Sensory Constraints on Perceptual Simulation During Sentence Reading

Xuefei Gao
Max Planck Institute for Psycholinguistics

Ting Jiang
Beijing Normal University

Resource-constrained models of language processing predict that perceptual simulation during language understanding would be compromised by sensory limitations (such as reading text in unfamiliar/difficult font), whereas strong versions of embodied theories of language would predict that simulating perceptual symbols in language would not be impaired even under sensory-constrained situations. In 2 experiments, sensory decoding difficulty was manipulated by using easy and hard fonts to study perceptual simulation during sentence reading (Zwaan, Stanfield, & Yaxley, 2002). Results indicated that simulating perceptual symbols in language was not compromised by surface-form decoding challenges such as difficult font, suggesting relative resilience of embodied language processing in the face of certain sensory constraints. Further implications for learning from text and individual differences in language processing will be discussed.

Public Significance Statement
Two studies with college-age young adults in simple sentence reading found that even though sensory challenges, such as those created by hard font, obviously slowed down the reading speed, but interestingly, did not affect readers' ability to infer (i.e., a reading process of predicting information not expressed by but that can be implied from the text) the physical shapes of the embedded concepts in sentences. Our results hold promises for individuals with either acquired (such as aging) or congenital (such as low vision/hearing, blindness or deafness) sensory disabilities, when implementing language-based training, intervention and/or rehabilitation programs to better their daily lives.

Keywords: perceptual simulation, reading, sensory constraints

Language processing encompasses multiple levels of computations (Kintsch, 1998; McNamara & Magliano, 2009). During language understanding, readers or listeners not only have to decode the written or spoken linguistic signals and map them onto individual words and concepts (i.e., word-level processing), but also have to combine multiple concepts together to form meanings (i.e., conceptual integration) and generate enriched and complex inferences about the literal world that single words or concepts could have failed to convey (i.e., situation model construction). Some aspects of these inferences about the literal world tend to be quite vivid and are grounded in human experiences through bodily

interactions with the physical world (Barsalou, 1999) and thus can be reprojected or be simulated partly or fully in one's mind during online language understanding. For example, Zwaan et al. (2002) asked participants first to read a very simple sentence (e.g., "The eagle was in the sky" vs. "The eagle was in the nest") and then to verify in a subsequent picture whether or not the object depicted in this picture had been mentioned in the sentence that was just read. The critical manipulation in this sentence-picture verification paradigm was whether or not the object in the picture was "matched" (an eagle flying with outstretched wings vs. a perching eagle with folded wings) or "mismatched" (a perching eagle with folded wings vs. an eagle flying with outstretched wings) with the implied shape that can be inferred from the sentence. Picture verification latency was faster for "matched" trials than "mismatched" trials (i.e., the classic "match effect"). The match effect implied that even during simple sentence comprehension, readers visually simulated the scenario in question and has been taken as evidence for obligatory activation of perceptual symbols during sentence understanding.

During the last approximately 15 years, the phenomenon of perceptual simulation in language processing has been widely and extensively studied by Zwaan and his colleagues (Stanfield & Zwaan, 2001; Zwaan et al., 2002; Zwaan & Pecher, 2012); however, one of the central issues in recent debates (Zwaan, 2014) and theoretical reflections (Barsalou, 2008, 2009) is around refining and defining the boundary conditions under which perceptual

simulation may or may not be obligatory. Some theorists (Barsa-lou, 1999; Glenberg, Witt, & Metcalfe, 2013; Zwaan & Pecher, 2012) have suggested that perceptual symbol processing in language might be intrinsic as it seems to remain intact and resilient in school-age children (Engelen, Bouwmeester, de Bruin, & Zwaan, 2011) as well as healthy aging adults (Dijk-stra, Yaxley, Madden, & Zwaan, 2004). In addition, the match effect has been recently well reproduced with larger-scale replication efforts (Zwaan & Pecher, 2012).

In contrast to this modal view of language processing, according to the resource-based models of language processing (Just & Carpenter, 1992; Fedorenko, 2014), readers' or listeners' ability to simulate perceptual symbols would be compromised by limited processing resources, either sensory or cognitive. One variation of these models that is most pertinent to this study is the effortfulness hypothesis (Rabbitt, 1968; Wingfield, Tun, & McCoy, 2005), which states that word-level decoding difficulty introduced either by experimental manipulations or individual differences in participants (e.g., sensory acuity or aging) during encoding could have downstream effects on higher levels of language processing (e.g., conceptual integration or perceptual simulation), because attentional resources devoted to decoding the linguistic signals would have consumed the resources that would otherwise have been available for higher levels of linguistic computations. Supporting this claim, studies that have operationalized sensory constraint either by embedding text in visual noise or by using harder font, have demonstrated that during sentence reading, simulated sensory declines impaired sentence understanding (i.e., conceptual integration), because of sensory constraints introduced during encoding (Gao, Stine-Morrow, Noh, & Eskew, 2011; Gao, Levinthal, & Stine-Morrow, 2012; Rayner, Reichle, Stroud, Williams, & Pollatsek, 2006). Speranza, Daneman, and Schneider (2000) manipulated both decoding difficulty and sentential predictability in a visual word recognition task among younger and older adults. For example, test sentences like, "The boat sailed across the BAY versus Mr. Smith knew about the BAY," with BAY as the target word for recognition test, were adopted. The authors found that (a) visual noise immediately impaired word recognition in sentential context; (b) sentence predictability effect on word recognition was reduced by noise; and interestingly, (c) older adults tended to benefit more from sentential context compared to the young, probably due to growth in world knowledge and reading experiences throughout the life course. The evidence implied that sensory decoding constraint might have negatively affected other higher-level processes in language understanding, such as inference-making—a form of situation model construction—beyond conceptual integration.

