COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 47 (2023) e13383

ISSN: 1551-6709 online
DOI: 10.1111/cogs.13383

Rational Sentence Interpretation in Mandarin Chinese

Meilin Zhan,[a] Sihan Chen,[a] Roger Levy,[a] Jiayi Lu,[b] Edward Gibson[a]

[a]*Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology*
[b]*Department of Linguistics, Stanford University*

Received 14 June 2023; received in revised form 3 November 2023; accepted 8 November 2023

Abstract

Previous work has shown that English native speakers interpret sentences as predicted by a noisy-channel model: They integrate both the real-world plausibility of the meaning—the prior—and the likelihood that the intended sentence may be corrupted into the perceived sentence. In this study, we test the noisy-channel model in Mandarin Chinese, a language taxonomically different from English. We present native Mandarin speakers sentences in a written modality (Experiment 1) and an auditory modality (Experiment 2) in three pairs of syntactic alternations. The critical materials are literally implausible but require differing numbers and types of edits in order to form more plausible sentences. Each sentence is followed by a comprehension question that allows us to infer whether the speakers interpreted the item literally, or made an inference toward a more likely meaning. Similar to previous research on related English constructions, Mandarin participants made the most inferences for implausible materials that could be inferred as plausible by deleting a single morpheme or inserting a single morpheme. Participants were less likely to infer a plausible meaning for materials that could be inferred as plausible by making an exchange across a preposition. And participants were least likely to infer a plausible meaning for materials that could be inferred as plausible by making an exchange across a main verb. Moreover, we found more inferences in written materials than spoken materials, possibly a result of a lack of word boundaries in written Chinese. Overall, the fact that the results were so similar to those found in related constructions in English suggests that the noisy-channel proposal is robust.

Keywords: Sentence processing; Noisy channel; Rational inference; Psycholinguistics; Mandarin

M.Z. and S.C. contributed to the article equally.

Correspondence should be sent to Sihan Chen, Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, 43 Vassar St., Cambridge, MA 02139, USA. E-mail: sihanc@mit.edu

1. Introduction

In normal language situations, people sometimes make errors in their productions (e.g., Garrett, 1975), and there is often noise in the environment. As a result of the ubiquity of such errors and potential misunderstandings, the language processing mechanism must be robust to noise.

Shannon (1948) was the first to formalize how language producers and comprehenders might behave in a noisy channel. Under such an approach, a rational comprehender is predicted to be sensitive to both *prior information*—including the kinds of things that are likely to be talked about (e.g., the plausibility of events in the world around us)—and the *noise model*: the possibility for production and comprehension errors (Levy, 2008; Levy, Bicknell, Slattery, & Rayner, 2009). Formally, the probability $p(s_i \mid s_p)$ for a rational comprehender to infer the intended sentence (s_i) from a perceived sentence (s_p) is proportional to both the prior probability of the intended sentence $p(s_i)$ and the likelihood of the intended sentence to be corrupted to the perceived sentence $p(s_i \rightarrow s_p)$, as is shown in Equation (1).

$$p(s_i|s_p) \propto p(s_i) \cdot p(s_i \rightarrow s_p) \tag{1}$$

In order to investigate the nature of the noise model in English, Gibson, Bergen, and Piantadosi (2013, GBP henceforth) presented participants with sentence materials from a range of syntactic alternations, in both plausible and implausible variants, and asked comprehension questions about the meanings of these sentences, as in (1)–(3):

(1) Active/Passive

(*a*) Plausible active: The girl kicked the ball.
(*b*) Plausible passive: The ball was kicked by the girl.
(*c*) Implausible active: The ball kicked the girl.
(*d*) Implausible passive: The girl was kicked by the ball.

Comprehension question: Did the girl kick something/someone?
Literal response: "Yes" for plausible, "No" for implausible

(2) Double object (DO)/Prepositional phrase object (PO):

(*a*) Plausible DO: The mother gave the daughter the candle.
(*b*) Plausible PO: The mother gave the candle to the daughter.
(*c*) Implausible DO: The mother gave the candle the daughter.
(*d*) implausible PO: The mother gave the daughter to the candle.

Comprehension question: Did the daughter receive something/someone?
Literal response: "Yes" for plausible, "No" for implausible

(3) Transitive/Intransitive

(*a*) Plausible transitive: The tax law benefited the businessman.
(*b*) Plausible intransitive: The businessman benefited from the tax law.
(*c*) Implausible transitive: The businessman benefited the tax law.
(*d*) Implausible intransitive: The tax law benefited from the businessman.

Comprehension question: Did the tax law benefit from anything?
Literal response: "No" for plausible, "Yes" for implausible

Participants' response to the comprehension question depends both on the sentence structure and the plausibility. For the plausible variants of the materials, the literal meaning of the presented word order results in a meaning that makes sense in the world, so participants answered the questions associated with the literal (plausible) meaning most of the time (over 90%). In contrast, for the implausible variants, the word order and grammatical structure suggest one meaning—an implausible meaning—whereas world knowledge suggests another more plausible meaning. Depending on the complexity of the edit—the noise—that it would take to get to the structure associated with a more likely meaning, people may make an inference that such a structure might have been intended. Behaviorally, a high inference rate is represented by a low probability of participants interpreting the test sentences literally, or in other words, a low literal interpretation rate. Throughout this paper, we say a participant "makes an inference" when they adopt a nonliteral interpretation of a sentence.

What kinds of edits were the participants considering? All of the implausible materials in these experiments can be formed by exchanging two noun phrases: for example, *the girl* and *the ball* in (1) or *the candle* and *the daughter* in (2). If the rates of literal interpretation across all the structures were the same, then one possible noise model would be one where two words or phrases of the same type were exchanged. This noise model would predict the same level of literal interpretation in the three types of alternations. But this was not the pattern that was observed. First, there was hardly any inference of such an exchange in examples like (1c) or (1d): people answered according to the literal implausible meaning over 90% of the time (in other words, less than 10% inference). Second, there was much less literal interpretation in materials like (2c), (2d), (3c), and (3d), where people often inferred the more plausible event, between 30% and 50% of the time (corresponding to 50–70% literal interpretations). And third, within the DO/PO (2) and transitive/intransitive (3) alternations, there was more inference for (2c) and (3c) than for (2d) and (3d).

Based on this pattern of data, GBP posited a noise model consisting of only insertion and deletion of function words and treating exchanges as a combination of insertions and deletions. When only one such insertion or deletion is needed—as in the examples in (2) and (3)—people will often infer the more plausible meaning. But when two such insertions or deletions are needed—as in (1)—then people will rarely infer the alternative structure with its more plausible meaning. Furthermore, they hypothesized that deletions were more likely than insertions, with the result that there would be more inference in (2c) and (3c) than in (2d) and (3d). For example, there was more inference for *The mother gave the candle the daughter* because it could be arrived at by a single deletion of the function word "to" from *The mother gave the candle to the daughter*. On the other hand, the error in (2d)—*The mother gave the daughter to the candle*—may have been caused by the accidental insertion of the function word "to," so people make fewer inferences to the more plausible event.

However, subsequent studies showed that comprehenders actually consider more than just insertions and deletions: Poppels and Levy (2016, P&L) replicated the results from GBP and

also tested how participants make inferences in materials like (4), where there are two post-verbal prepositional phrase adjuncts:

(4) Prepositional phrase adjuncts: canonical and noncanonical orders

(*a*) Plausible, canonical: The package fell from the table to the floor.
(*b*) Plausible, noncanonical: The package fell to the floor from the table.
(*c*) Implausible, canonical: The package fell from the floor to the table.
(*d*) Implausible, noncanonical: The package fell to the table from the floor.

Comprehension question: Did the package fall from the floor?
Literal response: "No" for plausible, "Yes" for implausible

Their results showed a limitation of GBP's proposed noise model, where only deletions and insertions are considered. Under the model, four such edits (two insertions and two deletions) are needed to go from the implausible versions to the plausible ones in (4a) and (4b). For example, to get from (4c) to (4a), we could delete "from" and "to," and then insert "to" and "from" in the same corresponding positions. If people rarely make edits corresponding to combinations of edits (as suggested by GBP's analysis), then participants should, therefore, rarely infer the more plausible meaning in sentences like (4c) and (4d) under this noise model. But this is not what P&L observed: participants often made an inference for materials like (4c) and (4d), and the effects are robust after taking plausibility into account.

