottom-up factors such as perceptual cues manipulated through linguistic devices are critical in event segmentation. In other words, the perceptual changes in mental simulation of linguistically depicted situational features could influence event

segmentation, which in turn influences the understanding and memory of narrative content.

To sum up, both the studies of situation model and event segmentation indicate the importance of linguistic factors in semantic processing. However, it remains unclear how the encoding of linguistic factors affects online dynamic semantic processing in natural reading. Additionally, narrative scene elements and details are crucial for semantic processing, as readers may mentally simulate these contents to construct a situation model. Previous situation model studies mainly explored the representation of explicitly introduced situation elements (e.g., time and space) during situation model construction (McNamara & Magliano, 2009), but it is not clear whether scene details not mentioned in narrative texts could be represented in the situation model by readers. The present study aims to fill these gaps under the narrative context with and without perspective shift, which is an important and common linguistic factor in narrative works such as literary fiction and may influence the coherence of situation model construction in narrative semantic processing. In particular, Chinese narrative text will be used to explore these issues since most previous studies regarding narrative perspective shift were conducted in alphabetic languages such as English, and little is known under non-alphabetic languages such as Chinese.

1.1. Encoding of the shifts in narrative perspective in reading

Narrative perspective is a basic linguistic component of narrative works as it determines how a narration is presented to readers (Miall & Kuiken, 2001; Rall & Harris, 2000). It is like a 'window' through which readers could 'see' and 'hear' the events happened in the narrative and track the progress of the narrative. Graesser et al. (2001) believe that readers will follow the perspectives of characters (i.e., perspective-taking, such as what characters do and feel) to construct and update the world of the narrative during natural reading. These thoughts were validated in previous empirical studies, which found that narrative perspective plays an important role in narrative engagement and reading comprehension (Brunyé et al., 2016; Child et al., 2018; Samur et al., 2021). For example, studies using personal pronouns as the indicator of narrative perspective showed that compared with the third-person pronouns ('he/she/they') that induced an external or observer's perspective of readers, the first- ('I/we') and second-person pronouns ('you') allowed readers to simulate the narrative world from an internal/actor's perspective, thus contributing more to readers' narrative comprehension and engagement in reading (Brunyé et al., 2016; Butler et al., 2016; Child et al., 2018; Ditman et al., 2010; Hartung et al., 2016; Samur et al., 2021).

Multiple narrative perspectives are typically found in narratives, with *perspective shift* occurring when the narrative changes from one viewpoint to another (Jin & Liu, 2023). Two types of perspective shifts have been proposed by previous studies. The first is the shift between *internal* and *external* perspectives (Millis, 1995). For example, 'Tom (*external perspective*) is washing dishes in the kitchen when he mutters to himself, "I'll (*internal perspective*) go for a video game after the dishes are done"'. In this example, the first clause described Tom's washing behavior from a third-person external perspective, while the second clause shifted to Tom's first-person internal perspective that described his conscious thought. In this example,

direct speech was used to manipulate the internal–external perspective shift. An internal–external perspective shift occurs when a direct speech is suddenly used in a third-person external narration, and the indirect speech would keep the original third-person external perspective consistent (Millis, 1995). The second type is the shift between two characters, that is, the *inter-role* perspective shift. An inter-role perspective shift could occur even when the third-person external perspective is used throughout the narrative (Black et al., 1979). For example, 'Tom (*Tom's perspective*) is washing dishes in the kitchen when his mother **goes** into (*mother's perspective*) the kitchen'. In this example, although Tom and his mother were both described from a third-person external perspective, the verb 'go' (vs. 'come') would elicit a perspective shift from Tom to his mother in readers' mind (vs. consistent to Tom's perspective) (Black et al., 1979). Therefore, the verb phrase 'go into' is used to manipulate an inter-role perspective shift, and the verb phrase 'come into' will keep the perspective consistent (Black et al., 1979). Previous studies suggested that information of both internal–external and inter-role perspective shift could be successfully encoded by readers in reading, though this process usually demanded more cognitive effort than that for a consistent perspective (Black et al., 1979; Cui, 2017; Jin & Liu, 2023; Millis, 1995; Schmid & Baccino, 2002).

1.2. Perspective shift, situation model's coherence and semantic processing

Many reading comprehension theories believe that the construction of a coherent situation model is a prerequisite for obtaining narrative semantics and thus reading comprehension (McNamara & Magliano, 2009). The encoding of the perspective shift should bring difficulties to the construction of a coherent situation model in natural reading, which may not be conducive to semantic processing.

Perspective shift encoding studies provide the indirect evidence that perspective shift impedes the coherence of situation model construction. These studies found that readers experienced greater difficulties and consumed more cognitive resources in comprehending text with shifted perspective than that with consistent perspective. For instance, Black et al. (1979) compared readers' representations of semantic coherence of English narratives with inter-role shifted and consistent perspectives. They found that the participants read the sentences with shifted perspectives more slowly, made more errors when recalling them and subjectively rated them as more difficult to understand, compared with that of the consistent perspective sentences. Following Black et al.'s (1979) study, some Western researchers further investigated the memory processing of perspective shift in children and found a similar recalling effect (Rall & Harris, 2000; Ziegler et al., 2005). A recent study by Jin and Liu (2023) explored whether the processing of inter-role perspective shifts in Chinese narratives affect readers' attentional focus. They used a dual-task paradigm (i.e., pressing keys to identify auditory tones [the secondary task] while reading Chinese narrative paragraphs [the primary task]) to investigate this question and found that the participants spent more attentional resources on the encoding of shifted than consistent perspective. Moreover, the cognitive demanding features of the perspective shift were also found in literary reading. For example, Millis (1995) used a rereading paradigm to investigate whether perspective shift would be encoded by readers in English literary reading. He found that the participants spent more time reading literary narratives with internal–external perspective shift during the first-round reading, but not the

second round. Millis (1995) explained that this was because the information of the perspective shift had already been successfully encoded by the participants from the first-round reading, resulting in less time in processing in the second round.

Overall, previous findings suggest that the encoding of perspective shift brings difficulties and consumes cognitive resources for readers to construct a coherent situation model. The shifts in perspective require readers to simulate the narrative content from a new perspective, which breaks the coherence of the mental simulation from the original perspective. In order to keep the coherence of situation model, we speculate that readers should spend more attentional resources on the semantic processing of the text after perspective shift; readers may also need to reallocate their attention to various elements in situation model following the new perspective, to effectively integrate the information from the new perspective into the existed situation model.

1.3. The details of situation model construction during semantic processing

Will any scene details not explicitly mentioned in a sentence be mentally imagined and incorporated into situation model during sentence semantic processing? If yes, how the scene details in situation model influence semantic processing? These questions have long existed in the study of text and discourse comprehension, but surprisingly few studies have empirically investigated them. Kintsch (1988) believes that readers' situation model construction involves generating inferences that lead to the incorporation of relevant background knowledge into mental representation (McNamara & Magliano, 2009). According to this view, the situation model includes all inferences that go beyond the concepts that are explicitly mentioned in the text (McNamara & Magliano, 2009). For example, when reading the sentence 'Xiao Fang was cooking in the kitchen', the readers may imagine the not-mentioned details such as pot, bowl and range hood of the given concept 'kitchen', in order to represent the semantics of the sentence. However, Zwaan and Radvansky (1998) believe that situation model construction is conducted in a rather fast and abstract way, without any concrete details (Marschark & Cornoldi, 1991). As can be seen, previous theories have no clear conclusions about whether readers construct the details of the scene in situation model. We attempt to explore this issue in the current study.

The detail construction in the situation model may not be 'all-or-nothing'. Conversely, detail construction may be dynamic and constrained by factors such as the importance of the detail to the coherence of situation model, as well as the availability of readers' cognitive resources. These possibilities are supported in some indirect evidence. For example, Sundermeier et al. (2005) found that readers did not encode the location of the object in a narrative unless the location information was important to build causal coherence of the narrative. Additionally, several studies on narrative memory found that individuals could encode gist (e.g., protagonist) and peripheral detail (e.g., scene details) information of the narrative into situation model, but the detail information fades faster (Adams et al., 1997; Sekeres et al., 2016). This is because fewer cognitive resources were spent on detail information since it was less important compared with gist information in narrative processing (Adams et al., 1997; Sekeres et al., 2016). In this light, gist information could be processed and stored at a deeper level than scene details not explicitly mentioned by the text. In natural reading, when a new gist information (e.g., a new perspective or a

new person) is presented, readers may prioritize the processing of this new gist information because it is more important to the coherence of situation model than scene details (at least in our text materials). As such, readers will focus their cognitive resources on the processing of the new gist information and thus less cognitive resources for the construction of scene details, resulting in unsuccessful construction of scene details. In sum, we assume that detail construction could be limited by the representation of new gist information such as a perspective shift.

1.4. Overview of the current study

The current study aims to investigate whether and how perspective shift and scene detail may influence Chinese readers' semantic processing during natural reading. Following the above literature review, we assume that perspective shift should break the coherence of situation model during semantic processing. Additionally, we assume that readers will construct scene details in situation model, but a new gist information such as a new perspective or a new person depicted in the text will hinder this process.

2. The rationale for the methodology

Sentence-picture verification (SPV) task and eye-tracking measures were combined to test the aforementioned assumptions. In the following section, we will explain why these two methods are a good fit for probing semantic processing during natural reading and how they work.

2.1. SPV task

SPV paradigm has been broadly used to investigate semantic processing features of the object (Yaxley & Zwaan, 2007; Zwaan et al., 2002) and the event (Brunyé et al., 2016; Just & Carpenter, 1971) depicted in single sentence. SPV paradigm is effective in assessing semantic representation of sentence because readers' response to the picture is made primarily based on their correct representation of sentence semantics (Carpenter & Just, 1972). Specifically, in a SPV task, participants were asked to judge whether the picture matched the semantic context of the sentence they just read. As such, participants should process the picture and then make a response based on the semantic information (e.g., perspective shift and scene detail) they just obtained in the previous sentence (Brunyé et al., 2016; Carpenter & Just, 1972; Chen et al., 2008). Reaction times (RTs) and response accuracy of picture verification are used to indicate the participants' semantic processing features of the sentence. Slower response and more errors in verifying the sentence-semantic-matched picture indicate harder semantic processing of the sentence (Chen et al., 2008). In the present study, RTs and response accuracy in SPV task are assumed to be caused by perspective shift manipulation in the sentence and scene detail manipulation in the picture. The relationships among these variables will be hypothesized in detail later. We will compare the participants' RTs and response accuracy between shifted versus consistent narrative perspective conditions and rich versus limited picture detail conditions, in order to examine the influence of perspective shift and scene detail on sentence semantic processing.

