

UC Merced

Cognitive Science Capstone Projects

Title

Alternative theories to (and evidence for) Gibson et al. (2013)'s Noisy-Channel Inference

Permalink

<https://escholarship.org/uc/item/1xm3v411>

Author

Fang, Xinzhu

Publication Date

2023-07-20

UNIVERSITY OF CALIFORNIA, MERCED

Alternative theories to (and evidence for) Gibson et al. (2013)'s Noisy-Channel Inference

A Thesis submitted in partial satisfaction of the requirements for the degree of Master of
Science

in

Cognitive and Information Sciences

by

Xinzhu Fang

Committee in charge:

Professor Paul Smaldino, Chair

Professor Colin Holbrook

Professor David Noelle

2023

Copyright
Xinzhu Fang, 2023
All rights reserved

The Thesis of Xinzhu Fang is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Colin Holbrook

David Noelle

Paul Smaldino, Chair

University of California, Merced

2023

Acknowledgements

I thank my mom for maintaining my living while I was working on this paper. I thank Graduate Group Chair Paul Smaldino for his comments on this paper.

I thank my ex-supervisor Ted Gibson for continuing to support me in ways he could in the past years. I thank people who helped me, many of whom, I think, prefer not to be acknowledged.

Abstract

Alternative theories to (and evidence for) Gibson et al. (2013)'s Noisy-Channel Inference

by Xinzhu Fang for the partial satisfaction of the requirements for the degree of Master of Science in Cognitive and Information Sciences, University of California, Merced, 2023

Dr. Paul Smaldino, Chair

In real life, language comprehension can be noisy, and sometimes the producer does not intend what they says. Under this consideration, Gibson et al. (2013) proposed a noisy-channel inference theory to account for participants' non-literal interpretation of sentences in language comprehension tasks. In this paper, I explain how key assumptions of this theory are inconsistent with results from more recent studies. Further, I argue that results presented as evidence for Gibson et al. (2013)'s noisy-channel inference do not reflect this process, as language comprehension in these studies was neither noisy nor communicative. I propose alternative theories to explain these data and point out alternative evidence for Gibson et al. (2013)'s noisy-channel inference from other studies.

Contents

Abstract	2
1 Introduction	2
2 Two noisy-channel inference theories	4
2.1 Levy (2008b)	4
2.2 Gibson et al. (2013)	5
3 “Noisy reading” studies	6
3.1 The transposed-word effect	6
3.2 The repeated “the” effect	8
3.3 Discussion	9
4 Non-literal interpretation of grammatical sentences	10
4.1 Noisy-channel studies on reading	10
4.2 A misperception explanation	14
4.3 (Which) noisy-channel inference during speech comprehension?	18
4.4 Evidence for Gibson et al. (2013)’s noisy-channel inference from language comprehension in “communicative” settings	19
5 The depth-charge illusion	19
5.1 Zhang et al. (2023)’s noisy-channel “explanation”	20
5.2 A rule deactivation theory	20
5.3 Relevant results from previous studies	24
6 Proposition: the core process of language comprehension is a kind of perception	25
6.1 Adaptation in object recognition	26
6.2 Adaptation in sentence processing	27
6.3 Predictive processing in sentence processing	28
6.4 Predictive processing and adaptation in basic visual scene analysis	29
6.5 From humans to machines	30
7 Final discussion	30
7.1 Evidence for a generalized Gibson et al. (2013)’s noisy-channel inference from language comprehension in challenging situations	32
References	33

Abstract

In real life, language comprehension can be noisy, and sometimes the producer does not intend what they says. Under this consideration, Gibson et al. (2013) proposed a noisy-channel inference theory to account for participants' non-literal interpretation of sentences in language comprehension tasks. In this paper, I explain how key assumptions of this theory are inconsistent with results from more recent studies. Further, I argue that results presented as evidence for Gibson et al. (2013)'s noisy-channel inference do not reflect this process, as language comprehension in these studies was neither noisy nor communicative. I propose alternative theories to explain these data and point out alternative evidence for Gibson et al. (2013)'s noisy-channel inference from other studies.

1 Introduction

Language comprehension is full of ambiguity. First, natural language is inherently ambiguous. A polysemous word has multiple meanings, and a sentence can too. Given the sentence "The women discussed the dogs on the beach", it is ambiguous whether the discussion happened on the beach or the dogs were on the beach (Jurafsky, 1996). Second, incremental processing causes temporal structural ambiguity while inputs are still unfolding. Reading a garden-path sentence (Frazier, 1979) such as "Sue doesn't know Edward Petherbridge is a librarian" (Fodor, 2017), likely, "Edward Petherbridge" is initially parsed as the object of "know" and then later re-parsed as the participant of the object clause when the reader reaches the disambiguating "is". Third, ambiguity is exacerbated by noise in the communication channel. "Utterances are often difficult to hear because of background noise; dialect and idiolect differences as well as competing sounds can make it difficult for the hearer to extract every word from an utterance; and speakers often produce utterances with disfluencies and outright errors" (Ferreira et al., 2002).