Based on this pattern of data, P&L argued that exchanges need to be included in the noise model. How then do we explain the lack of inference in the implausible active/passive materials in (1) (which P&L replicated)? This may be explained by a constraint on exchanges: exchanges may be easier when not across a main verb, into subject position, as would be needed in the active/passive materials in (1). This constraint may be driven by people's sensitivity to the plausibility of the subject-verb relation: people may not think it likely that others could produce a sentence like *The ball kicked the girl* because *ball* is obviously not a possible agent of *kicking* and we may be able to notice this before producing such a sequence. The local subject-verb cue does not block other examples that GBP or P&L investigated.

Furthermore, Ryskin, Futrell, Kiran, and Gibson (2018) provided evidence that English participants indeed treated exchanges as a separate noise category from insertions and deletions. In Ryskin et al.'s task, participants were told to correct errors in productions that they were told other people produced. Participants were assigned to different groups where they were exposed to sentences with various types of noise operations, including deletions, insertions, and exchanges. The results suggest that how participants corrected errors correlated with the noise operations they were exposed to. Interestingly, the authors found that participants were more likely to correct sentences by insertion and less likely to correct sentences by exchanging when they were exposed to insertions or deletions, possibly because participants treated exchanges as a different category of error from insertions and deletions, which might reflect the differences in the sources of these two types of error.

Thus, the resulting noise model that can account for the results to date has two components:

(a) exchanges of category-matched items (e.g., nouns, noun phrases, or prepositions);
(b) deletions or insertions of single function words.

Two limitations of current research in the noisy-channel framework (e.g., Chen, Nathaniel, Ryskin, & Gibson, 2023; Gibson et al., 2013; Liu, Ryskin, Futrell, & Gibson, 2020; Poliak, Ryskin, Braginsky, & Gibson, 2023; Poppels & Levy, 2016; Ryskin et al., 2018; Zhang et al., 2023; Zhang, Ryskin, & Gibson, 2023) are as follows. First, few studies have been conducted in languages other than English with the noisy-channel approach. Those that have been conducted outside of English (Keshev & Meltzer-Asscher, 2021 in Hebrew; Liu et al., 2020, in Mandarin Chinese; Poliak et al., 2023, in Russian) concern how the prior probability of sentences affects the way these sentences are interpreted. For example, Keshev and Meltzer-Asscher (2021) found that Hebrew speakers avoid interpretations that would lead to low-frequency structures, even if the interpretation with a higher structural frequency has subject-verb agreement errors. In addition, English and Mandarin speakers (Liu et al., 2020) and Russian speakers (Poliak et al., 2023) interpret implausible sentences based on their structural frequency. Specifically, sentences with high-frequency word order (e.g., "the boy threw the trash") are interpreted literally more often than those with low-frequency word order (e.g., "the trash, the boy threw"). The results seem to suggest comprehenders interpret sentences based on the probability of a sentence structure. This is consistent with the noisy-channel framework in that the prior probability of a sentence affects how the sentence is interpreted.

In addition, previous studies have generally been conducted using the written modality. Specifically, participants were instructed to read sentences and then to answer comprehension questions. A question that naturally follows is whether auditory communication between a speaker and a comprehender can be accounted for using the noisy-channel framework. In Gibson et al. (2017), participants listened to recordings of sentences in Gibson et al. (2013), narrated by the experimenters either with a foreign accent or without a foreign accent. First, they replicated the main findings in Gibson et al. (2013), suggesting that participants had the same noise model when listening as when reading. Furthermore, they found that participants were less likely to interpret implausible sentences spoken with a foreign accent literally, implying that participants assumed a higher noise rate for speech in a foreign accent, giving the speaker more benefit of the doubt. In our study, one group of participants read our materials, and a different group listened to recordings of the sentence stimuli, in order to see if the effects we observed in one modality generalized to the other, as they did in Gibson et al. (2017).

In this study, we seek to systematically test how different noise operations and modalities affect comprehenders' interpretation of sentences under the noisy-channel framework, in three syntactic alternations in Mandarin Chinese: active/passive, Double object (DO)/serial verb, and transitive/intransitive, which are parallel to the active/passive, DO/PO, and transitive/intransitive alternations in English as tested in GBP. A strong test of the noise model proposed in GBP and P&L is provided by examining this typologically different language.

2. Present research

When considering similar syntactic alternations across languages, the noisy-channel framework makes different predictions depending on the number of edits that are required to obtain a plausible sentence from an implausible one. Here, we test Mandarin Chinese in three syntactic alternations adapted from three of the English ones that GBP investigated: active/passive, DO/serial verb, and transitive/intransitive. First, consider the Mandarin active/passive:

(5) Mandarin active/passive:

(a) Plausible active:
奶奶 打碎了 这个碗。
Grandma break-ASP this-CL bowl
Grandma broke the bowl

(b) Plausible passive:
这个 碗 被 奶奶 打碎了。
This-CL bowl **bei** grandma break-ASP
The bowl was broken by Grandma.

(c) Implausible active:
这个 碗 打碎了 奶奶。
This-CL bowl break-ASP grandma **[NP exchange across main verb]**
The bowl broke Grandma.

(d) Implausible passive:
奶奶 被 这个 碗 打碎了。
Grandma **bei** this-CL bowl break-ASP **[NP exchange across *bei*]**
Grandma was broken by the bowl.

Comprehension question: 奶奶打碎了某个东西/某人了吗? (Did the grandma break something/someone?)
Literal response: "No" for implausible, "Yes" for plausible

The implausible active version (5c) requires a noun exchange across the main verb, hypothesized by P&L to be a low-probability noise operation. Hence, we expect people to often take the implausible active (5c) literally. In contrast, the implausible passive version (5d) can be repaired by a noun exchange across the passive marker *bei* into (5b), which reverses the semantic roles of the agent and the patient and make the sentence plausible. Meanwhile, Implausible passive sentences such as (5d) "奶奶被这个碗打碎了" (*the grandma was broken by the bowl*) can also be a result of a substitution from the marker *ba* into the passive marker *bei*, namely, from "奶奶把这个碗打碎了" to "奶奶被这个碗打碎了," although some suggestive evidence that substitutions may be somewhat unlikely is provided by Poliak et al. (in review), who explored simple potential misinterpretations in Russian. In order to most directly connect to previous research, we started by only considering exchanges as the

possible noise operations here. **Taken together, we expect a lower literal interpretation rate in the implausible passive compared to the implausible active**.

Next, consider the Mandarin DO/serial-verb alternation as in (6):

(6) Mandarin DO/Serial Verb

(a) Plausible DO:
老林 付了 清洁工 五十块 钱。
Laolin pay-ASP cleaner fifty-CL money
Laolin paid the cleaner fifty yuan.

(b) Plausible serial verb[1]:
老林 付了 五十块 钱 给 清洁工。
Laolin pay-ASP fifty-CL money **gei** cleaner
Laolin paid fifty yuan to the cleaner.

(c) Implausible DO:
老林 付了 五十块 钱 清洁工。
Laolin pay-ASP fifty-CL money cleaner **[deletion]**
Laolin paid fifty yuan the cleaner.

(d) Implausible serial verb:
老林 付了 清洁工 给 五十块 钱 。
Laolin pay-ASP cleaner **gei** fifty-CL money **[insertion]**
Laolin paid the cleaner to fifty yuan.

Comprehension question: 清洁工收到了某个东西/某人了吗? (Did the servant receive something/someone?)

Literal response: "No" for implausible, "Yes" for plausible

The DO construction in (6a, c) is similar to the English DO construction in (2a, c). Similar to the English prepositional phrase object (PO) construction (2b, d), the Mandarin serial-verb construction (6b, d) indicates transfer of possession with a main verb (e.g., "pay") and a co-verb *gei*. The implausible DO (6c) may arise from the exchange of post-verbal noun phrases, or from the deletion of the function word *gei*. The implausible serial-verb construction (6d) may arise from the exchange of post-verbal noun phrases, or from the insertion of the function word *gei*.

Since these are relatively likely edits according to GBP and P&L, the literal interpretation rate should be low in both constructions, similar to what was observed in the English DO/PO constructions. In addition, as hypothesized in Gibson et al. (2013) that deletions are more likely to happen than insertions, we would expect a higher literal interpretation rate in the serial-verb construction than in the DO construction.

Finally, consider the Mandarin transitive/intransitive alternation, somewhat related to the English transitive/intransitive alternations in (3):

(7) Mandarin transitive/intransitive

(a) Plausible transitive:
清水 溶解了 食盐。
Clear-water dissolve-ASP salt
The clear water dissolved the salt.

(b) Plausible intransitive:
食盐 在 清水 里 溶解了。
Salt **zai** clear-water **li** dissolve-ASP
The salt dissolved in the clear water.