2.2. Eye-tracking measures of sentence reading and picture viewing

The current study combines eye-tracking measures with SPV task to examine the semantic processing features of text with and without perspective shift. Eye-tracking is effective for online probing of readers' cognitive processing on the text (Reichle et al., 1998). This is because the cognitive processing of text is a time-sensitive process, and thus the increased cognitive processing should be detectable in fixation durations (in milliseconds) to the text elements that are thought to elicit the corresponding cognitive processes (Mak, 2022). In general, natural reading is fast, with a mean fixation duration of around 200–275 ms (Rayner, 1998). Due to its high temporal sensitivity, eye-tracking measures can be used to analyze the time course of text semantic processing within a few seconds of sentence reading (Barach et al., 2021). During natural reading, eye fixations on an individual word or phrase that depicts a single object reflect the mental simulation features during semantic processing of the object (Mak, 2022; Mak & Willems, 2021), and eye fixations on sentences reflect the semantic processing features of the sentences (Barach et al., 2021). Based on this, we examined the effect of perspective shift on text semantic processing by comparing eye-movement features on the text elements such as verb phrase and direct speech, which will be used to manipulate inter-role perspective shift and internal–external perspective shift separately in this study.

Additionally, the current study will record the participants' eye movement behaviors when they are viewing pictures in the SPV task. This method has been used in a few studies on text-picture semantics integration using SPV tasks (Carpenter & Just, 1972; Chen et al., 2008; Takacs & Bus, 2018). For example, Carpenter and Just (1972) have long demonstrated that in the SPV task, the participants' eye movement patterns when scanning a picture were determined by the semantic representation of the sentence preceding the picture. Takacs and Bus (2018) found that the participants view the elements in the picture that are congruent with the semantics of narrations in the same order as they appeared in the narrations. Since the participants need to make a judge on the basis of the semantic match between the sentence and the picture, the participants' eye movement pattern on viewing pictures should reflect the semantic validation process of the semantic information they obtained from the previous sentence. Based on this premise, how the participants processed the pictures which were highly semantic-related to the immediately previous presented sentences would reflect how the semantic processing of the sentences was conducted. In this light, the eye-movement recordings on the picture in SPV task are effective to assess the semantic representation process during narrative sentence reading.

Therefore, we can infer whether and how perspective shift information influence the participants' sentence semantic processing, as well as whether the participants construct scene details in their situation model, by analyzing the participants' eye-movement data during picture viewing. Specifically, if the participants encode shifted perspective differently from consistent perspective during sentence reading, their eye-movement patterns in viewing the picture under shifted perspective condition should be significantly different from the picture under consistent perspective condition. Similarly, if the participants incorporate scene details into situation model construction during sentence reading, they will fixate on picture details more often so as to verify the detail semantics they obtained from sentence reading. Therefore, the participants' eye-movement recordings on pictures with rich details should be

different from those with limited details. The specific eye-movement differences between different perspective or detail conditions will be proposed separately in the next section.

3. Hypotheses and analysis plan

Next, we will briefly summarize the main propositions from the literature review and propose the research hypotheses accordingly. The analysis plan for testing the hypotheses will also be made below.

We first consider the hypotheses regarding the effect of perspective shift. Inter-role perspective shift and internal–external perspective shift are both considered in this study because it is not clear whether these two types of perspective shift are also cognitively demanding under Chinese semantic processing. Since the encoding of shifted perspective is more cognitively demanding than that of consistent perspective, readers in shifted perspective condition would consume more cognitive resources in matching the shifted-perspective sentence and its semantically consistent pictures. Given the limited cognitive resources of readers, they should perform worse on the shifted than consistent perspective condition in the SPV task. Therefore, we propose a hypothesis regarding the behavioral indicators of perspective shift effect: the participants should response slower (i.e., longer RTs) and make more errors (i.e., lower accuracy) to the pictures under the shifted perspective condition than that under the consistent perspective condition (*Hypothesis 1*). We will test this hypothesis by comparing the participants' RTs and response accuracy to the pictures under the two perspective conditions, respectively.

Following the reasoning above, it is reasonable to infer that readers will spend more attention to understand the text that indicates a shifted perspective. Therefore, we propose two hypotheses regarding the eye-movement indicators of perspective shift effect during sentence reading: (1) *Hypothesis 2-1*: the participants would fixate longer on the phrases in the second sentence of a paragraph that indicates an inter-role perspective shift than a consistent perspective; (2) *Hypothesis 2-2*: the participants would fixate longer or more often on the third sentence of a paragraph that contains an internal–external perspective shift than a consistent perspective. We will test these hypotheses by comparing the eye fixation metrics on the text areas that depict a shifted and consistent perspective.

Additionally, since perspective shift could break the coherence of the constructed situation model, which brings difficulties for readers to integrate the new perspective information with the situation model constructed from the original perspective (Black et al., 1979; Jin & Liu, 2023; Millis, 1995). In other words, due to the shift in perspective, readers will confront a new narrative scene in which they are unfamiliar with the relations among the various elements in the scene. In this situation, the participants should have equal attentional priorities to each key element rather than giving attentional priority to specific elements, so as to effectively capture the relations among the elements. This is the prerequisite for obtaining the semantics of the new scene. As a result, the participants could effectively incorporate the extracted semantics of the new scene into the existing situation model. Based on these inferences, we speculate that the participants should have no attentional priorities on the various key elements of the pictures under shifted perspective condition; while under consistent perspective condition, the participants should give

attentional priority to specific components in the picture. Therefore, we propose a hypothesis about the eye-movement indicators of perspective shift effect during picture viewing: the participants will more evenly prioritize their first fixation on various elements in the pictures following the shifted perspective sentences, compared with that following the consistent perspective sentences (*Hypothesis 3*). We will test this hypothesis by comparing the first fixation time on the pictures following the sentences with or without perspective shift.

Finally, the hypotheses regarding scene detail effect are considered. Since the first sentence of the text provides the semantic foundation for the comprehension of the subsequent sentences (Gernsbacher, 1997; McNamara & Magliano, 2009), the first sentence is therefore critical to the coherence of situation model. In addition, readers have relatively rich cognitive resources at the start of text reading. Considering these two premises, the scene details depicted in the first sentence of a paragraph should be mentally represented. This possibility would present in the SPV task following the first sentence because a semantics match effect will occur if the picture semantics match the sentence semantics. If the detailed information is incorporated into sentence semantic processing, the participants should have a better performance to the pictures with rich details. Therefore, we propose a hypothesis about the behavioral indicators of detail effect during first sentence reading: in the SPV task following the first sentence of a paragraph, the participants should respond faster (i.e., shorter RTs) and make fewer errors (i.e., higher accuracy) to the pictures with rich details, compared with the pictures with limited details (*Hypothesis 4-1*). Moreover, consider that readers' eye fixation patterns on pictures are determined by the textual semantics they obtained during sentence reading in a SPV task (Carpenter & Just, 1972; Takacs & Bus, 2018). Therefore, if the participants mentally represent detail semantics during sentence reading, they should tend to fixate on the background area containing scene details in order to verify the detail semantics they obtained from sentence reading. A picture complexity issue should be noticed here since pictures should be more complex when it contains rich details, and this may confound the detail effect on eye fixation patterns. Fortunately, previous studies have found that in text-picture semantic verification tasks, participants were able to make a selective search of the picture to identify the referents and their relationships (Underwood et al., 2004), and they tend to focus their attention more on the elements they mentally represented or mentioned in the text than those not mentioned in the text (Glaser & Schwan, 2015). Considering these findings, although rich details increase the complexity of pictures, the participants mainly rely on existing semantic cues extracted from previous sentences to selectively process pictures. Therefore, picture complexity should not attract the participants' attention in our SPV task and therefore do not interfere with the effect of picture details. Based on these references, we propose the hypothesis about the eye-movement indicators of the detail effect during picture viewing following the first sentence: the participants fixated longer or more often on the detail area (i.e., background) of the pictures with rich details, compared to the pictures with limited details (*Hypothesis 4-2*). Once the perspective shift information is presented in the second sentence of a paragraph, this new gist information should occupy more attentional resources, and thus detailed information in the second sentence may not be considered in semantic processing. Therefore, all the behavioral and eye-movement effects speculated in Hypotheses 4-1 and 4-2 should disappear. In this light, we propose a general final hypothesis about the eye-movement indicators of the detail effect during the picture viewing following the second sentence of a

paragraph: there are no differences in behavioral response (RTs and response accuracy) nor eye fixation measures on the pictures with rich and limited details in the SPV task following the second sentence of a paragraph (*Hypothesis 5*). These three hypotheses regarding the detail issue will be tested by comparing the participants' behavioral and fixation-based measures on pictures with rich and limited details, respectively.

4. Methods

4.1. Participants

A priori power analysis was performed using G*Power 3.1.9 (Faul et al., 2007) and indicated that 40 participants were needed to provide an adequate power $(1 - \beta = 0.95)$ to detect an intermediate effect size (Cohen's $f = 0.30$) at a 0.05 significance level (α). As a result, a total of 52 undergraduate students (10 men) aged 19–26 years ($M = 21.4$ years, $SD = 1.80$ years) were randomly recruited from a local university in Wuhan. More participants than the theoretical sample size were recruited because there may be potentially invalid participants and we need to ensure that the valid participants reach the theoretical sample size after the invalid participants are removed. All the participants were native Chinese speakers with normal or corrected-to-normal vision. None of them majored in Chinese language or literature. All the participants voluntarily participated in this experiment, signed informed consent before the experiment, and received cash after the experiment as compensation for the time spent in this study. The experiment was reviewed and approved in advance by the research ethics committee of the university with which the authors are affiliated.