Therefore, there are good theoretical motivations to hypothesize that simulated sensory declines would penetrate into and negatively impact perceptual simulation during reading as well. In short, according to an obligatory view of language processing, simulating perceptual symbols in language may be robust and immune to proper levels of decoding challenges, whereas the effortfulness hypothesis would predict that sensory challenge would compromise higher-level perceptual simulation, the way it does for conceptual integration. In the present study, these two hypotheses were directly tested and contrasted in two experiments with a novel modified "sentence-picture verification" paradigm.

Experiment 1

Method

Participants. Young adult participants (18–30 years of age) from the Nijmegen area, The Netherlands were recruited via email to conduct an online experiment and were randomly assigned to either the easy (Arial) font condition or the hard (Bodoni MT) font condition. Thirty-nine subjects participated in the easy condition and 40 in the hard condition. Among those who participated, four of them from the easy condition and six of them from the hard condition dropped out at the beginning or in the middle of the experiment, resulting in 35 in the easy condition and 34 in the hard condition for the final analyses. Subjects gave informed consent and participated voluntarily without monetary rewards.

Materials and design. Twenty target sentences and 40 filler sentences were adopted and adapted from previous studies (Zwaan et al., 2002), resulting in four counterbalanced lists with 60 items in each. Based on a previous study[1] (Diemand-Yauman, Oppenheimer, & Vaughan, 2011), in which the font, size, grayscale, and italicization of the to-be-presented text were simultaneously varied, Arial (16-point and pure black) and Bodoni MT (12-point, italicized, and 60% grayscale) were chosen as easy and hard font, respectively. Furthermore, font (easy vs. hard) manipulation was crossed with the list, further resulting in 8 unique lists, to one of which participants were randomly assigned. Shape congruency was a with-subject factor whereas font was a between-subjects factor.

Procedure. Each participant completed the classic sentence-picture verification task (Zwaan & Pecher, 2012) in which the participant first read a short sentence (e.g., "An egg was in the pan" vs. "An egg was in the fridge") at their own speed and then after a 500-ms delay judged whether or not the presented picture contained the object that was mentioned in the sentence. For all the 20 target sentence-picture pairs, the pictures always depicted the object mentioned in the sentence (i.e., all "Yes" responses), but the shape of the object in the picture can either match (an egg cracked open vs. an egg intact) the implied shape from the sentential context or mismatch (an egg intact vs. an egg cracked open) that from the sentential context. There were 40 additional filler sentence-picture pairs with 10 requiring "yes responses and 30 requiring no responses so that the numbers of yes versus no responses were balanced across all trials. Each sentence could be presented in either easy Arial or hard Bodoni MT font and since font was a between-subjects factor, each participant read all the sentences in only one font.

Besides objective measures, which partially serve the purpose of our manipulation check, such as reading time and verification latency, a postreading four-item scale was developed to evaluate participants' subjective judgment of stimulus difficulty and processing fluency. The scale followed a 7-point Likert fashion ranging from 1 (*the least familiar*) and 7 (*the most familiar*) for

---

[1] As in Diemand-Yauman et al. (2011), the sole purpose of the current manipulation was to dichotomize the decoding difficulty at two reasonable levels that could potentially result in reading time difference between levels, without really addressing the question of which font type favors reading more, while carefully matching for other factors such as size, grayscale, and italicization.

familiarity ratings and 1 denoting "the easiest" and 7 denoting "the hardest" for difficulty ratings, respectively. Participants were prompted to rate both fonts in terms of their familiarity with them before participation and how hard they found either font to read after participation.

Results

For comparison purposes, key findings from both experiments were presented together, parallel to each other, in Tables 1 and 2 and Figures 1 and 2.

Sentence reading time. Sentence reading times[2] were first trimmed such that reading times shorter than 600 ms or longer than 10,000 ms were deleted, deleting 3.04% of the raw data. The resulting reading time data were submitted to a 2 (font: Arial vs. Bodoni MT) $\times$ 2 (congruency: Match vs. Mismatch) mixed-effects model.[3] There was main effect of font (see Table 1) on sentence reading time, $t = 1.79$, $p = .02$, but no main effect of congruency or interaction, both $ts < 1$.

This result confirmed that the font manipulation taxed sentence reading beyond single word recognition. The null effect of congruency was exactly as expected for sentence reading as the match/mismatch manipulation would not come into play until the picture verification phase in the current sentence-picture verification paradigm.

Picture verification.