(c) Implausible transitive:
食盐 溶解了 清水。
Salt dissolve-ASP clear-water **[exchange across main verb]**
The salt dissolved the clear water.

(d) Implausible intransitive:
清水 在 食盐 里 溶解了。
Clear-water **zai** salt **li** dissolve-ASP **[exchange across preposition *zai*]**
The clear water dissolved in the salt.

Comprehension question: 清水是否被什么东西溶解了? (Was the water dissolved by anything?)
Literal response: "Yes" for implausible, "No" for plausible

Unlike the English transitive/intransitive materials in (3), the implausible Mandarin versions are not easily edited by adding or deleting a single function word. Rather, they are likely most easily edited by exchanging content words. Like the active in (5c), the transitive version (7c) requires a noun exchange across a main verb, whereas the intransitive (7d) does not. The intransitive versions (7c) can be repaired by exchange across the preposition *zai* (unlike English). As suggested by P&L, exchanges across function words are potentially more local than exchanges across a main verb (7c). Furthermore, the speech error data from Garrett (1975) suggest that exchanges across a main verb are less common than exchanges across function words. Based on this observation, **we expect a lower interpretation rate in the intransitive (7d) compared to the transitive (7c)**.

In addition to the above within-alternation predictions, the noisy-channel theory makes the following between-alternation predictions. First, we expect **lower rates of literal interpretation in the DO/serial-verb alternation in (6c, d) than in the other two types of alternation**. This is because the implausible DO/serial-verb materials only require a single deletion or insertion of a function word to make them plausible, whereas the other implausible

materials require an exchange or substitution in order to make them plausible. Second, we expect **lower literal interpretation rate for (5d) and (7d), both of which require an exchange across a function word, compared to (5c) and (7c), which require an exchange across a main verb**. This is a more general comparison than (5d) versus (5c) and (7d) versus (7c) described above. If each of these is significant in the predicted direction, then this test should be too. A summary of noise operations and the corresponding plausible-implausible sentences as a result of these operations is listed in Table 1.

The rest of the paper is organized as follows. In Experiment 1, we tested the noisy-channel framework on Mandarin Chinese speakers, who read sentences under the three syntactic alternations: active/passive, DO/serial verb, and transitive/intransitive, listed in Table 1, testing the noise operations of insertions, deletions, and exchanges. Experiment 2 extended Experiment 1 to the auditory modality, in that participants listened to the same test sentences instead of reading them. Despite Mandarin Chinese being typologically different, we largely replicated the results in Gibson et al. (2013), and the differences were probably because participants arrived at interpretations that we did not anticipate (for the DO/serial-verb constructions), which we will elaborate further in the following sections.

3. Experiment 1

3.1. Methods

3.1.1. Ethics statement

This study (both experiments included) was approved by the Committee on the Use of Humans as Experimental Subjects at Massachusetts Institute of Technology. Participants completed the study through a web interface, which provided an informed consent form prior to beginning the study and indicated that initiating the study constituted their informed consent to participate.

3.1.2. Participants

Two hundred and nineteen self-reported Mandarin native speakers took part in one of the subexperiments. Participants were recruited from Witmart, a China-based crowd-sourcing platform.

3.1.3. Materials

Each of the three subexperiments had a 2×2 within-subject design, varying syntactic constructions (e.g., active vs. passive) and plausibility (plausible vs. implausible). Experiment 1A investigated the active/passive alternation as in (5); Experiment 1B investigated the DO/serial-verb alternation as in (6); Experiment 1C investigated the transitive/intransitive alternation as in (7). Each subexperiment had 20 target items[2] and 60 filler items. The filler items were shared across three subexperiments.

Each item contained a comprehension question (see above examples (5)–(7)). The answer to each question indicates whether literal syntax or semantic plausibility was used in sentence

Table 1
A summary of intended plausible sentences, noise operations, and the resulting perceived implausible sentences in Mandarin

Intended plausible sentence	Possible noise operation	Perceived implausible sentence
奶奶 打碎了 这个 碗 Grandma break-ASP this-CL bowl "Grandma broke the bowl." (5a)—plausible active	Exchange across main verb	这个 碗 打碎了 奶奶 This-CL bowl break-ASP grandma "The bowl broke grandma." (5c)—implausible active
这个 碗 被 奶奶 打碎了 This-CL bowl **bei** grandma break-ASP "This bowl was broken by Grandma" (5b)—plausible passive	Exchange across the passive marker *bei*	奶奶 被 这个 碗 打碎了 Grandma **bei** this-CL bowl break-ASP "Grandma was broken by the bowl." (with *bei* passive marker) (5d)—implausible passive
老林 付了 清洁工 五十块 钱 Laolin pay-ASP cleaner fifty-CL money "Laolin paid the cleaner 50 Yuan." (6a)—plausible DO	Insertion	老林 付了 清洁工 给 五十块 钱 Laolin pay-ASP cleaner **gei** fifty-CL money "Laolin paid the cleaner **to** 50 Yuan." (6d)—implausible serial verb
老林 付了 五十块 钱 给 清洁工 Laolin pay-ASP fifty-CL money **gei** cleaner "Laolin paid 50 Yuan **to** the cleaner." (6b)—plausible serial verb	Deletion	老林 付了 五十块 钱 清洁工 Laolin pay-ASP fifty-CL money cleaner "Laolin paid 50 Yuan the cleaner." (6c)—implausible DO
清水 溶解了 食盐 Clear-water dissolve-ASP salt "The clear water dissolved the salt" (7a)—plausible transitive	Exchange across the main verb	食盐 溶解了 清水 Salt dissolve-ASP clear-water "The salt dissolved the clear water." (7c)—implausible transitive
食盐 在 清水 里溶解了 Salt **zai** clear-water **li** dissolve-ASP "The salt dissolved in the clear water." (7b)—plausible intransitive	Exchange across preposition *zai*	清水 在食盐里溶解了 Clear-water **zai** salt **li** dissolve-ASP "The clear water dissolved in the salt" (7d)—implausible intransitive

Note. The color marks the NPs that are exchanged, and the underscore and bold font indicates characters that are inserted or deleted in the noise model.

interpretation. For example, in (5), for the implausible conditions, a "no" answer indicated the use of literal syntax in interpretation, whereas a "yes" answer indicated the reader relied on semantic plausibility in interpretation, and the opposite is true for (7). The items were

counterbalanced so that half of the questions were like the ones in (5) and (6), and the other half of the questions were like the ones in (7).

3.1.4. Procedure

Similar to GBP, participants were presented with 20 target items, interspersed with 60 filler items. Following a Latin square design, one of the four versions of each item was displayed with a yes-no comprehension question. The sentences were presented one by one in full. After reading each sentence, participants answered a comprehension question that indicated whether they had a literal or nonliteral interpretation of the sentence. The questions were counterbalanced for their a priori plausibility (e.g., Did the {grandma/bowl} break something/someone? In Chinese, {奶奶/这个碗} 打碎了某个东西/某人了吗). Half of the questions were plausible as written, and the other half were implausible. There was no time limit in reading each sentence, but they were not able to go back to the previous sentence once they proceeded to the next one.

3.1.5. Statistical analysis

The data were analyzed using mixed-effect logistic regression models powered by R (R Core Team, 2022), with the maximal random effect structure justified by the design of the experiment (Barr, Levy, Scheepers, & Tily, 2013). We adopted a Bayesian approach and carried out the analyses using the MCMCglmm package (Hadfield, 2010), under an uninformative prior, a common practice for MCMC-based mixed model fitting (Baayen et al., 2008). In the results, we reported $p_$MCMC, a Bayesian counterpart of p-value based on the posterior distribution of the regression model, and we said a result was significant if $p\_\text{MCMC} < 0.05$.

We were mainly interested in participants' interpretation of implausible sentences because in GBP and subsequent studies, participants overwhelmingly interpreted plausible sentences literally and hence the literal interpretation rate of those sentences were near ceiling.

We made three comparisons on participants' responses to implausible sentences (summarized in Table 2). We investigated **whether sentences made implausible by an exchange across a main verb were more likely to be interpreted literally than those made implausible by an exchange across a function word**. To do so, we compared the literal interpretation rate of implausible active (5c) and implausible transitive (7c) sentences and that of implausible passive (5d) and implausible intransitive (7d) sentences. We call the two groups of constructions in this analysis **construction groups**. The construction group was entered as a fixed effect. Items and participants were entered as random intercepts with random by-item and by-participant slopes for construction group. In addition, within the active/passive alternation and the transitive/intransitive alternation, we compared the literal interpretation rate of each construction. Here, the construction was entered as a fixed effect. Items and participants were entered as random intercepts with random by-item and by-participant slopes for construction.