4.2. Design

To detect the dynamic features of the influence of perspective shift and scene detail on narrative semantic processing, we used relatively long reading materials-paragraphs consist of three consecutive narrative sentences. The SPV tasks were arranged after the end of the first and the second sentence in each paragraph. Three independent variables were manipulated in these text materials and tasks, they are inter-role perspective shift (shifted vs. consistent), internal–external perspective shift (shifted vs. consistent) and picture detail (rich vs. limited). The Internal–external perspective shift is considered because it is not clear whether internal–external perspective shift is also cognitively demanding under Chinese semantic processing. The two types of perspective shift were manipulated as within-subject factors, with inter-role perspective shift manipulated in the second sentence and internal–external perspective shift manipulated in the third sentence in each paragraph. Picture detail was manipulated as a between-subject factor in the SPV tasks following the first and the second sentences in each paragraph. In other words, each participant received one of the two picture detail conditions (rich or limited detail) and all the four perspective shift conditions. We did not intend to compare the eye-tracking effect of the two types of perspective shift since they were manipulated in different ways, with inter-role perspective shift manipulated by action verbs and internal–external perspective shift manipulated by direct speech (Black et al., 1979; Millis, 1995). Considering this and the natural characteristics of the paragraphs we used as reading materials, we will

arrange for the two types of perspective shift to occur sequentially in the same trials, and correspondingly the two types of consistent perspective to occur in the other same trials. That is, for a given trial, it either includes both types of perspective shift or always maintains consistent perspectives. The specific settings of the two types of perspective shift will be introduced in detail in Section 4.4. The dependent variables were the participants' RTs and response accuracy in the SPV tasks and the eye movement patterns in the sentences and pictures.

4.3. Overview of eye-tracking measures

Based on previous studies (Barach et al., 2021; Massaro et al., 2012; Wang et al., 2015; Yan et al., 2013), the three eye-movement indicators selected for the area of interest (AOI) in the text reading were: (1) the *gaze duration* (i.e., the sum of all consecutive first-pass fixation durations on an AOI prior to moving onto a different AOI), which would reflect the early stage of text processing; (2) the *total fixation duration* (i.e., the sum of all the fixation durations on an AOI, including regressions back to the AOI), which is sensitive to slow and long-time cognitive processing and would reflect the late stage of text processing and (3) the *total fixation count* (i.e., the sum of the number of all the fixations on an AOI, including regression back to the AOI), which would effectively reflect the cognitive load in text processing and would be used for the analyses of certain sentences in the paragraphs. Additionally, the four eye-movement indicators selected for AOIs in picture viewing were: (1) the *gaze duration*; (2) the *total fixation duration*; (3) the *total fixation count* and (4) the *first fixation time* (i.e., the time from the beginning of image onset to the location of the first fixation in each AOI for the image), which would reflect how quickly an AOI is fixated (Huang et al., 2019; Neta et al., 2017) and is used as a measure of attentional priority (Thompson et al., 2019) of the participants. The shorter the first fixation time of an AOI, the stronger attentional priority of the AOI.

4.4. Materials

4.4.1. Paragraphs

The text materials we used in the current study are short paragraphs composed of three consecutive narrative sentences. We used longer text materials than those in previous studies because we intend to compare the changes of semantic processing and situation model construction before and after the perspective shift occurred in the second sentence of a paragraph. We prepared 45 sets of paragraphs and each set of paragraphs consists of two paragraphs, with one included shifted perspective information and the other kept original perspective consistent. Each paragraph consists of three sentences which described daily interactions between a man and a woman. We used a verb phrase pair '*go into*' versus '*come into*' to manipulate **inter-role** narrative perspective in the second sentence of each paragraph, with '*go into*' induces a shifted perspective and '*come into*' induces a consistent perspective (Black et al., 1979). Additionally, we also manipulated **internal–external** perspective shifts in the third sentence of each paragraph by using a direct or indirect speech, with direct speech induces a shifted perspective and indirect speech induces a consistent perspective (Millis, 1995). In summary, the shifted perspective paragraphs include both an inter-role perspective shift and an internal–external perspective shift, and the consistent

perspective paragraphs keep the inter-role perspective and internal–external perspective consistent throughout the paragraph. All the paragraphs under the two perspective conditions were controlled in similar lengths. The following is a set of sample paragraphs. In each example, a capital letter was used to mark the end of each sentence in the paragraph, and the asterisks were used to mark the positions where the pictures would occur. The manipulation of narrative perspective was underlined in the example.

Shifted perspective paragraph:

Xiao Fang was cooking in the kitchen (A)* when her husband Li Qiang <u>went into</u> the kitchen with a bunch of bananas (B)* and said, '<u>I want to make some banana milkshakes</u>'. (C)
Consistent perspective paragraph:

Xiao Fang was cooking in the kitchen (A)* when her husband Li Qiang <u>came into</u> the kitchen with a bunch of bananas (B)* and said <u>he wanted to make some banana milkshakes</u>. (C)

4.4.2. *Pictures*

Six black and white sketch target pictures were prepared for each set of paragraphs. Two of the pictures (i.e., Picture A, see Figure 1), either rich or limited in details, matched the semantic content of the first sentence (i.e., Sentence A) in each paragraph. The other four pictures (i.e., Picture B, see Figure 2) matched the semantic content of the second sentence (i.e., Sentence B) in each paragraph, corresponding to the four combinations of rich/limited details of shifted/consistent perspective conditions. The target picture stimuli always demonstrated what the sentences described, with rich/limited details. The correct responses to the target pictures should always be positive. If the participants correctly obtained the semantics of a sentence, they should make a positive response to the target picture following this sentence. The distribution of the six target pictures in each paragraph under different experimental conditions is shown in Table 1.

Figure 1 demonstrates how the rich/limited detailed pictures corresponding to sentence A ('*Xiao Fang was cooking in the kitchen*') were prepared and manipulated. Pictures with limited details depict only the person and the space mentioned by the sentence (e.g., Xiao Fang and the kitchen, see Figure 1b). However, pictures with rich details depict not only the person and the space mentioned in the sentence, but also
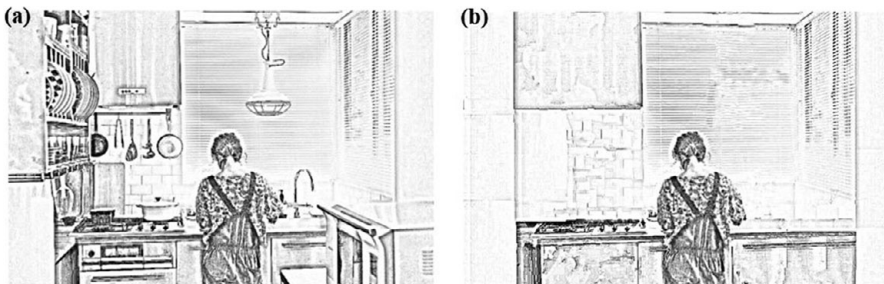

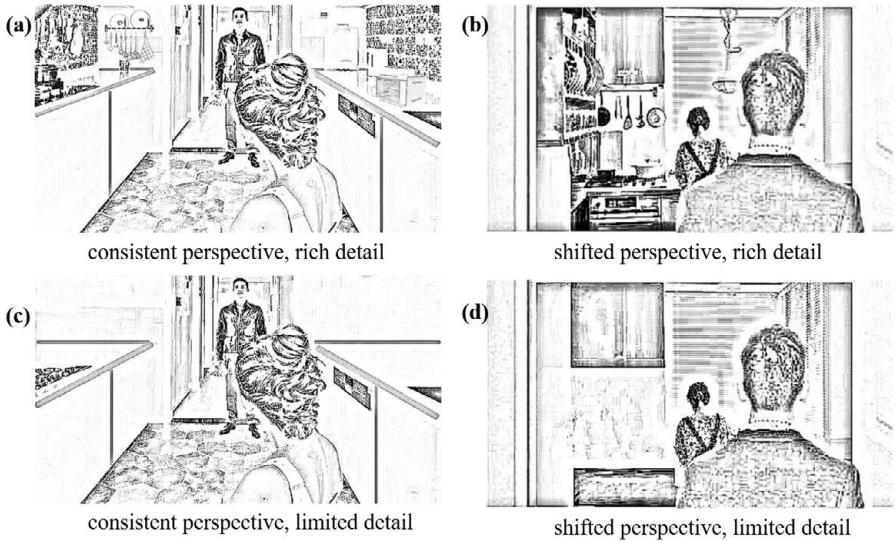**Figure 1.** Picture A with rich (a) and limited (b) details.

Figure 2. Picture B with rich or limited details demonstrated the contents of Sentence B under the consistent (a and c) and the shifted (b and d) perspective conditions.

Table 1. The six target pictures in each paragraph under different experimental conditions

Paragraph components	Narrative perspective	Picture	Picture details
Sentence A	/	Picture A (a)	Rich
Sentence B	Consistent	Picture B (a)	
	Shifted (inter-role)	Picture B (b)	
Sentence C	Consistent	/	
	Shifted (internal–external)		
Sentence A	/	Picture A (b)	Limited
Sentence B	Consistent	Picture B (c)	
	Shifted (inter-role)	Picture B (d)	
Sentence C	Consistent	/	
	Shifted (internal–external)		

Note: '/' means that there is no variable manipulation or picture materials under the corresponding condition. 'a', 'b', 'c' and 'd' in parentheses refer to the picture example in Figure 1 or Figure 2.

the spatial details (e.g., lights, dishes and faucets in the kitchen, see Figure 1a) not explicitly mentioned in the sentence.

Figure 2 shows how a target picture demonstrated what a sentence B described from a consistent perspective (a and c, in rich and limited details respectively) and from a shifted perspective (b and d, in rich and limited details respectively), with the consistent perspective sentence as '*when her husband Li Qiang* **came into** *the kitchen with a bunch of bananas*' and the shifted perspective sentence as '*when her husband Li Qiang* **went into** *the kitchen with a bunch of bananas*'.

In order to counterbalance the participants' positive responses to the target pictures, an additional 20 pictures that did not match the semantic context of the sentences were prepared as fillers in the experiment, with 10 fillers for Picture A and 10 others for Picture B. The correct responses to the filler pictures should always be negative. All the fillers were prepared in the same way as the target pictures in aspects
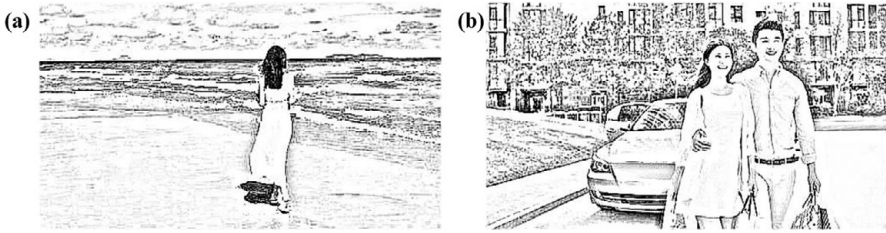
Figure 3. Fillers of Picture A (a) and Picture B (b).

of detail richness. However, the most distinct feature between the fillers and the targets is that the fillers do not depict the semantic context of the sentence. Specifically, fillers of Picture A (Figure 3a) do not depict the action and the location of the person described in Sentence A; fillers of Picture B (Figure 3b) do not depict the location and the inter-role perspective information between the two persons described in Sentence B.