Accuracy. Accuracy for picture verification was submitted to a 2 (font: Arial vs. Bodoni MT) $\times$ 2 (congruency: match vs. mismatch) mixed-effects model. Accuracy was higher for the match condition than for the mismatch condition (see Table 2), $z = 2.65$, $p < .01$, replicating previous findings (Zwaan et al., 2002); however, there was no effect of font or interaction, $z = 1.39$, $p < .16$; $z < 1$, $p > .05$, respectively.

Unstandardized latency. The top left panel of Figure 1 represents the result for unstandardized picture verification latency.[4]

Before being submitted to the 2 (font: Arial vs. Bodoni MT) $\times$ 2 (congruency: match vs. mismatch) mixed-effects model, raw latencies faster than 200 ms were deleted (0.14% of the total data) to control for inattentiveness. Verification latencies were analyzed for correct responses only. There were main effects of congruency and font, but no interaction, $t = 2.63$, $p < .01$; $t = 2.84$, $p < .001$; $t < 1$, $p > .05$, respectively. Replicating the match effect in picture verification (Zwaan et al., 2002),

Table 1
Sentence Reading Time (in ms) as a Function of Congruency and Processing Demand

Variable	Match		Mismatch	
	ms	SD	ms	SD
Experiment 1				
Arial	1,983	1,321	1,982	1,258
Bodoni MT	2,347	1,411	2,294	1,502
Experiment 2				
Easy_Slow	1,044	779	1,081	788
Easy_Fast	1,046	708	1,090	833
Hard_Fast	2,004	1,466	2,165	1,448

Table 2
Picture Verification Accuracy as a Function of Congruency and Processing Demand

Variable	Match		Mismatch	
	%	SD	%	SD
Experiment 1				
Arial	98	13	95	22
Bodoni MT	96	18	94	24
Experiment 2				
Easy_Slow	95	21	92	28
Easy_Fast	98	15	96	20
Hard_Fast	94	25	90	30

readers spent more time verifying an object in an incongruent shape than an object in a congruent one. Verification latencies for hard font were unexpectedly shorter than that for easy font.[5] This effect was unlikely to be due to speed–accuracy trade-off, as is shown by the zero correlation between verification speed and accuracy, but instead was at least partially due to the between-subjects design in this experiment, resulting in bigger between-condition (Arial vs. Bodoni MT) differences in general information processing speed, and partially due to the online Web-based testing environment, resulting in less vigilance and attentiveness and more sloppy responses in easy font condition (see cross-experiment comparisons and t-statistics in Footnote 5). Critically, it was found that the match effect was reliable under easy and hard conditions and this effect did not interact

---

[2] In Experiment 1, a lexical-decision task was administered during a pilot study as a manipulation check. Accuracies for easy and hard font conditions ($M_{Arial} = 88.54\%$, $SE = 6.51\%$; $M_{Bodoni\ MT} = 86.72\%$, $SE = 6.94\%$) were comparable, $z = 0.36$, $p > .05$. Before analyzing the latency data, raw data were first trimmed. Lexical decision latencies longer than 2,500 ms were deleted and furthermore within each participant response, latencies longer than 3SDs above the mean of that participant were deleted, resulting in a total deletion of 3.56% of the raw data. Response latencies were analyzed for the correct responses only. As predicted, latency for harder Bodoni MT font, $M_{Bodoni\ MT} = 1,058$ ms, $SE = 74$ ms, was clearly longer than that for the easier Arial font, $M_{Arial} = 979$ ms, $SE = 66$ ms, $t = 3.59$, $p < .001$, suggesting that our font manipulation was taxing enough to demand attentional allocation such that readers took more time to decode isolated words in hard font.

[3] Unless otherwise noted, analyses for this task and all the following tasks were conducted in the liner mixed-effects models in R.

[4] A power analysis of the congruency effect for both experiments revealed a power of 84.39% with a Cohen's $d = 0.2214$ for Experiment 1, and a power of 77.50% with a Cohen's $d = 0.2405$ for Experiment 2, respectively.

