


ARTICLE

# Semantic integration of multidimensional perceptual information in L1 sentence comprehension

Bing Bai , Caimei Yang\*  and Jiabao Fan

Department of English, Soochow University, China

\*Corresponding author. Email: [cmyang@suda.edu.cn](mailto:cmyang@suda.edu.cn)

(Received 17 June 2021; Revised 28 October 2021; Accepted 28 October 2021)

## Abstract

Many studies have substantiated the perceptual symbol system, which assumes a routine generation of perceptual information during language comprehension, but little is known about the processing format in which the perceptual information of different dimensions is conveyed simultaneously during sentence comprehension. The current study provides the first experimental evidence of how multidimensional perceptual information (color and shape) was processed during online sentence comprehension in Mandarin. We designed three consecutive sentence–picture verification tasks that only differed in the delay of the display of pictures preceded by declarative sentences. The processing was analyzed in three stages based on time intervals (i.e., 0ms, +750ms, +1500ms). The response accuracy and response time data were reported. The initial stage (i.e., ISI=0ms) attested the match effect of color and shape, but the simulated representation of color and shape did not interact. In the intermediate stage (i.e., ISI=750ms), the routinely simulated color and shape interacted, but the match facilitation was found only in cases where one perceptual information was in mismatch while the other was not. In the final stage (i.e., ISI=1500ms), the match facilitation of one particular perceptual property was influenced by a mismatch with the other perceptual property. These results suggested that multiple perceptual information presented simultaneously was processed in an additive manner to a large extent before entering into the final stage, where the simulated perceptual information was integrated in a multiplicative manner. The results also suggested that color and shape were comparable to object recognition when conjointly conveyed. In relation to other evidence from behavioral and event-related potential studies on sentence reading in the discussion, we subscribed to the idea that the full semantic integration became available over time.

**Keywords:** semantic integration; L1; color; shape; sentence comprehension; embodied cognition

## 1. Introduction

The mental work in representing concepts and events in high-level cognitive processes (e.g., language, memory) remains intriguing in cognitive science. Semantic

integration in language comprehension subsumes the mapping of the perceptual concepts in memory and linguistic input. A popular theoretical account of the embodied comprehension is the perpetual symbol system, which assumes that the representing and processing of linguistic expressions are grounded in a routine activation of perceptual information that is implicitly or explicitly conveyed (Barsalou, 1999). This idea has been widely substantiated by testing one particular perceptual property during language comprehension across languages, including color (e.g., for English, Connell, 2007; Connell & Lynott, 2009; Mannaert et al., 2017; Richter & Zwaan, 2009; for Dutch, de Koning et al., 2017; Huettig et al., 2020; Mannaert et al., 2021; Redmann et al., 2019), shape (e.g., for English, Briner et al., 2014; Gershkoff-Stowe & Smith, 2004; Hupp et al., 2020; Kang et al., 2020; Lincoln et al., 2007; Zeng et al., 2016; Zwaan et al., 2002; Zwaan & Yaxley, 2004; for Japanese, Sato et al., 2013), and other properties such as orientation (e.g., Holman & Gîrbă, 2019; Stanfield & Zwaan, 2001). The mental simulation of perceptual information in semantic integration has established two important findings. One is the match effect, whereby the object property matching the implied property in the linguistic input (e.g., words or sentences) can be identified faster than when they mismatch. The match effect has been widely demonstrated in the sentence–picture verification (SPV) task, where participants are asked to judge whether the perceptual information in the picture matches the simulated perceptual information (e.g., shape, size, and color) implied in the preceding sentence they had read before. The other is the difference in the match advantage of different object properties, reported in both behavioral (e.g., de Koning et al., 2016; Rommers et al., 2013; Zwaan & Pecher, 2012) and neuroimaging studies (e.g., Harris & Dux, 2005; Michie et al., 1999; Proverbio et al., 2004). However, what still remains unclear is how perceptual information of different dimensions is integrated when conjointly presented in sentences. As far as we know from the literature, following Bransford’s idea that the comprehension of linguistic input is a construction of the situation in the real or imaginary world (Bransford & Franks, 1971), Garnham and Oakhill (1992) argued that the incrementally constructed mental representation of linguistic input had close connection with the linguistic context that has been available before. This theoretical account suggests that semantic integration is incrementally constructed during sentence reading, in line with the proposal in Pickering et al. (1999) that a major constraint on the architectures and mechanisms underlying language processing is that they must support extremely incremental processing during comprehension. More recently, Richter and Zwaan (2010) have advanced two views to account for the representation of multiple perceptual information: the additive combination view, which holds that object features are represented independently, and the multiplicative combination view, which holds that representation cues are integrated interactively. The current research seeks to investigate the processing format in which Mandarin-speaking adults integrate color and shape properties conjointly conveyed during sentence comprehension in three consecutive SPV tasks.

Object properties (including color, shape, size, and orientation) are characterized by different affordance to object recognition. In the empirical philosophy tradition, a distinction was drawn between primary properties (e.g., shape, size, motion) and second properties (e.g., color) based on the interaction between human perceptual experience and the external environment (e.g., Jackson, 1977). This distinction recently has been further re-organized into intrinsic properties (e.g., color, shape, size) and extrinsic properties (e.g., orientation), since the intrinsic properties are

dominant in object recognition (de Koning et al., 2017; Harris & Dux, 2005; Tanaka et al., 2001; Zwaan & Pecher, 2012). The divergent recognition of the status of color and shape suggests a need to clarify the mental organization of different visual properties. In our view, when color and shape are conjointly conveyed during sentence comprehension, the time course of the semantic integration can indicate whether color takes primacy over shape, and whether the processing is combinative or additive. The following section starts with a review of the representation of color and shape in the embodied cognitive process.

Color differs from other object properties in visual processing and semantic representation. A line of evidence indicates that color processing is featured by higher memory workload (Aginsky & Tarr, 2000; Vandenberg & Rensink, 2003), and by faster attentional selection (Duncan, 1984; Eimer, 1996; Karayanidis & Michie, 1997; Michie et al., 1999) than other object properties in comprehension. Another line of work resolves around the match effect, but the results are mixed. Connell (2007) reported a counter-intuitive effect in an SPV task, wherein the response time was shorter in the mismatch condition than in the match condition, and the author attributed the effect to a difference in the representation of stable and unstable properties. To be specific, when the input mismatched an unstable property (e.g., color), the cognitive cost was low to ignore the already simulated property; by comparison, when the input mismatched a stable property (e.g., orientation), the processing cost was high in order to ignore the mismatch. A plausible account in our view is that the inhibitory control ability is at play in processing perceptual information in conflict with the simulated representation. In other words, the comprehender can handle object-related knowledge in a selective way, since certain features can suffice to suppress irrelevant information for decision-making. Following this vein, it can be predicted that, even though both color and shape are relevant to object recognition, the varied statuses of object properties in attentional selection determine the degree to which either color or shape is sufficient to support the decision-making. Recently, Zwaan and Pecher (2012) found the larger match effect of color than orientation and shape in object-verification tasks; de Koning et al. (2017) reported that color showed the strongest effect, while orientation showed no such effect. They found a significant correlation between the three intrinsic visual properties (color, shape, and size), in line with previous works (e.g., Rommers et al., 2013). They also found that the extrinsic visual property (i.e., orientation) was not significantly correlated with any intrinsic properties, suggesting that object properties were organized differently in mind. In contrast, Proverbio et al. (2004) found that color did not affect categorization, wherein the reaction time of color–shape pairings (identical color but different shape vs. different color and shape) was measured. These divergent results of color representation call for a re-evaluation of the mental organization of different perceptual properties, especially during real-time language comprehension on the one hand, and necessitate a well-manipulated task – by which the differences in match effects are investigated when multiple perceptual properties of an object are presented – on the other.