All the pictures were 2400 × 1350 px-sized with a horizontal and vertical resolution of 300 dpi. The elements in the pictures were collected, edited and synthesized from the Internet by one of the authors using an open-source image editor GIMP2.10.14 (https://www.gimp.org/).

4.4.3. The assessment of the paragraphs and the pictures

Thirty-one undergraduates were recruited online to assess the paragraphs and the pictures. Previous studies suggest that comprehensibility, emotional valence and semantic plausibility of the text should influence reading comprehension (Ballenghein et al., 2019; Zhang et al., 2022). Therefore, these factors were evaluated and matched for the paragraphs to avoid them potentially confounding the results: (1) the comprehensibility (1 = *very hard* to 5 = *very easy*); (2) the semantic appropriateness (1 = *inappropriate* to 5 = *appropriate*) and (3) the emotional valence (1 = *negative* to 5 = *positive*). For the assessment of the pictures, the participants were asked to evaluate how suitable the pictures demonstrate the contents of the sentences (1 = *not suitable* to 5 = *suitable*) and to what extent the picture is rich of details that are not explicitly mentioned in the sentences (1 = *limited* to 5 = *rich*). In the material assessment, each participant received both shifted and consistent perspective sentences, and half of the participants (16) received rich detail pictures and the other half (15) received limited detail pictures. As a result, 40 sets of the paragraphs were found to be qualified for the experiment, and there were no significant differences between the two perspective versions of the paragraphs in the above-mentioned dimensions or in the length ($ps \geq 0.187$). All the paragraphs have an approach-to-neutral emotion. The results of the picture details assessment indicated that for all the three pairs of the pictures in each paragraph, the participants rated the pictures with rich details as richer in details than their counterparts with limited details ($ps \leq 0.015$). Those few pictures with problematic feedback from the participants were slightly modified to meet the requirement. To summarize, the materials used for the experiment included 40 sets of paragraphs, 240 pictures matched to the paragraphs and 20 filler pictures.

4.5. Apparatus and stimuli formats

An EyeLink 1000 eye-tracker (SR Research, Canada) in a desktop mount configuration with a sampling rate of 1000 Hz was used to record the participants' monocular (right) eye movements. The stimuli were presented on a 19 inches Lenovo LCD computer screen with a 1280 × 1080 resolution and a 60 Hz refresh rate. The participants placed their head on a chin and forehead rest at approximately 72 cm in distance from the computer screen. The texts were displayed in 25-point black bold KaiTi font at the center of the screen on a white background, with each character forming a viewing angle of about 0.70°. The pictures were presented at the center of the white background of the screen with a scaled size of 720 × 405 px.

4.6. Procedure

The experiment was conducted in a quiet laboratory with a low-light condition. After the participants were seated, the chin and forehead rest were slightly adjusted to a suitable height for the participants to ensure a comfort and optimal reading posture. The participants were advised to read the texts on the screen in a natural way. After a short practice in a separate program, the participants began the formal experiment. A 9-point calibration and validation procedure was performed before the experiment to ensure the accuracy of the tracking. In each trial, a 60 px red fixation cross was first presented at the center of the screen for 500 ms. After the fixation cross disappeared, Sentence A, Picture A, Sentence B, Picture B and Sentence C were presented one by one. The presentation time for Sentences A, B and C was 4, 4.2 and 4.5, respectively. Pictures A and B were both presented for 5 s. The presentation time of the sentences was determined in advance based on a pre-test so that the participants could complete the sentence reading within the time limit. When a picture was presented, the participants were required to judge whether the picture matched the semantic context of the sentence they just read as quickly and accurately as possible, by pressing 'F' or 'J' on the keyboard. The picture disappeared immediately after the participants pressed the button. After all the sentences and the pictures of a paragraph had been presented, a single-item, close-ended comprehension check of the paragraph content was presented (e.g., 'Li Qiang wanted to make some apple salad') and the participants were asked to judge whether the statement was correct or not by clicking a 'Yes' or 'No' button below the statement on the screen. Throughout a trial, the interval before a sentence was presented was 300 ms, and there was no interval between a sentence and its following picture. The inter-trial interval was 1000 ms. The buttons for the two judgments were counter-balanced across the participants. After the experiment, the participants' demographic information (such as gender and age) was collected. The schematic diagram for the procedure of a trial is shown in Figure 4.
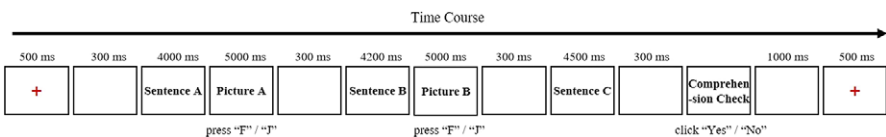

Figure 4. The procedure of a single trial.

5. Results

We used the '*lme4*' package (Bates et al., 2015) within the R studio environment (R version 4.2.0) to analyze the data with linear mixed-effects models (LMMs). The response accuracy data were analyzed with the logistic generalized linear mixed models (GLMMs). For each model, perspective shift or picture detail (or both) was entered as a fixed effect, and subjects and items were treated as random effects. If an initial full model failed to converge then the random structure was systematically trimmed until the model converges. Results from the best-fitting model justified by the data were reported. We report regression coefficients (*b*), standard errors (*SE*), and *t* values (for liner mixed models) or *z* values (for logistic mixed models) for the best-fitting model. *P* values for the coefficient significance tests were estimated using the '*lmerTest*' package (Kuznetsova et al., 2017).

5.1. Descriptions of the performance of reading comprehension and picture verification

On average, the accuracy of the comprehension questions was about 97% (90–100%), suggesting that all the participants read the paragraphs carefully and understood the meaning of the paragraphs correctly. On average, 98% (90–100%) of the responses to Picture A and 97% (87–100%) to Picture B were correct, suggesting that the participants could efficiently judge the semantic match between sentences and pictures.

5.2. Data preprocessing before analyses

Three types of data are included in this study, they are RTs and response accuracy in the SPV task, and the eye movement measures of sentence reading and picture viewing. Before analyzing any type of data, the trials that contain incorrect responses to the comprehension questions were excluded. For RTs analyses, we further excluded the incorrect responses to picture verification. In addition, the outliers of the RTs smaller than 1.5 times the interquartile of the first quartile (Q1–1.5IQR) or larger than 1.5 times the interquartile of the third quartile (Q3 + 1.5IQR) of boxplot distribution (Schwertman et al., 2004) were replaced by the mean RT. As a result, the outliers of 5.6% of the total responses to Picture A and 5.1% of the total responses to Picture B were replaced by the means. For eye movement data analyses, we first manually aligned the fixations systematically deviated from the AOI due to fixation drifts via the Data Viewer software. The incorrect responses to picture verification were further excluded from the analyses. Additionally, fixations shorter than 80 ms or more extreme than three standard deviations from the population mean of each fixation indicator on the pictures (≤8.0% of trials) or the phrases (≤12.2% of trials) were removed from fixation-based analyses (Barach et al., 2021). Sentences which received less than three fixations were deleted (2.6% of trials) from sentence eye-movement analyses (Yan et al., 2021).

5.3. Effects of perspective shift

We first examined whether the participants respond slower or make more errors to pictures under shifted than consistent perspective condition (Hypothesis 1).
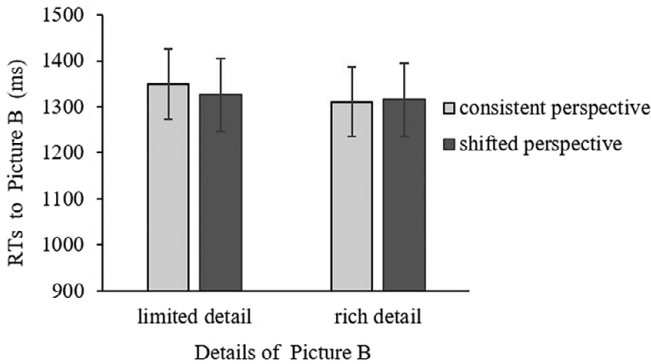
Figure 5. RTs to Picture B under different experimental conditions.
Note: The error bar represents 95% CI of the standard error (*SE*) of the mean.

Response accuracy to Picture B was first analyzed to test this hypothesis. Logistic LMMs with perspective shift, picture detail, and their interaction as fixed effect were built and the results found no main effect for perspective shift ($b = 0.35$, $SE = 0.32$, $z = 1.09$, $p = 0.277$), nor the interaction ($b = -0.67$, $SE = 0.64$, $z = -1.05$, $p = 0.294$). RTs to Picture B were then analyzed with LMMs in a similar manner. The results (Figure 5) showed no main effect for perspective shift ($b = -9.21$, $SE = 42.10$, $t = -0.22$, $p = 0.828$), nor the interaction ($b = 28.42$, $SE = 50.89$, $t = 0.56$, $p = 0.577$). These results did not support Hypothesis 1.

Next, we examined whether the participants fixate longer on the verb phrases in Sentence B that indicate an inter-role perspective shift (i.e., 'went into') than that indicate a consistent perspective (i.e., 'came into') (Hypothesis 2-1). LMMs with fixed factors for perspective shift were built to test this hypothesis. Picture A's detail was considered as a covariate because the participants have already viewed Picture A before they read Sentence B, and thus any effects that might be found after reading the verb phrases may be due to the influences from the previous detail manipulation of Picture A. The results did not show the main effects of perspective shift on gaze duration ($b = 21.50$, $SE = 26.73$, $t = 0.80$, $p = 0.421$) and total fixation duration ($b = 3.21$, $SE = 24.07$, $t = 0.13$, $p = 0.894$). These results did not support Hypothesis 2-1.