[5] Verification latencies for hard font were unexpectedly shorter than that for easy font, while correlation analysis of latency and accuracy for picture verification, $r(67) = .00$, $p = 1.00$, for overall; $r(33) = -.09$, $p = .61$; $r(32) = .02$, $p = .93$, for Arial and Bodoni MT condition, respectively, revealed that this effect is unlikely due to speech-accuracy trade off. As a matter of fact, according to Table 1 and top panel of Figure 1, compared to Experiment 2, Experiment 1 which has adopted a between-participants design, has resulted in both longer sentence reading time (see Table 1) and longer picture verification latency (top panel of Figure 1), for participants in the Arial (essentially, Easy_Slow) condition than those for the participants from the same condition in Experiment 2, with independent t test statistics of $t(69) = 6.14$, $p < .001$; $t(69) = 3.81$, $p < .001$, respectively, from the cross-experiment comparisons at both baselines. Standardized Z-scores were adopted to further control for individual differences either between experiments or between conditions within an experiment.
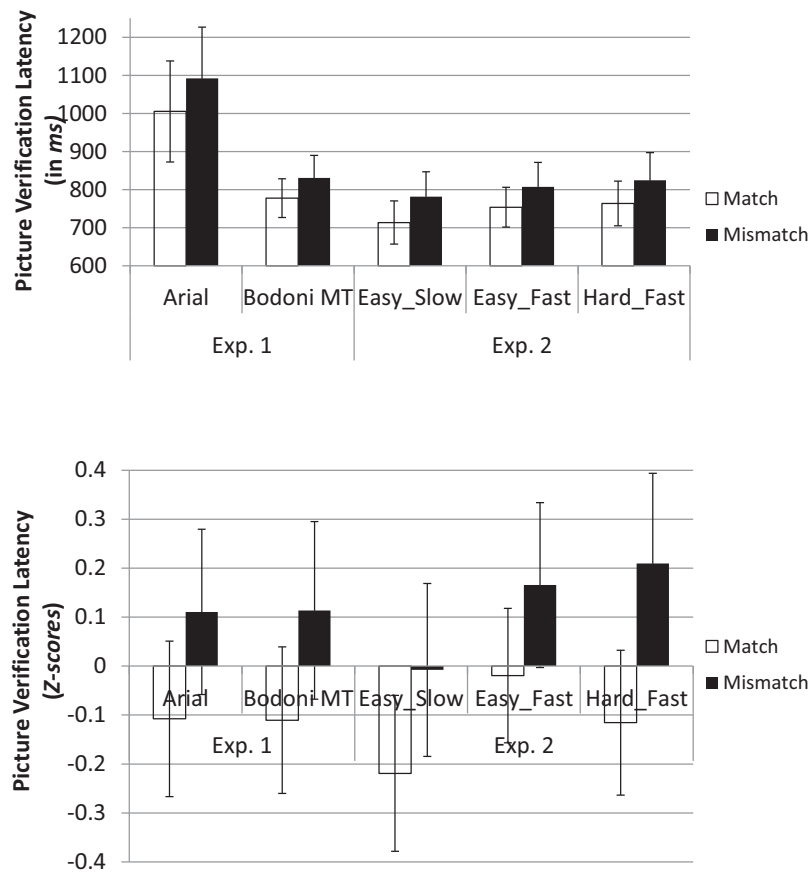
Figure 1. Picture verification latency in ms (top panel) and *Z*-scores (bottom panel). Error bars represent 1 standard error.

with decoding challenges introduced during encoding, consistent with the idea that perceptual simulation is particularly resilient even under developmentally distinct cognitive situations (e.g., younger child and older adults, Engelen et al., 2011; Dijkstra et al., 2004).

Standardized Z-scores.   Because of the nature of the design and the experimental implementation (i.e., font as between-subjects factor and online home-based experimentation), to control for individual differences in processing speed and apparatus configurations (e.g., desktop vs. laptop; operation systems; monitor size; refreshing rate; resolution; keyboard scanning rate; and many others, just to name a few), within-participant standardized z-scores are ideal to examine the Font × Congruency interaction more precisely. The effect of congruency was still robust (bottom left panel of Figure 1), $t = 4.44$, $p < .001$, and crucially, congruency did not interact with font, $t < 1$, $p > .05$, even with standardized scores, confirming our above findings with unstandardized scores and providing further evidence for resilient perceptual simulation under adverse conditions, despite considerable variabilities between individual participants.

Familiarity and difficulty ratings.   Because of technical failures, rating data from seven participants in the Arial and 7 in the Bodoni MT condition were not collected, resulting in the remaining 28 and 27 participants in either condition, respectively. Participants found Arial font more familiar (top left panel of Figure 2) than Bodoni MT font, $t = 9.42$, $p < .001$, and concomitantly, Bodoni MT font harder to read than Arial font (top right panel of Figure 2), $t = 4.74$, $p < .001$, for familiarity and difficulty, respectively, irrespective of the font in which they had just read the sentences, ts < 1, for the Font (Read) × Font (Rated) interaction. Combined with reading time data, these subjective ratings from participants justified our font manipulation as veridical.