Moving to the special status of shape property in the retrieval of semantic memory and decision-making during decoding linguistic input, the results are far more mixed. It has been widely recognized that shape information is essential for lexical learning (Hupp et al., 2020; Landau et al., 1998; Samuelson & Smith, 2005). Zwaan et al. (2002) were among the first to have found a faster response to objects that matched the shape of the object implied by the sentence (e.g., *eagle in the sky/nest*) in an SPV task and a

picture-naming task. Given that color and shape are intrinsic properties and that both show the match effect, some researchers claim that shape is comparable to color in object identification (e.g., Biederman & Cooper, 1991; Graham & Diesendruck, 2010). Recently, in a categorization task, children aged four were less likely than adults to represent the variation of shape (Sera & Millett, 2011). In our view, these confounding results indicate that shape can not easily be considered less resource-demanding than other properties in mental simulation, and that the difference in match advantage can be testified when color and shape are conveyed conjointly. In addition, the activation of shape information is connected with the structure of input that implies the shape property explicitly or implicitly (e.g., Briner et al., 2014; van Weelden et al., 2014). For instance, the match effect was found in the left hemisphere when shape was implicitly described in a sentence (e.g., *there was a tomato on the pizza*), and in both left and right hemispheres when shape was explicitly described in a sentence (e.g., *there was a slice of tomato on the pizza*) (Briner et al., 2014). These findings suggest that the hemispheric activation during shape processing depends on how explicit information is conveyed, calling for a well-manipulated sentence structure when shape is investigated in embodied cognition.

So far, we have reviewed the theoretical recognition and the empirical evidence of color and shape in mental representation, further pondering the dynamic processing format of color and shape when they are conjointly conveyed during sentence comprehension. Up to now, behavioral data have shown that the actual state of affairs depicted in a sentence becomes available over time, and this has also been widely demonstrated by the comprehension of negation vs. affirmation (e.g., Kaup, 2001; Kaup et al., 2006; Kaup & Zwaan 2003). A often-cited study by Hasson and Glucksberg (2006) investigated how participants represented affirmative and negated assertions where metaphor was expressed (e.g., *this lawyer is / is not a shark*). Participants were asked to read affirmative and negated expressions and decide whether a given affirmative-related word (e.g., *vicious*) or negated-related word (e.g., *gentle*) matched the sentence. Three kinds of delay were created between the endpoint of the presented metaphoric sentences and the lexicons, i.e., 150ms, 500ms, and 1000ms. The result showed that both the negative and affirmative assertions facilitated access to the affirmative-related words in the early stages. The result also showed that only affirmative assertions facilitated the affirmative-related words after 1000ms delay. This study suggested a shift from affirmation to negation during the comprehension of negated metaphor. Similarly, Kaup et al. (2006) also reported a time course of sentence comprehension in an SPV task wherein the actual state of affairs depicted in negative sentences was available (e.g., *the door is not open* is finally represented as *the door is closed*). They found a variation of the match effect in different time intervals. Specifically, when the picture and sentence were presented without delay, the match advantage was found in a picture matching the context implied by both types of sentences; when the delay was 750ms between sentence and picture, the match effect was found only in pictures depicting the actual state implied in the affirmative sentences; when the delay was 1500ms, the match effect was found only in pictures depicting the actual state of the affairs occurring in the negative sentences. The above studies on the processing of negation bring insight to the dynamics of sentence comprehension, and further remind us to use well-manipulated sentence structures to investigate the match effect of different perceptual information, especially when the possible hierarchically organized perceptual information is involved during sentence comprehension. More importantly, the time

course of semantic integration has also been supported by event-related potentials (ERPs) (Lüdtke et al., 2008).

Turning to the current study, this paper first asks whether different types of perceptual information take a different status in semantic integration at the sentence level. Considering the less stable internal structure of color (e.g., saturation, brightness) (e.g., de Koning et al., 2017; Mannaert et al., 2017), it seems fair to claim volatility in the match advantage when the internal structure of a perceptual property changes. For example, Mannaert et al. (2017) did not find a match advantage of color under reduced saturation, but a match advantage at the normal level. This finding demonstrated the activation of perceptual information rich in detail, a reminder for us to compare the process of representing multiple properties that differ in the richness of detail (e.g., color vs. shape). Specifically, data are sparse on whether perceptual information rich in details is activated earlier or later or synchronously as opposed to other conjointly presented properties, for the sake of efficiency in comprehension. Further, it is also unknown whether the earlier activated information influences the incoming simulated information in the whole process. One often-cited research paper showed that certain perceptual information did not necessarily contribute to comprehension (Proverbio et al., 2004), when the reaction time of color–shape pairings (identical color but different shape vs. different color and shape) was measured in a categorization task. They found that the attentional selection of color and shape occurred in parallel though not independently, and affected the amplitude of temporal N1 and N2 components as well as that of P300. More importantly, their ERP evidence also indicated that the selection of color depended on object shape, but not vice versa. The representation asymmetry in the activation dependence between color and shape necessitates an investigation into whether the attentional selection of color and shape influence each other at different time intervals. Although Proverbio et al. provided evidence of the parallel selection in representing multiple perceptual information, their claim rests only on a categorization task. In fact, we argue that the results in Proverbio et al. are weakened by the experimental design in two respects. For one thing, they did not manipulate the typicality effect. According to the literature, whether the instance can be recognized and categorized depends on its typicality (Posner, 1970), according to which a typical item is easier to be identified as a member than an atypical item in categorization tasks. As a result, when the typicality is not manipulated, the reaction time in object identification will be influenced. For another, they did not manipulate the presentation order of the instance and category. In fact, the categorization task is influenced by the presentation order of category and instance, since both instance (e.g., banana) and category (e.g., fruit) can be cues to categorization. Collins and Quillian (1969) classified three kinds of presentation order based on the typicality effect,<sup>1</sup> and they argued that only those cases that are characterized by high instance dominance and high category dominance are independent of presentation order. Compared with categorization, sentence comprehension is characterized by far more complexity,

<sup>1</sup>In the cases where stimulus pairs were characterized by high category dominance (e.g., BUTTERFLY-INSECT), the instance was presented first. In the cases where stimulus pairs were characterized by high instance dominance (e.g., SEAFOOD-SHRIMP), the category was presented first. With regard to the measure of instance dominance and category dominance, see Loftus and Scheff (1971), and Shipiro and Palermo (1970).

since handling sentences requires the storage and manipulation of perceptual information, and the mapping of perceptual representation with linguistic input.

This study also considers whether the representation of multiple perceptual information is mechanically combined additively or integrated interactively. Either the additive combination view or the multiplicative combination view suggests that multiple perceptual information conjointly presented becomes fully integrated over time during sentence reading. Empirically, the multiplicative view predicts that the match effect is greater in those cases where both color and shape match the object, than in the cases where only dimension matches the object. For instance, according to the multiplicative view, a *banana* in the sentence *a ripe banana tastes sweet* activates the representation of yellow color and a typical shape, and afterwards both simulated perceptual activations interact and serve as strong cues in categorizing an instance of banana as a ripe banana. By comparison, according to the additive view, when encountering a *banana* in the sentence *a ripe banana tastes sweet*, the comprehender retrieves the color of yellow and a typical shape independently from one another, and are then additively combined to establish the referent. In our view, the additive view has to answer whether the first activated perceptual information benefits efficient comprehension if the activated perceptual information is combined in a linear manner (since properties are not equally sufficient for decision-making); the multiplicative view needs to answer whether particular perceptual information influences the interaction between perceptual information in different dimensions. Both issues can, in our view, be resolved by investigating the time course of representation during sentence comprehension. As a result, we are curious about how the dynamics happen during online reading according to empirical data.

Recently, only Sato et al. (2013) have investigated the dynamic representation of shape during sentence processing. Participants were given paired sentences, one with a verb that forced comprehenders to reconsider the shape, and the other with a verb that indicated a canonical shape (see their Experiment 2). Since Japanese is a verb-final language, the time course of the processing can reveal whether the mental simulation is formed incrementally along with the reading process or is integrated until the verb appears. They found that the mental simulation immediately shifted from the original according to the final sentential meaning as soon as the final verb was processed. In line with the situation model (e.g., Zwaan & Radvansky, 1998), their work suggested that all details were activated and updated along the decoding process. However, they examined only one particular perceptual dimension, leaving room for investigating when and what properties of an object are activated in the time course of online language comprehension. Of the most relevance is Richter and Zwaan (2010), where the representation of color (match vs. mismatch) and shape (match vs. mismatch) was investigated during word access. They found that responses in the three tasks (a categorization task, a lexical-decision task, and a word-naming task) were facilitated when both the shape and the color matched the actual fruits and vegetables implied by words, lending support to a combinative representation of multiple perceptual representation. However, their design in our view is weakened by the variance in the nature or the extent of visual experience with particular target objects, which can affect participants' prototypes in object recognition according to the typicality effect (e.g., Posner, 1970). Richter and Zwaan (2010) provided evidence neither for the time course of the representation of multiple perceptual dimensions, nor for the dynamic semantic integration at the sentence level.