We then checked **whether sentences made implausible by insertions were more likely to be interpreted literally than those made implausible by deletions**. We compared the literal interpretation rate of DO sentences with that of serial-verb sentences. Here, the construction

Table 2
A summary of the analyses done in Experiment 1, grouped by the noise operations each analysis compares

Comparison 1: Exchange across a main verb versus exchange across a function word	
active (5c) + transitive (7c) versus passive (5d) + intransitive (7d)	response ~ construction group + (1 + construction group | participant) + (1 + construction group | item)
active versus passive transitive versus intransitive	response ~ construction + (1 + construction | participant) + (1 + construction | item)
Comparison 2: Insertion versus deletion	
serial verb (6d) versus DO (6c)	response ~ construction + (1 + construction | participant) + (1 + construction | item)
Comparison 3: Exchange versus insertion/deletion	
active/passive versus DO/serial verb transitive/intransitive versus DO/serial verb	response ~ alternation + (1 | participant) + (1 | item)

Note. The left column specifies the fixed effect variables in comparison and the right column shows the fixed effect and random effect structure in lme4 syntax (Bates, Maechler, Bolker, & Walker, 2015).

was entered as a fixed effect. Items and participants were entered as random intercepts with random by-item and by-participant slopes for construction.

Finally, we examined **whether sentences made implausible by exchanges were more likely to be interpreted literally than those made implausible by insertions or deletions**, by comparing the literal interpretation rate across alternations. There were two analyses in this part: one comparing DO/serial-verb sentences with active/passive sentences, and another comparing DO/serial-verb sentences with transitive/intransitive sentences. In each analysis, the alternation was entered as a fixed effect. Items and participants were entered as random intercepts.

3.2. Results

The literal interpretation rate of sentences in each construction is presented in Fig. 1. Consistent with GBP and other previous studies, in all three syntactic alternations, the plausible materials were interpreted literally much more often than the implausible materials (ps $<$ .001), and furthermore, the rates of literal interpretation of the plausible materials were near ceiling.

Consequently, we omitted the plausible materials from further analyses (similar to the analysis procedure that GBP followed).

The results from subsequent analyses are summarized in Table 3. We found that active sentences (5c) and transitive sentences (7c) were more likely to be interpreted literally, compared with passive sentences (5d) and intransitive sentences (7d, p_MCMC < 0.001). This is still true when the construction groups were broken down, in that active sentences were more likely to be interpreted literally than passive sentences (p_MCMC $= 0.042$), and transitive sentences were more likely to to be interpreted literally than intransitive sentences (p_MCMC < 0.001). All three analyses were consistent with our predictions that exchanges across a

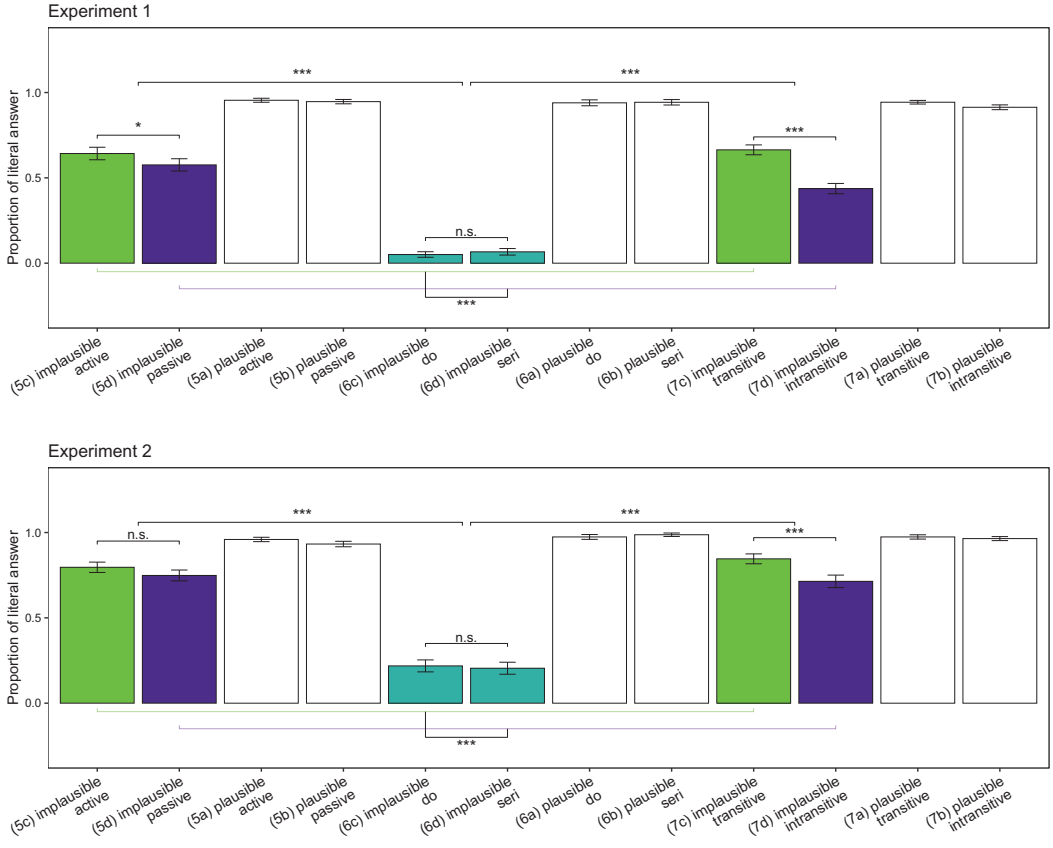

Fig. 1. Percentage of trials where participants relied on the literal syntax for interpretation of the sentences presented in: written form (Experiment 1, top panel) and spoken form (Experiment 2, bottom panel). The symbols ***, *, and n.s. indicate $p<.001$, $p<.05$ and not significant, respectively. Error bars indicate standard errors.

function word are more likely than exchanges across a main verb. Nevertheless, we did not find a significant difference between the literal interpretation rate of DO (6c) sentences and that of serial-verb (6d) sentences (p_MCMC $= 0.617$). Across alternations, we found participants were more likely to adopt a nonliteral interpretation of the implausible DO/serial-verb materials than of implausible materials under other alternations, as predicted (p_MCMCs $<$ 0.001).

3.3. Discussion

Overall, these results were broadly as predicted by the noisy-channel framework under the noise model that was proposed based on English data.

First, implausible passive sentences and implausible intransitive sentences were interpreted nonliterally more often than implausible active sentences and implausible transitive sentences,

Table 3
Results of the mixed-effect logistic regression in Experiment 1 analyzing the effects of different noise operations on literal interpretation, namely: effects of exchanges across a main verb versus across a function word (Comparison 1), effects of insertion versus deletion (Comparison 2), and effects of exchange versus insertion or deletion (Comparison 3)

Experiment 1[4]				
Comparison 1: Exchange across a main verb versus exchange across a function word				
Variables	2.5% CI	Mean	97.5% CI	*p*_MCMC
Active (5c) + Transitive (7c) versus Passive (5d) + Intransitive (7d)	0.698	1.076	1.472	<0.001***
Active versus Passive	−4.500	−0.256	−0.005	0.042*
Transitive versus Intransitive	−1.187	−0.861	−0.542	<0.001***
Comparison 2: Insertion versus deletion				
Variables	2.5% CI	Mean	97.5% CI	*p*_MCMC
DO (6c) versus Serial verb (6d)	−0.504	0.156	0.796	0.617
Comparison 3: Exchange versus insertion/deletion				
Variables	2.5% CI	Mean	97.5% CI	*p*_MCMC
Active/Passive versus DO/Serial verb	4.457	5.619	6.800	<0.001***
Transitive/Intransitive versus DO/Serial verb	−5.841	−4.677	−3.639	<0.001***

Note. *, and *** indicate the *p*-value is below .05, below .01, and below .001, respectively.

possibly because the former two require an exchange across a function word, whereas the latter two require an exchange across a main verb.

More specifically, implausible active sentences were interpreted literally more often than implausible passive sentences, because the edit to form a plausible version of the active sentence—an exchange across a main verb—is less likely than the edit to form a plausible version of the passive sentence—an exchange across the particle *bei*.

Similarly, implausible transitive sentences were interpreted literally more often than implausible intransitive sentences, because the edit to form a plausible version of the transitive sentence—an exchange across a main verb—is less likely than the edit to form a plausible version of the intransitive sentence—an exchange across the particle *zai*.