In Experiment 1, it was found that sensory challenge as created by harder font during encoding did not penetrate into perceptual simulation, suggesting that resource-based models of sentence comprehension may only apply to text-based processes (e.g., understanding whatever is explicitly conveyed by the sentence, that is, understanding the location of the egg in our example), but not apply to situation-based processes (e.g., inferring the shape of the object implied by the sentence). However, it could have been argued that in the paradigm introduced in Experiment 1, the 500-ms delay between the sentence and the image might have created a scenario that encourages situation model construction, which tends to get activated later in the time course of sentence understanding (Kintsch, Welsch, Schmalhofer, & Zimny, 1990; Madden &
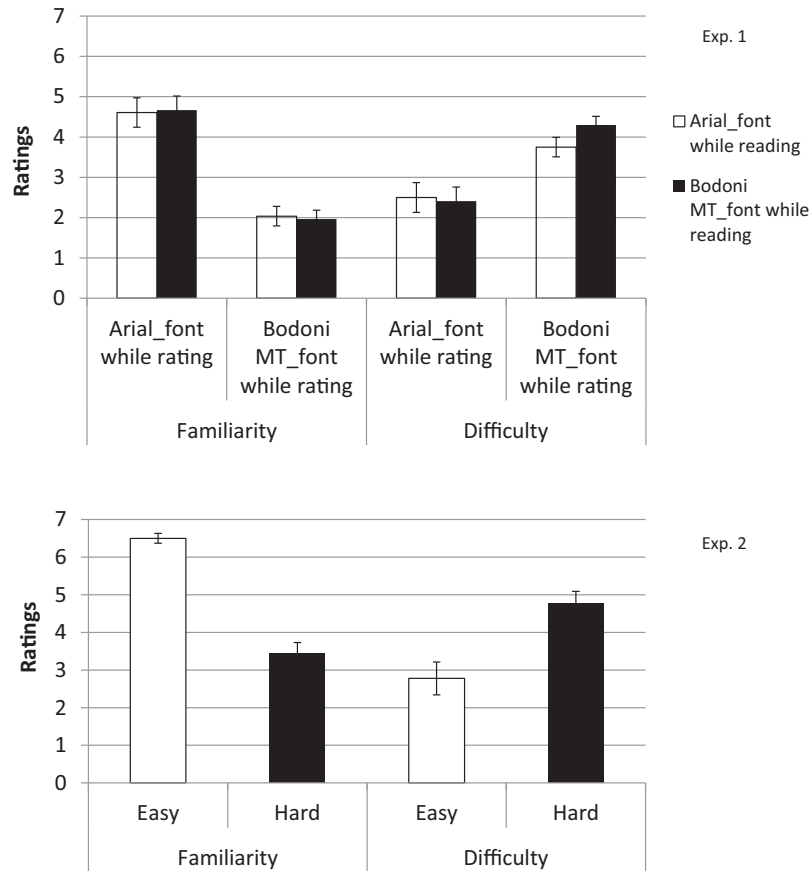
Figure 2. Subjective ratings of familiarity (left panel) and difficulty (right panel) with both fonts. Error bars represent 1 standard error.

Zwaan, 2006). Experiment 2 was conducted to rule out this possibility, involving both short (0-ms) and long (500-ms) delays between the sentence and the image.

Experiment 2

Method

Participants. Thirty-six young adult participants (18–30 years of age), who gave their informed consent, from Beijing, China were recruited to conduct a lab experiment and were randomly assigned to one of our six experimental lists (see list description below). Subjects participated in this one-session in-lab experiment (~5–10 min) with a monetary reward.

Materials, design, and procedure. SongTi and KaiTi were chosen as easy and hard font in Chinese, respectively, as previous literature (Lv & Wang, 2012) suggested that such manipulation has successfully created processing difficulty in Chinese word recognition. In addition, to make the font size comparable across experiments and across languages,[6] 15-point a pure black 宋体 (SongTi) and 8-point, italicized and 60% grayscale 楷体 (KaiTi), were adopted such that each Chinese character subtended approximately 1.5 times of the visual angle which was subtended by corresponding western character (let-

ter; Liversedge, Drieghe, Li, Yan, Bai, & Hyönä, 2016), for easy/hard font, respectively. Moreover, processing difficulty was further manipulated by combining font and delay, posing three unique levels (Easy_Slow, Easy_Fast, and Hard_Fast) of processing demand for readers in the new modified paradigm. Furthermore, both congruency and processing difficulty (six treatment conditions) were counterbalanced across lists, resulting in six unique lists, to one of which participants were randomly assigned. Thirty-six target sentences and 36 filler sentences were adapted and translated into Chinese from previous experiment, resulting in 72 total items within each list, among which there were six target items for each of the six conditions. Both shape congruency and processing demand were within-subject factors. Except for the short/long delay, all the other aspects of procedure in Experiment 2 was exactly the same as that in Experiment 1.

---

[6] More recent literature have revealed that language reading in general, as measured by eye-tracking, is universal and comparable across language systems (Liversedge et al., 2016) and literature have also suggested that perceptual simulation in specific, as measured by response latency, is reliable and robust across language systems as well (Zwaan & Pecher, 2012; Li, Wang, & Mo, 2005).

Results

Sentence reading time. As in Experiment 1, sentence reading times[7] were first trimmed such that reading times shorter than 300 ms or longer than 8,000 ms[8] were deleted, deleting 2.16% of the raw data. The resulting reading time data were submitted to a 3 (difficulty: Easy_Slow, Easy_Fast, vs. Hard_Fast) × 2 (congruency: match vs. mismatch) mixed-effects model. Again, there was a main effect of difficulty, $t = 11.15$, $p < .001$, although there was no main effect of congruency or an interaction, $t = 1.58$, $p = .11$; $t < 1$, respectively. Sentences under the Hard_Fast condition took longer to read than those under the Easy_Fast condition, $t = 11.37$, $p < .001$, whereas sentences under the Easy_Slow and the Easy_Fast conditions did not distinguish themselves from each other in reading time, $t < 1$ (see Table 1). Once again, these results confirmed that the difficulty manipulation was taxing enough for sentence reading. The null effect of congruency was exactly expected for sentence reading as the match/mismatch manipulation was not expected to impact initial sentence reading but rather the later picture verification in the current paradigm.

Picture verification.