To sum up, different object properties are very possibly hierarchically accessed in mental representation during language comprehension, and the final representation of a scenario depicted in sentences becomes available over time. In fact, more evidence is needed concerning the dynamic processing of affirmative sentences where multidimensional perceptual information is explicitly or implicitly conveyed, to elucidate the format in which the multidimensional perceptual information is represented in an additive or multiplicative manner. Thus, the main goal of this work is to compare the two alternative theoretical accounts of how perceptual information in different dimensions (i.e., color and shape) is integrated at different time intervals. Another goal is to reconsider the organization of the simulated perceptual information in object recognition during language comprehension.

## 2. Methods

As we mentioned previously, object properties are characterized by their different status in mental organization, and behavioral and event-related potentials showed that the actual state affairs depicted in sentences become available over time. Following the most relevant studies (e.g., Hasson & Glucksberg, 2006; Kaup et al., 2006; Proverbio et al., 2004; Richter & Zwaan, 2010), we designed three consecutive SPV tasks differing in the display of pictures (i.e., ISI = 0ms/750ms/1500ms). In the early stage of sentence comprehension, since time is limited for semantic processing, we predicted that the perceptual properties of an object might only activate the simulation of the corresponding perceptual dimension conveyed in the sentence. In the intermediate stage, we predicted that color and shape information might interact but get partially integrated in an incremental processing. In the final stage, we predicted that, due to sufficient time available for semantic integration, sentence–picture pairing in the match or mismatch condition could be processed thoroughly. Thus, we have two empirical questions below regarding the time course of semantic integration:

*Empirical question 1:* Are the different dimensions of perceptual information (i.e., color and shape) represented and processed parallelly or independently?

*Empirical question 2:* Does the match condition of one aspect of perceptual information (i.e., either color or shape) influence the processing of the other aspect of perceptual information?

### 2.1. Experiment 1

#### 2.1.1. Participants

Thirty-five Mandarin-speaking undergraduate students living in mainland China were recruited with their consent (mean age 23.5 years). All participants had normal or corrected-to-normal vision. None of the participants suffered from psychiatric or neurological disorders. Informed consent was obtained from each participant before the experiment. Participants confirmed their willingness for their involvement in the experiment.

#### 2.1.2. Materials

We designed sixty-six simple Chinese declarative sentences with high imagery, of which ten were used for training and the remaining were critical items.




The vocabularies used in the sentences were taken from the list of common words in the *Modern Chinese Dictionary* issued in 2016 by the Chinese Academy of Social Sciences. The sentence length was between 6 and 12 words. The comprehension difficulty of the sentences, as well as the degree to which the picture and the sentence matched or mismatched, was assessed by three linguists independently.

Initially, we prepared seventy pictures, out of which fifty-six pictures were chosen as critical items and fourteen pictures were excluded based on the rating results. We had two types of object properties (i.e., color and shape) and two conditions (i.e., match and mismatch), and thus we had four types of pictures: both color and shape matched the object (ColorMatch–ShapeMatch), the color matched but the shape mismatched (ColorMatch–ShapeMismatch), the shape matched but the color mismatched (ColorMismatch–ShapeMatch), and neither color nor shape matched (ColorMismatch–ShapeMismatch) (see an exemplar in Table 1). The picture pairing with the sentence included prototypical objects and their variants that were modified by techniques.

With regard to the four pictures corresponding to each object, we manipulated the typicality of the objects, since the variance of color and shape might influence the categorization and object identification and further influence the sentence reading time. The structure of each category was composed of distributions of attributes, and typicality was unbalanced among instances (Posner, 1970). High-typicality instances were identified faster than low-typicality instances (e.g., Glass & Holyoak, 1974; Rosch, 1973; Smith et al., 1974). Reaction time (RT) data in receptive language have revealed that the varied typicality in L1 comprehension leads to different speeds of decision-making in typicality judgments. In other words, if the typicality is not manipulated, sentence reading will be indirectly influenced due to the time variance of the object identification, particularly when multidimensional perceptual information is conveyed. Thus, thirty Mandarin-speaking undergraduates were recruited to rate the prototype object depicted in the picture.

After the rating task, we decided on the presentation order of picture and sentence, since either instance or category could be a cue to categorization. Although three types of presentation order of instance and category are articulated in Collins and

**Table 1.** An exemplar of the current 2 (Color and Shape) × 2 (Match and Mismatch) design

Color	Shape		
	Match	Mismatch	
Match			
Mismatch			
Sentence	熟透的 shutou-de ripe	香蕉 xiangjiao banana	真好吃。 zhen haochi very good taste 'The ripe banana tastes very good.'



Quillian (1969), the presentation order is independent of category and instance (e.g., ORANGE–FRUIT) in cases of high category dominance and high instance dominance. Since our stimuli are characterized by high category and high instance dominance, the presentation order is independent of category and instance.

With regard to shape modification, one prototype derived four variants by changing the prototype shape into four kinds of shapes in Adobe Photoshop. For example, the typical shape of a banana was modified into orange-shaped, winter melon-shaped, cantaloupe-shaped, and onion-shaped. The prototype and four variants were rated by another twenty undergraduates on a five-point scale (1 indicating the lowest deviation and 5 indicating the highest deviation), to make sure that the picture variant did not influence the recognition of the object. With regard to color modification, Adobe Photoshop CS6 was used to change the hue of the picture with a threshold of 60 units. After four variants were produced by deviating 60 units, 120 units, 180 units, and 240 units, respectively, both the variants and the prototype were rated in the same way as the objects in shape were rated.

The rating result showed that the average recognition accuracy of the deviated shape was 90.2%, and that the average degree to which shape deviated was 2.21. The rating result showed that the average recognition accuracy of the deviated color was 89.83%, and that the average degree to which color deviated was 2.19. The independent sample *t*-test on the degree of shape and color deviation indicated that there was no significant difference between the degree to which shape and color deviated ( $t(108) = -1.63, p > .05$ ).

Finally, the pictures required for experimental materials were constructed according to the shape and color deviation index. Fifty-six were chosen as critical items based on rating results. Experimental items were rotated across 4 lists in a Latin square design. Each list had fourteen items in four types of items (i.e., ColorMatch–ShapeMatch, ColorMatch–ShapeMismatch, ColorMismatch–ShapeMatch, and ColorMismatch–ShapeMismatch). Fifty-six sentences which mismatched the picture stimuli but had the same syntax as critical items were added as fillers.

### 2.1.3. Procedure

The experiment was a 2 (two types of perceptual information: color and shape) × 2 (two sentence–picture pairing conditions: match and mismatch) design. The reaction time to the pictures was the dependent variable. Programmed by E-prime 3.0., the experiments were conducted in a sound-attenuated lab. Participants sat in front of the screen with their eyes about 40 cm from the screen, and placed the index finger of the right hand on the ‘K’ key, and the index finger of the left hand on the ‘D’ key, and the thumbs of both hands on the ‘space’ key. Participants were randomly assigned to one of the four lists, and were tested individually. Participants pressed the ‘Q’ key to start the familiarization task, before the experiment began. After the training period, the participants felt able to join the experiment.