How do humans comprehend language in the face of such fierce ambiguity? What do humans comprehend? Assuming the language comprehension system takes in noise-free perceptual input, previous theories addressed the first two kinds of ambiguity by modeling sentence processing as incremental, probabilistic, and parallel process: multiple potential syntactic structures are assigned nonzero probabilities, which fluctuate as the sentence unfolds (Hale, 2001; Jurafsky, 1996; Levy, 2008a). Jurafsky (1996) modeled the preference for a potentially partial sentence structure consistent with the linguistic input received thus far (e.g., having read "The women discussed the dogs on", does the reader prefers to attach "on" to the noun "dog" versus to the verb "discussed") as the probability of that partial structure under a probabilistic context-free grammar (PCFG). The probability of a partial structure is calculated by marginalizing over all full structures that begin with that partial structure. Hale (2001) modeled the difficulty of processing the k th word in a sentence w_k after having processed preceding words $w_1...w_{k-1}$ as its surprisal: $-\log P(w_k|w_1...w_{k-1}) = -\log \frac{P(w_1...w_k)}{P(w_1...w_{k-1})} = \log P(w_1...w_{k-1}) - \log P(w_1...w_k)$, where the probability of a sequence of words is the sum of the probabilities of all (partial) structures consistent with that (partial) string under a PCFG. Later, Levy (2008a) proved that the surprisal of w_k is equal to the relative entropy of the probability distribution over all structures consistent with $w_1...w_{k-1}$ with respect to the probability distribution over all structures consistent with $w_1...w_k$. The surprisal theory is broadly compatible with the view that predictive processing is a core feature of language comprehension (more discussion on predictive processing in Section 6). The effects of word predictability on sentence processing have been demonstrated by numerous eye-tracking reading studies (Altarriba et al., 1996; Balota et al., 1985; Ehrlich & Rayner, 1981; Inhoff, 1984; Rayner et al., 2011; Rayner & Well, 1996; Schustack et al., 1987; Staub & Clifton, 2006; Vitu, 1991; Zola, 1984; for reviews see Rayner, 1998; Staub, 2015), event-related potential (ERP) studies (Federmeier et al., 2007; Kutas & Hillyard, 1980, 1984; for a review see Kutas & Federmeier, 2011), and pupillometry studies (Winn, 2016). More predicted words given the context are fixated for less duration and more likely to be skipped, and elicit smaller N400. Sentences that are more semantically constraining (e.g., "The lion gave an angry roar") are processed with less effort as indexed by less pupil dilation than sentence that are less semantically constraining (e.g., "They thought about the roar"). Specifically, surprisal somewhat predicts fixation durations and regression probabilities from eye-movement data (Boston et al., 2008; Demberg & Keller, 2008; Smith & Levy, 2008). In "The horse raced past the barn fell", a classic example of garden-path sentence, the disambiguating "fell" has a very high surprisal when considering the probabilities of structure alone (reduced relative clauses are rare) (Hale, 2001).

The surprisal of “fell” will be even higher when considering that “raced” is more frequently intransitive (not followed by a direct object) than transitive, or when considering that “horse” is more often the participant than the object of “raced” (Jurafsky, 1996; Levy, 2011).

Addressing the third type of ambiguity, Levy (2008b) extended the probabilistic approach from inferring the structure of a sentence to inferring the identity of the sentence. In his noisy-channel inference model, “the comprehender takes perceptual noise into account when making inferences about sentence form and structure, so that at all times the comprehender has a probability distribution over the sequence of words comprising the current sentence, taking into account perceptually similar and grammatically permissible variants of the sentence read thus far” (Levy et al., 2009). Later, Gibson et al. (2013) proposed a similar noisy-channel inference theory. Levy (2008b)’s comprehender and Gibson et al. (2013)’s comprehender makes inferences about noise from different parts of the noisy-channel. Under Levy (2008b), comprehenders maintain uncertainty about what they have perceived knowing their own perception may be noisy. Under Gibson et al. (2013), comprehenders maintain uncertainty about what the producer intended knowing the producer’s production may be noisy, making this theory a theory about language comprehension in communicative settings. Different from Levy (2008b), Gibson et al. (2013) assume that sentence perception is veridical, and thus any non-literal interpretation of a sentence results exclusively from their noisy-channel inference rather than (potentially beneficial) errors in sentence perception. As Huang & Staub (2021b) pointed out, this assumption is inconsistent with recent experimental results suggesting that reading itself is noisy (Huang & Staub, 2022, 2021a; Liu et al., 2022, 2020; Mirault et al., 2022, 2018; Snell & Grainger, 2019; Staub et al., 2019; Wen et al., 2021).