Accuracy. As in Experiment 1, accuracy for picture verification was submitted to a 3 (difficulty: Easy_Slow, Easy_Fast, vs. Hard_Fast) × 2 (congruency: match vs. mismatch) mixed-effects model. Accuracy was higher for match condition than mismatch condition (see Table 2), $z = 2.20$, $p < .05$, replicating Experiment 1; however, there was no effect of difficulty or difficulty by congruency interaction, $z = 1.03$, $p = .30$; $z < 1$, $p = .39$, respectively. Despite a lack of overall effect of difficulty, post hoc contrasts revealed that verification accuracy for pictures under the Hard_Fast condition was lower than that for pictures under the Easy_Fast condition, $z = 2.65$, $p < .01$, whereas accuracies for pictures under Easy_Fast and Easy_Slow conditions did not differ from each other, $z = 1.79$, $p > .05$.

Unstandardized latency. The top right panel of Figure 1 represents the result for unstandardized picture verification latency. Before being submitted to the 3 (difficulty: Easy_Slow, Easy_Fast, vs. Hard_Fast) × 2 (congruency: match vs. mismatch) mixed-effects model, raw latencies faster than 200 ms or longer than 3,000 ms were deleted (1.54% of the total data) to control for inattentiveness. Verification latencies were analyzed for correct responses only. There were main effects of congruency and difficulty, without any interaction, $t = 3.84$ ($p < .001$), $t = 2.07$ ($p < .05$), and $t = 1.07$ ($p = .31$), respectively, suggesting that both incongruent picture and cognitive demand increased verification latency. Pictures under the Easy_Slow condition took shorter to verify than those under the Easy_Fast condition, $t = 2.73$, $p < .01$, whereas pictures under the Hard_Fast and the Easy_Fast conditions did not distinguish themselves from each other in verification latency, $t < 1$.

Critically, it was found that the match effect was reliable at varying levels of processing demand, either sensory challenge during encoding or responding under time pressure, replicating exactly what we reported in Experiment 1.

Standardized Z-scores. As in Experiment 1, to further control for individual differences such as in processing speed, within-participant standardized z-scores are recommended to weigh in the Difficulty × Congruency interaction. By standardized score, the effects of congruency and difficulty remained significant (bottom right panel of Figure 1), $t = 3.79$, $p < .001$; $t = 2.91$, $p < .01$, respectively. Crucially, congruency did not interact with difficulty, $t = 1.43$, $p = .18$, confirming our findings with unstandardized scores given variabilities between individual participants. Further contrasts revealed that pictures under the Easy_Slow condition took shorter to verify than those under the Easy_Fast condition, $t = 3.53$, $p = .001$, whereas pictures under the Hard_Fast and the Easy_Fast conditions did not distinguish themselves from each other in verification latency, $t < 1$.

Familiarity and difficulty ratings. Thanks to the within-design in which each participant read the sentences and rate the questions in both fonts, mean ratings were presented in four, instead of eight, bars in Figure 2. Participants found SongTi font more familiar (bottom left panel of Figure 2) than KaiTi font, $t = 9.99$, $p < .001$, and concomitantly, KaiTi font harder to read than SongTi font (bottom right panel of Figure 2), $t = 3.73$, $p < .001$. Combined with reading time data, these subjective ratings from participants justified our font manipulation as veridical in this experiment.

Discussions and Conclusion

Results from both experiments indicated that simulating perceptual symbols in language was not compromised by surface-form decoding challenges, such as difficult font, suggesting that attentions devoted to effortful decoding did not have a downstream effect on perceptual simulation. Our findings are consistent with and extend previous literature on obligatory perceptual symbol processing (Barsalou, 1999; Zwaan et al., 2002; Zwaan & Pecher, 2012; Glenberg et al., 2013) by further demonstrating relative resilience of perceptual symbol activation under certain less-than-ideal communicative situations. This indicated that resource-based models of language processing (Just & Carpenter, 1992; Fedorenko, 2014; Rabbitt, 1968; Wingfield, Tun, & McCoy, 2005) may only be applicable to word recognition (Pelli, Farell, & Moore, 2003) and conceptual integration (Gao et al., 2011, 2012), but not to perceptual simulation.

Although the hard font introduced in this current study obviously hindered reading, as reflected in sentence reading time in both experiments, readers' ability to make perceptual inferences (shape in this case) of the embedded concepts in sentences remained reliable and robust under both easy and hard conditions. This implied that certain aspects of linguistic computations (per-

---

[7] In Experiment 2, every measure (see Materials and Design & Procedure section of Experiment 2) has been taken to ensure the reproducibility of the key manipulation without another pilot and the cross-experiment comparisons possible. In fact, the manipulation in Experiment 2 produced similar effects as did in Experiment 1, as reflected in sentence reading time, and accuracy and latency of picture verification; cross-experiment comparisons also confirmed that sentence reading time, and accuracy and latency of picture verification were comparable across experiments, except for those under easy (Arial) condition due to the between-subjects design in Experiment 1 (see also Footnote 5).