The experiment began with instructions in Chinese on the screen. A fixation cross appeared in the center of the screen for 1000ms, followed by the sentence (see the exemplar in Mandarin in Figure 1; the English gloss is shown in Table 1) presented in the center of the screen. The sentence disappeared immediately (i.e., ISI = 0ms) when participants pressed the space bar to indicate that they had finished reading, and then a picture appeared at the center of the screen. The task for participants was to accurately and quickly decide whether the object in the picture was mentioned in

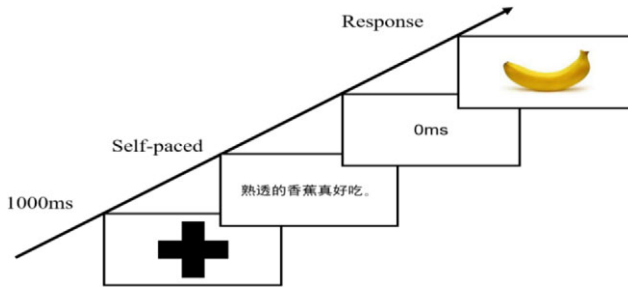


Fig. 1. The stimulus sequence within a trial of the sentence-based picture recognition task when ISI was 0ms.

the preceding sentence by pressing 'D' for YES and 'K' for NO (see Figure 1). The judgment was counterbalanced across participants. The reaction time was recorded from the appearance of the picture to the time the participants made a judgment, and was processed by R (R Core Team, 2013). The following is an example of the critical item and the presentation procedure. We recorded the accuracy data and response times. Following Baayen et al. (2008), Bates et al. (2015), and Dirix et al. (2017), the current study reported the statistical results. To be specific, the reports of main effects and interaction effects included Chisq value, Df, and *p*-value; the function of "glht" was used for post comparison, and the statistical results replaced the Estimate ( $\beta$ ), Std. Error(SE), *z*-value, and  $\text{Pr}( > |z| )$  (Dirix et al., 2017). The use format of the function of 'Anova' we used was `Anova(model, type="II")`, according to which we assessed the main effects and interaction effects.

#### 2.1.4. Accuracy data

The accuracy rate of thirty participants was higher than 80% in all four types of conditions, while data from five participants were excluded due to an accuracy rate of less than 80% in at least one of the conditions. Reaction time data exceeding  $\pm 2.5$  standard deviations were excluded. As for the thirty participants whose accuracy rates were over 80%, incorrect responses to trials were also excluded. The response accuracy of ColorMatch–ShapeMatch, ColorMatch–ShapeMismatch, ColorMismatch–ShapeMatch, and ColorMismatch–ShapeMismatch was 97.67%, 92.8%, 98.11%, and 99.03%, respectively. The mixed-effects regression model showed that the main effect of color was significant ( $\chi^2(1) = 46.02, p < .001$ ), i.e., the response was faster when the implied color matched the color in the picture. It was also observed that the main effect of shape was significant ( $\chi^2(1) = 15.98, p < .001$ ), i.e., the response was faster when the implied shape matched the shape in the picture. The results showed that the interaction between color and shape was significant ( $\chi^2(1) = 34.93, p < .001$ ).

Post-hoc comparisons showed that, when the object color in the picture matched the implied color in the sentence, the response accuracy of the shape information in the matched condition was significantly different from that in the mismatched condition ( $\beta = 4.84, SE = 0.69, z = 7.00, p < .001$ ). Specifically, the response accuracy of shape in the matched condition was significantly higher than that in the mismatched condition. In contrast, when the object color in the picture mismatched the color implied in the sentence, the response accuracy of the shape

information in the match condition was not significantly different from that in the mismatched condition ( $\beta=0.93$ ,  $SE=0.69$ ,  $z=1.35$ ,  $p=.53$ ). These results suggested that the match condition of color improved the processing accuracy of shape, and that the mismatch condition of color did not influence the processing accuracy of shape.

By comparison, post-hoc comparisons showed that, when the object shape in the picture matched the implied shape in the sentence, the response accuracy of the color information in the matched condition was not significantly different from that in the mismatched condition ( $\beta=0.43$ ,  $SE=0.69$ ,  $z=0.62$ ,  $p=.93$ ). In contrast, when the object shape in the picture mismatched the shape implied in the sentence, the response accuracy of the color information in the match condition was significantly different from that in the mismatched condition ( $\beta=6.20$ ,  $SE=0.69$ ,  $z=8.98$ ,  $p < .001$ ). Specifically, the response accuracy of color in the matched condition was significantly higher than that in the mismatched condition. These results suggested that the match condition of shape did not influence the processing accuracy of color, and that the mismatch condition of shape degraded the processing accuracy of color.

#### 2.1.5. Response time

The mean value and standard error of response time in four conditions when ISI was 0ms were presented in Table 2. A mixed-effects regression model was used to fit the reaction time of participants, and the  $\beta$ -value was reported. The reaction time was analyzed by lme4-package LMER function (Bates et al., 2015). When we examined the stochastic structure of the mixed model, we also considered the random slope and random intercept of the participants, as suggested in Barr et al. (2013). The fitted regression model for the mixed effects was *logged RTs~color+shape + (1 | participants) + (1 | items)*, SPdata) (the base of the log was 10). The consistency of color and shape with the sentence was a fixed factor, and participants and items were random factors. Orthogonal coding was used for the two independent variables, and ANOVA (type= III) from the car package was used to determine the main effect and the interaction effect.

The mixed-effects model showed a significant main effect of color information ( $\chi^2(1) = 217.30$ ,  $p < .001$ ), as well as a match effect, i.e., responses were faster when the implied color in the sentence matched the color in the picture than when they mismatched. The main effect of shape information was also significant ( $\chi^2(1) = 86.96$ ,  $p < .001$ ), and subjects also showed a match effect, i.e., responses were faster when implied shape in the sentence matched the shape in the picture than when they mismatched. There was no interaction between color information and shape information ( $\chi^2(1) = 0.70$ ,  $p=.40$ ). Overall, these results support the embodied view on language comprehension, but the activation of the multiple perceptual dimensions has not yet interacted at the very early stage of semantic integration.

**Table 2.** ISI=0ms: Chinese speakers' Mean and SE of RTs in the four conditions

	ColorMatch- ShapeMatch	ColorMatch- ShapeMismatch	ColorMismatch- ShapeMatch	ColorMismatch- ShapeMismatch
Mean	953	1189	1321	1541
SE	27	26	30	30

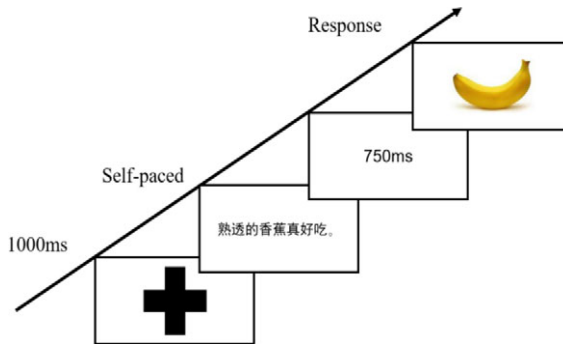


Fig. 2. The stimulus sequence within a trial of the sentence-based picture recognition task when ISI was 750ms.

## 2.2. Experiment 2

### 2.2.1. Participants

We recruited thirty-one Mandarin-speaking undergraduates with consent (mean age 22.4 years) who had not been recruited for Experiment 1. All participants had normal or corrected-to-normal vision. None of the participants suffered from psychiatric or neurological disorders. Informed consent was obtained from each participant before the experiment. Participants confirmed their willingness for their involvement in the experiment.

### 2.2.2. Materials and procedure

Materials, as well as the experimental design, were the same as those in Experiment 1. The only difference was that pictures appeared 750ms later when the participants pressed the space bar to indicate that they had finished reading (see Figure 2 as a demonstration).

### 2.2.3. Accuracy data

Data of two participants were excluded. The response accuracy of ColorMatch–ShapeMatch, ColorMatch–ShapeMismatch, ColorMismatch–ShapeMatch, and ColorMismatch–ShapeMismatch was 97.53%, 93.46%, 98.02%, and 99.10%, respectively. The mixed-effects regression model showed that the main effect of color was significant ( $\chi^2(1) = 49.37, p < .001$ ), i.e., the response was faster when the implied color matched the color in the picture. It was also observed that the main effect of shape was significant ( $\chi^2(1) = 11.70, p < .001$ ), i.e., the response was faster when the implied shape matched the shape in the picture. The results showed that the interaction between color and shape was significant ( $\chi^2(1) = 34.79, p < .001$ ).

