Language and Cognition (2026), 18, e71, 1–23
doi:10.1017/langcog.2026.10120

ARTICLE

Task-sensitive semantic-visual alignment of motion schema during incremental Mandarin sentence comprehension

Menghan Wang[1] and Helen Zhao[2]

[1]School of International Studies, University of International Business and Economics, Beijing, China and [2]School of Languages and Linguistics, The University of Melbourne, Melbourne, Australia
Corresponding author: Menghan Wang; Email: menghan.w@uibe.edu.cn

(Received 05 April 2026 UTC; Revised 30 July 2026 UTC; Accepted 18 September 2026 UTC)

Abstract

Language comprehension involves dynamically integrating linguistic input into coherent event representations. A common approach to investigating this process has been to examine behavioural congruency effects between linguistic and visual information. However, it remains unclear under what conditions these effects emerge and whether they reflect embodied processing or alternative cognitive mechanisms. The present study examined semantic-visual alignment of motion schema during incremental processing of Mandarin motion sentences, focusing on whether congruency effects are modulated by task demands and visual probe regions. 149 native Mandarin speakers completed either a self-paced sentence-diagram verification task or a self-paced sentence-diagram viewing task. In both tasks, participants processed segmented Chinese motion sentences, while schematic diagrams were presented either after the directional verb (e.g., *jìn*, 'into') or the landmark (e.g., *bàn-gōng-shì*, 'office'). Results revealed overall congruency effects across tasks and diagram regions. Crucially, effects were stronger in the verification task than in the viewing task but remained stable across diagram regions. These findings suggest that semantic-visual alignment is sensitive to task demands but does not necessarily become stronger as the sentence unfolds. These congruency effects are open to both embodied and amodal semantic interpretations and are therefore not uniquely diagnostic of embodied processing.

Keywords: congruency effect; Mandarin; semantic-visual alignment; sentence processing; task sensitivity

1. Introduction

Embodied approaches to language comprehension propose that conceptual meaning is grounded, at least in part, in perceptual-motor systems (e.g., Fischer & Zwaan, 2008; Gallese & Lakoff, 2005; Glenberg & Kaschak, 2002; Pulvermüller, 2013). From this perspective, understanding language may involve the partial re-enactment of

© The Author(s), 2026. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/langcog.2026.10120 Published online by Cambridge University Press

sensorimotor experiences associated with the described event (Barsalou, 1999, 2008), such that perceptual-motor representations contribute to the construction of meaning in language comprehension. This theoretical framework has been particularly influential in explaining action, spatial and motion language, where meaning is closely related to embodied experience (Ahlberg et al., 2018; Kaschak et al., 2005; Stanfield & Zwaan, 2001; Zwaan et al., 2004). Embodied accounts therefore converge in the argument that perceptual-motor information is activated and contributes to mental simulation in language comprehension (Barsalou, 1999, 2008; Bergen & Chang, 2005; Zwaan, 2004). Although embodied approaches have motivated extensive language comprehension research, questions remain regarding how observed behavioural evidence should be interpreted and the conditions under which they emerge during sentence processing.

One of the most widely used behavioural indices for investigating mental simulation has been the response time (RT). Across a range of studies (e.g., Chen et al., 2020; de Koning et al., 2017; Rommers et al., 2013; Zwaan et al., 2002; Zwaan & Pecher, 2012), responses are typically faster when linguistic descriptions are paired with semantically congruent rather than incongruent visual stimuli, a phenomenon commonly referred to as the *compatibility effect* (Bergen, 2007). Such effects have often been interpreted as being consistent with embodied accounts because they may indicate that comprehenders construct perceptually grounded representations during language comprehension, facilitating behavioural performance when these representations align with subsequently presented visual information.

However, critics have argued that compatibility effects should be treated as suggestive evidence for the embodied cognition hypothesis rather than unequivocally demonstrate that language comprehension relies on perceptual simulation (Mahon & Caramazza, 2008; Meteyard et al., 2012). This is not only because mental simulation can also be revealed through interference effects (e.g., Bergen et al., 2003, 2010; Edmiston & Lupyan, 2017; Horchak & Garrido, 2022; Kaschak et al., 2005; Ostarek & Huettig, 2017; Richardson et al., 2003), but also because this behavioural pattern may sometimes reflect amodal semantic processing, task-induced mental imagery or decision-related comparison processes between sentence and visual stimuli, rather than mental simulation per se (Weiskopf, 2010). Consequently, compatibility effects remain interpretively ambiguous. For terminological clarity, the term *congruency effect* is used instead throughout the study to refer to the behavioural advantage for semantically congruent over incongruent semantic-visual pairings, which reflects semantic-visual alignment.

Moreover, the magnitude of congruency effects varies substantially across different studies. A growing body of research suggests that their magnitude depends on experimental context, particularly task demands (Bechtold et al., 2023; Ostarek & Huettig, 2019; Wang, 2026). Tasks requiring explicit comparison between visual and linguistic information generally produce stronger congruency effects than tasks that do not require such mapping (Rommers et al., 2013). In addition, previous studies have suggested that the point at which visual information is encountered during sentence comprehension may modulate the magnitude of congruency effects, with effect sizes increasing as additional semantic information accumulates (Nishide et al., 2025; Sato et al., 2013). Together, these findings suggest that the magnitude of congruency effects is not fixed across all contexts but is modulated by the processing goals and the availability of semantic information. However, previous studies have typically examined task demands and visual probe timing separately, making it

https://doi.org/10.1017/langcog.2026.10120 Published online by Cambridge University Press

difficult to determine how these two factors may jointly influence the emergence and magnitude of congruency effects during sentence comprehension.

Accordingly, the present study investigates the conditions under which congruency effects become behaviourally observable during incremental sentence processing. Specifically, the study compares two behavioural paradigms with different processing goals (verification vs. viewing) while systematically manipulating the point at which visual probes are presented. By jointly manipulating task demands and informational availability while holding the linguistic materials constant, the study aims to clarify when congruency effects emerge during incremental sentence comprehension and how they are shaped by task demands. The study contributes to the ongoing debate about the semantic-visual alignment by specifying the conditions under which congruency effects emerge.

1.1 Task demands and congruency effects in semantic-visual alignment

Congruency effects have been reported across a wide range of studies examining semantic-visual alignment in language comprehension (e.g., Redmann et al., 2019; Stanfield & Zwaan, 2001; Yang, 2016; Zwaan et al., 2002; Zwaan & Taylor, 2006), but their magnitude varies considerably across behavioural paradigms (Chen et al., 2020; Rommers et al., 2013). One of the influential explanations for the variability in congruency effects concerns task demands. Task demands have been interpreted as reflecting differences in attentional focus and cognitive resource allocation devoted to semantic-visual information alignment during language comprehension (Bechtold et al., 2023; Ostarek & Huettig, 2017, 2019).

Congruency effects have been widely observed in sentence-picture verification tasks (e.g., de Koning et al., 2017; Stanfield & Zwaan, 2001; Zwaan et al., 2002; Zwaan & Pecher, 2012). In this paradigm, participants read or listen to a sentence and judge whether a subsequently presented visual stimulus matches the described event. The task often yields robust congruency effects because it places a strong emphasis on the perceptual-semantic correspondence by demanding explicit judgment on the alignment between the interpreted sentence meaning and the visual stimulus.