Besides contesting the validity of Gibson et al. (2013)’s noisy-channel inference theory itself, I also contest the validity of Gibson and colleagues applying this theory to explain several phenomena in sentence processing including non-literal interpretation of “ordinary” grammatical sentences in reading (Gibson et al., 2013) and listening (Gibson et al., 2017) – ordinary in the sense that these sentences are not of some rare structure – and the depth-charge illusion, an illusion in comprehending sentences of a rare structure (Zhang et al., 2023). In these studies, sentences were presented not embedded in any communicative context. In addition, Gibson et al. (2013)’s noisy-channel inference theory’s noise model, the model that they assume comprehenders use to judge what kind of production error is more likely, which has let the theory explain results from Gibson et al. (2013), is inconsistent with results from follow-up studies (Poppels & Levy, 2016; Ryskin et al., 2018). As to the depth-charge illusion, an illusion is not a communicative phenomenon in the first place. I will provide alternative theories to explain these two phenomena: the misinterpretation of ordinary grammatical sentences and the depth-charge illusion, and demonstrate how my theories are supported by existing results and can be further tested in future studies. I further argue that the core process of language comprehension is a kind of perception, not cognition (e.g., reasoning), and therefore phenomena resulting from the core process of language comprehension cannot possibly be caused by some post-perceptual process, such as Gibson et al. (2013)’s noisy-channel inference.

In Section 2, I summarize Levy (2008b)’s noisy-channel inference theory and Gibson et al. (2013)’s noisy-channel inference theory. In Section 3, I review studies demonstrating “noisy reading”. In Section 4, I review studies conducted by Levy’s and Gibson’s groups testing Levy (2008b)’s noisy-channel inference theory and Gibson et al. (2013)’s noisy-channel inference theory (Gibson et al., 2013; Levy, 2011; Levy et al., 2009; Poppels & Levy, 2016; Ryskin et al., 2018), question whether some of these results that have been regarded as evidence for Gibson et al. (2013)’s noisy-channel inference actually reflect Gibson et al. (2013)’s noisy-channel inference or rather noisy reading, propose studies to differentiate these two hypotheses, propose a verbal model of rational reading that could potentially exhibit both “noisy reading” and non-literal interpretation of “ordinary” grammatical sentences, discuss how speech comprehension study that has been regarded as evidence for Gibson et al. (2013)’s noisy-channel inference (Gibson et al., 2017) is susceptible to similar criticisms as the reading studies are, and point out potential evidence for Gibson et al. (2013)’s noisy-channel inference from existing studies of language comprehension in “communicative” settings. In Section 5, I quickly show that Zhang et al. (2023)’s noisy-channel explanation of the depth-charge illusion is impossible, present my rule deactivation theory and its predictions, explain how experimental results from previous studies (Paape et al., 2020; Zhang et al., 2023) either confirm my theory’s predictions or are at least compatible with my theory. In Section 6, I argue that the core process of language comprehension is a kind of perception, not

cognition, mostly through analogy between language comprehension and visual scene analysis. In Section 7, I summarize the paper, discuss the implication of my perception proposition for and beyond Gibson et al. (2013)’s noisy-channel inference, and point out potential evidence for a generalized Gibson et al. (2013)’s noisy-channel inference from existing studies of language comprehension in challenging situations.

2 Two noisy-channel inference theories

2.1 Levy (2008b)

Following the bayesian principles of previous application-oriented noisy-channel inference models (for character recognition: Raviv, 1967; for spelling correction: Kernighan et al., 1990; Mays et al., 1991) in dealing with potential noise rise in a noisy communication channel (Shannon, 1948), Levy (2008b) proposed a noisy-channel inference model of human sentence processing.

Under Levy (2008b)’s theory, the comprehender is presented with some string of words \mathbf{w}^* , which gets processed by the comprehender’s sensory system and arrives at the comprehender’s language comprehension system as some noisy evidence I (potentially corrupted by noise in the communication channel and in the sensory system), from which the comprehender infers the sentence identity \mathbf{w} and its structure T . Here is how Levy (2011) recaps Levy (2008b)’s noisy-channel inference mathematically:

For a true sentence \mathbf{w}^* which yields perceptual input I , joint inference on sentence identity \mathbf{w} and structure T marginalizing over I yields:

$$P(T, \mathbf{w}|\mathbf{w}^*) = \int_I P(T, \mathbf{w}|I, \mathbf{w}^*)P(I|\mathbf{w}^*)dI \quad (1)$$

\mathbf{w} and \mathbf{w}^* must be conditionally independent given I since \mathbf{w}^* is not observed by the comprehender, giving us (through Bayes’ Rule)

$$P(T, \mathbf{w}|\mathbf{w}^*) = \int_I \frac{P(I|T, \mathbf{w})P(T, \mathbf{w})}{P(I)}P(I|\mathbf{w}^*)dI \quad (2)$$

What Equation (2) basically says is that the more probable a potential sentence and its structure is (represented by $P(T, \mathbf{w})$), and that the more similar a potential sentence is to the sentence presented (represented by $P(I|T, \mathbf{w})$ and $P(I|\mathbf{w}^*)$), the more likely it will be inferred.