Moreover, there were more inferences in the implausible DO/Serial alternation than in the other alternations, plausibly because the edit to a plausible version consists of a single deletion or insertion.

However, in contrast to the prediction of the noisy-channel framework, the implausible serial-verb construction was generally interpreted nonliterally. The implausible serial-verb construction requires either the insertion of a function word, or an exchange (but not across a

main verb). Consequently, the noisy-channel framework predicts some literal interpretations, above floor, contrary to observation.

After obtaining these results, our subsequent discussion with several other native Mandarin speakers revealed some variability in whether people think that our "implausible" materials are actually ungrammatical for the DO/serial-Verb constructions. In some dialects—possibly more often in colloquial Mandarin—the supposedly implausible (6c) is acceptable and yields the same plausible interpretation as (6a) and (6b). The same applies to the implausible (6d), especially if there is a pause following the second noun phrase ("cleaner" in (6d)), perhaps with an aspect marker deleted following the verb "gei" (give). This turns (6d) into (8).

(8) 老林 付了 清洁工, 给了 五十块 钱。
Laolin pay-ASP cleaner **gei-ASP** fifty-CL money
Laolin paid the cleaner. (Laolin) gave (the cleaner) fifty yuan.

This observation suggests that some Mandarin speakers may have conventionalized supposedly implausible DO strings (6c) as plausible sentences without any noise inference. If so, the results from DO sentences would not test the noisy-channel hypothesis for these speakers.[3] Meanwhile, (8) is a plausible and grammatical sentence under a different construction: here, *gei* becomes a main verb instead of a serial verb, making the sentence a compound. It is possible that when participants were reading implausible serial-verb sentences like (6d), they were actually inferring (8) as the plausible interpretation, instead of (6a) as we intended. Since the plausible alternative (8) has a high prior, and deletions that turn (8) to (6d) are likely to happen, participants might indeed have inferred a deletion from (8) instead of an insertion from (6a), which could potentially explain the low literal interpretation rate and a lack of difference between the results in the DO sentences and the serial-verb sentences.

Another potential question that arises from comparing the results from these experiments to results from English in Gibson et al. (2013) is that the percentage of literal interpretation in Mandarin is lower than what was found in English for similar construction pairs. There are many possible sources of such between-language differences. One possible source is that Mandarin orthography—with no spaces between the words—may make local word order permutations harder to detect by readers. In English, there are white spaces between words to indicate word boundaries, but there are no such spaces in written Mandarin. There have been numerous studies on how spacing affects Chinese reading (e.g., Bai, Yan, Liversedge, Zang, & Rayner, 2008; Hsu & Huang, 2000a, 2000b; Inhoff, Liu, Wang, & Fu, 1997; Liu & Li, 2014; Liu & Lu, 2018). These studies suggest that for native speakers, adding space between *words* does not affect reading, whereas adding space between *characters* affects reading. In addition, Gu and Li (2015) show that sentences with a transposition of characters between words (e.g., 庄严肃穆 vs. 庄肃严穆) take longer for readers to comprehend than control sentences, but those with a transposition of characters within words (e.g., 剑拔弩张 vs. 剑弩拔张) do not. Taken together, these findings suggest that Chinese readers actively segment sentences into words (Li & Pollatsek, 2020; Li, Rayner, & Cave, 2009), a process not present among English readers, and the Chinese character order within a word seems to not affect reading (Gu & Li, 2015; see Rayner et al., 2006 for a similar finding in English). In addition, when reading alphabet-based writing scripts such as English scripts, readers are able

to decide where to land their next saccade based on cues such as word length, but in contrast, Chinese readers do not have such cues due to a lack of spacing between words. It has been suggested that Chinese readers first obtain as much information from a fixation as possible and then saccade to the next chunk of text containing completely novel information. The saccade length may then be determined by factors such as word frequency and word length (Li, Liu, & Rayner, 2011; Wei, Li, & Pollatsek, 2013). Therefore, it is possible that due to the lack of space between Chinese characters, noise operations such as deletions, especially those after a high-frequency word, might be unnoticed by the reader. If so, the reader could have the impression of comprehending a plausible sentence despite seeing an implausible sentence. This might make the overall literal interpretation rate in Mandarin lower than in English.

To evaluate this idea, we ran Experiment 2 with auditory versions of the same materials. We suspected that in the auditory modality, speakers could use intonational cues associated with constituent boundaries, such as pauses (or lengthening, and shifts in pitch), to let listeners interpret *gei* as a serial verb, instead of a main verb. Then, if participants hear sentences such as (6d) without any boundary between "cleaner" (清洁工) and the serial verb "gei," they might be less likely to interpret the sentence like (8), where a boundary is needed between "cleaner" and "gei." In addition, we suspected that the intonational cues would also help participants interpret materials like (6c) as DO, in order to rule out the alternative plausible interpretation due to dialectal differences or failure to notice the deletion of the serial verb.

A previous study (Gibson et al., 2017) tested the noisy-channel framework in the auditory modality, among native English speakers. The results from the study follow the same patterns as GBP, suggesting that participants might have assumed a similar noise model when given an auditory input, compared to when given a written input. Therefore, in Experiment 2 of our study, we would expect noisy-channel inference from participants but probably with less misinterpretation due to the additional intonational cues.

4. Experiment 2

There were two goals to Experiment 2. First, we wanted to replicate the results from Experiment 1 on a new group of participants. Second, we wanted to test whether switching the modality of the stimuli might alter the literal interpretation rate for implausible sentences in Mandarin. This might be especially useful for testing the noisy-channel theory's predictions regarding the implausible DO/Serial materials: they might be interpreted implausibly more in auditory versions, thus establishing them as reasonable baselines for the other comparisons.

4.1. Methods

4.1.1. Participants

A total of 288 participants were recruited. All participants were self-reported native Mandarin Chinese speakers recruited on Amazon Mechanical Turk. Chinese proficiency of all participants was screened through three open-ended questions in Mandarin at the beginning of the experiment. The three screening questions that we used were as follows:

"What is your favorite food and why/你最喜欢的食物是什么?为什么?"
"Please briefly introduce your hometown/请简短介绍一下你的家乡"
"What's your favorite sport/你最喜欢什么运动?"

Participants were asked to type in their answers in the text boxes provided below each question. Participants who did not answer the screening questions or did not answer them in Mandarin Chinese were excluded from the experiment. One hundred and ninety four participants who passed the screening and did not have duplicated submissions were included in the analysis.

4.1.2. Materials

As in Experiment 1, Experiment 2 had three subexperiments: (2A) active/passive sentences as in (5); (2B) DO/serial-verb sentences as in (6); and (2C) transitive/intransitive sentences as in (7). Each subexperiment had 20 test items and 60 filler items. The materials were audio-recorded versions of the materials from Experiment 1 narrated by one of the authors as natural speech. In each subexperiment, syntactic construction and plausibility were manipulated as independent factors to form a 2×2 factorial design. We employed a latin-square design to form four lists for each subexperiment, and all lists were randomized.

4.1.3. Procedure

All 80 test items for each subexperiment (2A, 2B, 2C) were presented on a single page. There were 80 rows in total, with each row containing a play button for an audio recording, a comprehension question (written), and a Yes/No choice bubble for the comprehension question. Participants clicked the "play" button on each row of the list to play the recording for each sentence. Comprehension questions were presented visually. Participants could see the comprehension question as they were listening to the recordings. There was no time limit in completing each comprehension question, as long as participants submitted their responses within 2 h.

4.1.4. Statistical analysis

The statistical analyses were largely identical to those conducted in Experiment 1 (see Table 2 for the summary).

4.2. Results

Two-thirds of the results of Experiment 2 (shown in the bottom panel of Fig. 1) replicated the results from Experiment 1.

The results are graphed in Fig. 1 and summarized in Table 4. First, in all three alternations, the plausible materials were interpreted literally much more often than the implausible materials ($ps < .001$), and the rates of literal interpretation of the plausible materials were near ceiling. As in Experiment 1, we, therefore, omitted the plausible materials from further analyses.