Post-hoc comparisons showed that, when the object color in the picture matched the implied color in the sentence, the response accuracy of the shape information in the matched condition was significantly different from that in the mismatched condition ( $\beta = 4.07, SE = 0.62, z = 6.59, p < .001$ ). Specifically, the response accuracy of shape in the matched condition was significantly higher than that in the mismatched condition. In contrast, when the object color in the picture mismatched the color implied in the sentence, the response accuracy of the shape information in the match condition was not significantly different from that in the mismatched

condition ( $\beta=1.08$ ,  $SE=0.62$ ,  $z=1.75$ ,  $p=0.30$ ). These results replicated the results in Experiment 1, i.e., the match condition of color improved the processing accuracy of shape, and the mismatch condition of color did not influence the processing accuracy of shape.

By comparison, post-hoc comparisons showed that when the object shape in the picture matched the implied shape in the sentence, the response accuracy of the color information in the matched condition was not significantly different from that in the mismatched condition ( $\beta=0.49$ ,  $SE=0.62$ ,  $z=0.80$ ,  $p=.86$ ). In contrast, when the object shape in the picture mismatched the shape implied in the sentence, the response accuracy of the color information in the match condition was significantly different from that in the mismatched condition ( $\beta=5.65$ ,  $SE=0.62$ ,  $z=9.14$ ,  $p<.001$ ). Specifically, the response accuracy of color in the matched condition was significantly higher than that in the mismatched condition. These results were a replication of those in Experiment 1, i.e., the match condition of shape did not influence the processing accuracy of color, and the mismatch condition of shape degraded the processing accuracy of color.

#### 2.2.4. Response time

The mean and standard error of response time in the four conditions when ISI was 750ms are presented in Table 3. The fitted regression model for the mixed effects was  $\logged RTs \sim color * shape + (1 | participants) + (1 | items)$ ,  $SPdata$  (the base of log was 10). The mixed-effects model showed a significant main effect of color information ( $\chi^2(1) = 97.66$ ,  $p < .001$ ), as well as a match effect, i.e., responses were faster when implied color implied in the sentence matched the color in the picture than when they mismatched. The main effect of shape information was also significant ( $\chi^2(1) = 106.47$ ,  $p < .001$ ), and subjects also showed a match effect, i.e., responses were faster when implied shape in the sentence matched the shape in the picture than when they mismatched. There was a significant interaction effect between color information and shape information ( $\chi^2(1) = 6.51$ ,  $p=.01$ ).

Post-hoc comparisons showed that, when the object color in the picture matched the implied color in the sentence, the response time to the shape information in the match condition was not significantly faster than when they mismatched ( $\beta=0.002$ ,  $SE=0.01$ ,  $z=0.42$ ,  $p=.12$ ). By contrast, when the object color in the picture mismatched the color implied in the sentence, the response time to shape information in the match condition was significantly faster than when they mismatched ( $\beta=0.07$ ,  $SE=0.01$ ,  $z=5.49$ ,  $p < .001$ ). In addition, when the object shape in the picture matched the implied shape in the sentence, the response to color in the match condition was not significantly faster than when they mismatched ( $\beta=0.009$ ,  $SE=0.01$ ,  $z=0.59$ ,  $p=.18$ ). By comparison, when the object shape in the picture mismatched the implied shape in the sentence, the response time to color information in the match condition

**Table 3.** ISI=750ms: Chinese speakers' Mean and SE of RTs in the four conditions

	ColorMatch- ShapeMatch	ColorMatch- ShapeMismatch	ColorMismatch- ShapeMatch	ColorMismatch- ShapeMismatch
Mean	1294	1331	1340	1551
SE	48	49	63	59

was significantly faster than when they mismatched ( $\beta=0.07$ ,  $SE=0.01$ ,  $z=5.18$ ,  $p < .001$ ).

Overall, Experiment 2 replicated the match effect for color and shape reported in Experiment 1; the results in Experiment 2 indicated that color and shape partially interacted, and that the identification of one property was not necessarily influenced by the other in both conditions. These results also suggested that both color and shape were comparable when conjointly presented in real-time comprehension.

### 2.3. Experiment 3

#### 2.3.1. Participants

We recruited thirty-four Mandarin-speaking undergraduates with consent (mean age 22.7 years) who did not participate in Experiment 1 or Experiment 2. All participants had normal or corrected-to-normal vision. None of the participants suffered from psychiatric or neurological disorders. Informed consent was obtained from each participant before the experiment. Participants confirmed their willingness for their involvement in the experiment.

#### 2.3.2. Materials and procedure

Materials, as well as the experimental design, were the same as those in Experiment 1 and Experiment 2. The only difference was that the picture appeared 1500ms later when participants pressed the space bar to indicate that they had finished reading (see Figure 3 as a demonstration).

#### 2.3.3. Accuracy data

Four participants whose data was below 80% in one of the four conditions were excluded. The response accuracy of ColorMatch–ShapeMatch, ColorMatch–ShapeMismatch, ColorMismatch–ShapeMatch, and ColorMismatch–ShapeMismatch was 97.53% , 93.46% , 98.02%, and 99.10%, respectively. The mixed-effects regression model showed that the main effect of color was significant ( $\chi^2(1) = 58.78$ ,  $p < .001$ ), i.e., the response was faster when the implied color matched the color in the picture.

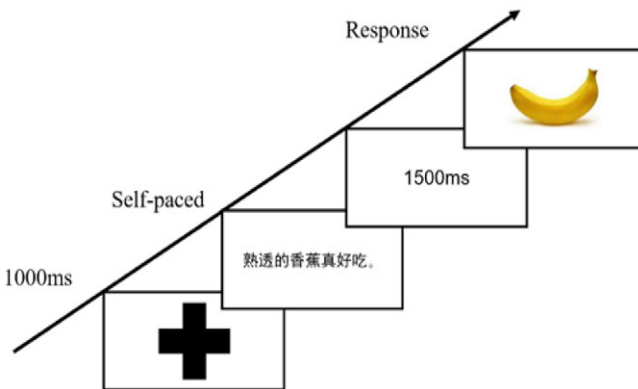


Fig. 3. The stimulus sequence within a trial of the sentence-based picture recognition task when ISI was 1500ms.

It was also observed that the main effect of shape was significant ( $\chi^2(1) = 14.33, p < .001$ ), i.e., the response was faster when the implied shape matched the shape in the picture. The results showed that the interaction between color and shape was significant ( $\chi^2(1) = 39.64, p < .001$ ).

Post-hoc comparisons showed that, when the object color in the picture matched the implied color in the sentence, the response accuracy of the shape information in the matched condition was significantly different from that in the mismatched condition ( $\beta=4.07, SE=0.57, z=7.13, p < .001$ ). Specifically, the response accuracy of shape in the matched condition was significantly higher than that in the mismatched condition. In contrast, when the object color in the picture mismatched the color implied in the sentence, the response accuracy of the shape information in the match condition was not significantly different from that in the mismatched condition ( $\beta=1.10, SE=0.57, z=1.78, p=.29$ ). These results again replicated the results in Experiment 1, i.e., the match condition of color improved the processing accuracy of shape, and the mismatch condition of color did not influence the processing accuracy of shape.

By comparison, post-hoc comparisons showed that, when the object shape in the picture matched the implied shape in the sentence, the response accuracy of the color information in the matched condition was not significantly different from that in the mismatched condition ( $\beta=0.55, SE=0.57, z=0.97, p=.77$ ). In contrast, when the object shape in the picture mismatched the shape implied in the sentence, the response accuracy of the color information in the match condition was significantly different from that in the mismatched condition ( $\beta=5.64, SE=0.57, z=9.87, p < .001$ ). Specifically, the response accuracy of color in the matched condition was significantly higher than that in the mismatched condition. These results again were a replication of Experiment 1, i.e., the match condition of shape did not influence the processing accuracy of color, and the mismatch condition of shape degraded the processing accuracy of color.