[8] Experiment 2 has adopted smaller lower and upper limits for data trimming in both sentence reading time and picture verification latency, as participants in Experiment 2 tended to read and respond faster, with independent t test statistics of $t(103) = 5.31$, $p < .001$; $t(103) = 2.69$, $p < .01$, respectively, than participants in Experiment 1, probably because of their overall faster information processing speed in the current design that has included laboratory-based testing.

ceptual simulation vs. conceptual integration) were resistant to superficial variations in the processing environment. This may be desirable as our daily communication environment tends to be rather disorganized, if not distracting, and as there are considerable adult individual differences in sensory acuity (Wingfield et al., 2005). This relative fault-tolerant feature of perceptual symbols in the face of degraded linguistic inputs and/or individual variations in sensory acuity might have been adaptive, as going beyond the text and making contextually appropriate inferences might be one of most transformative powers of and yet one of the greatest enjoyments of reading, at all times.

It could be argued that the fault-tolerant perceptual symbol system might have benefited from self-paced reading introduced in the current sentence-picture verification paradigm as participants were able to control their own reading speed under challenges in order to develop a more full-fledged understanding of the scenarios in question. It is not unreasonable to speculate that given time pressure or under experimental instructions that emphasize speed, perceptual simulation could be compromised; for example, weakened but not completely diminished (Madden & Zwaan, 2006). However, our Experiment 2 directly tested this idea and demonstrated, clearly and unequivocally, that even under time pressured situations, the combination of sensory and cognitive demands did not pose threats to the perceptual symbol system. Despite ample data in sentence processing suggesting that speeded response impairs many other levels of language processing (Gernsbacher, Varner, & Faust, 1990; Kintsch et al., 1990), this finding has provided convincing evidence for obligatory perceptual symbol activation in sentence understanding, even under resource (either sensory, cognitive or their combination) constrained situations.

Contrasted with Speranza et al.'s (2000) finding of reduced sentence context effect under noise, our finding of preserved "match effect" with harder font suggested that traditional (amodal) models of language could only explain semantically driven sentential context effect (Graesser, Millis, & Zwaan, 1997; Kintsch, 1998; Just & Carpenter, 1992) but not perceptually driven mental simulation phenomenon. One of the most likely explanations available for this discrepancy could be that initially, like semantic information, perceptual properties of the to-be-integrated concepts in sentence are activated nonselectively in normal reading (Hickok & Poeppel, 2007), and what makes perceptual activation and simulation so special is that unlike blocked or at least deterred semantic/associative activation (Speranza et al., 2000; Kintsch et al., 1990), perceptual symbols are obligated to get activated in the presence of moderate noisy linguistic input, because of its primitive nature (Barsalou, 1999). The integration of perceptual symbol system with traditional symbolic language processing models (Barsalou, 2008, 2009), which did not necessarily incorporate a perceptual component (Zwaan et al., 2002), should pose itself as a fruitful avenue for future research, by generating more precise and more testable hypotheses, a call that was quickly responded to in recent hybrid models of the human language faculty (Dell & Chang, 2013; Hickok & Poeppel, 2007).

With the choice of controlling one's own pace, readers' perceptual symbol system is rather adaptive to cope with surface-form decoding difficulties embedded in text, as literature in many other domains of learning demonstrates that learners across the age spectrum are self-regulated and can prioritize their own learning goals and select their own learning materials in the face of chal-

lenging/difficult items to optimize and maximize their learning outcomes (Metcalfe, 2009; Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013; Stine-Morrow, Miller, & Hertzog, 2006). Whether middle-aged and/or older adults would have demonstrated similar resilience in perceptual symbol processing faced with gradual sensory declines and decoding challenges is under current investigation.

References

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences, 22,* 577–609.

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

Barsalou, L. W. (2009). Simulation, situated conceptualization, and prediction. *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences, 364,* 1281–1289. http://dx.doi.org/10.1098/rstb.2008.0319

Dell, G. S., & Chang, F. (2013). The P-chain: Relating sentence production and its disorders to comprehension and acquisition. *Philosophical Transactions of the Royal Society of London.* Series B, Biological Sciences, *369,* 20120394. http://dx.doi.org/10.1098/rstb.2012.0394

Diemand-Yauman, C., Oppenheimer, D. M., & Vaughan, E. B. (2011). Fortune favors the bold (*and the Italicized*): Effects of disfluency on educational outcomes. *Cognition, 118,* 111–115. http://dx.doi.org/10.1016/j.cognition.2010.09.012

Dijkstra, K., Yaxley, R. H., Madden, C. J., & Zwaan, R. A. (2004). The role of age and perceptual symbols in language comprehension. *Psychology and Aging, 19,* 352–356. http://dx.doi.org/10.1037/0882-7974.19.2.352

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest, 14,* 4–58. http://dx.doi.org/10.1177/1529100612453266

Engelen, J. A. A., Bouwmeester, S., de Bruin, A. B. H., & Zwaan, R. A. (2011). Perceptual simulation in developing language comprehension. *Journal of Experimental Child Psychology, 110,* 659–675. http://dx.doi.org/10.1016/j.jecp.2011.06.009

Fedorenko, E. (2014). The role of domain-general cognitive control in language comprehension. *Frontiers in Psychology, 5,* 335. http://dx.doi.org/10.3389/fpsyg.2014.00335

Gao, X., Levinthal, B. R., & Stine-Morrow, E. A. L. (2012). The effects of ageing and visual noise on conceptual integration during sentence reading. *The Quarterly Journal of Experimental Psychology, 65,* 1833–1847. http://dx.doi.org/10.1080/17470218.2012.674146