In addition, congruency effects have also been investigated in tasks imposing relatively lower demands on explicit semantic-visual alignment, including sensibility judgment tasks (Hu & Zhao, 2026; Yang, 2016; Zwaan & Taylor, 2006), picture-naming tasks (Bub & Masson, 2006; Redmann et al., 2019; Rommers et al., 2013; Zwaan et al., 2002) and delayed picture-recognition tasks (Pecher et al., 2007; Richardson et al., 2003), although these effects are sometimes attenuated or absent in these paradigms (Chen et al., 2020; Rommers et al., 2013). This variability suggests that congruency effects reflect not only the correspondence between linguistic and visual information, but also the extent to which such correspondence becomes behaviourally expressed under different attentional focus towards integrating linguistic and visual information. However, task demands have been frequently invoked as a post-hoc explanation for inconsistent magnitude of congruency effects across studies. Given most studies have adopted only a single task while differing substantially in their materials and experimental paradigms (Pecher et al., 2009; Stanfield & Zwaan, 2001; Yang, 2016; Zwaan et al., 2004), the specific contribution of task demands remains difficult to isolate.

https://doi.org/10.1017/langcog.2026.10120 Published online by Cambridge University Press

A few studies have demonstrated that the behavioural detectability of congruency effects is sensitive to task requirements by manipulating task demands while holding visual stimuli constant. Rommers et al. (2013) reported stronger effects when participants were instructed to create a visual image of the situation described by the sentence in the picture-naming task, or when they performed a sentence-picture verification task, than when they completed the picture-naming task without any explicit emphasis on semantic-visual correspondence. These findings suggest that the behavioural detectability of semantic-visual alignment is enhanced when tasks direct attention towards the correspondence between linguistic and visual representations and support a goal-adaptive view of congruency effects. More broadly, similar task-sensitive patterns have also been reported in other behavioural congruency paradigms. For example, Bub and Masson (2006) observed faster gesture initiation only when participants were required to identify and name the object, whereas no reliable congruency effects were found when participants merely viewed the objects. Moreover, existing viewing paradigms have primarily investigated lexical-level processing, leaving it unclear whether semantic-visual alignment can be behaviourally detected during viewing the visual stimuli in sentence comprehension. In addition, although these studies have compared performance across behavioural paradigms while holding the materials constant (Bub & Masson, 2006; Rommers et al., 2013; Zwaan et al., 2002), these comparisons have predominantly contrasted processing-based tasks (e.g., sentence-picture verification, viewing) with (speech) production-based tasks (e.g., picture-naming). As a result, it is difficult to determine whether observed differences reflect genuine task modulation or broader differences between responding across processing and (speech) production modalities.

To address these gaps, controlled experimental comparisons are needed, in which task demands are systematically manipulated while linguistic and visual materials, as well as response modality (e.g., button pressing, speech production), are held constant across conditions. Such designs enable a more precise evaluation of how online processing task demands (i.e., whether explicit sentence-visual mapping is required) influence the behavioural expression of semantic-visual alignment during sentence processing. Nevertheless, even under comparable task requirements, the amount of semantic information available at the point of visual presentation may also influence the behavioural detectability of semantic-visual alignment. This factor is referred to here as the informational context of semantic-visual alignment.

1.2 Informational context of semantic-visual alignment in incremental sentence processing

Sentence comprehension can involve constructing situation models – dynamic mental representations that integrate information about entities, actions, spatial relations and temporal structure – which are assumed to be incrementally updated as linguistic input unfolds (Zwaan, 2008; Zwaan & Radvansky, 1998). However, much of the existing evidence has relied on probes presented after sentence completion (Bai et al., 2022; Borreggine & Kaschak, 2006; Liu et al., 2024). Such designs make it difficult to determine whether semantic-visual alignment emerges immediately upon encountering early diagnostic cues or instead develops gradually as additional linguistic information accumulates across sentence constituents.

https://doi.org/10.1017/langcog.2026.10120 Published online by Cambridge University Press

Self-paced reading paradigms interleaved with visual probes have been particularly informative in addressing this question because they allow researchers to test the behavioural detectability of semantic-visual alignment at different stages of sentence processing. Sato et al. (2013), for example, employed a self-paced sentence-picture verification paradigm to examine incremental semantic-visual alignment of object shape information in Japanese. By inserting pictures at different sentence positions, they observed reliable congruency effects at multiple stages of sentence comprehension, with stronger effects when pictures were presented later in the sentence (i.e., after the verb) than at earlier positions. They interpreted this pattern as evidence that mental representations are progressively refined as additional linguistic constraints accumulate, supporting an incremental strengthening account in which early event representations remain underspecified and gradually converge on a more detailed interpretation.

More recently, Nishide et al. (2025) extended this insertion-based logic to motion event comprehension by examining how learners incrementally construct locational versus directional interpretations of English prepositional phrases (e.g., *The bird flew in the cage*). Critically, they manipulated the informational context by presenting a target video either immediately after the preposition or after sentence completion, thereby testing whether motion construal becomes available as soon as an ambiguous cue is encountered or becomes more behaviourally detectable after later integration. Their results showed that directional interpretations were more difficult than locational ones, and this difference was larger when verification occurred after sentence completion, suggesting that directional construal became increasingly salient as additional linguistic information accumulated.

Additional evidence for the role of informational context comes from Hu and Zhao (2026), who used a self-paced sensibility judgment task interleaved with schematic diagrams to examine native English speakers' processing of English phrasal verbs. They found that diagram congruency facilitated sensibility judgments only when the diagram appeared after sentence completion rather than immediately after the phrasal verb. They therefore argued that congruency effects in first language (L1) processing are sensitive to the amount of linguistic information available, with later probe positions supporting stronger congruency effects when earlier linguistic input is not yet sufficient to support a stable event representation.

Together, these findings indicate that congruency effects can exhibit temporal build-up, consistent with the view that semantic-visual alignment may strengthen as comprehenders accumulate constraining information and progressively commit to a specific event representation (Bergen & Chang, 2005, 2013). Under such an account, early representations may remain provisional or underspecified and become increasingly refined as additional cues narrow the space of possible interpretations.

Nevertheless, caution is needed in interpreting these findings. First, larger congruency effects observed at sentence-final probe positions may partially reflect sentence wrap-up processes, during which comprehenders consolidate and integrate information across the entire sentence, making the event representation more stable and easier to align with a visual stimulus. Second, the increasing pattern reported in previous studies has typically been observed in contexts where early linguistic cues provide only partially diagnostic information, requiring later constituents to resolve ambiguity or fully specify event structure. It therefore remains

https://doi.org/10.1017/langcog.2026.10120 Published online by Cambridge University Press

unclear whether congruency effects would necessarily exhibit the same temporal build-up when early cues already provide highly diagnostic information about the event schema. In such contexts, congruency effects may stabilise earlier during processing, resulting in comparable effect magnitude across incremental stages rather than progressively increasing effects. Clarifying how the temporal profile of congruency effects varies as a function of informational context is therefore important for understanding how semantic-visual alignment unfolds during incremental sentence comprehension.

1.3 The present study

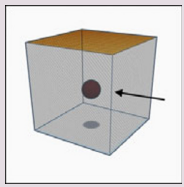
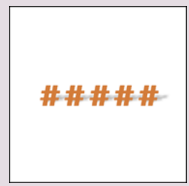
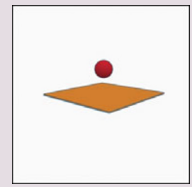
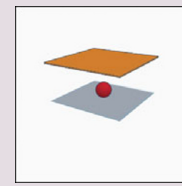
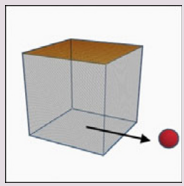
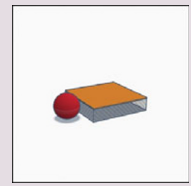
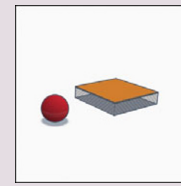
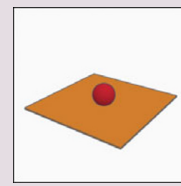
Motivated by the above research gaps, the present study examines how semantic-visual alignment is shaped by task demands and the temporal locus of visual probes during incremental processing of Mandarin motion sentences. Motion events provide a suitable domain because spatial relations become available incrementally as linguistic input unfolds. According to image schema accounts, the evolving representation of a motion event is organised around components such as the trajector (TR), landmark (LM), path, source and goal (Johnson, 1987; Talmy, 2000). For example, the spatial relation conveyed by the preposition *into* can be characterised by a constellation of three schemas including TR-LM, Source-Path-Goal and Container (Bergen & Chang, 2005). In a sentence like '*A worker walked into the office*', the worker functions as the TR (a movable figure), the office as the LM (a stable reference object) and the motion trajectory progresses towards a goal location (i.e., *office*) that also functions as a container – a bounded region with an interior and exterior. Therefore, motion language comprehension requires the incremental integration of spatial information into evolving situation models.

The present study was designed to investigate online semantic-visual alignment of schematic motion representations during incremental sentence comprehension, rather than perceptual inference following sentence completion. Whereas traditional sentence-picture verification tasks often ask participants to verify perceptual properties that are implied (e.g., de Koning et al., 2017; Stanfield & Zwaan, 2001; Zwaan et al., 2002), the present diagrams correspond to motion relations that are explicitly encoded by the linguistic input (e.g., *into*, *out of*) in the sentence stimuli. Unlike the realistic pictures commonly used in traditional sentence-picture verification tasks, the present study employed schematic diagrams as the visual stimuli presented during sentence processing. Whereas pictures typically preserve rich perceptual features of objects and events, the target diagrams used here are graphic representations composed of simple symbols (see Table 1), such as lines, dots, arrows and geometric shapes (Bryant & Tversky, 1999; Tversky et al., 2000), which should not be regarded as pictorial equivalents of the lexical items *into* or *out of*. Rather, they were heuristic representations of underlying image-schema relations, abstracting away from object identity and perceptual detail while preserving only the relational structure of the motion event (Wang, 2026; Wang & Zhao, 2024b).