#### 2.3.4. Response time

The mean and standard error of response time in the four conditions when ISI was 1500ms were presented in Table 4. The fitted regression model for the mixed effects was  $\text{logged RTs} \sim \text{color} * \text{shape} + (1 | \text{participants}) + (1 | \text{items}), \text{SPdata}$  (the base of log was 10). The mixed-effects model showed a significant main effect of color information ( $\chi^2(1) = 405.26, p < .001$ ), as well as a match effect, i.e., responses were faster when the implied color in the sentence matched the color in the picture than when they mismatched. The main effect of shape information was also significant ( $\chi^2(1) = 230.97, p < .001$ ), and subjects also showed a match effect, i.e., responses were faster when the implied shape in the sentence matched the shape in the picture than when they mismatched. There was a significant interaction effect between color information and shape information ( $\chi^2(1) = 8.84, p=.03$ ).

**Table 4.** ISI=1500ms: Chinese speakers' Mean and SE of RTs in the four conditions

	ColorMatch- ShapeMatch	ColorMatch- ShapeMismatch	ColorMismatch- ShapeMatch	ColorMismatch- ShapeMismatch
Mean	937	1301	1426	1774
SE	31	39	37	44

**Table 5.** ISI=0ms/750ms/1500ms: Chinese speakers' Mean and SE of RTs in the four conditions

	Experiment 1 (0ms)		Experiment 2 (750ms)		Experiment 3 (1500ms)	
	Mean	SE	Mean	SE	Mean	SE
ColorMatch–ShapeMatch	953	27	1294	48	937	31
ColorMatch–ShapeMismatch	1189	26	1331	49	1301	39
ColorMismatch–ShapeMatch	1321	30	1340	63	1426	37
ColorMismatch–ShapeMismatch	1541	30	1551	59	1774	44

Post-hoc comparisons showed that, when the object color in the picture matched the implied color in the sentence, the response time to the shape information in the match condition was significantly faster than that in the mismatch condition ( $\beta=0.14$ ,  $SE=0.01$ ,  $z=12.85$ ,  $p<.001$ ). Similarly, when the object color in the picture mismatched the color implied in the sentence, the response to shape information in the match condition was significantly faster than that in the mismatch condition ( $\beta=0.10$ ,  $SE=0.01$ ,  $z=8.64$ ,  $p<.001$ ). In addition, when the object shape in the picture matched the implied shape in the sentence, the response to color information in the match condition was significantly faster than that in the mismatch condition ( $\beta=0.18$ ,  $SE=0.01$ ,  $z=16.34$ ,  $p<.001$ ). Similarly, when the object shape in the picture mismatched the implied shape in the sentence, the response to color information in the match condition was significantly faster than that in the mismatch condition ( $\beta=0.14$ ,  $SE=0.01$ ,  $z=12.13$ ,  $p<.001$ ).

Now we included the response time in four conditions to compare the response time in the consecutive tasks (Table 5). Taken together, responses were faster in the cases where both color and shape matched the object than in the cases where only dimension matched the object, suggesting that the selection of shape and color was not linear but multiplicative. The RTs also showed that the mental representation of color and shape became fully integrated in the final stage of online comprehension. The accuracy data in three stages showed that the match condition of color significantly improved the processing accuracy of shape. The accuracy data in the beginning and intermediate stage shown that the mismatched shape significantly degraded the processing of color, but that the mismatched color did not affect the processing of shape. However, our data substantially evidenced neither which type of perceptual information was more important at the display delay of 1500ms, nor the equal influence exerted by color and shape, but told us only that the activated simulations reached full semantic integration over time.

### 3. Discussion

The current work first compared the multiplicative and additive views on how multiple perceptual information was represented during real-time sentence comprehension, and reminded us to re-evaluate the mental organization of color and shape when they were conjointly conveyed. Our results showed that the multiplicative integration of multiple perceptual information was apparent in the final stage but less apparent in the intermediate stage, and that color and shape were not differentiated in object identification regarding the processing advantage when conjointly conveyed.



First, compared with the behavioral data on sentence reading that requires the construction of a mental and situational model, the current work has enriched the embodied cognition of color and shape in sentence reading in Mandarin. Compared with previous studies on the representation of perceptual information, the current study elucidated the dynamics in processing color and shape conjointly conveyed at the sentence level. Contrary to the finding of only perceptual information (i.e., shape) in Zwaan et al. (2002), the current data supported the match effect of shape reported in many studies (e.g., Briner et al., 2014; Hupp et al., 2020; Kang et al., 2020; Sato et al., 2013; van Weelden et al., 2014), and the match effect of shape that is conjointly conveyed with orientation (e.g., Pecher et al., 2009; Rommers et al., 2013) across languages. Our findings also ran counter to Connell (2007), where no match effect of color was reported, but accorded with many works on only color (e.g., Berndt et al., 2020; Connell & Lynott, 2009; Mannaert et al., 2021; Redmann et al., 2019), as well as color and other object properties conjointly conveyed (e.g., de Koning et al., 2017; Richter & Zwaan, 2010; Zwaan & Pecher, 2012). By manipulating the typicality of object features (i.e., color and shape) our design following the the offline data at the word level (e.g., Richter & Zwaan, 2010) again lent support to the combinative view accounting for the processing of multiple perceptual information. Comparing the time course of processing reported in shape representation in Japanese (e.g., Sato et al., 2013) and in the current study, we argued that, although Japanese and Mandarin are different in word order that influences the incremental semantic integration, sentence reading at the final point can reach full semantic integration, in line with the incremental comprehension reported across Indo-European languages.

More importantly, our behavioral data also elucidated the dynamic processing of multidimensional information based on empirical data. This dynamic representation can be in relation to the event-related potential studies on the selection of color and other features (e.g., orientation, color, and size), wherein the timing of attentional modulation for such selective processing is over time (Anllo-Vento & Hillyard, 1996; Karayanidis & Michie, 1997; Zani & Proverbio, 1995). The theoretical account claimed that the perceptual information from different senses can be processed in an additive manner or a multireplicative manner (Richter & Zwaan, 2010), but little is known about dynamic representation during sentence reading. The current study elucidated the processing manner in three stages. At the initial stage, although the processing of both color and shape showed a match effect, there was no interaction between the perceptual information conjointly conveyed. The match effect we found strongly substantiated the embodied language comprehension, in line with the previous work (e.g., de Koning et al., 2017; Richter & Zwaan, 2010; Rommers et al., 2013; Sato et al., 2013; Zwaan & Pecher, 2012), but the match effect of color ran counter to the shorter response time in an SPV task, where people's responses were faster when the color mismatched the implied color than when they matched (Connell, 2007). The distinction does not hold between color and shape during object recognition, according to the empirical philosophical view (i.e., shape is a primary property but color is a secondary property). Instead, we argued that color and shape were comparable, at least in the initial stage of perceptual simulation, in contrast to previous works (e.g., de Koning et al., 2017; Rommers et al., 2013; Tanaka et al., 2001).

In order to address whether the match or mismatch referent of one particular property in color–shape pairing can influence the processing of the other (e.g.,

when color information in the sentences matches the color in the picture, whether the match effect of shape can be found), we designed Experiments 2 and 3. The results showed that the activation of color and shape interacted partially at the intermediate stage, wherein one of the perceptual representations in the match-condition will cause a significant match facilitation in the other perceptual information. Thus, we proposed a plausible process of object recognition: object-related knowledge (e.g., visual properties; tactile properties, etc.) can be automatically retrieved; the comprehender automatically filters the perceptual information mismatching the input, as long as the other perceptual dimensions satisfy the match effect. In other words, although successful language comprehension counts on simulation of all perceptual information assembled in the situation model, whole objects can be identified even when one visual attribute sufficiently meets the object identification. At the cognitive level, this process depends on inhibitory control, by which the comprehenders suppress information in conflict with other possible referents, and the automatic attentional inhibition ability helps resist interference (e.g., Nigg, 2017; Stahl et al., 2014), i.e., the comprehender relies on only a certain feature to fulfill object identification. This proposal is also supported by well-cited neuroimaging evidence that object recognition is supported by automatic activation of object-related knowledge (e.g., Gerlach et al., 2002; O'Craven et al., 1999). In addition, the dominant role of either color or shape, as well as the mutual correlation between shape and color, is not confirmed. As reported, when color mismatched the simulated color in the sentence, the difference in response time for shape in the match condition was not significantly faster than when they mismatched, but when the mismatched shape definitely degraded the processing of color. Thus, it is hard to conclude whether one of the pairing properties depends on the other in the intermediate stage, in contrast to the ERP evidence in a categorization task that the selection of color depends on object shape, but not vice versa (Proverbio et al., 2004). Thus, it seems fair to support the idea that both color and shape are comparable in object representation.