Gao, X., Stine-Morrow, E. A. L., Noh, S. R., & Eskew, R. T., Jr. (2011). Visual noise disrupts conceptual integration in reading. *Psychonomic Bulletin & Review, 18,* 83–88. http://dx.doi.org/10.3758/s13423-010-0014-4

Gernsbacher, M. A., Varner, K. R., & Faust, M. E. (1990). Investigating differences in general comprehension skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 430–445. http://dx.doi.org/10.1037/0278-7393.16.3.430

Glenberg, A. M., Witt, J. K., & Metcalfe, J. (2013). From the revolution to embodiment: 25 years of cognitive psychology. *Perspectives on Psychological Science, 8,* 573–585. http://dx.doi.org/10.1177/1745691613498098

Graesser, A. C., Millis, K. K., & Zwaan, R. A. (1997). Discourse comprehension. *Annual Review of Psychology, 48,* 163–189. http://dx.doi.org/10.1146/annurev.psych.48.1.163

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience, 8,* 393–402. http://dx.doi.org/10.1038/nrn2113

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99,* 122–149. http://dx.doi.org/10.1037/0033-295X.99.1.122

Kintsch, W. (1998). *Comprehension: A paradigm for cognition.* Cambridge, UK: Cambridge University Press.

Kintsch, W., Welsch, D., Schmalhofer, F., & Zimny, S. (1990). Sentence memory: A theoretical analysis. *Journal of Memory and Language, 29,* 133–159. http://dx.doi.org/10.1016/0749-596X(90)90069-C

Li, Y., Wang, R., & Mo, L. (2005). Wù Tǐ Yǐn Hán De Xíng Zhuàng Xìn Xī Duì Tú Piàn Zài Rèn De Yǐng Xiǎng [The effect of implied shape of objects on picture recognition]. *Psychological Science, 28,* 588–590.

Liversedge, S. P., Drieghe, D., Li, X., Yan, G., Bai, X., & Hyönä, J. (2016). Universality in eye movements and reading: A trilingual investigation. *Cognition, 147,* 1–20. http://dx.doi.org/10.1016/j.cognition.2015.10.013

Lv, C., & Wang, Q. (2012). Font effects of Chinese characters and pseudo-characters on the N400: Evidence for an orthographic processing view. *Brain and Cognition, 80,* 96–103. http://dx.doi.org/10.1016/j.bandc.2012.05.002

Madden, C. J., & Zwaan, R. A. (2006). Perceptual representation as a mechanism of lexical ambiguity resolution: An investigation of span and processing time. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32,* 1291–1303. http://dx.doi.org/10.1037/0278-7393.32.6.1291

McNamara, D. S., & Magliano, J. P. (2009). Towards a comprehensive model of comprehension. In B. Ross (Ed.), *The psychology of learning and motivation* (pp. 297–384). New York, NY: Elsevier Science. http://dx.doi.org/10.1016/S0079-7421(09)51009-2

Metcalfe, J. (2009). Metacognitive judgments and control of study. *Current Directions in Psychological Science, 18,* 159–163. http://dx.doi.org/10.1111/j.1467-8721.2009.01628.x

Pelli, D. G., Farell, B., & Moore, D. C. (2003). The remarkable inefficiency of word recognition. *Nature, 423,* 752–756. http://dx.doi.org/10.1038/nature01516

Rabbitt, P. M. A. (1968). Channel-capacity, intelligibility and immediate memory. *The Quarterly Journal of Experimental Psychology, 20,* 241–248. http://dx.doi.org/10.1080/14640746808400158

Rayner, K., Reichle, E. D., Stroud, M. J., Williams, C. C., & Pollatsek, A. (2006). The effect of word frequency, word predictability, and font difficulty on the eye movements of young and older readers. *Psychology and Aging, 21,* 448–465. http://dx.doi.org/10.1037/0882-7974.21.3.448

Speranza, F., Daneman, M., & Schneider, B. A. (2000). How aging affects the reading of words in noisy backgrounds. *Psychology and Aging, 15,* 253–258. http://dx.doi.org/10.1037/0882-7974.15.2.253

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

Stine-Morrow, E. A. L., Miller, L. M. S., & Hertzog, C. (2006). Aging and self-regulated language processing. *Psychological Bulletin, 132,* 582–606. http://dx.doi.org/10.1037/0033-2909.132.4.582

Wingfield, A., Tun, P. A., & McCoy, S. L. (2005). Hearing loss in older adulthood. *Current Directions in Psychological Science, 14,* 144–148. http://dx.doi.org/10.1111/j.0963-7214.2005.00356.x

Zwaan, R. A. (2014). Embodiment and language comprehension: Reframing the discussion. *Trends in Cognitive Sciences, 18,* 229–234. http://dx.doi.org/10.1016/j.tics.2014.02.008

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

Zwaan, R. A., Stanfield, R. A., & Yaxley, R. H. (2002). Language comprehenders mentally represent the shapes of objects. *Psychological Science, 13,* 168–171. http://dx.doi.org/10.1111/1467-9280.00430