Then, as in Experiment 1, we compared the literal interpretation rate among sentences that are made implausible by different types of exchanges. We found that participants are more

Table 4
Results of the mixed-effect logistic regression in Experiment 2, analyzing the effects of different noise operations on literal interpretation (same as in Experiment 1)

Experiment 2				
Comparison 1: Exchange across a main verb versus exchange across a function word				
Variables	2.5% CI	Mean	97.5% CI	p_MCMC
Active (5c) + Transitive (7c) versus Passive (5d) + Intransitive (7d)	−1.189	−0.813	−0.365	< 0.001***
Active versus Passive	−0.154	0.415	0.932	0.120
Transitive versus Intransitive	−2.470	−1.701	−0.965	< 0.001***
Comparison 2: Insertion versus deletion				
Variables	2.5% CI	Mean	97.5% CI	p_MCMC
DO (6c) versus Serial verb (6d)	−0.298	0.541	1.495	0.206
Comparison 3: Exchange versus insertion/deletion				
Variables	2.5% CI	Mean	97.5% CI	p_MCMC
Active/Passive versus DO/Serial verb	−5.115	−4.505	−3.856	< 0.001***
Transitive/Intransitive versus DO/Serial verb	3.770	4.334	4.860	< 0.001***

Note. We report the mean of the posterior distribution and the boundaries of the 95% credible interval. *** indicate p_MCMC is below 0.05, below 0.01, and below 0.001, respectively.

likely to make inferences of sentences made implausible by an exchange across function words than those made implausible by an exchange across main verbs (p_MCMC < 0.001).

We also compared the literal interpretation rates for each alternation. First, active implausible sentences (5c) were interpreted literally numerically more than the passive implausible sentences (5d) but this comparison did not reach significance (p_MCMC = 0.120). Second, there was no significant difference between the implausible DO (6c) and serial-verb constructions (6d, p_MCMC = 0.206). Third, transitive implausible sentences (7c) were interpreted literally more than intransitive implausible sentences (7d, p_MCMC < 0.001), as predicted.

In our cross-alternation comparisons, participants were more likely to adopt a nonliteral interpretation of the implausible DO/serial-verb materials than for the implausible materials in other syntactic alternations (p_MCMC < 0.001 for Active-Passive vs. DO-Serial Verb and p_MCMC < 0.001 for Transitive-Intransitive vs. DO-Serial Verb). Unlike for the written DO-Serial verb materials, the literal interpretation of the auditory DO-Serial verb materials was well above the baseline, suggesting that these are in fact implausible materials for many participants (as originally designed).

Table 5
A summary of the analyses done combining results from both experiments, grouped by the noise operations each analysis compares

Comparison 1: Exchange across a main verb versus exchange across a function word	
active + transitive versus passive + intransitive	response ~ construction group * modality + (1 + construction group | participant) + (1 + construction group * modality | item)
active versus passive transitive versus intransitive	response ~ construction * modality + (1 + construction | participant) + (1 + construction * modality | item)
Comparison 2: Insertion versus deletion	
serial verb (6d) versus DO (6c)	response ~ construction + (1 + construction | participant) + (1 + construction | item)
Comparison 3: Exchange versus insertion/deletion	
active/passive versus DO/serial verb transitive/intransitive versus DO/serial verb	response ~ alternation * modality + (1 | participant) + (1 + modality | item)

Note. The left column specifies the fixed effect variables in comparison and the right column shows the fixed effect and random effect structure in lme4 syntax (Bates et al., 2015).

4.3. Analyzing the results from both experiments

To analyze the effect of different noise operations, modality, and their interactions altogether, we polled the data from Experiments 1 and 2 and conducted a mixed-effect logistic regression. The comparisons remained the same, except in each analysis, we added a fixed effect of modality and an interaction term. The random effect structure also varied accordingly in order to achieve the most complex structure permitted by experimental design (Barr et al., 2013). A summary of the analyses is shown in Table 5.

The results are presented in Table 6. Participants were much less likely to interpret sentences literally when they were presented with passive or intransitive sentences, than when they were presented with active or transitive sentences (p_MCMC < 0.001), as predicted. They also made significantly fewer literal interpretations when they were presented with written sentences (p_MCMC < 0.001). Again, we also found no interaction between type of exchanges and modality (p_MCMC $= 0.243$). The more fine-grained analyses also showed a main effect of modality (p_MCMCs < 0.001) and no interactions between construction and modality (p_MCMCs > 0.3). However, they did not replicate the results in Experiment 1, in that we found only an effect of construction among transitive/intransitive sentences (p_MCMC < 0.001) but not among active/passive sentences (p_MCMC $= 0.134$).

Similar to Experiments 1 and 2, we did not find a significant difference between the literal interpretation rate of DO sentences and that of serial-verb sentences (p_MCMC $= 0.297$), but

Table 6
Results of the mixed-effect logistic regression on data pooled from both experiments, analyzing the effects of different noise operations and modality on literal interpretation (same as in Experiments 1 and 2), along with their interactions

Pooled data from Experiments 1 and 2					
Comparison 1: Exchange across a main verb versus exchange across a function word					
Variable	Fixed effect	2.5% CI	Mean	97.5% CI	p_MCMC
Active+Transitive versus Passive+Intransitive	Construction group	−1.135	−0.751	−0.317	< 0.001***
	Modality	−1.860	−1.288	−0.650	< 0.001***
	Interaction	−0.850	−0.330	0.236	0.243
Active versus passive	Construction	−0.100	0.428	1.020	0.134
	Modality	0.499	1.318	2.192	< 0.001***
	Interaction	−0.684	0.030	0.867	0.940
Transitive versus intransitive	Construction	−1.870	−1.298	−0.628	< 0.001***
	Modality	1.233	1.996	2.882	< 0.001***
	Interaction	−1.282	−0.434	0.394	0.300
Comparison 2: Insertion versus deletion					
DO versus Serial verb	Construction	−0.409	0.463	1.313	0.297
	Modality	1.419	2.761	4.005	< 0.001***
	Interaction	−1.450	−0.124	1.187	0.854
Comparison 3: Exchange versus insertion/deletion					
Active/Passive versus DO/Serial verb	Alternation	−5.757	−4.970	−4.215	< 0.001***
	Modality	−2.135	−1.512	−0.811	< 0.001***
	Interaction	−1.559	−0.481	0.399	0.320
DO/Serial verb versus Transitive/Intransitive	Alternation	3.740	4.469	5.158	< 0.001***
	Modality	−3.090	−2.235	−1.226	< 0.001***
	Interaction	−0.652	0.361	1.565	0.523

Note. We report the mean of the posterior distribution and the boundaries of the 95% credible interval. *, **, and *** indicate p_MCMC is below 0.05, below 0.01, and below 0.001, respectively.

we did find a significant difference between modalities (p_MCMC < 0.001) and no interactions between modality and construction (p_MCMC = 0.854).

Across alternations, as predicted, the participants made significantly less literal interpretation in DO-Serial Verb alternation compared to Active-Passive alternation and Transitive-Intransitive alternation (p_MCMC < 0.001 for Active-Passive vs. DO-Serial Verb and p_MCMC < 0.001 for Transitive-Intransitive vs. DO-Serial Verb). Compared with those in the written experiment, participants in the auditory experiment made significantly more literal interpretation (p_MCMC < 0.001 in both analyses). Similar to the within-subexperiment analysis above, we found no interaction between alternation and modality (p_MCMCs > 0.320).

4.4. Discussion

The results from Experiment 2 are broadly consistent with what was found in Experiment 1. The lack of interaction between modality and construction in all three syntactic alternations suggests that the level of literal sentence interpretation derived from different syntactic constructions does not depend on the modality. Although the difference in Active/Passive did not reach significance in Experiment 2, the direction of this effect was the same as in Experiment 1, and there was no interaction across modality in the two experiments.

In addition, we found a main effect of modality in all three subexperiments, such that there was less literal interpretation in implausible sentences in reading compared to the auditory versions. This is compatible with our speculation that the lower rate of literal interpretation of implausible sentences in Mandarin Chinese (as seen in Experiment 1) compared to English might be partially due to the lack of white space between written words in Chinese. Further cross-linguistic studies testing potential effects of writing system properties (such as whether white spaces are used to mark word boundaries, as in Gu and Li (2015), for example) on literal interpretation rate would be necessary to test our speculation that the lack of white spaces between written words lowers the proportion of literal interpretation.

Interestingly, in the auditory modality, the literal interpretation rate in DO sentences is also about the same as that in serial-verb sentences, suggesting that even with the intonation cues, participants still tended to interpret implausible serial-verb materials like (6d) as sentences like (8). To better test the effect of deletions and insertions, future studies should consider other syntactic alternations where implausible sentences are less likely to be interpreted as plausible ones with a different meaning. In addition, collecting data such as acceptability or asking participants what the speaker might have intended to say would also be helpful in making sure that participants are indeed interpreting sentences the way we intended.