Levy posits that the human language comprehension system “perform(s) fully joint inference on sentence identity and structure given perceptual input, using linguistic knowledge both prospectively and retrospectively in drawing inferences as to how raw input should be segmented and recognized as a sequence of linguistic tokens, and about the degree to which each input token should be trusted during grammatical analysis.” (Levy, 2011) This theory offered new explanations for two puzzling findings on sentence processing by Christianson et al. (2001) and Tabor et al. (2004). Probing participants’ comprehension of garden-path sentences using a question-answering paradigm, Christianson et al. (2001) found that in response to the question “Did the deer run into the woods?” participants were more likely to answer yes to 1a than to 1b, despite the fact that, under the only grammatical construction of 1a, “the deer” has to be the participant of “ran” and thus cannot be the object of “hunted”:

1a While the man hunted the deer ran into the woods. (garden-path)

1b While the man hunted the pheasant the deer ran into the woods. (not garden-path)

Investigating participants’ incremental processing of sentences containing superficial part-of-speech ambiguity (unlike *thrown*, which can only be used as a past participle, *tossed* can be used as both a past tense verb and a past participle) using the self-paced reading paradigm, Tabor et al. (2004) found that response times to the critical word was the highest in the ambiguous, reduced relative clause condition, followed by the

unambiguous, reduced relative clause condition, and then the two unreduced relative clause conditions.

2a The coach smiled at the player *tossed* a frisbee by the opposing team. (ambiguous, reduced relative clause)

2b The coach smiled at the player *thrown* a frisbee by the opposing team. (unambiguous, reduced relative clause)

2c The coach smiled at the player who was *tossed* a frisbee by the opposing team. (unambiguous, unreduced relative clause)

2d The coach smiled at the player who was *thrown* a frisbee by the opposing team. (unambiguous, unreduced relative clause)

While the main effect of reduced relative clause versus unreduced relative clause is likely caused by difference in surprisal/predictability, its interaction with part-of-speech ambiguity is puzzling: the sentence-initial context in 2a should rule out the possibility of *tossed* being a main verb and forces it to be a past participle, why the part-of-speech ambiguity of *tossed* is only superficial.

Levy (2008b)'s noisy-channel inference theory predicts that reading 1a, participants' language comprehension system assigns a fairly high probability to its near-neighbor such as 3a, the structure of which is more probable than 1a and where "the deer" would be the object of "hunted". When reading 2a, participants' language comprehension system initially assigns a very high probability for "The coach smiled at" whereas a very low probability for "The coach smiled as"; but once they reaches *tossed*, the probability for "The coach smiled at" decreases whereas the probability for "The coach smiled as" increases, since the latter allows *tossed* to be a main verb while the earlier forces *tossed* to be part of a reduced relative clause, like "raced" in "The horse raced past the barn fell". "This change in beliefs about the past is treated as an error identification signal (EIS). In reading, a sensible response to an EIS would be a slowdown or a regressive saccade; in spoken language comprehension, a sensible response would be to allocate more working memory resources to the comprehension task."

3a While the man hunted the deer it ran into the woods.

3b The coach smiled as the player tossed a frisbee.

Levy also provided a computational model of his noisy-channel inference theory. Modeling Equation (2) is difficult because I is not linguistic but neural representation. To circumvent this problem, Levy modeled noisy-channel inference end-to-end from \mathbf{w}^* to \mathbf{w} :

$$P(\mathbf{w}|\mathbf{w}^*) = P(\mathbf{w}) \int_I \frac{P(I|\mathbf{w})P(I|\mathbf{w}^*)}{P(I)} dI \quad (3)$$

$$\propto P(\mathbf{w})Q(\mathbf{w}, \mathbf{w}^*) \quad (4)$$

The prior of \mathbf{w} is modeled as its probability under a PCFG. The negative log likelihood of inferring \mathbf{w} given \mathbf{w}^* is modeled as the Levenshtein distance between \mathbf{w} and \mathbf{w}^* . Intersecting a weighted context-free grammar (WCFG) representation of prior with a weighted finite-state automaton (wFSA) representation of likelihood generates a new WCFG representing unnormalized posterior. As this model reads a sentence incrementally, its prior (and posterior) of what past input could be changes. Levy experimented on his noisy-channel inference model using material adapted from Christianson et al. (2001) and Tabor et al. (2004). By taking into account the possibility that the input has been corrupted, the parser finds more alternative parses for 1a than for 1b (see Figure 3 of Levy (2008b)), and is more likely to revise how it has parsed past input (higher EIS) as it incrementally parses 2a compared to 2b (see Figure 5 of Levy (2008b)), verifying predictions discussed above.

2.2 Gibson et al. (2013)

In Gibson et al. (2013)'s communication system, the comprehender encodes m_i , the abstract non-linguistic message they intends, which gets encoded into the sentence they intends s_i , which travels through a noisy-channel and then gets perceived by the comprehender as s_p , from which the perceived message m_p is extracted.

s_p and m_p are the comprehender’s inference about s_i and m_i respectively. They verbally modeled their noisy-channel inference as bayesian inference:

$$P(s_i|s_p) \propto P(s_i)P(s_p|s_i) \tag{5}$$

Following Levy (2008b), they estimated the likelihood $P(s_p|s_i)$ as the Levenshtein distance between s_p and s_i . This noise model is not optimal for modeling comprehenders’ misperception or for modeling comprehenders’ mental model of producers’ misproduction, as I will demonstrate in the next two sections.