Then, we examined whether the participants fixated longer or more often on Sentence C that contains an internal–external perspective shift than a consistent perspective (Hypothesis 2-2). LMMs with fixed factors for perspective shift and covariate for picture detail were built to test this hypothesis. The results (Figure 6) found that the participants fixated more often (i.e., more fixation count) on Sentence C with internal–external perspective shift than that with consistent perspective ($b = 0.48$, $SE = 0.12$, $t = 3.87$, $p < 0.001$). No significant differences of gaze duration ($b = 57.73$, $SE = 42.40$, $t = 1.36$, $p = 0.173$) nor total fixation duration ($b = 10.44$, $SE = 24.64$, $t = 0.42$, $p = 0.672$) on Sentence C was found between the two perspective conditions. These results partially supported Hypothesis 2-2.

Finally, we examined whether the participants more evenly prioritize their first fixation to different elements when viewing Picture B with inter-role perspective shift, compared with that of consistent perspective (Hypothesis 3). Three AOIs were selected in Picture B: person 1, person 2 and the background. Person 1 was the person
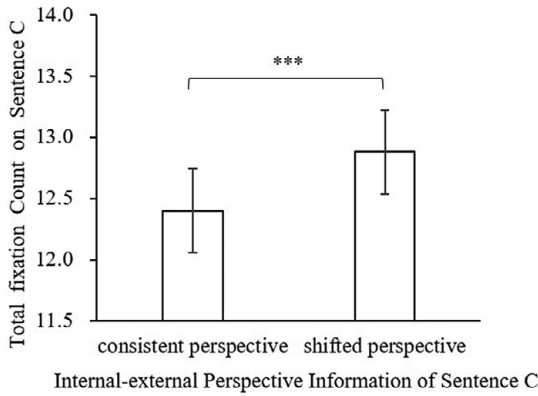
Figure 6. Total fixation count on Sentence C under the two perspective conditions.
Note: The error bar represents 95% CI of the standard error (*SE*) of the mean. *** *p* < .001

who looked closer and larger in the picture and shared the same perspective with the participants (such as Xiao Fang in Figure 2a and Li Qiang in Figure 2b). Person 2 was the other person who looked farther away and smaller and interacted with person 1 (such as Li Qiang in Figure 2a and Xiao Fang in Figure 2b). LMMs with fixed effects for perspective shift, picture detail and AOIs were built to test this hypothesis. The results (Figure 7) showed that regardless of picture details, the participants evenly prioritized their first fixation to person1, person2 and the background in Picture B under shifted perspective condition ($\Delta Ms = 6.49 \sim 26.21$, $ps = 1.000$). However, under consistent perspective condition, the participants preferentially allocated their first fixation to person 1, and then to person 2 and the background ($\Delta M_{person1-person2} = -239.76$, $SE = 21.20$, $t = -11.31$, $p < 0.001$; $\Delta M_{person1-background} = -298.398$, $SE = 26.13$, $t = -11.42$, $p < 0.001$). No difference of first fixation prioritization was found between person 2 and the background ($\Delta M_{person2-background} = -58.63$, $SE = 25.83$, $t = -2.27$, $p = 0.349$).

In addition, the results (Figure 8) showed that (regardless of picture details) the participants fixated longer on person 1 under shifted than consistent perspective
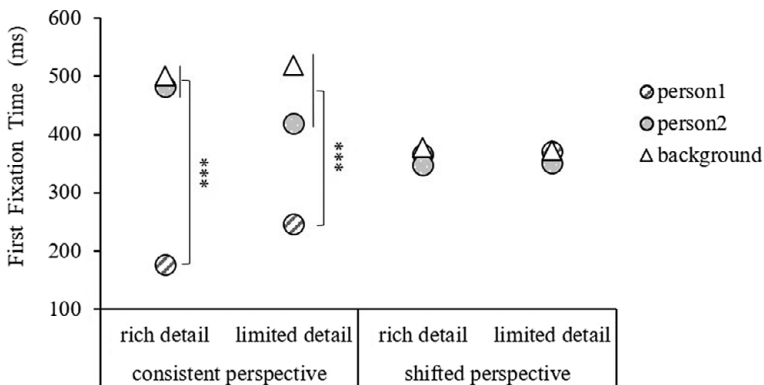

Figure 7. First fixation time on AOIs of Picture B. *** *p* < .001
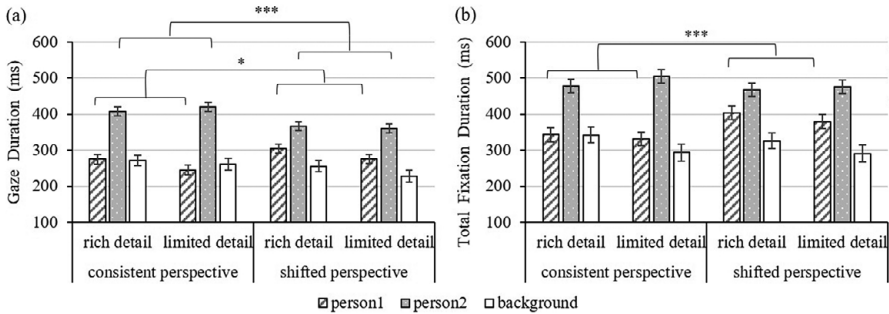
Figure 8. Gaze duration (a) and total fixation duration (b) on AOIs of Picture B.
Note: The error bar represents 95% CI of the standard error (*SE*) of the mean. * *p* < .05, *** *p* < .001

condition (gaze duration: $\Delta M_{\text{person1|consistent-shifted}} = 30.17$, $SE = 9.24$, $t = 3.26$, $p = 0.017$; total fixation duration: $\Delta M_{\text{person1|consistent-shifted}} = 54.78$, $SE = 12.11$, $t = 4.52$, $p < 0.001$); on the contrary, the participants gazed shorter on person 2 ($\Delta M_{\text{person2|consistent-shifted}} = -50.65$, $SE = 8.99$, $t = -5.64$, $p < 0.001$) under shifted than consistent perspective condition. Moreover, the participants fixated longer on person 1 than background under shifted perspective condition (gaze duration: $\Delta M_{\text{person1-background|shifted}} = 38.45$, $SE = 14.11$, $t = 2.72$, $p = 0.007$; total fixation duration: $\Delta M_{\text{person1-background|shifted}} = 83.20$, $SE = 18.11$, $t = 4.59$, $p < 0.001$), while there are no significant difference of fixation duration between person 1 and background under consistent perspective condition (gaze duration: $\Delta M_{\text{person1-background|consistent}} = -2.92$, $SE = 13.81$, $t = -0.21$, $p = 0.832$; total fixation duration: $\Delta M_{\text{person1-background|consistent}} = 20.70$, $SE = 17.86$, $t = 1.16$, $p = 0.246$). Both gaze duration and total fixation duration on the background of Picture B did not differ between the two perspective conditions ($ps \geq 0.891$). These results were not speculated in the hypotheses but were explainable with the perspective shift manipulation and will be discussed later.

5.4. Effects of picture detail

We first examined whether the participants response faster and made fewer errors to the pictures with rich details than that with limited details in the SPV task following Sentence A (Hypothesis 4-1). Response accuracy to Picture A was analyzed with logistic LMMs and no main effect of picture detail was found ($b = 0.41$, $SE = 1.83$, $z = 0.22$, $p = 0.825$). RTs to Picture A was analyzed with LMMs and the result showed that the participants responded 107.8 ms faster to Picture A with rich than limited details, but this did not reach a significant level ($b = -107.8$, $SE = 58.73$, $t = -1.84$, $p = 0.072$).

Next, we examined whether the participants fixated longer or more often on the detail area (i.e., background) of Picture A with rich details than that with limited details (Hypothesis 4-2). The person and the background in Picture A were selected as the two AOIs. LMMs with fixed effects for picture detail and AOIs were built to test this hypothesis. The results found no significant differences of fixation duration or fixation count measures on the background of Picture A with rich versus limited details (gaze duration: $\Delta M_{\text{background|rich-limited}} = 3.18$, $SE = 17.31$, $t = 0.18$, $p = 1.000$; total fixation

duration: $\Delta M_{\text{background|rich-limited}} = -12.75$, $SE = 28.19$, $t = -0.45$, $p = 1.000$; total fixation count: $\Delta M_{\text{background|rich-limited}} = 0.03$, $SE = 0.13$, $t = 0.26$, $p = 1.000$).

Finally, we examined whether there are no any differences of behavioral response nor fixation measures on the Picture B with rich and limited details in the SPV task following Sentence B (Hypothesis 5). Response accuracy to Picture B was already analyzed in the perspective shift effect analyses with logistic LMMs and no main effect of picture detail was found ($b = 0.23$, $SE = 0.34$, $z = 0.69$, $p = 0.493$). RTs to Picture B was also analyzed in the perspective shift effect analyses with LMMs (see Figure 5) and no main effect of picture detail was found ($b = -24.63$, $SE = 95.48$, $t = -0.26$, $p = 0.798$). Additionally, the eye-movement data to Picture B was also analyzed in the analyses of perspective shift effect (see Figure 8) and the results did not show any main effects of picture detail on gaze duration ($b = 14.62$, $SE = 13.54$, $t = 1.08$, $p = 0.286$), total fixation duration ($b = 14.15$, $SE = 20.78$, $t = 0.68$, $p = 0.499$), and total fixation count ($b = 0.03$, $SE = 0.08$, $t = 0.31$, $p = 0.755$). These results supported Hypothesis 5.

5.5. Other effects in the eye-movement analyses

Some other significant results regarding the semantic processing of Sentence A, Picture A and Picture B beyond the hypotheses of perspective shift and picture detail were found. Specifically, we found that the participants fixated longer or more often on the person than on the background in the pictures ($ps \leq 0.025$); the participants gazed longer on person-related words (i.e., proper name and action verb) than location words ($ps \leq 0.018$), while the total fixation duration on location words was longer to person-related words ($ps < 0.001$). Additionally, we found that the participants fixated longer on person 2 than person 1 ($ps < 0.001$) and background ($ps < 0.001$) when processing Picture B, regardless of perspective shift or picture detail. Please see https://osf.io/rfn95/?view_only=0fecd031c4ef48cc9a411e1f91f584c9 for the detailed results and their discussion.