The present study focuses on Mandarin motion expressions, which provides a theoretically informative test case as motion events are typically expressed through resultative verb compounds (RVCs). An RVC consists of two verbal elements (e.g., *dǎ pò* 'hit-break', *kàn wán* 'read-finish'), with the second element describing the resulting state or path associated with the event denoted by the first element

https://doi.org/10.1017/langcog.2026.10120 Published online by Cambridge University Press

Table 1. Target, distractor and filler diagrams

Target diagram	Distractor diagram	Filler diagram	
			
into		*HIGHER THAN*	*LOWER THAN*
			
out of		*ADJACENCY*	*PROXIMITY*
			
		SUPPORT	

(Li, 2013; Wang et al., 2015). The present study focuses on motion-related RVCs (Li & Thompson, 2009; Wang & Zhao, 2024b), which combine a displacement verb with a directional verb (e.g., *zǒu-jìn* 'walk-into' and *bān-chū* 'move-out'). In these constructions, the displacement verb encodes the manner of motion (e.g., *zǒu* 'walk', *pǎo* 'run', *bān* 'move'), whereas the directional verb specifies the path of motion (e.g., *jìn* 'into', *chū* 'out'). Because the directional verb provides the earliest explicit cue to motion path, this linguistic property enables us to test whether effects emerge once critical path information becomes available and remain stable as subsequent linguistic information unfolds, or whether congruency effects strengthen as additional sentence information accumulates, as suggested by previous studies using self-paced reading paradigms interleaved with visual probes (Hu & Zhao, 2026; Nishide et al., 2025; Sato et al., 2013). Building on the congruency effects observed at the sentence-final diagram region in Mandarin processing (Wang & Zhao, 2024b), the present study aims to examine whether these effects become behaviourally detectable during incremental comprehension of Mandarin sentences.

To investigate these issues, we adopt a self-paced reading paradigm interleaved with schematic diagrams in which diagrams are inserted at two different regions of segmented sentences. Two factors are manipulated. First, task demands are varied by comparing a sentence-diagram verification task with a sentence-diagram viewing task. Importantly, the two tasks employ identical linguistic and visual stimuli,

https://doi.org/10.1017/langcog.2026.10120 Published online by Cambridge University Press

enabling a controlled comparison of whether congruency effects reflect task-sensitive enhancement of semantic-visual alignment rather than differences in materials and response modes. Second, the informational context of visual probing is manipulated by presenting the diagram either immediately after the directional verb or after the LM noun. These regions are selected because the directional verb provides the earliest reliable cue to the motion path in Mandarin RVCs, whereas the LM noun completes the spatial relation of motion. Comparing these two regions therefore allows us to examine whether semantic-visual alignment is established immediately after path information becomes available, or whether they become stronger as additional spatial information is incrementally integrated. Diagrams are adopted as they provide controlled and effective probes of semantic-visual alignment of schemas during language processing (Wang, 2026; Wang & Zhao, 2024a, 2024b).

The study addresses two guiding research hypotheses. First, we examine whether congruency effects are task-sensitive and predict stronger effects in the verification task than in the viewing task. Second, we test whether congruency effects vary as a function of diagram insertion region. If early path cues in Mandarin allow rapid establishment of semantic-visual alignment, congruency effects should emerge early and remain relatively stable across regions. The present study therefore provides a stringent assessment of whether congruency effects under different task conditions necessarily strengthen as linguistic input unfolds or instead reflect temporally stable semantic-visual alignment once diagnostic motion information becomes available during incremental Mandarin motion sentence processing.

2. Method

2.1 Participants

149 Chinese adults (48 males and 101 females) were recruited from a WeChat official platform (*mean age* = 23.64, *SD* = 5.10). The participants were from different universities in China, including 119 undergraduates and 30 postgraduates. All participants were native Mandarin Chinese speakers with an average self-rated L1 proficiency score of 89.09 (*SD* = 11.57) out of 100[1]. In addition to Mandarin and English, some participants reported knowledge of other languages, such as Cantonese ($n = 17$), Japanese ($n = 16$), Korean ($n = 6$), Spanish ($n = 1$) and Thai ($n = 1$). Written informed consent was obtained from each participant prior to participation. Upon completion of the tasks, participants received monetary compensation for their time. Demographic questionnaire is provided in Supplementary Material 1.

2.2 Materials and design

The study employed a self-paced reading interleaved with diagram paradigm (Sato et al., 2013), using a 2 Task (verification, viewing) × 2 Congruency (congruent, incongruent) × 2 Diagram Region (after directional verb, after LM noun) factorial Latin square design. By manipulating the semantic congruency between sentences and diagrams, the study aimed to examine congruency effects, whereby congruent

[1]Self-rated proficiency scale adopted from Wang and Zhao (2024b): 10 = extremely poor, 20 = very poor, 30 = poor, 40 = limited, 50 = average, 60 = standard, 70 = good, 80 = very good, 90 = excellent 100 = native-like.

https://doi.org/10.1017/langcog.2026.10120 Published online by Cambridge University Press

sentence-diagram pairings were expected to facilitate behavioural responses relative to incongruent pairings in both verification and viewing tasks. RT, defined as the latency from diagram onset to the participant's key press, served as the dependent measure in both tasks. In the verification task, RT reflected the time participants spent making a diagram verification judgment, whereas in the viewing task it reflected the time they spent viewing the diagram before advancing to the next segment.

The task was manipulated between participants to avoid strategic carry-over effects that might arise if the same participants performed both tasks within a single session. The verification and viewing tasks shared identical sentence and visual stimuli and differed only in task instructions. Experiencing the verification task could encourage participants to actively evaluate diagram-sentence relations even during the viewing task, thereby reducing the intended contrast in task demands. The diagram region was manipulated between participants to minimise the salience of the timing manipulation. A within-participant manipulation would have made variation in diagram position more apparent, potentially encouraging participants to anticipate when the diagrams would appear or adopt task-specific strategies directed at diagram presentation. Such strategic monitoring could have influenced online sentence processing and complicated the interpretation of the observed effects across diagram regions.

2.2.1 Visual stimuli

The visual stimuli were adapted from Wang and Zhao (2024b) and consisted of two target diagrams (*into* diagram, *out-of* diagram), one distractor diagram ('######') and five filler diagrams (See Table 1). The target diagrams represented spatial relations corresponding to the meanings of the directional verb *jìn* ('into') and *chū* ('out of'). The *into* diagram depicted a red ball moving into a box, whereas the *out-of* diagram depicted a red ball moving out of a box. The distractor diagram, consisting of a string of hashtags (######), was comparable in size and colour to the target diagrams but did not contain spatial information. It was included to reduce the salience and predictability of the target stimuli, thereby discouraging diagram verification strategies.

Five filler diagrams depicted other spatial configurations, including SUPPORT, ADJACENCY, PROXIMITY and VERTICALITY (HIGHER THAN and LOWER THAN). These filler diagrams were crafted based on descriptions of TR-LM relations such as *on*, *close to*, *near*, *over/above* and *under/below*, as discussed in previous cognitive linguistic studies (Lindstromberg, 2010; Tyler & Evans, 2003). The inclusion of filler diagrams was intended to obscure the study focus and minimise strategic responding.

2.2.2 Sentence stimuli

The target Chinese directional verbs were 进 (*jìn*, 'into') and 出 (*chū*, 'out of'). A total of 78 sentence stimuli were used (See Supplementary Material 2), including 48 target sentences and 30 filler sentences. All target sentences were divided into ten segments (e.g., 早上九点,/一位/勤奋的/员工/走/进/办公室/然后/开始/工作。English Translation: *At 9 a.m., /a/diligent/worker/walked/into/the office/and/started/working.*). Target sentences were randomly divided into three subsets, in which 16 were paired with the *into* diagram, 16 with the *out-of* diagram

https://doi.org/10.1017/langcog.2026.10120 Published online by Cambridge University Press

and the remaining 16 with the distractor diagram. Sentence-to-diagram assignments were counterbalanced across two diagram regions, resulting in six stimulus lists.