Despite stating in their abstract that all three kinds of errors “a noisy environment, producer errors, or perceiver errors” could reside in the noisy-channel, when illustrating their theory using their experiment design, Gibson and colleagues assumed that the comprehender always perceives what is presented. For example, given the sentence “The mother gave the candle the daughter”, they assumed that participants always perceived this sentence as it was presented. In the language of Levy (2008b), Gibson and colleagues equated the noisy input I with the sentence presented w^* . Consequently, the only kind of noise that exists in their communication system is production noise. While environmental noise rarely applies to reading, perceptual errors do exist in reading.

3 “Noisy reading” studies

3.1 The transposed-word effect

It had been long established that reading is noisy in the sense that intra-word errors¹ sometimes go unnoticed, especially when such a typo is the result of a letter transposition rather than a random replacement (e.g., “JUGDE” as opposed to “JUNGE”), a phenomenon often referred to as the transposed-letter effect (Bruner & O’Dowd, 1958; Chambers, 1979; Forster et al., 1987; Perea & Lupker, 2003). Exploring whether there exists a transposed-word effect similar to the transposed-letter effect, Mirault et al. (2018) had French participants perform speeded grammaticality judgment on French sentences presented in standard parallel visual presentation (PVP) where all words are presented simultaneously. Here are English translations of examples of their four conditions:

4a The white cat was big. (grammatical)

4b The white cat was slowly. (ungrammatical)

4c The white was cat big. (ungrammatical, transposed-word/TW, resulted from transposing “cat” and “was” in 4a.)

4d The white was cat slowly (ungrammatical, control, resulted from transposing “cat” and “was” in 4b.)

Mirault and colleagues analyzed error rate and response time to grammatical, TW, and control conditions. Subjects’ accuracy was very high for both grammatical and control, but lower for TW². Response time was longer for TW than for control, and longer for control than for grammatical. This transposed-word effect was replicated in Chinese by Liu et al. (2020)³ following Mirault et al. (2018)’s procedure, and in English by Huang & Staub (2021a) using a slight modification of Mirault et al. (2018)’s procedure in their Experiment 1 and a more natural procedure in their Experiment 2. Mirault et al. (2018) hypothesized that the transposed-word effect they discovered “points to a noisy encoding of word-order information during sentence reading” that results from “parallel processing of words during written sentence comprehension combined with top-down constraints from sentence-level structures.”

¹The errors I refer to as intra-word errors are commonly referred to as character-level, letter-level, sub-word-level, or word-level errors; the errors I refer to as inter-word errors are commonly referred to as word-level or sentence-level errors. As you can see, the meaning of word-level errors are ambiguous, so I coined these new terms

²Comparing the grammatical condition to the other conditions is not so informative since “the fact that they belonged to different response categories constitutes a confound” (Mirault et al., 2018). This applies to Snell & Grainger (2019) as well. I report such comparison for both studies because I consider it a good sanity check.

³Liu et al. (2020) found the transposed-word effect in all three of their experiments. Experiment 1: transposing a 1-character word with a 1-character word; Experiment 2: transposing a 2-character word with a 2-character word; Experiment 3: transposing a 1-character word with a 3-character word

Investigating the spatial constraint of the supposed parallel word processing, Snell & Grainger (2019) found that participants performed grammaticality judgment more accurately and faster for intact sentence than for outer-transposed sentence, and more accurately and faster for outer-transposed sentence than for inner-transposed sentences. The study is, again, in French. Here are English translations of examples of their three conditions:

5a the man can run (grammatical, intact)

5b the can man run (ungrammatical, inner-transposition of 5a.)

5c ran man can the (ungrammatical, outer-transposition of 5a.)

Investigating the potential contribution of “top-down syntactic constraints” to the transposed-word effect, Wen et al. (2021) found that participants performed grammaticality judgment more accurately and faster for within-phrase transpositions than for across-phrase transpositions. The study is, again, in French. Here are English translations of examples for each of their three conditions. There are two examples for each condition, separated by ‘/’, one example from each of their two lists:

6a the bells have rung long / often the river flows gently (grammatical)

6b the bells rung have long / often river the flows gently (ungrammatical, transposition within a syntactic phrase/within-phrase)

6c the have bells rung long / often the flows river gently (ungrammatical, transposition across syntactic phrases/across-phrase)

They used two lists to ensure that the crossing of a syntactic boundary was not confounded by the position of the transposed words. In list 1, a transposition of the words at positions 2 and 3 generated across-phrase transposed sequences, while transposing the words at positions 3 and 4 generated within-phrase transposed sequences, which was reversed for list 2.