6. Discussion

The present study explored whether and how perspective shift and scene detail may influence readers' narrative semantic processing during natural reading. The eye-movement results found that compared with consistent perspective condition, the participants under inter-role shifted perspective condition allocated their first fixations more evenly to different elements in Picture B (Hypothesis 3, see Figure 7). Additionally, the internal–external perspective shift in Sentence C increased the participants' fixation count on the sentence (Hypothesis 2-2, see Figure 6). However, the behavioral results did not show any effects of perspective shift. Contrary to expectations, neither the behavioral nor eye-movement results showed any significant effects of scene detail depicted in the picture. Below we will make a further explanation and reflection on these results.

6.1. Perspective shift and situation model's coherence during semantic processing

How do the shifts in narrative perspective influence text semantic processing as far as cognitive function is concerned? From the eye-movement results on Picture B, it

seemed that the participants made this happen by first fixation priority and attention allocation mechanism on the various elements in the new scene. Specifically, under the shifted perspective condition, the participants tended to evenly prioritize their first fixation to the two persons and the background depicted in Picture B when they first saw it; while under the consistent perspective condition, the participants prioritized their first fixation to person 1 over person 2 and the background, showing a first fixation priority effect (Hypothesis 3, see Figure 7). This is because person 1 looks closer and appears more saliant than person 2 and background in Picture B, and there are no significant spatial changes in the constructed situation model under consistent perspective condition. In this situation, the participants' first fixation priority features on Picture B may follow a 'physical salience' rule, which makes they fixated earlier on saliant person 1 and then on less saliant person 2 and the background. On the contrary, although person 1 in Picture B is still physically saliant under the shifted perspective condition, there are significant spatial changes brought by inter-role perspective shift. This breaks the coherence of the participants' situation model construction process. In this situation, the participants need to follow a 'situation model's refreshing' rule, in which they should evenly allocated their first fixation to the different elements in the new scene depicted by Picture B, so as to effectively recognize the relations among the elements and finally match the semantics between Picture B and Sentence B.

Comparing the attentional resources distributed to the two different persons in Picture B, the participants paid more attention to person 1 (see Figure 8a,b) and less attention to person 2 (see Figure 8a) under the shifted perspective condition than they did under the consistent perspective condition. This is because a shifted perspective presents the participants with an unfamiliar person 1 (e.g., Li Qiang, see Figure 2b) in Picture B and thus they need more attention on the processing of person 1; while the image of person 2 (e.g., Xiao Fang, see Figure 2b) had already been processed before and thus the participants allocate less attention on person 2. Conversely, under consistent perspective condition, the mental image of person 1 (e.g., Xiao Fang, see Figure 2a) had already been simulated in previous processing and person 2 (e.g., Li Qiang, see Figure 2a) needed more attention as a new piece of information. Moreover, we also found that the participants paid more attention to person 1 than background under shifted perspective condition; while no attention difference between person 1 and background under consistent perspective condition (see Figure 8a,b). This is aligned with the above-mentioned fact that person 1 under shifted perspective condition is new and therefore it receives more attention than background from the participants; whereas person 1 under consistent perspective condition is old (i.e., already introduced in Sentence A and Picture A) and therefore it received the same amount of attention as the background. These results show that when readers carry out semantic processing of a new narrative situation (such as Figure B), they will strictly rely on the previously established situation model and try their best to reconcile the relationship between the new situation and the old one, so as to construct a coherent situation model.

To put in another way, shifts of perspective in sentences may disrupt the coherence of situation model based on previous information, which caused a change of previous cognition input pattern based on the narrative context. As a result, the participants had to temporarily give up previous plan in semantic processing, and established a new focus plan in this process. This may be a strategy to facilitate the

merge of new information into the previously established situation models and to eventually form a coherent storyline (Zwaan & Radvansky, 1998).

6.2. Perspective shift and cognitive cost during semantic processing

Another cognitive function of perspective shift is its cognitive demanding feature. The same as previous studies (Black et al., 1979; Jin & Liu, 2023; Millis, 1995), we found the participants showed a larger cognitive cost on reading Sentence C with internal–external perspective shift than that with consistent perspective (see Figure 6, Hypothesis 2-2). However, we must be cautious in interpreting this result, as we cannot directly test whether the more fixation count on Sentence C is caused by the internal–external perspective shift perse in Sentence C or by a prolonged effect of the inter-role perspective shift that already happened in sentence B. A recent study by Jin and Liu (2023) could provide an interpretation for our result. They found that the effect of inter-role perspective shift in Chinese narrative text disappears rather quickly and does not extend to the end of the next sentence following the perspective shift. In this light, we are more confident that the more fixation count is most probably caused by the inter-external perspective shift in Sentence C. Therefore, a cognitive demanding feature of internal–external perspective shift was found out in Chinese narrative reading. However, the cognitive demanding feature of inter-role perspective shift was not supported by the fixation measures on the phrase ('came into' vs. 'went into') that manipulate inter-role perspective in Sentence B (Hypothesis 2-1). This may be because there is only one character difference between '*came into*' ('走进来') and '*went into*' ('走进去') in Chinese, and eye-tracking measures may not be sensitive enough to effectively detect such subtle difference in rapid reading.

How does cognitive cost relate to semantic processing and reading performance? The present study cannot answer this question because we did not systematically manipulate the cognitive cost of text reading, nor did we measure the participants' reading performance (e.g., reading scores, reading engagement). Previous studies have conducted preliminary exploration on this issue but did not reach a consensus. Some studies found that a more disfluent text leads to a poorer attention and reading engagement, which is detrimental to text comprehension and learning (Feng et al., 2013; Walter et al., 2021). However, other studies have found that higher cognitive cost is usually associated with higher attentional focus and deeper reading engagement (Faber et al., 2017; Nielsen & Escalas, 2010). For example, Nielsen and Escalas (2010) explored the impact of processing difficulties (manipulated by changing text formatting features such as color and font size) in narrative advertisement on reading engagement and the participants' preference for the brand described in the advertisement. They found that the participants felt more engaged when reading advertisement with processing difficulties compared with that was easy to process, which in turn led to an increased liking for the brands depicted in the advertisement. Another strand of studies supported this view by revealing that readers preferred fictional novels with more complex linguistic features such as higher lexical richness/complexity, longer words/sentences or lower readability (Ashok et al., 2013; Jin & Liu, 2022; Lin & Hsieh, 2019; Maharjan et al., 2017). Consider these positive effects of cognitive cost, perspective shift should be conductive to reader' semantic processing under certain circumstances. For example, in a given length of narratives, there

should be an optimal number of perspective shift that is most conducive to readers' semantic processing.

6.3. Null effects of scene detail on semantic processing

The realization of the richness of the meaning of a language depends partly on the rich details that could be extracted and mentally imagined on the basis of the literal expression of texts (Hayakawa & Keysar, 2018; Keogh & Pearson, 2017). In this study, we investigated whether the participants could extract and imagine the detail information implied by the narrative sentence through the manipulation of picture details in a SPV task. However, we did not find any significant results regarding picture details. Behavioral and eye-movement results both did not support our picture detail effect hypotheses (Hypothesis 4-1, 4-2 and 5). We speculate that this may be largely caused by the properties of the text materials we used. All the sentences we used describe an individual's behavior occurred in a specific space (e.g., '*Xiao Fang was cooking in the kitchen*', Sentence A), and his/her interaction with another one in the space (e.g., '*when her husband Li Qiang came into the kitchen with a bunch of bananas*', Sentence B). To understand these sentences, readers may only need to construct information about the characters and the abstract space, rather than imagining a vivid, detailed space (such as a kitchen with lights, dishes and faucets). There are some specific situations in which space details are important. For example, when reading a mystery novel, readers should vividly imagine the details of the scene described in the novel. Here, the detailed elements of the scene may be crucial to solving the suspense and helping readers understand the plot development of the novel.

In many of our daily narrative experiences (e.g., literary reading, episodic memory), scene details may be less important than gist information such as character, time and location in the experiences. We usually tend to encode gist information about who, when and where at a higher cognitive level than details in the narrative context. As a result, much of gist information in narrative experiences is firmly remembered a long time later, but the details fade quickly (Sekeres et al., 2016). In short, we can conclude that readers will not extract and imagine the detail information implied by narrative text during natural reading.

Overall, this study discovered some of the fundamental cognitive mechanisms that is involved in the effect of perspective shifts on semantic processing. That is, shifts in perspective that are cued by very simple linguistic features could break the coherence of readers' situation model construction, which make readers arrange their attentional priority, take a new perspective in the imagined scene, and pay different amount of attention to the various elements of the scene during semantic processing. This suggests that low-level linguistic factors in narratives could influence readers' high-level semantic processing. However, scene detail could not be extracted and imagined by readers unless it is important for the coherent situation model construction that leads to a full interpretation of narrative context.

6.4. Limitations

There are certain limitations in our study. First, we introduced two types of perspective shift and manipulated them in our text materials, but we did not

compare their effects. This is limited by the design and the manipulation of the two types of perspective shift in our study. We manipulated inter-role perspective shift through the verb phrase, while we manipulated internal–external perspective shift through the whole sentence. Moreover, an inter-role perspective shift was always combined with an inter-external perspective shift. These flaws in experimental manipulation and design make it impossible to directly compare the eye-movement effects of the two types of perspective shift. Additionally, since it is hard to arrange a SPV task following the final sentence with internal–external perspective shift, we cannot compare the behavioral effects of the two types of the perspective shift. Inter-role perspective shift is linked with spatial mental simulation, while internal–external perspective shift is more likely linked with mental activity simulation, they may be processed in different way. Therefore, although previous studies found a common cognitive demanding feature of them, it is meaningful to compare them within a more comprehensive experiment in future studies, in which they were manipulated simultaneously and compared directly. Second, we did not randomize the order of presentation of the two types of perspective shift, which arises the problem that we cannot well isolate the possible after-effect of inter-role perspective shift in Sentence B on the semantic processing of Sentence C. We referred to Jin and Liu's (2023) finding and believed that the increased fixation count on Sentence C was caused by the internal–external perspective shift in Sentence C, our experiment perse did not validate this possibility. Specifically, an inter-role perspective shift is always followed by an internal–external perspective shift, and no other perspective shift combinations or presentation orders in our text materials. This made us unable to exclude the effect of inter-role perspective shift on the following Sentence C. Third, the present study explored the basic cognitive processes during semantic processing of the sentences with perspective shift, but we did not investigate the after-effects of the influence of perspective shift on semantic processing. Therefore, we cannot answer whether perspective shift benefits semantic processing or not. Future research may examine the circumstances under which perspective shift can promote readers' semantic processing, and the circumstances under which it is detrimental to semantic processing. Finally, we conducted our study in Chinese language and to our knowledge, little studies explored the relationship between linguistic features and semantic processing with the use of SPV task and eye-tracking measures simultaneously. This brings a question of the generalization of our results to alphabetic languages such as English. However, we found a cognitive demanding feature of internal–external perspective shift as those found in English studies (Millis, 1995). As far as we know, although Chinese and alphabetic language (e.g., English) are two sharply different writing systems, previous studies found more similar rather than different eye-movement patterns between them in adult readers (Feng et al., 2009; Rayner et al., 2007; Sun et al., 1985; Sun & Feng, 1999). Moreover, some studies found that Chinese and English reading both follow a serial processing pattern, that is, readers' attention allocated to only one word at any time during reading (i.e., E–Z reader model) (Liu & Reichle, 2018; Rayner et al., 2007). These general results on Chinese and English reading behavior do not directly answer the similarities or differences of semantic processing between Chinese and English. But what they suggest to us is that there is certain possibility that our results can be generalized to English context. This possibility will be tested in future cross-culture studies.