To ensure that the sentence sets were comparable, frequencies of RVCs and RVC-LM collocations were checked using the Centre for Chinese Linguistics (CCL) corpus (Zhan et al., 2003, 2019). After log-transformation, one-way ANOVAs revealed no significant differences across three target sentence subsets for the following measures: RVC frequency (F (2, 21) = 0.07, MSE = 1.68, p = 0.94), RVC-LM collocation frequency (F (2, 21) = 0.03, MSE = 2.56, p = 0.97) and sentence length (F (2, 21) = 2.14, MSE = 0.003, p = 0.14) for *jìn*; RVC frequency (F (2, 21) = 0.29, MSE = 2.22, p = 0.75), RVC-LM collocation frequency (F (2, 21) = 2.01, MSE = 1.30, p = 0.16) and sentence length (F (2, 21) = 0.35, MSE = 0.003, p = 0.71) for *chū*.

Filler sentences were adapted from a Chinese grammar book (Ross & Ma, 2014), and did not include the two target Chinese directional verbs. The filler sentences matched in length to target sentences and segmented into ten regions. They were randomly paired with five filler diagrams and were consistent across six stimulus lists.

2.2.3 Sentence-diagram verification task

The sentence-diagram verification task was adapted from the self-paced sentence-picture verification task in Sato et al. (2013) and implemented using a noncumulative linear (i.e., moving window) display with phrase-by-phrase segmentation. Each trial began with a fixation cross (+) indicating the onset position of the first sentence segment. Subsequently, each segment of the sentence was displayed on the screen one at a time. A diagram was interleaved with the sentence and appeared at one of two diagram regions: after the directional verb (Region 7) or after the LM (Region 8). Table 2 shows the specific regions where diagrams were inserted in a sample sentence stimulus.

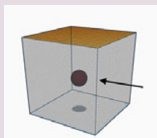
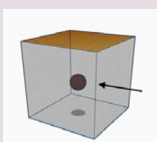
In each trial, participants pressed the space bar to read the sentence segment by segment. When a diagram appeared, they were asked to make a binary verification judgment on whether the diagram depicted a spatial configuration implied by the sentence. Participants pressed 'F' for 'no' or 'J' for 'yes' as quickly as possible within a 5-second time limit. After each trial, including both target and filler trials, participants answered a reading comprehension question presented in Chinese. The questions targeted either information from the sentence (e.g., the adverbial clause, LM or TR) or the accompanying diagram to encourage attention to both linguistic and visual information. For sentence-based questions (e.g., *Who went into the office?*), two Chinese answer options were presented (e.g., *the worker* vs. *the boss*). For diagram-based questions (e.g., *Does either of the following diagrams appear in this sentence?* or *Which of the following diagrams appears in this sentence?*), participants responded by selecting between two Chinese options (e.g., *yes* vs. *no*; *left* vs. *right*) presented together with two of the diagram stimuli (see Table 1). A complete list of comprehension questions and answer options is provided in Supplementary Material 2. Participants responded by pressing the 'F' or 'J' key within a 10-second time limit. The inclusion of comprehension questions following both target and filler trials was intended to discourage superficial or inattentive responding and enhance the reliability of the RT data.

Diagram verification RTs and accuracy rates (ARs), as well as reading comprehension question ARs, were automatically recorded online. Verification responses were coded as correct if participants responded 'yes' to diagrams that were semantically congruent with the Chinese directional verb in the sentence (e.g., *into* diagram

https://doi.org/10.1017/langcog.2026.10120 Published online by Cambridge University Press

https://doi.org/10.1017/langcog.2026.10120 Published online by Cambridge University Press

Table 2. Regions of sentence segments and diagrams

Region	1	2	3	4	5	6	7	8	9	10	11
After directional verb	早上九点, *zǎo-shàng jiǔ-diǎn* nine AM	一位 *yī-wèi* a	勤奋的 *qín-fèn-de* diligent	员工 *yuán-gōng* worker	走 *zǒu* walk	进 *jìn* into		办公室 *bàn-gōng-shì* office	然后 *rán-hòu* and then	开始 *kāi-shǐ* start	工作 *gōng-zuò* work
After LM	早上九点, *zǎo-shàng jiǔ-diǎn* nine AM	一位 *yī-wèi* a	勤奋的 *qín-fèn-de* diligent	员工 *yuán-gōng* worker	走 *zǒu* walk	进 *jìn* into	办公室 *bàn-gōng-shì* office		然后 *rán-hòu* and then	开始 *kāi-shǐ* start	工作 *gōng-zuò* work

Note: LM, landmark. The English translations are literal, word-for-word translations of the original Mandarin stimuli.

– *jìn* sentences) or if 'no' to diagrams that were semantically inconsistent with the directional verb in the sentence (e.g., *into* diagram – *chū* sentences). These criteria were validated by semantic consistency ratings between *jìn* sentences and *into* diagrams, as well as *chū* sentences and *out-of* diagrams (Wang, 2026; Wang & Zhao, 2024b). Trials involving distractor and filler diagrams were excluded from data analysis because they did not instantiate the motion relations central to the experimental manipulation and served only to reduce stimuli predictability.

2.2.4 Sentence-diagram viewing task

The sentence-diagram viewing task procedure closely resembled the sentence-diagram verification task, adopting a self-paced reading interleaved with diagram paradigm. The diagram regions were the same as the sentence-diagram verification task (see Table 2), appearing after the directional verb (Region 7) or after the LM (Region 8). Participants read each sentence segment-by-segment by pressing the spacebar. The key procedural difference was that, when a diagram appeared, participants were instructed to view it and press the spacebar when they finish, but they were not required to make any explicit verification judgment about the sentence-diagram relationship. To discourage superficial reading and maintain participants' attention to both the linguistic and visual stimuli during the viewing task, sentence- and diagram-based comprehension questions were randomly interleaved across target and filler trials. The comprehension questions were identical to those used in the verification task, so that the two tasks differed only in whether participants verified the diagram or simply viewed it. Diagram RTs and comprehension ARs in the target trials were automatically recorded online. Only trials involving target diagrams were included in the analysis.

2.3 Procedure

Data were collected online using PsyToolkit (version 3.4.4; Stoet, 2010, 2017). Participants first completed a demographic questionnaire and were then randomly assigned to either the sentence-diagram verification task (n = 82) or the sentence-diagram viewing task (n = 67). Within each task, participants were further randomly assigned to one of six counterbalanced stimulus lists. Consequently, 71 participants were presented with diagrams after the directional verb (verification: n = 40; viewing: n = 31), and 78 were presented with diagrams after the LM (verification: n = 42; viewing: n = 36). In both tasks, participants completed six practice trials before the formal phase. During the practice phase of the verification task, participants received corrective feedback on both diagram verification responses and comprehension question responses. During the practice phase of the viewing task, corrective feedback was provided for comprehension question responses only. In the formal phase of both tasks, participants completed 78 trials without receiving any feedback. Each participant completed the experiment only once and was not allowed to revise previous responses. The entire experimental session lasted approximately 15 to 20 minutes.

2.4 Data analysis

Data analyses were conducted in *R* (version 4.4.0; R Core Team, 2024). Anonymised data are provided in Supplementary Material 3. Prior to analysis, standard data

https://doi.org/10.1017/langcog.2026.10120 Published online by Cambridge University Press

trimming procedures were applied. Three participants with comprehension ARs below 60% (verification: $n = 2$; viewing: $n = 1$) were excluded because such performance suggests insufficient attention to sentence content and approaches chance level on two-choice comprehension questions. In addition, ten participants in the verification task with verification ARs below 60% were excluded because this indicates unreliable engagement with the diagram-sentence relation, making the verification responses unsuitable for analysing congruency effects (Chen et al., 2020). The final sample comprised 136 participants, including 70 in the verification task (diagram after directional verb: $n = 33$; diagram after LM: $n = 37$) and 66 in the viewing task (diagram after directional verb: $n = 31$; diagram after LM: $n = 35$).

For both tasks, analyses were restricted to trials containing target diagrams from the formal task phase. Trials with incorrect verification responses or incorrect comprehension question responses were excluded (11.7%). RTs exceeding ±3 SDs from each participant's mean were removed (1.9%) to reduce the influence of extreme outliers while preserving the majority of the response distribution (Chen et al., 2020).