Questioning the parallel word processing account of the transposed-word effect, Liu et al. (2022) presented half of the Chinese material from Liu et al. (2020) in PVP while the other half in word-by-word serial visual presentation (SVP). The transposed-word effect generalized from PVP to SVP in terms of error rate, but with a smaller magnitude in SVP. Compared to PVP, under SVP participants responded equally fast to sentences in the critical ungrammatical condition (4c) and to sentences in the control ungrammatical condition (4d). These results were replicated in English by Huang & Staub (2022) and in French by Mirault et al. (2022). Individual words were presented for 250 ms, 250 ms, and 300 ms in Liu et al. (2022), Huang & Staub (2022), and Mirault et al. (2022), respectively. Dufour et al. (2022) generalized the transposed-word effect to the auditory domain. Presenting the French material from Mirault et al. (2018) auditorily and having participants do grammaticality judgment, they found the transposed-word effect in both response time and error rate. Audio were synthesized using some text-to-speech program that is meant to sound naturally. Each sentence had five words. Mean sentence duration was 1372 ms. Mean word duration was 274 ms. Since parallel word processing is not possible under SVP, Huang & Staub (2022) (and previously Huang & Staub (2021a)⁴) proposed “instead, that words are recognized serially from left to right, but that integration of a recognized word into a higher-level representation of sentence structure and meaning may sometimes be delayed, so that two words are available for integration simultaneously; a reader’s syntactic knowledge is then brought to bear in making an inference about word order.” But when integration is delayed and when it is not delayed? If it is always delayed, that would not predict the transposed-word effect. A strong piece of evidence for the parallel word processing hypothesis comes from Experiment 2 of Snell & Grainger (2017). In this experiment, participants judged the syntactic category of a target word (noun or verb) flanked by words of the same or different category. Participants responded more slowly and less accurately in the incongruent condition than in the congruent condition, suggesting that they were involuntarily processing the flankers as they were trying to process the target. To succeed at the task, participants could simply focus at the center of the screen throughout the task: the target word was always presented at the center of the screen, and there

⁴Huang & Staub (2021a) proposed “that word recognition is serial, but post-lexical integration of each word into its context may not be perfectly incremental.”

were centralized vertical fixation bars guiding participants to fixate the target location throughout the task. It is also unlikely that participants shifted attention to flankers given that each row of words was displayed for only 170 ms. I will present a relaxed version of the parallel word processing hypothesis to account for the transposed-word effect at the end of this section.

One may argue that the speeded grammaticality judgment paradigm’s emphasis on speed, especially under PVP, makes participants more likely to miss errors than they would in normal reading and have led these studies to overestimate the transposed-word effect. Although one could also argue that the paradigm’s emphasis on grammaticality makes participants more vigilant about grammatical errors and have led these studies to underestimate the transposed-word effect. At the end of the day, the speeded grammaticality judgment paradigm is quite unnatural: most of the time people read to comprehend not to proofread. Within this paradigm, SVP is even more unnatural than PVP. To investigate the mechanism underlying the transposed-word effect, Huang & Staub (2021a) conducted an eye-tracking reading study. In Experiment 1, they had participants perform grammaticality judgment following previous studies. In Experiment 2, they adopted a more natural procedure from Staub et al. (2019) by embedding experimental sentences followed by the question, “Was there anything wrong with that sentence?” among a larger number of filler sentences that were followed by comprehension questions. Participants did not know which type of questions they would be asked until they have finished reading⁵. All sentences were seven words long. Each item had a grammatical version and a transposed version created by transposing word 3 and word 4 in the grammatical version, where word 4 is the point of ungrammaticality:

7a I walk my fat dog every morning (grammatical).

7b I walk fat my dog every morning (transposed).

There were two key findings. First, although across both experiments readers more frequently failed to notice transposition when their eyes skipped either word 3 or word 4 than when they fixated both words, transpositions were still missed sometimes in the latter case. Second, “The transposed words caused disruption in the eye movement record only on trials when participants ultimately judged the sentence to be ungrammatical, not when they judged the sentence to be grammatical.” Compared to when reading transposed-sentences that participants ultimately judged to be grammatical, when reading transposed-sentences that they ultimately judged to be ungrammatical, participants had longer first fixation duration, gaze duration, go-past time, and total viewing time⁶ for both word 3 and word 4 across both experiments. In contrast, these differences did not exist when comparing how participants read transposed-sentences that they ultimately judged to be grammatical and how participants read grammatical filler sentences.

3.2 The repeated “the” effect

Staub et al. (2019) conducted an eye-tracking experiment to investigate another phenomenon of inter-word errors getting autocorrected by readers, the common experience of English readers to miss duplications of “the”. Staub et al. (2019) developed the paradigm of mixing error detection questions with comprehension questions that was later used in Huang & Staub (2021a)’s Experiment 2. Here are examples of Staub et al. (2019)’s four conditions:

8a Amanda jumped off the swing and landed on her feet. (G/grammatical)

8b Amanda jumped off the the swing and landed on her feet. (RT/repeated “the”)

8c Amanda jumped off the swing swing and landed on her feet. (RN/repeated noun)

8d Amanda jumped off swing and landed on her feet. (OT/omitted “the”)

⁵Another improvement they made for both experiments was to have participants fixate at the left edge of the screen, “prompting participants to read from left to right”. Previous studies presented sentence centrally and had participants fixated at the center before sentence presentation, which Huang & Staub (2021a) suspected had made participants more likely to miss errors.