By contrast, the ending stage (i.e., ISI=1500ms) of semantic integration showed robust evidence of a multiplicative manner for representing perceptual information in different dimensions. According to the results, the match or mismatch in one dimension influenced the process by which the other dimension gets integrated (e.g., the initial shape activation would not become deactivated in a mental simulation whenever the color matched or mismatched the implied color), and vice versa. For one thing, this result indicates that the simulated color and shape information interact, and that either color or shape mismatching the implied perceptual dimension would cause an interference with comprehension. For another, results suggest that mental simulations can be actively updated to map the unfolding language when entering the final stage, in line with results in online SPV tasks (e.g., Sato et al., 2013) and offline SPV tasks (e.g., Mannaert et al., 2019). The match effect we found in the final stage suggests that the prior activated information remains active and interacts with other simulated representations. Recent studies on one particular perceptual property on the contrary found that the initial activation would be deactivated and only the final simulation stayed active (e.g., Kang et al., 2020; Mannaert et al., 2019). For example, the match facilitation is not found at 1500ms in affirmative syntax, and not at 750ms in negative syntax (Kaup et al., 2006). We assume these divergent findings may be due to the syntax itself. In addition, no evidence showed that either

color or shape was dominant in the final stage of semantic integration, during which participants had enough cognitive resources to handle the perceptual information.

In addition, our findings also enriched the evidence of how the actual scenario became available over time by examining the multidimensional perceptual information in Mandarin, in line with previous works on the processing of negation and metaphor in Indo-European languages, lending support to embodied cognition in incremental sentence comprehension. Many behavioral and event-related potential studies have contended that to understand negation in English we have to shift from a factual situation to a counterfactual situation according to the online reading data (e.g., Giora et al., 2005; Hasson & Gluckberg, 2006; Kaup, 2001; Kaup et al., 2006; Lüdtke et al., 2008), but these studies only evidenced the dynamic comprehension of factual vs. counterfactual situations. For example, in Lüdtke et al. (2008) participants were presented with an affirmative or a negative sentence (e.g., *in the front of the tower there is a/no ghost*) followed by a matching or mismatching picture that was presented after a delay of 250ms or 1500ms. When the delay was 250ms, verification latencies and ERPs showed a priming effect independent of whether the sentence contained a negation or not. By comparison, when the delay was 1500ms, they observed a main effect of truth value and negation in addition to the priming effect already in the N400 time window. These results suggested that negation was fully integrated into sentence meaning only at a later point in the comprehension process. The current work, based on comprehension of declarative sentences where color and shape were conveyed, again indicated that language comprehension uses a mental model which guides interpretation of the sentences, and controls inference making, in line with many theoretical accounts claiming that language comprehension involves construction of the situational or perceptual information conveyed (e.g., Barsalou, 1999; Bransford & Franks, 1971; Glenberg et al., 1987; Zwaan & Radvansky, 1998). In other words, information from different senses is integrated interactively at the final stage where an event described by a sentence thoroughly maps with the activated perceptual information.

To conclude, we did not find the multiplicative effect throughout the process of representing color and shape conjointly conveyed in sentence comprehension, but we found a multiplicative effect in the final stage. The processing advantage of color and other intrinsic object properties can be further clarified by other approaches, such as event-related potentials, which is highly suggestive of covert steps in information processing and decision-making. Further inquiries about the perceptual symbol system also lie in testing the efficiency of semantic integration during L2 semantic development across different proficiency levels, i.e., to what extent L2 learners can automatically integrate one particular or even multiple types of perceptual information at both the word and sentence level. Future research can also compare the processing of multiple perceptual information conjointly conveyed in relative clauses in head-final and head-initial languages, since the head-initial languages may initiate the whole processing of the incoming sequences based on partial semantic retrieval.

**Acknowledgments.** We are thankful to two anonymous reviewers who helped us improve this paper.

**Funding statement.** This study is supported by the Postgraduate Research & Practice Innovation Program of Jiangsu Province (Grant No. KYCX 20\_2642).

## References

- Aginsky, V. & Tarr, M. J. (2000). How are different properties of a scene encoded in visual memory? *Visual Cognition* 7(1), 147–162.
- Anllo-Vento, L. & Hillyard, S. A. (1996). Selective attention to the color and direction of moving stimuli: electrophysiological correlates of hierarchical feature selection. *Perception & Psychophysics* 58(2), 191–206.
- Baayen R. H., Davidson D. J. & Bates D. M., (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* 59(4), 390–412.
- Barr, D. J., Levy, R., Scheepers, C. & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: keep it maximal. *Journal of Memory and Language* 68(3), 255–278.
- Barsalou, L. W. (1999). Perceptions of perceptual symbols. *Behavioral and Brain Sciences* 22(4), 637–660.
- Bates, D., Kliegl, R., Vasishth, S. & Baayen, H. (2015). Parsimonious mixed models. Online <<http://arxiv.org/abs/1506.04967>>.
- Berndt, E., Dudschig, C. & Kaup, B. (2020). Green as a cbemcuru: modal as well as amodal color cues can help to solve anagrams. *Psychological Research* 84(2), 491–501.
- Biederman, I. & Cooper, E. E. (1991). Priming contour-deleted images: evidence for intermediate representations in visual object recognition. *Cognitive Psychology* 23(3), 393–419.
- Bransford, J. D. & Franks, J. J. (1971). The abstraction of linguistic ideas. *Cognitive Psychology* 2(4), 331–350.
- Briner, S. W., Virtue, S. M. & Schutzenhofer, M. C. (2014). Hemispheric processing of mental representations during text comprehension: evidence for inhibition of inconsistent shape information. *Neuropsychologia* 61, 96–104.
- Collins, A. M. & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior* 8(2), 240–247.
- Connell, L. (2007). Representing object colour in language comprehension. *Cognition* 102(3), 476–485.
- Connell, L. & Lynott, D. (2009). Is a bear white in the woods? Parallel representation of implied object color during language comprehension. *Psychonomic Bulletin and Review* 16(3), 573–577.
- de Koning, B. B., Wassenburg, S. I., Bos, L. T. & van der Schoot, M. (2016). Size does matter: implied object size is mentally simulated during language comprehension. *Discourse Processes* 54(7), 493–503.
- 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.
- Dirix, N., Cop, U., Drieghe, D. & Duyck, W. (2017). Cross-lingual neighborhood effects in generalized lexical decision and natural reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 43(6), 887–915.
- Duncan, J. (1984). Selective attention and the organization of visual information. *Journal of Experimental Psychology: General* 113(4), 501–517.
- Eimer, M. (1996). The N2pc component as an indicator of attentional selectivity. *Electroencephalography and Clinical Neurophysiology* 99(3), 225–234.
- Garnham, A. & Oakhill, J. (1992). Discourse processing and text representation from a ‘Mental Models’ perspective. *Language and Cognitive Processes* 7(3/4), 193–204.
- Gerlach, C., Aaside, C. T., Humphreys, G. W., Gade, A., Paulson, O. B. & Law, I. (2002). Brain activity related to integrative processes in visual object recognition: bottom-up integration and the modulatory influence of stored knowledge. *Neuropsychologia* 40(8), 1254–1267.
- Gershkoff-Stowe, L. & Smith, L. B. (2004). Shape and the first hundred words. *Child Development* 75(4), 1–17.
- Giora, R., Fein, O., Ganzi, J., Levi, N. A. & Sabah, H. (2005). On negation as mitigation: the case of negative irony. *Discourse Processes* 39(1), 81–100.
- Glass, A. L. & Holyoak, K. J. (1974). Alternative conceptions of semantic theory. *Cognition* 3(4), 313–339.
- Glenberg, A. M., Meyer, M. & Lindem, K. (1987). Mental models contribute to foregrounding during text comprehension. *Journal of Memory and Language* 26(1), 69–83.
- Graham, S. A. & Diesendruck, G. (2010). Fifteen-month-old infants attend to shape over other perceptual properties in an induction task. *Cognitive Development* 25(2), 111–123.
- Harris, I. M. & Dux, P. E. (2005). Orientation-invariant object recognition: evidence from repetition blindness. *Cognition* 95(1), 73–93.