All RTs were log-transformed prior to modelling. Linear mixed-effects models were fitted to the log-transformed RTs using the *lme4* package (Bates et al., 2022). An initial model was fitted with task, congruency and diagram region as fixed effects (Linck & Cunnings, 2015). Random-effects structures were evaluated by comparing the model with random intercepts only and the model including random slopes for congruency by participants and items using likelihood ratio tests. The latter showed improved model fit and was therefore retained for reported analyses. To examine whether the congruency effect changed over the course of the experiment, an additional linear mixed-effects model was fitted including trial order, congruency and their interaction as fixed effects. Statistical significance was assessed using the *lmerTest* package (Kuznetsova et al., 2017). Model convergence was verified, and statistical assumptions were checked. Pairwise comparisons and estimated marginal means (EMMs) were computed using the *emmeans* package (Lenth et al., 2023). EMMs were back-transformed to the response scale for reporting and visualisation. In addition, post-hoc simulation-based power analyses were conducted using the *simr* package (Green & MacLeod, 2016). Based on 1000 simulations, the estimated power to detect the target fixed effects and interactions reached approximately 80%, indicating adequate sensitivity for the effects of interest. Effect sizes were estimated using Cohen's *d*. Following the guideline (Cohen, 1988), values of approximately 0.2, 0.5 and 0.8 were interpreted as small, medium and large effects, respectively. Visualisations were generated using the *ggplot2* package (Wickham, 2016). R scripts used for data analysis are provided in Supplementary Material 4.

3. Results

First, descriptive statistical analysis was conducted. Results showed the mean comprehension accuracy was 0.89 ($SD = 0.16$) in the verification task and 0.92 ($SD = 0.12$) in the viewing task. The high comprehension ARs across both tasks indicate that participants attended to sentence meaning during the experiment. Descriptive statistical results of RTs by diagram region, congruency and task are reported in Table 3.

https://doi.org/10.1017/langcog.2026.10120 Published online by Cambridge University Press

Table 3. Descriptive statistics of response times by diagram region, congruency and task

		Verification task (ms)		Viewing task (ms)	
Diagram region	Congruency	*M*	*SD*	*M*	*SD*
After directional verb	Congruent	1383.56	616.48	697.89	615.45
	Incongruent	1607.25	714.83	706.31	573.87
After landmark	Congruent	1504.76	593.06	675.41	351.79
	Incongruent	1668.24	760.01	709.22	399.88

Next, a linear mixed-effects model[2] was built to investigate RTs. Results revealed significant fixed effects of task and congruency on RTs, as well as a significant interaction between congruency and task (see Table 4). No significant fixed effect of diagram region was observed; nor did diagram region interact significantly with any other fixed effects.

Post-hoc comparisons of the fixed effect of congruency indicated that mean RTs in the congruent condition (*EMM* = 888, *SE* = 27.3, *df* = 133, 95% CI [828, 952]) were estimated to be 68 ms shorter than those in the incongruent condition (*EMM* = 956, *SE* = 29.8, *df* = 132, 95% CI [891, 1026]), with a small congruency effect ($t = 5.28$, $p < 0.001$, Cohen's $d = 0.23$, *SE* = 0.04, *df* = 132, 95% CI [0.15, 0.32]). Post-hoc comparisons of the fixed effect of task showed that mean RTs in the verification task (*EMM* = 1432, *SE* = 60.1, *df* = 134, 95% CI [1302, 1575]) were substantially longer than those in the viewing task (*EMM* = 593, *SE* = 25.5, *df* = 132, 95% CI [538, 653]), with a large effect ($t = 14.71$, $p < 0.001$, Cohen's $d = 2.78$, *SE* = 0.19, *df* = 132, 95% CI [2.40, 3.16]).

Table 4. Linear mixed-effects model results of response times

Fixed effects	*β*	*SE*	*95% CI*	*t*	*p*
Intercept	7.17	0.06	[7.05, 7.29]	115.27	<0.001***
Congruency	0.13	0.03	[0.08, 0.19]	4.56	<0.001***
Diagram region	0.09	0.09	[−0.08, 0.25]	1.01	0.312
Task	−0.84	0.09	[−1.014, −0.66]	−9.43	<0.001***
Congruency × diagram region	−0.05	0.04	[−0.13, 0.03]	−1.29	0.195
Congruency × task	−0.11	0.04	[−0.19, 0.03]	−2.72	0.007**
Diagram region × task	−0.02	0.12	[−0.26, 0.22]	−0.16	0.870
Congruency × diagram region × task	0.08	0.06	[−0.03, 0.18]	1.36	0.172
Random effects	*Variance*	*S.D.*	*Correlation*		
Participant (intercept)	0.11	0.35			
Item (intercept)	<0.001	0.02			
Congruency (slope): Participant	0.01	0.10	−0.12		
Congruency (slope): Item	<0.001	0.002	−1.00		
Residual	0.10	0.32			
Model fit	R^2		Marginal		Conditional
			0.47		0.76

Note: ***$p < 0.001$, **$0.001 < p < 0.01$, *$0.01 < p < 0.05$.

[2]Model equation: log(RT) ~ Congruency*Diagram Region*Task + (1 + Congruency | Participant) + (Congruency | Item).

https://doi.org/10.1017/langcog.2026.10120 Published online by Cambridge University Press

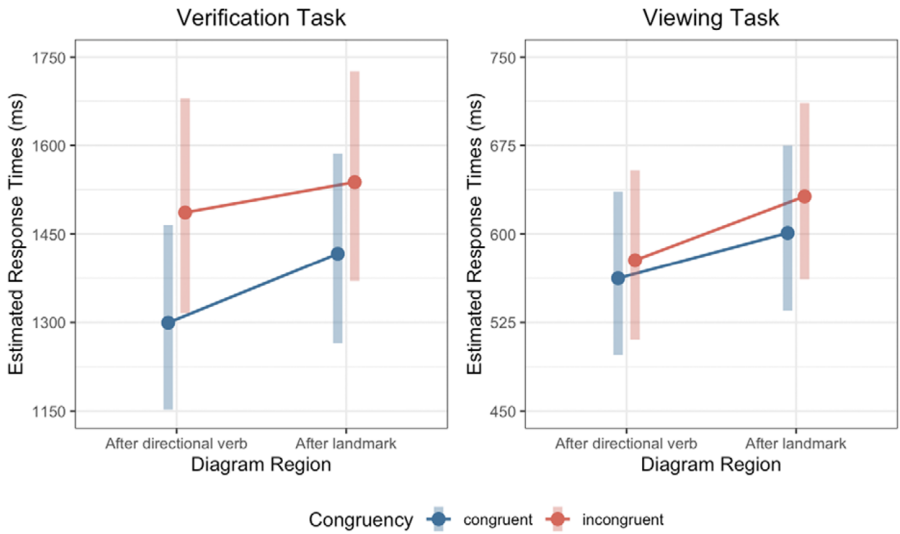


Figure 1. Estimated response times (ms) for congruent and incongruent diagram conditions across diagram regions in the verification task (left panel) and viewing task (right panel). The *x*-axis shows the two diagram insertion regions: after the directional verb and after the landmark. Blue symbols and lines represent congruent trials, and red symbols and lines represent incongruent trials. Points indicate model-estimated mean response times, and error bars represent 95% confidence intervals.

To further examine the congruency × task interaction, follow-up analyses were conducted. In the verification task, responses in the congruent condition (*EMM* = 1357, *SE* = 58.1, *df* = 135, 95% CI [1232, 1495]) were estimated to be 155 ms faster than those in the incongruent condition (*EMM* = 1512, *SE* = 65.9, *df* = 135, 95% CI [1370, 1669]), corresponding to a small congruency effect ($t = 5.37$, $p < 0.001$, Cohen's $d = 0.34$, *SE* = 0.06, *df* = 135, 95% CI [0.22, 0.47]). In the viewing task, a smaller congruency effect was observed ($t = 2.02$, $p = 0.046$, Cohen's $d = 0.12$, *SE* = 0.06, *df* = 130, 95% CI [0.002, 0.24]): responses in the congruent condition (*EMM* = 581, *SE* = 25.5, *df* = 132, 95% CI [526, 642]) were estimated to be 23 ms faster than those in the incongruent condition (*EMM* = 604, *SE* = 26.8, *df* = 130, 95% CI [546, 668]). Figure 1 presents the estimated RTs by task, congruency and diagram region.

To check whether the congruency effect reflected progressively learned verb-diagram associations over the course of the experiment, an additional linear mixed-effects model[3] was fitted. Results showed a significant fixed effect of trial order ($\beta = -0.01$, $p < 0.001$), indicating participants became significantly faster over the course of the tasks. Crucially, the interaction between congruency and trial order was not significant ($\beta = -0.001$, $p = 0.24$), suggesting that the magnitude of the congruency effect remained stable throughout the tasks rather than increasing with repeated exposure.