5. General discussion

There have been numerous studies investigating the effects of noise in English sentence comprehension under a noisy-channel theory (e.g., Bergen, Levy, & Gibson, 2012; Cai, Zhao, & Pickering, 2022; Gibson et al., 2013; Kane & Slevc, 2019; Levy, 2008; Poppels & Levy, 2016; Ryskin et al., 2018; Staub, Dodge, & Cohen, 2018), but few have tested the theory in other languages (e.g., Keshev & Meltzer-Asscher, 2021; Liu et al., 2020; cf. related work in German by Bader & Meng, 2018; and Meng & Bader, 2021). In this work, we tested the noisy-channel theory on three different alternations in Mandarin: Active/Passive, DO/Serial Verb, and Transitive/Intransitive, each with both plausible controls and implausible versions, which could be made plausible with some minimal edits, as presented in Table 1.

Specifically, the noise model developed from work on English interpretation predicts that the implausible DO materials might be interpreted as having an accidental insertion of a function word; and the implausible serial-verb materials might be interpreted as having an accidental deletion of a function word. These edits have been shown to drive high inference rates in English, and hence we expected high inference in Mandarin, which is what we observed.

However, we also expected more inference for the implausible serial-verb materials than for the implausible DO materials because they involve a deletion, which has been shown to drive higher inference rates in English than insertions (as in the implausible DO materials), but we did not see the predicted difference here. Our results could have been because participants made types of inferences that we did not expect when we designed the stimuli. In particular, participants might have conventionalized implausible DO sentences such as (6c) as plausible. The conventionalization might have been a result of noisy-channel inference, in that if comprehenders always interpreted sentences like (6c) as a deletion from plausible PO sentences like (6b), they might have interpreted sentences like (6c) as plausible on its own. On the other hand, the low literal interpretation rate in serial-verb sentences like (6d) might have been because participants inferred a deletion from plausible sentences like (8), instead of an insertion from plausible DO sentences like (6a).

Each of the other four implausible constructions involved exchanges in order to make them plausible. The noise model developed from English suggests that materials that were correctable by single deletions and insertions of function words were more likely to drive inference than exchanges of nouns. Thus, we expected more inference in implausible DO and serial-verb constructions than in the other four implausible constructions, and this is exactly what we saw.

For the implausible passive (5d) and the implausible intransitive (7d) structures, the minimal exchange that is needed is an exchange across a function word. For the implausible active (5c) and the implausible transitive (7c) structures, the minimal exchange that is needed is an exchange across a main verb. The noise model developed from English suggests that exchanges across function words are more likely than those across main verbs, so we expected more inference in implausible passive sentences and in implausible intransitive sentences, relative to implausible active sentences and implausible transitive sentences. We found this pattern of data in general in both experiments. First, it was robustly the case that implausible exchange-across-function word materials were inferred as their plausible variants more than the implausible exchange-across-verb materials, as a whole. When the constructions were considered individually, many of these tests were significant: only the implausible active versus implausible passive comparison in Experiment 2 was not.

In addition to finding support for the noisy-channel model in these constructions, we also found that there was more inference with written Mandarin materials compared to auditory materials. We speculate that this difference may be partly dependent on a lack of spaces between characters in Mandarin Chinese orthography, which may make it difficult for participants to detect implausible sentences, similar to the observation that English speakers have a hard time noticing transposed characters in written words, such as in Rayner et al.'s classic sentence materials, "raeding wrods with jubmled lettres."

Although the results of our experiments suggest that Mandarin speakers seem to be sensitive to the same kinds of noise when interpreting implausible materials as English speakers, the inference rate in the Mandarin implausible active and passive constructions appears to be higher in Mandarin (e.g., 64.3% literal for implausible active in E1, 57.6% literal for implausible passive in E1, 79.6% literal for implausible active in E2, and 74.9% literal for implausible passive in E2) compared to the results among English speakers reported in two studies

conducted in different time periods, namely, Gibson et al. (2013) (98.6% literal for implausible active; 96.8% literal for implausible passive), and Gibson et al. (2017) (93.6% literal for implausible active; 93.5% literal for implausible passive). This difference in inference rate may be due to additional cues that Mandarin speakers use in sentence processing, such as the passive marker *bei* and noun animacy (Li, Bates, & MacWhinney, 1993). That is, Mandarin speakers rely heavily on both noun animacy and word order when interpreting simple sentences, as opposed to English speakers who appear to rely only on word order when interpreting simple sentences (Su, 2001). There are many situations where an animacy cue and a word-order cue conflict (such as when there is a post-verbal animate object, or a pre-verbal inanimate subject, as in "the rock fell on the hiker"). English speakers rely overwhelmingly on word-order, whereas Mandarin speakers pay close attention to animacy cues to meaning. Thus, English speakers will make few inferences on an implausible sentence like "the pizza ate the boy," whereas Mandarin speakers will make many inferences in such situations.

Although we present these results in terms of noisy-channel theory, our results are also broadly consistent with the "good enough" proposal of language processing (e.g., Christianson, Williams, Zacks, & Ferreira, 2006; Ferreira, 2003; Ferreira & Lowder, 2016; Ferreira & Patson, 2007). According to the good-enough framework, there are two routes to interpretation: an exact "algorithmic" route; and an approximate "heuristic" route. We may use one or the other of these routes depending on our goals in the task at hand. One possibility is that the noisy-channel theory is a principled formalization of the idea that interpretations are faithful to the grammar but not entirely faithful to the input, a core idea in the good-enough framework. Compared to the good-enough approach, the noisy-channel approach provides a more fine-grained, quantitative account on which syntactic structure is more likely to be overwhelmed by the initial analysis, in the form of a noise model, and the results so far generally agree with the prediction. Another account broadly consistent with our results is the prediction account proposed in Cai et al. (2022). When the comprehender reads an implausible DO/PO sentence, such as "the mother gave the candle the daughter," the main verb "gave" in the sentence predicts two analyses: one that is literal but implausible (in this case, a DO analysis), and another that is nonliteral but semantically plausible (a PO analysis). The comprehender is likely to form a nonliteral interpretation of DO/PO sentences because the nonliteral analysis has a relatively strong activation. In contrast, when the comprehender reads an implausible Active/Passive sentence, such as "the ball kicked the boy," the main verb "kicked" only predicts one analysis: the active analysis, which explains why the comprehender is much more likely to interpret this sentence literally than in the previous case. We leave it for future works to tease these accounts apart.

There are a number of limitations of the current study. First, we started with the assumption that the noise model in the Chinese script is similar to that in the English script, in that we considered the same types of noise operations (insertions, deletions, and exchanges) as in previous studies in English. However, differences in these writing systems could indeed give rise to differences in the noise model. For instance, speculatively, substitutions in Chinese script could potentially occur more frequently than in an alphabet-based script, especially nowadays as more and more Chinese scripts are generated electronically with pinyin-based input methods. To type a Chinese character, a user first types its *pinyin*, a representation

of the character's pronunciation in Latin script, and then chooses the desired character in the drop-down menu. Hence, substitution of a character of the same pronunciation could occur if the user chooses the wrong character. However, according to the authors who are native Chinese speakers, this sort of substitution is still not likely to result in syntactically licit while semantically implausible sentences. Second, our predictions are all only regarding the noise model, or in other words, differences in the literal interpretation rate in sentences under various constructions are a result of differences in the likelihood of noise operations. In fact, this is only one of the factors affecting how people interpret implausible sentences: the other one being the prior, including a meaning prior and a structural prior (Liu et al., 2020; Poliak et al., 2023; Poppels & Levy, 2016). For example, it is possible that the difference in literal interpretation rate in Active/Passive sentences and in Transitive/Intransitive sentences are at least partially due to differences in structural frequency, causing a difference in the prior term. To further investigate the effect of prior, including the structural frequency within a syntactic alternation, future studies shall gather the frequency of both constructions for each main verb in the experimental material in a corpus, as well as the plausibility norming data from participants.

In conclusion, this work demonstrates that similar to English speakers, Mandarin speakers also interpret implausible sentences in a rational approach, integrating both the likelihood of various ways of corruption and their world knowledge. Compared with English speakers, Mandarin speakers tend to make more inference when hearing an implausible sentence, possibly due to the additional cues they rely on when interpreting sentences.

Data availability statement

All the raw data and analysis scripts are available online at https://osf.io/mx8b3/?view_only=1e6d9a91784a4eeba803419f761bf96e.

Conflict of interest

The authors declare no conflict of interest.