7. Conclusions

To conclude, we found that: (1) inter-role perspective shift disrupts the coherence of the constructed situation model; (2) the encoding of internal–external perspective shift is cognitively demanding and (3) readers did not extract and imagine the details of the scene during narrative semantic processing.

References

Adams, C., Smith, M. C., Nyquist, L., & Perlmutter, M. (1997). Adult age-group differences in recall for the literal and interpretive meanings of narrative text. *Journals of Gerontology, Series B: Psychological Sciences and Social Sciences*, 52(4), 187–195. https://doi.org/10.1093/geronb/52B.4.P187

Ashok, V. G., Feng, S., & Choi, Y. (2013). Success with style: Using writing style to predict the success of novels. In *Proceedings of the 2013 conference on empirical methods in natural language processing* (pp. 1753–1764). Association for Computational Linguistics. https://www.researchgate.net/publication/286941537

Bailey, H. R., Kurby, C. A., Sargent, J. Q., & Zacks, J. M. (2017). Attentional focus affects how events are segmented and updated in narrative reading. *Memory & Cognition*, 45, 940–955. https://doi.org/10.3758/s13421-017-0707-2

Ballengehein, U., Megalakaki, O., & Baccino, T. (2019). Cognitive engagement in emotional text reading: Concurrent recordings of eye movements and head motion. *Cognition and Emotion*, 33(7), 1448–1460. https://doi.org/10.1080/02699931.2019.1574718

Barach, E., Feldman, L. B., & Sheridan, H. (2021). Are emojis processed like words? Eye movements reveal the time course of semantic processing for emojified text. *Psychonomic Bulletin & Review*, 28, 978–991. https://doi.org/10.3758/s13423-020-01864-y

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

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

Bestgen, Y., & Vonk, W. (2000). Temporal adverbials as segmentation markers in discourse comprehension. *Journal of Memory and Language*, 42(1), 74–87. https://doi.org/10.1006/jmla.1999.2670

Black, J. B., Turner, T. J., & Bower, G. H. (1979). Point of view in narrative comprehension, memory, and production. *Journal of Verbal Learning and Verbal Behavior*, 18, 187–198. https://doi.org/10.1016/S0022-5371(79)90118-X

Brunyé, T. T., Ditman, T., Giles, G. E., Holmes, A., & Taylor, H. A. (2016). Mentally simulating narrative perspective is not universal or necessary for language comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 42(10), 1592–1605. https://doi.org/10.1037/xlm0000250

Brunyé, T. T., Ditman, T., Mahoney, C. R., & Taylor, H. A. (2011). Better you than I: Perspectives and emotion simulation during narrative comprehension. *Journal of Cognitive Psychology*, 23(5), 659–666. https://doi.org/10.1080/20445911.2011.559160

Butler, A. C., Rice, H. J., Wooldridge, C. L., & Rubin, D. C. (2016). Visual imagery in autobiographical memory: The role of repeated retrieval in shifting perspective. *Consciousness and Cognition*, 42, 237–253. https://doi.org/10.1016/j.concog.2016.03.018

Carpenter, P. A., & Just, M. A. (1972). Semantic control of eye movements in picture scanning during sentence-picture verification. *Perception & Psychophysics*, 12, 61–64. https://doi.org/10.3758/BF03212843

Chen, Q. R., Deng, Z., & Tan, D. L. (2008). Sentence-picture integration in Chinese: Evidence from the measurement of eye movement. *Acta Psychologica Sinica*, 40(5), 543–551. https://doi.org/10.3724/SP.J.1041.2008.00543

Child, S., Oakhill, J., & Garnham, A. (2018). You're the emotional one: The role of perspective for emotion processing in reading comprehension. *Language, Cognition and Neuroscience*, 33(7), 878–889. https://doi.org/10.1080/23273798.2018.1431397

Cui, Y. (2017). Reader responses to shifts in point of view: An empirical study. *Language and Literature*, 26 (2), 122–136. https://doi.org/10.1177/0963947017704727

Dam, W. O. V., Rueschemeyer, S. A., Lindemann, O., & Bekkering, H. (2010). Context effects in embodied lexical-semantic processing. *Frontiers in Psychology*, 1, 150. https://doi.org/10.3389/fpsyg.2010.00150

Ditman, T., Brunyé, T. T., Mahoney, C. R., & Taylor, H. A. (2010). Simulating an enactment effect: Pronouns guide action simulation during narrative comprehension. *Cognition*, 115(1), 172–178. https://doi.org/10.1016/j.cognition.2009.10.014

Faber, M., Mills, C., Kopp, K., & D'mello, S. (2017). The effect of disfluency on mind wandering during text comprehension. *Psychonomic Bulletin & Review*, 24, 914–919. https://doi.org/10.3758/s13423-016-1153-z

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175–191. https://doi.org/10.3758/BF03193146

Feng, G., Miller, K., Shu, H., & Zhang, H. (2009). Orthography and the development of reading processes: An eye-movement study of Chinese and English. *Child Development*, 80(3), 720–735. https://doi.org/10.1111/j.1467-8624.2009.01293.x

Feng, S., D'Mello, S., & Graesser, A. C. (2013). Mind wandering while reading easy and difficult texts. *Psychonomic Bulletin & Review*, 20, 586–592. https://doi.org/10.3758/s13423-012-0367-y

Flores, S., Bailey, H. R., Eisenberg, M. L., & Zacks, J. M. (2017). Event segmentation improves event memory up to one month later. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 43(8), 1183–1202. https://doi.org/10.1037/xlm0000367

García, A. M., Moguilner, S., Torquati, K., García-Marco, E., Herrera, E., Muñoz, E., Castillo, E. M., Kleineschay, T., Sedeño, L., & Ibáñez, A. (2019). How meaning unfolds in neural time: Embodied reactivations can precede multimodal semantic effects during language processing. *Neuroimage*, 197, 439–449. https://doi.org/10.1016/j.neuroimage.2019.05.002

Gernsbacher, M. A. (1997). Two decades of structure building. *Discourse Processes*, 23(3), 265–304. https://doi.org/10.1080/01638539709544994

Givón, T. (1992). The grammar of referential coherence as mental processing instructions. *Linguistics*, 30(1), 5–55. https://doi.org/10.1515/ling.1992.30.1.5

Glaser, M., & Schwan, S. (2015). Explaining pictures: How verbal cues influence processing of pictorial learning material. *Journal of Educational Psychology*, 107(4), 1006–1018. https://doi.org/10.1037/edu0000044

Graesser, A. C., Bowers, C., Bayen, U. J., & Hu, X. (2001). Who said what? who knows what? Tracking speakers and knowledge in narratives. In W. van Peer & S. Chatman (Eds.), *New perspectives on narrative perspective* (pp. 255–272). SUNY University Press.

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.

Hayakawa, S., & Keysar, B. (2018). Using a foreign language reduces mental imagery. *Cognition*, 173, 8–15. https://doi.org/10.1016/j.cognition.2017.12.010

Huang, P., Cai, B., Zhou, C., Wang, W., Wang, X., Gao, D., & Bao, B. (2019). Contribution of the mandible position to the facial profile perception of a female facial profile: An eye-tracking study. *American Journal of Orthodontics and Dentofacial Orthopedics*, 156(5), 641–652. https://doi.org/10.1016/j.ajodo.2018.11.018

Jin, J., & Liu, S. (2022). An analysis of the linguistic features of popular Chinese online fantasy novels, *Discourse Processes*, 59(4), 326–344. https://doi.org/10.1080/0163853X.2022.2028432

Jin, J., & Liu, S. (2023). Shifts in narrative perspectives consume attentional resources and facilitate reading engagement. *Scientific Studies of Reading*, 27, 393–407. https://doi.org/10.1080/10888438.2023.2179922

Just, M. A., & Carpenter, P. A. (1971). Comprehension of negation with quantification. *Journal of Verbal Learning and Verbal Behavior*, 10(3), 244–253. https://doi.org/10.1016/S0022-5371(71)80051-8

Keogh, R., & Pearson, J. (2017). The perceptual and phenomenal capacity of mental imagery. *Cognition*, 162, 124–132. https://doi.org/10.1016/j.cognition.2017.02.004

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction–integration model. *Psychological Review*, 95(2), 163–182. https://doi.org/10.1037/0033-295X.95.2.163

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

Lin, Y. J., & Hsieh, S. K. (2019). The secret to popular Chinese web novels: A corpus-driven study. In *Proceedings of the 2nd conference on language, data and knowledge (LDK 2019)* (24:1–24:8). Dagstuhl Publishing. https://doi.org/10.4230/OASIcs.LDK.2019.24

Liu, Y., & Reichle, E. D. (Eds.). (2018). Eye-movement evidence for object-based attention in Chinese reading. *Psychological Science*, 29(2), 278–287. https://doi.org/10.1177/0956797617734827

Maharjan, S., Arevalo, J., Montes, M., González, F. A., & Solorio, T. (2017). A multi-task approach to predict likability of books. In *Proceedings of the 15th conference of the European chapter of the association for computational linguistics: Volume 1, long papers* (pp. 1217–1227). Association for Computational Linguistics. https://www.aclweb.org/anthology/E17-1114

Mak, M. (2022). *What's on your mind: Mental simulation and aesthetic appreciation during literary reading* [Doctoral dissertation]. Radboud University Nijmegen. https://hdl.handle.net/2066/284932