4. Discussion

The present study examined whether semantic-visual alignment of motion schema during incremental Mandarin sentence processing is modulated by task demands

[3]Model equation: log(RT) ~ Congruency*Trial Order + (1 + Congruency | Participant) + (Congruency | Item).

https://doi.org/10.1017/langcog.2026.10120 Published online by Cambridge University Press

and the informational context in which perceptual probes are encountered. Overall, congruency effects were observed across both tasks and diagram regions, indicating that participants processed the spatial relations conveyed by the sentences in a manner that aligned with the schematic structure represented in the diagrams. Importantly, larger congruency effects were found in the verification task compared to the viewing task, but the congruency effects did not vary regardless of whether the diagram was after the directional verb or the LM noun. These task-sensitive congruency effects provide a behavioural index of how task demands modulate semantic-visual alignment during incremental sentence comprehension.

4.1 Task-sensitive semantic-visual alignment of motion schema

First, we found a significant fixed effect of task, with substantially faster RTs in the viewing task than the verification task. This finding is consistent with the experimental manipulation of different processing demands on the two tasks. In the verification task, RTs reflect not only diagram processing but also the explicit evaluation of the sentence-diagram correspondence, decision formation and response selection. By contrast, participants in the viewing task were required to inspect the diagram. Consequently, the viewing task placed lower immediate cognitive demands on evaluating sentence-diagram correspondence, although the task encouraged participants to maintain attention to both linguistic and visual information throughout the experiment. Therefore, absolute RT differences between tasks should be interpreted as reflecting differences in overall task demands.

The congruency × task interaction provides a more direct index of how specific processing demands modulate the behavioural expression of semantic-visual alignment. Specifically, the larger congruency effect observed in the verification task suggests that requiring participants to explicitly evaluate whether the diagram matched the sentence may strengthen the mapping between linguistic meaning and schematic diagram representations. By contrast, participants in the viewing task could inspect the diagram without being required to explicitly compare it with the preceding sentence, reducing the need for detailed evaluation of the sentence-diagram correspondence. Consequently, a substantially weaker congruency effect was observed in the viewing task, suggesting a modest sensitivity to congruent directional information in sentence-diagram pairings.

Notably, the use of schematic diagrams may have contributed to the stronger congruency effects observed in the verification task by providing a structured basis for mapping linguistic meaning onto visual spatial relations. In the present stimuli, circles represented TRs, boxes represented LMs and arrows encoded motion direction, enabling participants to rapidly align sentence meaning with an abstract spatial event structure. Such alignment may function as an external representational scaffold that facilitates the construction of a coherent situation model (Hu & Zhao, 2026), particularly when participants are required to verify semantic congruency. In contrast, the viewing task imposed no requirement to establish this mapping, allowing diagrams to be processed more superficially and reducing the behavioural expression of semantic-visual alignment between sentence meaning and spatial structure.

The present findings can further be interpreted from multiple accounts. From the perspective of embodied accounts, the stronger congruency effects observed in the verification task are consistent with the view that language comprehension may

https://doi.org/10.1017/langcog.2026.10120 Published online by Cambridge University Press

involve the partial reactivation of modality-specific sensorimotor traces associated with prior perceptual experience (Barsalou, 1999, 2008). Under this interpretation, such traces are not assumed to be obligatorily recruited during comprehension but may be engaged to a greater extent when explicit sentence-diagram verification makes perceptual information task-relevant, while remaining relatively weakly activated under conditions that permit shallower processing (Ostarek et al., 2019). This interpretation is broadly consistent with previous behavioural studies reporting stronger congruency effects when perceptual or motor information is directly enhanced by task requirements (Bub & Masson, 2006; Chen et al., 2020; Rommers et al., 2013; Tomasino et al., 2007).

An alternative interpretation is provided by amodal accounts of semantic processing. From this perspective, the stronger congruency effects in the verification task may reflect enhanced semantic comparison between linguistic and visual information, without requiring the recruitment of modality-specific sensorimotor representations (Mahon & Caramazza, 2008; Meteyard et al., 2012; Ostarek et al., 2019). Explicit verification encourages participants to evaluate the correspondence between sentence meaning and diagrammatic spatial relations, thereby increasing the behavioural expression of semantic-visual alignment.

The present behavioural findings do not allow these interpretations to be distinguished or establish the causal cognitive mechanisms underlying the semantic-visual alignment (Horchak & Garrido, 2020; Ostarek et al., 2019). Rather, they demonstrate that the behavioural expression of such alignment is sensitive to task demands. Future research combining congruency paradigms with more diagnostic approaches such as interference or neurophysiological measures (Horchak & Garrido, 2022; Szűcs & Soltész, 2008; Yao & Vanduffel, 2022), will be necessary to determine the representational format and causal mechanisms underlying these task-sensitive congruency effects.

4.2 Congruency effects across incremental sentence regions

A second key finding of the present study is that congruency effects did not differ as a function of diagram region. Specifically, presenting the diagram immediately after the directional verb or after the LM noun yielded comparable congruency effects, and no significant interaction between congruency and diagram region was observed. One interpretation is that Mandarin directional verbs such as *jìn* ('into') and *chū* ('out of') provide sufficiently strong schematic constraints to establish a basic motion representation early in comprehension. Once the directional verb is encountered, comprehenders may already have sufficient information for the level of diagram-sentence alignment measured in the present task, with the LM noun primarily contributing to referential specification rather than altering the core spatial relation.

At the same time, this null interaction should be interpreted cautiously. The present data do not show that processing was identical across the two probe positions; rather, they indicate that the behavioural congruency effect remained stable across the specific within-sentence regions tested here. This contrasts with earlier studies reporting stronger congruency effects at later probe positions (Nishide et al., 2025; Sato et al., 2013). One plausible explanation concerns the informational structure of the linguistic input. In those studies, earlier sentence regions provided only partial

https://doi.org/10.1017/langcog.2026.10120 Published online by Cambridge University Press

constraints on the intended interpretation, so additional linguistic material was needed before a stable event representation could be established. By contrast, in the present Mandarin materials, the directional verb may already have supplied the critical schematic information needed to establish a motion representation at an early stage. Because the schematic diagrams transparently encoded the relevant spatial relation, this early representation could be readily aligned with the visual stimulus.

A related methodological consideration is that later congruency effects in some studies may partly reflect wrap-up processes (Hu & Zhao, 2026; Nishide et al., 2025; Sato et al., 2013). That is, integrative processing becomes stronger when comprehenders reach a clause or sentence boundary and consolidate preceding information. In the present design, however, both probe positions occurred within the sentence rather than at sentence-final wrap-up points, which makes a wrap-up-based explanation less likely as the primary source of the observed pattern.

More broadly, the present findings suggest that congruency effects do not necessarily increase across sentence regions during incremental comprehension. Their temporal profile may depend on how strongly early linguistic cues constrain the relevant representation and on how much later information is needed for revision or disambiguation. On this view, increasing congruency effects are more likely when early interpretations remain underspecified, whereas stable effects may arise when key schematic information is available early.

4.3 Limitations and future directions

The present study provides novel evidence regarding the conditions under which congruency effects emerge during incremental sentence comprehension, but several limitations should be acknowledged when interpreting these findings and considering future research. First, although the present study employed only two highly frequent Mandarin directional verbs and two corresponding motion schemas, the trial-order analysis indicated that the congruency effect remained stable across the experiment rather than increasing with repeated exposure, arguing against an explanation based primarily on progressively learned verb-diagram mappings. Nevertheless, future studies should include a broader range of directional verbs, motion schemas and probe locations to further minimise the possibility of strategic mapping and to determine when congruency effects strengthen over time and when they remain stable.

Another methodological consideration concerns the between-participant manipulation of both task and diagram region. This design was adopted to minimise the salience of the experimental manipulations and reduce the likelihood that participants would develop strategic expectations regarding task requirements or the timing of diagram presentation. Nevertheless, manipulating both factors within participants in future research may increase statistical power and provide a more sensitive test of how task demands and informational context jointly modulate semantic-visual alignment. To minimise potential carry-over or strategy effects associated with repeated exposure to different task conditions and diagram positions, future studies could incorporate additional filler items or counterbalancing procedures to reduce the salience of the experimental manipulations while retaining the advantages of a within-participant design.

https://doi.org/10.1017/langcog.2026.10120 Published online by Cambridge University Press

5. Conclusion

The present study examined how task demands and informational context shape the behavioural detectability of semantic-visual alignment of motion schema during incremental processing of Mandarin motion sentences. Three main conclusions emerge.