⁶Here is a brief explanation of standard measures in eye-tracking reading studies: gaze duration would be longer than first fixation duration if the word was refixated during first-pass reading; go-past time would be longer than gaze duration if there was any regression to words to its left before the word was exited to the right; total viewing time would be longer than gaze duration if the word was refixated during any regression from words to its right.

While participants rarely reported errors for G (a sanity check passed), they reported errors more often for RN, followed by OT, then RT, confirming that the duplication of “the” is hard to notice. Eye-tracking results showed that when reading RT, when participants fixated one of the two “the” only but not both, they more frequently missed than noticed the duplication error; in contrast, when they fixated both “the”, they more frequently noticed than missed the duplication error. Although participants rarely missed the duplication error when reading RN, they were more likely to miss the error when fixating one noun only than when fixating both nouns. Re-analyzing these eye-tracking data – analyzing not only whether people fixated but also when and for how long they fixated a word – Huang & Staub (2021b) found that “when readers did not notice the repeated the, fixation durations and regression probabilities on each of the two instances of the were statistically indistinguishable from eye movements on the single the in grammatical sentences.”

3.3 Discussion

I reviewed studies investigating the transposed-word effect and the repeated “the” effect, as well as two accounts of the transposed-word effect: the parallel word processing account put forth by Grainger’s group and the delayed integration account put forth by Staub’s group. The parallel word processing account does not explain the transposed-word effect when words are presented in serial, whereas the delayed integration account is rather vague. However, both accounts agree that some kind of simultaneous processing, be it recognition or integration, is key to the transposed-word effect. Here I present a relaxed version of the parallel word processing hypothesis. Since word processing takes time, however little it may be, processing of adjacent words could still overlap even if the two words are not processed fully in parallel as required by the parallel word processing account for the transposed-word effect to happen. When two words are mistakenly transposed during production, the more parallelly they are processed by the comprehender, the more likely the transposition goes unnoticed by the comprehender. For example, compared to when both words are fixated, two words are processed more parallelly when the reader fixates the first word but not the second word, which means that the second word is processed mostly in parafovea when the first word is fixated. This prediction is somewhat⁷ confirmed by Huang & Staub (2021a). Similarly, when a word is duplicated, the more parallelly two instances are processed, the more likely the duplication goes unnoticed. This prediction is somewhat confirmed by Staub et al. (2019).

Under the relaxed parallel word processing hypothesis, the noisiness of sentence perception would further increase when word processing is prolonged, which would allow comprehenders to maintain uncertainty about what they have perceived as posited by Levy (2008b). The reason why participants in the transposed-word effect studies responded more slowly to sentences containing transposition errors may be that they engaged in Levy (2008b)’s noisy-channel inference. Perhaps they were double-checking that the transposition error was indeed present in the sentence and was not the result of their own misperception, before judging the sentence to be ungrammatical.

The implications of the eye-tracking results from Huang & Staub (2021a) and Staub et al. (2019) go beyond the phenomena these studies set out to investigate. These results establish bidirectional causal effects between error detection and eye-movement. An error is more likely to be missed if it was not fixated – eye-movement is predictive of error detection. An error when noticed is likely to be revisited – eye-movement is postdictive of error detection. These empirical linkages between error detection and eye-movement provide linking hypotheses that will become useful when discriminating between competing hypotheses on what causes participants to misinterpret ordinary grammatical sentences in Gibson et al. (2013) in the next section.

⁷My account does not explain why a transposition error is also more likely to be missed when the second word is fixated and the first word is processed mostly in parafovea when the word preceding it being fixated.

4 Non-literal interpretation of grammatical sentences

4.1 Noisy-channel studies on reading

4.1.1 Levy et al. (2009)

Levy et al. (2009) had participants read sentences while recording their eye-movement. Every sentence was followed by a yes/no comprehension question. Levy and colleagues kept the superficial part-of-speech ambiguity manipulation from Tabor et al. (2004) (e.g., **tossed** versus **thrown**). In addition, they varied whether the preceding preposition had near-neighbors that would license the potentially ambiguous verb (the critical word) to be a main verb: *at* has neighbors “as” and “and” whereas *toward* none:

9a The coach smiled *at* the player **tossed** the frisbee by the opposing team. (*at*+ambiguous)

9b The coach smiled *at* the player **thrown** the frisbee by the opposing team. (*at*+unambiguous)

9c The coach smiled *toward* the player **tossed** the frisbee by the opposing team. (*toward*+ambiguous)

9d The coach smiled *toward* the player **thrown** the frisbee by the opposing team. (*toward*+unambiguous)

Levy (2008b) predicts that “strong coherence of current input with a perceptual neighbor of previous input may induce confusion in comprehenders as to the identity of that previous input” (Levy, 2011), which in this study means that reading 9a, by the time readers reach **tossed** which more likely acts as a main verb, participants may wonder if the word they just perceived as *at* may actually have been “as” or “and”. Indeed, during first-pass reading of the critical word, participants made more regressions in the *at*+ambiguous condition than in the other three conditions (reflected in go-past time and proportion of regressions out, see Fig. 2A and B of Levy et al. (2009)). Specifically, they fixated the preceding preposition more often in the *at*+ambiguous condition than in the other three conditions (reflected in proportion of go-past regressions, see their Fig. 2C). These patterns suggest that participants maintained uncertainty about their perception of past linguistic input, and when one of the past words has its near neighbor(s) “postdicted” by words further down the sentence, participants backtrack to that word to confirm that they perceived it correctly.