- Hasson, U. & Glucksberg, S. (2006). Does understanding negation entail affirmation? *Journal of Pragmatics* 38(7), 1015–1032.
- Holman, A. C. & Gírbá, A. E. (2019). The match in orientation between verbal context and object accelerates change detection. *Psihologija* 52(1), 93–105.
- Huetig, F., Guerra, E. & Helo, A. (2020). Towards understanding the task dependency of embodied language processing: the influence of colour during language-vision interactions. *Journal of Cognition* 3(1), 41.
- Hupp, J. M., Jungers, M. K., Porter, B. L. & Plunkett, B. A. (2020). The implied shape of an object in adults' and children's visual representations. *Journal of Cognition and Development* 21(3), 368–382.
- Jackson, F. (1977). *Perception: a representative theory*. Cambridge: Cambridge University Press.
- Kang, X., Joergensen, G. H. & Altmann, G. T. (2020). The activation of object-state representations during online language comprehension. *Acta Psychologica* 210, 103–162.
- Karayanidis, F. & Michie, P. T. (1997). Evidence of visual processing negativity with attention to orientation and color in central space. *Electroencephalography and Clinical Neurophysiology* 103(2), 282–297.
- Kaup, B. (2001). Negation and its impact on the accessibility of text information. *Memory and Cognition* 29(7), 960–967.
- Kaup, B., Lüdtke, J. & Zwaan, R. A. (2006). Processing negated sentences with contradictory predicates: Is a door that is not open mentally closed? *Journal of Pragmatics* 38(7), 1033–1050.
- Kaup, B. & Zwaan, R. A. (2003). Effects of negation and situational presence on the accessibility of text information. *Journal of Experimental Psychology: Learning, Memory and Cognition* 29(3), 439–446.
- Landau, B., Smith, L. & Jones, S. (1998). Object shape, object function, and object name. *Journal of Memory and Language* 38(1), 1–27.
- Lincoln, A. E., Long, D. L. & Baynes, K. (2007). Hemispheric differences in the activation of perceptual information during sentence comprehension. *Neuropsychologia* 45(2), 397–405.
- Loftus, E. F. & Scheff, R. W. (1971). Categorization norms for fifty representative instances. *Journal of Experimental Psychology* 91(2), 355–364.
- Lüdtke, J., Friedrich, C. K., De Filippis, M. & Kaup, B. (2008). Event-related potential correlates of negation in a sentence-picture verification paradigm. *Journal of Cognitive Neuroscience* 20(8), 1355–1370.
- Mannaert, L. N. H., Dijkstra, K. & Zwaan, R. A. (2017). Is color an integral part of a rich mental simulation? *Memory and Cognition* 45(6), 974–982.
- Mannaert, L. N. H., Dijkstra, K. & Zwaan, R. A. (2019). How are mental simulations updated across sentences? *Memory and Cognition* 47(6), 1201–1214.
- Mannaert, L. N. H., Dijkstra, K. & Zwaan, R. A. (2021). Is color continuously activated in mental simulations across a broader discourse context? *Memory and Cognition* 49(1), 127–147.
- Michie, P. T., Karayanidis, F., Smith, G. L., Barrett, N. A., Large, M. M., O'Sullivan, B. T. & Kavanagh, D. J. (1999). An exploration of varieties of visual attention: ERP findings. *Cognitive Brain Research* 7(4), 419–450.
- Nigg, J. T. (2017). Annual research review: on the relations among self-regulation, self-control, executive functioning, effortful control, cognitive control, impulsivity, risk-taking, and inhibition for developmental psychopathology. *Journal of Child Psychology and Psychiatry* 58(4), 361–383.
- O'Craven, K. M., Downing, P. E. & Kanwisher, N. (1999). fMRI evidence for objects as the units of attentional selection. *Nature* 401(6753), 584–587.
- Pecher, D., van Dantzig, S., Zwaan, R. A. & Zeelenberg, R. (2009). Language comprehenders retain implied shape and orientation of objects. *Quarterly Journal of Experimental Psychology* 62(6), 1108–1114.
- Pickering, M., Crocker, M. W. & Clifton, C. (1999). Architectures and mechanisms in sentence comprehension. In M. Pickering, M. W. Crocker & C. Clifton (eds), *Architectures and mechanisms for language processing* (pp. 1–28). Cambridge: Cambridge University Press.
- Posner, M. I. (1970). Abstraction and the process of recognition. In G. H. Bower & J. T. Spence (eds), *Psychology of Learning and Motivation* (Vol. 3, pp. 43–100). New York: Academic Press.
- Proverbio, A. M., Burco, F., del Zotto, M. & Zani, A. (2004). Blue piglets? Electrophysiological evidence for the primacy of shape over color in object recognition. *Cognitive Brain Research* 18(3), 288–300.
- Redmann, A., FitzPatrick, I. & Indefrey, P. (2019). The time course of colour congruency effects in picture naming. *Acta Psychologica* 196, 96–108.
- Richter, T. & Zwaan, R. A. (2009). Processing of color words activates color representations. *Cognition* 111(3), 383–389.

- Richter, T. & Zwaan, R. A. (2010). Integration of perceptual information in word access. *Quarterly Journal of Experimental Psychology* 63(1), 81–107.
- 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.
- Rosch, E. H. (1973). Natural categories. *Cognitive Psychology* 4(3), 328–350.
- R Core Team (2013). R: a Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna. Online <<http://www.R-project.org/>>.
- Samuelson, L. K. & Smith, L. B. (2005). They call it like they see it: spontaneous naming and attention to shape. *Developmental Science* 8(2), 182–198.
- 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.
- Sera, M. D. & Millett, K. G. (2011). Developmental differences in shape processing. *Cognitive Development* 26(1), 40–56.
- Shipiro, S. I. & Palermo, D. S. (1970). Conceptual organization and class membership: normative data for representatives of 100 categories. *Psychonomic Monograph Supplements* 3(11), 107–127.
- Smith, E. E., Shoben, E. J. & Rips, L. J. (1974). Structure and process in semantic memory: a featural model for semantic decisions. *Psychological Review* 81(3), 214–241.
- Stahl, C., Voss, A., Schmitz, F., Nuszbaum, M., Tüscher, O., Lieb, K. & Klauer, K. C. (2014). Behavioral components of impulsivity. *Journal of Experimental Psychology* 143(2), 850–886.
- 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.
- Tanaka, J., Weiskopf, D. & Williams, P. (2001). The role of color in high-level vision. *Trends in Cognitive Sciences* 5(5), 211–215.
- Vandenbeld, L. A. & Rensink, R. A. (2003). The decay characteristics of size, color, and shape information in visual short-term memory. *Journal of Vision* 3(9), 682.
- van Weelden, L., Schilperoord, J. & Maes, A. (2014). Evidence for the role of shape in mental representations of similes. *Cognitive Science* 38(2), 303–321.
- Zani, A. & Proverbio, A. M. (1995). ERP signs of early selective attention effects to check size. *Electroencephalography and Clinical Neurophysiology* 95(4), 277–292.
- Zeng, T., Zheng, L. & Mo, L. (2016). Shape representation of word was automatically activated in the encoding phase. *PloS One* 11(10), e0165534.
- Zwaan, R. A. & Pecher, D. (2012). Revisiting mental simulation in language comprehension: six replication attempts. *Plos One* 7(12), e51382.
- Zwaan, R. A. & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin* 123(2), 162–185.
- Zwaan, R. A., Stanfield, R. A. & Yaxley, R. H. (2002). Language comprehenders mentally represent the shapes of objects. *Psychological Science* 13(2), 168–171.
- Zwaan, R. A. & Yaxley, R. H. (2004). Lateralization of object-shape information in semantic processing. *Cognition* 94(2), B35–B43.

---

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