First, congruency effects were observed across both tasks, indicating that comprehenders systematically aligned sentence meaning with the motion relations encoded in the diagrams. At the same time, these effects were stronger in the verification task than in the viewing task, showing that congruency effects are not a fixed consequence of sentence comprehension but are modulated by task demands. This finding is consistent with a goal-adaptive account, whereby semantic-visual alignment becomes more behaviourally salient when explicit sentence-diagram mapping is encouraged.

Second, congruency effects remained stable across the two within-sentence probe locations tested in the present design. This pattern suggests that congruency effects do not necessarily strengthen as linguistic input unfolds. Rather, their temporal profile may depend on the informativeness of early linguistic cues and the extent to which subsequent input requires the developing interpretation to be refined.

Third, the study contributes methodologically by demonstrating the value of combining task manipulation with incremental visual probing while holding the linguistic and visual materials constant. By contrasting verification and viewing and by probing semantic-visual alignment at two theoretically motivated regions, the present design offers a more controlled basis for interpreting congruency effects than comparisons across paradigms that differ in response modes or rely solely on sentence-final probes.

Taken together, these findings contribute to a more nuanced understanding of congruency effects during language comprehension. Rather than treating congruency effects as a uniform index of a particular representational mechanism, the present study demonstrates that their behavioural expression is shaped by task demands and linguistic informativeness. Although the present findings can be interpreted from an embodied perspective, they are also compatible with the alternative account of amodal semantic comparison. Future research employing more diagnostic methodologies will be important for distinguishing between these interpretations and clarifying the mechanisms underlying semantic-visual alignment during incremental sentence comprehension.

Data availability statement. Data and supplementary materials are accessible from https://osf.io/5hbjy/

Acknowledgements. We would like to express our gratitude to the editors and anonymous reviewers for their valuable comments and constructive suggestions.

Funding statement. The work is supported by the Fundamental Research Funds for the Central Universities in UIBE (grant number 26QD25).

Ethics statement. The study involving humans was approved by the Office of Research Ethics and Integrity of the University of Melbourne. All participants provided their written informed consent to participate in this study.

References

Ahlberg, D. K., Bischoff, H., Kaup, B., Bryant, D., & Strozyk, J. V. (2018). Grounded cognition: Comparing language × space interactions in first language and second language. *Applied PsychoLinguistics*, 39(2), 437–459. https://doi.org/10.1017/S014271641700042X

Bai, B., Yang, C., & Fan, J. (2022). Semantic integration of multidimensional perceptual information in L1 sentence comprehension. *Language and Cognition*, 14(1), 109–130. https://doi.org/10.1017/langcog.2021.24

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, 22(4), 577–660. https://doi.org/10.1017/S0140525X99002149

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., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., Dai, B., Scheipl, F., Grothendieck, G., Green, P., Fox, J., Bauer, A., & Krivitsky, P. N. (2022). *Package 'lme4'* (Version 1.1–29) [Computer software]. Comprehensive R Archive Network. https://cran.r-project.org/web/packages/lme4/lme4.pdf

Bechtold, L., Cosper, S. H., Malyshevskaya, A., Montefinese, M., Morucci, P., Niccolai, V., Repetto, C., Zappa, A., & Shtyrov, Y. (2023). Brain signatures of embodied semantics and language: A consensus paper. *Journal of Cognition*, 6(1), 1–40. https://doi.org/10.5334/joc.237

Bergen, B. K. (2007). Experimental methods for simulation semantics. In M. Gonzalez-Marquez, Mittelberg, S. Coulson, & M. J. Spivey (Eds.), *Human cognitive processing* (Vol. 18, pp. 277–301).John Benjamins Publishing Company.

Bergen, B. K., & Chang, N. (2005). Embodied construction grammar in simulation-based language understanding. In J.-O. Östman & M. Fried (Eds.), *Constructional approaches to language* (pp. 147–190). John Benjamins Publishing Company.

Bergen, B. K., & Chang, N. (2013). Embodied constructional grammar. In T. Hoffmann & G. Trousdale (Eds.), *The Oxford handbook of construction grammar* (pp. 168–190). Oxford University Press.

Bergen, B. K., Lau, T.-T. C., Narayan, S., Stojanovic, D., & Wheeler, K. (2010). Body part representations in verbal semantics. *Memory & Cognition*, 38(7), 969–981. https://doi.org/10.3758/MC.38.7.969

Bergen, B. K., Narayan, S., & Feldman, J. (2003). Embodied verbal semantics: Evidence from an image-verb matching task. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 25. https://escholarship.org/uc/item/1hg3n38m

Borreggine, K. L., & Kaschak, M. P. (2006). The action-sentence compatibility effect: It's all in the timing. *Cognitive Science*, 30(6), 1097–1112. https://doi.org/10.1207/s15516709cog0000_91

Bryant, D. J., & Tversky, B. (1999). Mental representations of perspective and spatial relations from diagrams and models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25(1), 137–156. https://doi.org/10.1037/0278-7393.25.1.137

Bub, D. N., & Masson, M. (2006). Gestural knowledge evoked by objects as part of conceptual representations. *Aphasiology*, 20(9), 1112–1124. https://doi.org/10.1080/02687030600741667

Chen, D., Wang, R., Zhang, J., & Liu, C. (2020). Perceptual representations in L1, L2 and L3 comprehension: Delayed sentence–picture verification. *Journal of Psycholinguistic Research*, 49(1), 41–57. https://doi.org/10.1007/s10936-019-09670-x

Chen, S.-C., de Koning, B. B., & Zwaan, R. A. (2020). Does object size matter with regard to the mental simulation of object orientation? *Experimental Psychology*, 67(1), 56–72. https://doi.org/10.1027/1618-3169/a000468

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Routledge.

de Koning, B. B., Wassenburg, S. I., Bos, L. T., & van der Schoot, M. (2017). Mental simulation of four visual object properties: Similarities and differences as assessed by the sentence–picture verification task. *Journal of Cognitive Psychology*, 29(4), 420–432. https://doi.org/10.1080/20445911.2017.1281283

Edmiston, P., & Lupyan, G. (2017). Visual interference disrupts visual knowledge. *Journal of Memory and Language*, 92, 281–292. https://doi.org/10.1016/j.jml.2016.07.002

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

Gallese, V., & Lakoff, G. (2005). The brain's concepts: The role of the sensory-motor system in conceptual knowledge. *Cognitive Neuropsychology*, 22(3–4), 455–479. https://doi.org/10.1080/02643290442000310

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

Green, P., & MacLeod, C. J. (2016). simr: An R package for power analysis of generalised linear mixed models by simulation. *Methods in Ecology and Evolution*, 7(4), 493–498. https://doi.org/10.1111/2041-210X.12504

Horchak, O. V., & Garrido, M. V. (2020). Is complex visual information implicated during language comprehension? The case of cast shadows. *Cognitive Science*, 44(7), e12870. https://doi.org/10.1111/cogs.12870

https://doi.org/10.1017/langcog.2026.10120 Published online by Cambridge University Press

Horchak, O. V., & Garrido, M. V. (2022). Simulating background settings during spoken and written sentence comprehension. *Psychonomic Bulletin & Review*, 29(4), 1426–1439. https://doi.org/10.3758/s13423-022-02061-9

Hu, Y., & Zhao, H. (2026). Mental imagery effects in bilingual phrasal verb comprehension. *Language and Cognition*, 18(e13), 1–29. https://doi.org/10.1017/langcog.2026.10062

Johnson, M. (1987). *The body in the mind: The bodily basis of meaning, imagination, and reason*. University of Chicago Press.

Kaschak, M. P., Madden, C. J., Therriault, D. J., Yaxley, R. H., Aveyard, M., Blanchard, A. A., & Zwaan, R. A. (2005). Perception of motion affects language processing. *Cognition*, 94(3), B79–B89. https://doi.org/10.1016/j.cognition.2004.06.005

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

Lenth, R. V., Buerkner, P., Giné-Vázquez, I., Herve, M., Jung, M., Love, J., Miguez, F., Riebl, H., & Singmann, H. (2023). *Package 'emmeans'* (Version 1.8.4–1) [Computer software]. https://cran.r-project.org/web/packages/emmeans/emmeans.pdf

Li, C. (2013). Mandarin resultative verb compounds: Simple syntax and complex thematic relations. *Language Sciences*, 37, 99–121. https://doi.org/10.1016/j.langsci.2012.11.001

Li, C. N., & Thompson, S. A. (2009). *Mandarin Chinese: A functional reference grammar* (1st ed.). University of California Press.