Notes

1 There is actually some uncertainty as to the correct analysis of sentences like (6b) and (6d). One anonymous reviewer suggested they could be analyzed as either serial-verb constructions or double-object dative constructions, whereas another reviewer suggested that they could be prepositional phrase dative constructions. For consistency, we will treat sentences like (6b) and (6d) as serial-verb constructions.
2 An item is a group of sentences that share the same nouns and verbs but vary in construction and plausibility. For example, all four sentences in (1) are considered from the same item.
3 We thank two anonymous reviewers for their suggestions regarding this topic.

4 An anonymous reviewer pointed out that sentence length could also be a potential explanation to our results, since among our stimuli, DO/serial sentences are on average longer than transitive/intransitive sentences, which are in turn longer than active/passive sentences on average, and the literal interpretation rate also decreases in this order (Fig. 1). To investigate sentence length as a potential factor, we reran all the analyses listed in Table 2, except this time we included stimulus length as an additional fixed effect and added by-participant and by-item random slopes for sentence length. We found the results in Table 3 stayed largely the same, except the difference among active/passive sentences is no longer significant (p_MCMC $= 0.074$). In addition, we found sentence length does not have a significant effect on participants' literal interpretation rate (p_MCMCs > 0.103). However, it is still possible that sentence length may also have an impact: as sentences become longer and longer, the likelihood that there will be at least one noise operation happening increases. When we designed our stimuli, we did not intend to investigate how sentence length affects a comprehender's sentence interpretation, and, therefore, the sentence length in our study heavily covaries with sentence plausibility, construction, and alternations.

References

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. https://doi.org/10.1016/j.jml.2007.12.005

Bader, M., & Meng, M. (2018). The misinterpretation of noncanonical sentences revisited. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(8), 1286.

Bai, X., Yan, G., Liversedge, S. P., Zang, C., & Rayner, K. (2008). Reading spaced and unspaced Chinese text: Evidence from eye movements. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(5), 1277.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

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.

Bergen, L., Levy, R., & Gibson, E. (2012). Verb omission errors: Evidence of rational processing of noisy language inputs. In *Proceedings of the 34th Annual Conference of the Cognitive Science Society*.

Cai, Z. G., Zhao, N., & Pickering, M. J. (2022). How do people interpret implausible sentences? *Cognition*, *225*, 105101.

Chen, S., Nathaniel, S., Ryskin, R., & Gibson, E. (2023). The effect of context on noisy-channel sentence comprehension. *Cognition*, *238*, 105503.

Christianson, K., Williams, C. C., Zacks, R. T., & Ferreira, F. (2006). Younger and older adults' "good-enough" interpretations of garden-path sentences. *Discourse Processes*, *42*(2), 205–238.

Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive Psychology*, *47*(2), 164–203.

Ferreira, F., & Patson, N. D. (2007). The 'good enough' approach to language comprehension. *Language and Linguistics Compass*, *1*(1–2), 71–83.

Ferreira, F., & Lowder, M. W. (2016). Prediction, information structure, and good-enough language processing. In B. H. Ross (Ed.), *Psychology of learning and motivation* (Vol. 65, pp., 217–247). Academic Press.

Garrett, M. F. (1975). The analysis of sentence production. In Gordon H. Bower (Ed.), *Psychology of learning and motivation* (Vol. 9, pp. 133–177). Academic Press.

Gibson, E., Bergen, L., & Piantadosi, S. (2013). The rational integration of noisy evidence and prior semantic expectations in sentence interpretation. *Proceedings of the National Academy of Sciences*, *110*(20), 8051–8056.

Gibson, E., Tan, C., Futrell, R., Mahowald, K., Konieczny, L., Hemforth, B., & Fedorenko, E. (2017). Don't underestimate the benefits of being misunderstood. *Psychological Science*, *28*(6), 703–712.

Gu, J., & Li, X. (2015). The effects of character transposition within and across words in Chinese reading. *Attention, Perception, & Psychophysics*, *77*(1), 272–281.

Hadfield, J. D. (2010). MCMC methods for multi-response generalized linear mixed models: The MCMCglmm R package. *Journal of Statistical Software*, *33*, 1–22.

Hsu, S. H., & Huang, K. C. (2000a). Effects of word spacing on reading Chinese text from a video display terminal. *Perceptual and Motor Skills*, *90*(1), 81–92.

Hsu, S. H., & Huang, K. C. (2000b). Interword spacing in Chinese text layout. *Perceptual and Motor Skills*, *91*(2), 355–365.

Inhoff, A. W., Liu, W., Wang, J., & Fu, D. J. (1997). Use of spatial information during the reading of Chinese text. In Peng, D., Shu, H., Chen, H. (Eds.), *Cognitive research on Chinese language, Shandong Educational Publishing*, 296–329.

Kane, E., & Slevc, L. R. (2019). Evidence for integration of noisy linguistic evidence and prior expectations depends on the task. In *Poster presented at the 32nd CUNY Conference on Human Sentence Processing*, Boulder, CO.

Keshev, M., & Meltzer-Asscher, A. (2021). Noisy is better than rare: Comprehenders compromise subject-verb agreement to form more probable linguistic structures. *Cognitive Psychology*, *124*, 101359.

Levy, R., Bicknell, K., Slattery, T., & Rayner, K. (2009). Eye movement evidence that readers maintain and act on uncertainty about past linguistic input. *Proceedings of the National Academy of Sciences of the United States of America*, *106*, 21086–21090.

Levy, R. (2008). A noisy-channel model of rational human sentence comprehension under uncertain input. In *Proceedings of the 13th Conference on Empirical Methods in Natural Language Processing* (pp. 234–243).

Li, P., Bates, E., & MacWhinney, B. (1993). Processing a language without inflections: A reaction time study of sentence interpretation in Chinese. *Journal of Memory and Language*, *32*(2), 169–192.

Li, X., Liu, P., & Rayner, K. (2011). Eye movement guidance in Chinese reading: Is there a preferred viewing location? *Vision Research*, *51*(10), 1146–1156.

Li, X., & Pollatsek, A. (2020). An integrated model of word processing and eye-movement control during Chinese reading. *Psychological Review*, *127*(6), 1139.

Li, X., Rayner, K., & Cave, K. R. (2009). On the segmentation of Chinese words during reading. *Cognitive Psychology*, *58*(4), 525–552.

Liu, P., & Li, X. (2014). Inserting spaces before and after words affects word processing differently in Chinese: Evidence from eye movements. *British Journal of Psychology*, *105*(1), 57–68.

Liu, P., & Lu, Q. (2018). The effects of spaces on word segmentation in Chinese reading: Evidence from eye movements. *Journal of Research in Reading*, *41*(2), 329–349.

Liu, Y., Ryskin, R., Futrell, R., & Gibson, E. (2020). Structural frequency effects in comprehenders' noisy-channel inferences. In *Proceedings of the 26th Architectures and Mechanisms for Language Processing Conference*. Poster presentation.

Meng, M., & Bader, M. (2021). Does comprehension (sometimes) go wrong for noncanonical sentences? *Quarterly Journal of Experimental Psychology*, *74*(1), 1–28.

Poliak, M., Ryskin, R., Braginsky, M., & Gibson, E. (2023). It's not what you say but how you say it: Evidence from Russian shows robust effects of the structural prior on noisy channel inferences. *Journal of Memory and Language: Learning, Memory, and Cognition*, https://dx.doi.org/10.1037/xlm0001244

Poppels, T., & Levy, R. P. (2016). Structure-sensitive noise inference: Comprehenders expect exchange errors. In *Proceedings of the 38th Annual Meeting of the Cognitive Science Society* (pp. 378–383).

R Core Team. (2022). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing.

Rayner, K., White, S. J., Johnson, R. L., & Liversedge, S. P. (2006). Raeding wrods with jubmled lettres: There is a cost. *Psychological Science*, *17*, 192–193.

Ryskin, R., Futrell, R., Kiran, S., & Gibson, E. (2018). Comprehenders model the nature of noise in the environment. *Cognition*, *181*, 141–150.

Shannon, C. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*, 379–423.

Staub, A., Dodge, S., & Cohen, A. L. (2018). Failure to detect function word repetitions and omissions in reading: Are eye movements to blame? *Psychonomic Bulletin & Review*, *26*, 340–346. https://doi.org/10.3758/s13423-018-1492-z

Su, I. R. (2001). Transfer of sentence processing strategies: A comparison of L2 learners of Chinese and English. *Applied Psycholinguistics*, *22*(1), 83–112.

Wei, W., Li, X., & Pollatsek, A. (2013). Word properties of a fixated region affect outgoing saccade length in Chinese reading. *Vision Research*, *80*, 1–6.

Zhang, Y., Ryskin, R., & Gibson, E. (2023). A noisy-channel approach to depth-charge illusions. *Cognition*, *232*, 105346.