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

Mak, M., & Willems, R. M. (2021). Mental simulation during literary reading. In D. Kuiken & A. M. Jacobs (Eds.), *Handbook of empirical literary studies* (pp. 63–84). De Gruyter. https://doi.org/10.1515/9783110645958-004

Mariola, A., Fountas, Z., Barnett, L., & Roseboom, W. (2022). Event segmentation in continuous, naturalistic videos from model-based, data-driven, and human perspectives. *PsyArXiv*. https://doi.org/10.31234/osf.io/t6fkg

Marschark, M., & Cornoldi, C. (1991). Imagery and verbal memory. In C. Cornoldi & M. A. McDaniel (Eds.), *Imagery and cognition* (pp. 41–56). Springer. https://doi.org/10.1007/978-1-4684-6407-8_5

Massaro, D., Savazzi, F., Di Dio, C., Freedberg, D., Gallese, V., Gilli, G., & Marchetti, A. (2012). When art moves the eyes: A behavioral and eye-tracking study. *PLoS One*, 7(5), e37285. https://doi.org/10.1371/journal.pone.0037285

McNamara, D. S., & Magliano, J. (2009). Toward a comprehensive model of comprehension. In B. Ross (Ed.), *The psychology of learning and motivation* (Vol. 51, pp. 297–384). Elsevier Science. https://doi.org/10.1016/S0079-7421(09)51009-2

Miall, D. S., & Kuiken, D. (2001). Shifting perspectives: Readers' feelings and literary response. In W. van Peer & S. Chatman (Eds.), *New perspectives on narrative perspective* (pp. 289–301). SUNY University Press.

Millis, K. (1995). Encoding discourse perspective during the reading of a literary text. *Poetics*, 23(3), 235–253. https://doi.org/10.1016/0304-422X(94)00028-5

Morrow, D. G. (1986). Grammatical morphemes and conceptual structure in discourse processing. *Cognitive Science*, 10(4), 423–455. https://doi.org/10.1207/s15516709cog1004_2

Neta, M., Tong, T. T., Rosen, M. L., Enersen, A., Kim, M. J., & Dodd, M. D. (2017). All in the first glance: first fixation predicts individual differences in valence bias. *Cognition and Emotion*, 31(4), 772–780. https://doi.org/10.1080/02699931.2016.1152231

Newberry, K. M., & Bailey, H. R. (2019). Does semantic knowledge influence event segmentation and recall of text?. *Memory & Cognition*, 47, 1173–1187. https://doi.org/10.3758/s13421-019-00926-4

Nielsen, J. H., & Escalas, J. E. (2010). Easier is not always better: The moderating role of processing type on preference fluency. *Journal of Consumer Psychology*, 20(3), 295–305. https://doi.org/10.1016/j.jcps.2010.06.016

Phillips, C. I., Sears, C. R., & Pexman, P. M. (2012). An embodied semantic processing effect on eye gaze during sentence reading. *Language and Cognition*, 4(2), 99–114. https://doi.org/10.1515/langcog-2012-0006

Radvansky, G. A., & Zacks, J. M. (2017). Event boundaries in memory and cognition. *Current Opinion in Behavioral Sciences*, 17, 133–140. https://doi.org/10.1016/j.cobeha.2017.08.006

Rall, J., & Harris, P. L. (2000). In Cinderella's slippers? Story comprehension from the protagonist's point of view. *Developmental Psychology*, 36(2), 202–208. https://doi.org/10.1037/0012-1649.36.2.202

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

Rayner, K., Li, X., Williams, C. C., Cave, K. R., & Well, A. D. (2007). Eye movements during information processing tasks: Individual differences and cultural effects. *Vision Research*, 47(21), 2714–2726. https://doi.org/10.1016/j.visres.2007.05.007

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. https://doi.org/10.1037/0033-295x.105.1.125

Samur, D., Tops, M., Slapšinskaitė, R., & Koole, S. L. (2021). Getting lost in a story: How narrative engagement emerges from narrative perspective and individual differences in alexithymia. *Cognition and Emotion*, 35(3), 576–588. https://doi.org/10.1080/02699931.2020.1732876

Schmid, S., & Baccino, T. (2002). Perspective shift and text format: An eye-tracking study. *Current Psychology Letters: Behaviour, Brain & Cognition*, 9, 73–87. https://doi.org/10.4000/cpl.388

Schwertman, N. C., Owens, M. A., & Adnan, R. (2004). A simple more general boxplot method for identifying outliers. *Computational Statistics & Data Analysis*, 47(1), 165–174. https://doi.org/10.1016/j.csda.2003.10.012

Sekeres, M. J., Bonasia, K., St-Laurent, M., Pishdadian, S., Winocur, G., Grady, C., & Moscovitch, M. (2016). Recovering and preventing loss of detailed memory: Differential rates of forgetting for detail types in episodic memory. *Learning & Memory*, 23(2), 72–82. https://doi.org/10.1101/lm.039057.115

Shin, Y. S., & DuBrow, S. (2021). Structuring memory through inference-based event segmentation. *Topics in Cognitive Science*, 13(1), 106–127. https://doi.org/10.1111/tops.12505

Sun, F., & Feng, D. (1999). Eye movements in reading Chinese and English text. In J. Wang, A. W. Inhoff, & H.-C. Chen (Eds.), *Reading Chinese script: A cognitive analysis* (pp. 189–206). Erlbaum. https://psycnet.apa.org/record/1999-02259-009

Sun, F., Morita, M., & Stark, L. W. (1985). Comparative patterns of reading eye movement in Chinese and English. *Perception & Psychophysics*, 37, 502–506. https://doi.org/10.3758/BF03204913

Sundermeier, B. A., van der Broek, P., & Zwaan, R. A. (2005). Causal coherence and the availability of locations and objects during narrative comprehension. *Memory & Cognition*, 33(3), 462–470. https://doi.org/10.3758/BF03193063

Takacs, Z. K., & Bus, A. G. (2018). How pictures in picture storybooks support young children's story comprehension: An eye-tracking experiment. *Journal of Experimental Child Psychology*, 174, 1–12. https://doi.org/10.1016/j.jecp.2018.04.013

Thompson, S. J., Foulsham, T., Leekam, S. R., & Jones, C. R. (2019). Attention to the face is characterised by a difficult to inhibit first fixation to the eyes. *Acta Psychologica*, 193, 229–238. https://doi.org/10.1016/j.actpsy.2019.01.006

Underwood, G., Jebbett, L., & Roberts, K. (2004). Inspecting pictures for information to verify a sentence: Eye movements in general encoding and in focused search. *Quarterly Journal of Experimental Psychology Section A*, 57(1), 165–182. https://doi.org/10.1080/02724980343000189

Walter, N., Bilandzic, H., Schwarz, N., & Brooks, J. J. (2021). Metacognitive approach to narrative persuasion: The desirable and undesirable consequences of narrative disfluency. *Media Psychology*, 24(5), 713–739. https://doi.org/10.1080/15213269.2020.1789477

Wang, Q. D., Li, Q. G., Chen, K. K., & Fu, G. Y. (2015). Attentional bias to untrustworthy faces: Evidence from eye tracking data. *Advances in Psychological Science*, 23(9), 1508–1513. https://doi.org/10.3724/SP.J.1042.2015.01508

Yan, G., Lan, Z., Meng, Z., Wang, Y., & Benson, V. (2021). Phonological coding during sentence reading in Chinese deaf readers: An eye-tracking study. *Scientific Studies of Reading*, 25(4), 287–303. https://doi.org/10.1080/10888438.2020.1778000

Yan, G. L., Xiong, J. P., Zang, C. L., Yu, L. L., Cui, L., & Bai, X. J. (2013). Review of eye-movement measures in reading research. *Advances in Psychological Science*, 21(4), 589–605. https://doi.org/10.3724/SP.J.1042.2013.00589

Yaxley, R. H., & Zwaan, R. A. (2007). Simulating visibility during language comprehension. *Cognition*, 105(1), 229–236. https://doi.org/10.1016/j.cognition.2006.09.003

Zacks, J. M., Speer, N. K., & Reynolds, J. R. (2009). Segmentation in reading and film comprehension. *Journal of Experimental Psychology: General*, 138(2), 307–327. https://doi.org/10.1037/a0015305

Zacks, J. M., & Swallow, K. M. (2007). Event segmentation. *Current Directions in Psychological Science*, 16(2), 80–84. https://doi.org/10.1111/j.1467-8721.2007.00480

Zhang, G., Yao, P., Ma, G., Wang, J., Zhou, J., Huang, L., Xu, P., Chen, L., Chen, S., Gu, J., Wei, W., Cheng, X., Hua, H., Liu, P., Lou, Y., Shen, W., Bao, Y., Liu, J., Lin, N., & Li, X. (2022). The database of eye-movement measures on words in Chinese reading. *Scientific Data*, 9(1), 411. https://doi.org/10.1038/s41597-022-01464-6

Ziegler, F., Mitchell, P., & Currie, G. (2005). How does narrative cue children's perspective taking? *Developmental Psychology*, 41(1), 115–123. https://doi.org/10.1037/0012-1649.41.1.115

Zwaan, R. A. (2009). Mental simulation in language comprehension and social cognition. *European Journal of Social Psychology*, 39(7), 1142–1150. https://doi.org/10.1002/ejsp.661

Zwaan, R. A. (2016). Situation models, mental simulations, and abstract concepts in discourse comprehension. *Psychonomic Bulletin & Review*, 23, 1028–1034. https://doi.org/10.3758/s13423-015-0864-x

Zwaan, R. A., Langston, M. C., & Graesser, A. C. (1995). The construction of situation models in narrative comprehension: An event-indexing model. *Psychological Science*, 6(5), 292–297. https://doi.org/10.1111/j.1467-9280.1995.tb00513.x

Zwaan, R. A., & Pecher, D. (2012). Revisiting mental simulation in language comprehension: Six replication attempts. *PLoS One*, 7(12), e51382. https://doi.org/10.1371/journal.pone.0051382

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, 123(2), 162–185. https://doi.org/10.1037/0033-2909.123.2.162

Zwaan, R. A., Stanfield, R. A., & Yaxley, R. H. (2002). Language comprehenders mentally represent the shapes of objects. *Psychological Science*, 13(2), 168–171. https://doi.org/10.1111/1467-9280.0043