Linck, J. A., & Cunnings, I. (2015). The utility and application of mixed-effects models in second language research: Mixed-effects models. *Language Learning*, 65(S1), 185–207. https://doi.org/10.1111/lang.12117

Lindstromberg, S. (2010). *English prepositions explained*. John Benjamins Publish Company.

Liu, D., Wang, L., & Han, Y. (2024). Mental simulation of colour properties during language comprehension: Influence of context and comprehension stages. *Cognitive Processing*, 25(4), 587–600. https://doi.org/10.1007/s10339-024-01201-4

Mahon, B. Z., & Caramazza, A. (2008). A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. *Journal of Physiology-Paris*, 102(1–3), 59–70. https://doi.org/10.1016/j.jphysparis.2008.03.004

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

Nishide, T., Zhao, H., & De Deyne, S. (2025). Cross-linguistic L1 dis/similarity effect on mental imagery in incremental motion event processing. *Bilingualism: Language and Cognition*, 1–17. https://doi.org/10.1017/S1366728925100801

Ostarek, M., & Huettig, F. (2017). A task-dependent causal role for low-level visual processes in spoken word comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 43(8), 1215–1224. https://doi.org/10.1037/xlm0000375

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

Ostarek, M., Joosen, D., Ishag, A., De Nijs, M., & Huettig, F. (2019). Are visual processes causally involved in "perceptual simulation" effects in the sentence-picture verification task? *Cognition*, 182, 84–94. https://doi.org/10.1016/j.cognition.2018.08.017

Pecher, D., van Dantzig, S., Zwaan, R. A., & Zeelenberg, R. (2009). Short article: Language comprehenders retain implied shape and orientation of objects. *Quarterly Journal of Experimental Psychology*, 62(6), 1108–1114. https://doi.org/10.1080/17470210802633255

Pecher, D., Zanolie, K., & Zeelenberg, R. (2007). Verifying visual properties in sentence verification facilitates picture recognition memory. *Experimental Psychology*, 54(3), 173–179. https://doi.org/10.1027/1618-3169.54.3.173

Pulvermüller, F. (2013). Semantic embodiment, disembodiment or misembodiment? In search of meaning in modules and neuron circuits. *Brain and Language*, 127(1), 86–103. https://doi.org/10.1016/j.bandl.2013.05.015

R Core Team. (2024). *R: A language and environment for statistical computing*. [Computer software]. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Redmann, A., FitzPatrick, I., & Indefrey, P. (2019). The time course of colour congruency effects in picture naming. *Acta Psychologica*, 196, 96–108. https://doi.org/10.1016/j.actpsy.2019.04.005

https://doi.org/10.1017/langcog.2026.10120 Published online by Cambridge University Press

Richardson, D., Spivey, M. J., Barsalou, L. W., & McRae, K. (2003). Spatial representations activated during real-time comprehension of verbs. *Cognitive Science*, 27(5), 767–780. https://doi.org/10.1207/s15516709cog2705_4

Rommers, J., Meyer, A. S., & Huettig, F. (2013). Object shape and orientation do not routinely influence performance during language processing. *Psychological Science*, 24(11), 2218–2225. https://doi.org/10.1177/0956797613490746

Ross, C., & Ma, J. S. (2014). *Modern Mandarin Chinese grammar: A practical guide* (2nd ed.). Routledge.

Sato, M., Schafer, A. J., & Bergen, B. K. (2013). One word at a time: Mental representations of object shape change incrementally during sentence processing. *Language and Cognition*, 5(4), 345–373. https://doi.org/10.1515/langcog-2013-0022

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

Stoet, G. (2010). PsyToolkit: A software package for programming psychological experiments using Linux. *Behavior Research Methods*, 42(4), 1096–1104. https://doi.org/10.3758/BRM.42.4.1096

Stoet, G. (2017). PsyToolkit: A novel web-based method for running online questionnaires and reaction-time experiments. *Teaching of Psychology*, 44(1), 24–31. https://doi.org/10.1177/0098628316677643

Szűcs, D., & Soltész, F. (2008). The interaction of task-relevant and task-irrelevant stimulus features in the number/size congruency paradigm: An ERP study. *Brain Research*, 1190, 143–158. https://doi.org/10.1016/j.brainres.2007.11.010

Talmy, L. (2000). *Toward a cognitive semantics* (Vol. 1). MIT Press.

Tomasino, B., Werner, C. J., Weiss, P. H., & Fink, G. R. (2007). Stimulus properties matter more than perspective: An fMRI study of mental imagery and silent reading of action phrases. *NeuroImage*, 36, T128–T141. https://doi.org/10.1016/j.neuroimage.2007.03.035

Tversky, B., Zacks, J., Lee, P., & Heiser, J. (2000). Lines, blobs, crosses and arrows: Diagrammatic communication with schematic figures. In M. Anderson, P. Cheng, & V. Haarslev (Eds.), *Theory and application of diagrams* (Vol. 1889, pp. 221–230). Springer.

Tyler, A., & Evans, V. (2003). *The semantics of English prepositions: Spatial scenes, embodied meaning and cognition*. Cambridge University Press.

Wang, M. (2026). *Mental simulation in bilingual language processing: Task dependency, temporal dynamics and learner modulation* [Doctoral dissertation, The University of Melbourne]. https://hdl.handle.net/11343/370501

Wang, M., & Zhao, H. (2024a). Are schematic diagrams valid visual representations of concepts? Evidence from mental imagery in online processing of English prepositions. *Language and Cognition*, 16(4), 1564–1587. https://doi.org/10.1017/langcog.2024.18

Wang, M., & Zhao, H. (2024b). Perceptual representations in L1 and L2 spatial and abstract language processing: Applying an innovative sentence-diagram verification paradigm. *Frontiers in Human Neuroscience*, 18(1425576), 1–18. https://doi.org/10.3389/fnhum.2024.1425576

Wang, W. S.-Y., Sun, C., & Tham, S. W. (2015). Resultative verb compounds in Mandarin. In W. S.-Y. Wang & C. Sun (Eds.), *The Oxford handbook of Chinese linguistics* (pp. 307–321). Oxford University Press.

Weiskopf, D. A. (2010). Embodied cognition and linguistic comprehension. *Studies in History and Philosophy of Science Part A*, 41(3), 294–304. https://doi.org/10.1016/j.shpsa.2010.07.005

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis* [Computer software]. Springer-Verlag New York. https://ggplot2.tidyverse.org

Yang, T. (2016). Image schemas in verb–particle constructions: Evidence from a behavioral experiment. *Journal of Psycholinguistic Research*, 45(2), 379–393. https://doi.org/10.1007/s10936-015-9354-6

Yao, T., & Vanduffel, W. (2022). Neuronal congruency effects in macaque prefrontal cortex. *Nature Communications*, 13(1), 4702–4713. https://doi.org/10.1038/s41467-022-32382-1

Zhan, W., Guo, R., Chang, B., Chen, Y., & Chen, L. (2019). The building of the CCL corpus: Its design and implementation. *Corpus Linguistics*, 6(1), 71–86.

Zhan, W., Guo, R., & Chen, Y. (2003). *The CCL Corpus of Chinese Texts: 700 million Chinese characters, the 11th Century B.C. - present.* Available online at the website of Center for Chinese Linguistics (abbreviated as CCL) of Peking University, http://ccl.pku.edu.cn:8080/ccl_corpus

Zwaan, R. A. (2004). The immersed experiencer: Toward an embodied theory of language comprehension. In B. H. Ross (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 44, pp. 35–62). Elsevier.

https://doi.org/10.1017/langcog.2026.10120 Published online by Cambridge University Press

Zwaan, R. A. (2008). Time in language, situation models, and mental simulations. *Language Learning*, 58(s1), 13–26. https://doi.org/10.1111/j.1467-9922.2008.00458.x

Zwaan, R. A., Madden, C. J., Yaxley, R. H., & Aveyard, M. E. (2004). Moving words: Dynamic representations in language comprehension. *Cognitive Science*, 28(4), 611–619. https://doi.org/10.1207/s15516709cog2804_5

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/10.1111/1467-9280.00430

Zwaan, R. A., & Taylor, L. J. (2006). Seeing, acting, understanding: Motor resonance in language comprehension. *Journal of Experimental Psychology: General*, 135(1), 1–11. https://doi.org/10.1037/0096-3445.135.1.1

Cite this article: Wang, M., & Zhao, H. (2026). Task-sensitive semantic-visual alignment of motion schema during incremental Mandarin sentence comprehension, *Language and Cognition*, 18, e71, 1–23. https://doi.org/10.1017/langcog.2026.10120

https://doi.org/10.1017/langcog.2026.10120 Published online by Cambridge University Press