4.1.2 Levy (2011)

Levy (2011) had participants do self-paced reading on grammatically unambiguous sentences, where the subordinate clause (e.g., “as the soldiers marched”) may or may not have a prepositional phrase (+PP versus -PP), and the main clause may or may not have a locative inversion, where the locative prepositional phrase and the participant trade places. The comma was presented with the word it followed.

10a As the soldiers marched, toward the tank *lurched* an injured enemy combatant. (inverted, -PP)

10b As the soldiers marched into the bunker, toward the tank *lurched* an injured enemy combatant. (inverted, +PP)

10c As the soldiers marched, an injured enemy combatant *lurched* toward the tank. (uninverted, -PP)

10d As the soldiers marched into the bunker, an injured enemy combatant *lurched* toward the tank. (uninverted, +PP)

Levy (2008b) predicts that “comprehenders might under some circumstances adopt a grammatical analysis inconsistent with the true raw input comprising a sentence they are presented with, but consistent with a slightly perturbed version of the input that has higher prior probability. If this is the case, then subsequent input strongly disconfirming this ‘hallucinated’ garden-path analysis might be expected to induce the same effects as seen in classic cases of garden-path disambiguation traditionally studied in the psycholinguistic literature.” (Levy, 2011) In this particular study, the prediction would be that reading 10a, by the time participants reach the main clause preposition “toward”, they may write off the comma following the subordinate clause verb “marched” as a mistake, but when they reach the critical word *lurched*, they will

realize that the comma is actually not a mistake – correcting their previous correction⁸. Indeed, response times to the critical word were longer in inverted, -PP condition than in the other conditions⁹.

While Levy (2008b)’s noisy-channel inference theory explains the eye-tracking results from Levy et al. (2009) and the reading time results from Levy (2011), it does not explain the question-answering results. In Levy et al. (2009) and Levy (2011), question-answering was less accurate in the critical condition than in the other three conditions. In both studies, every sentence had only one grammatical interpretation, just as sentences in Christianson et al. (2001) and Tabor et al. (2004) did. After potential backtracking in Levy et al. (2009) and pausing for re-reanalysis in Levy (2011), why did participants still misinterpret the sentence? I will resume this discussion in Subsection 4.4.

4.1.3 Gibson et al. (2013)

Besides proposing a new noisy-channel inference theory, Gibson et al. (2013) conducted a question-answering experiment to test their theory, where they probed participants’ comprehension of ordinary grammatical sentences. Each experimental item is of one of five construction pairs and has four versions, two plausible, two implausible, all grammatical. 11d-t in Table 1 are five example items, one item each of construction active/passive, construction uninverted/inverted, construction double-object(DO)/prepositional phrase object(PO) goal, transitive/intransitive, and DO/PO benefactive. Every sentence came with a yes/no comprehension question. Given 11b and the prompt “Did the girl kick something/someone?”, if a participant responded no, Gibson and colleagues would interpret this response as the participant interpreted 11b literally; if a participant responded yes, Gibson and colleagues would interpret this response as that the participant interpreted the sentence non-literally, that the participant inferred that the producer intended 11d but made a production error by inserting “was” and “by”. Why 11d not 11c when both indicate that the girl kicked the ball? Gibson and colleagues assumed that participants consider only two types of string edits: insertion and deletion, as in Levenshtein distance. Under this assumption, although 11b is only one exchange away from 11c, that exchange is equivalent to 2 deletions plus 2 insertions totaling 4 edits, whereas 11b is two insertions away from 11d. Therefore, the likelihood of producing 11b, given that 11d is intended, is higher than the likelihood of producing 11b, given that 11c is intended. This noise model, which let Gibson et al. (2013)’s noisy-channel inference theory explain human participants results in Gibson et al. (2013), is inconsistent with data from Poppels & Levy (2016) and Ryskin et al. (2018), as I will demonstrate later in this section.

Gibson et al. (2013)’s noisy-channel inference theory makes four predictions: “in a sentence comprehension task: (i) semantic cues should pull sentence interpretation towards plausible meanings, especially if the wording of the more plausible meaning is close to the observed utterance in terms of the number of edits; (ii) this process should asymmetrically treat insertions and deletions due to the Bayesian ‘size principle’; such non-literal interpretation of sentences should (iii) increase with the perceived noise rate of the communicative situation and (iv) decrease if semantically anomalous meanings are more likely to be communicated.” They conducted three experiments to test these predictions. All experiments tested prediction 1 and 2. Experiments 2 and 3 tested prediction 3 and 4 respectively. In Experiment 1.1 to Experiment 1.5, every participant read 20 experimental sentences of one of the five construction pairs (e.g., active/passive), with 10 plausible and 10 implausible sentences, distributed equally between the two constructions in the pair (e.g., active or passive). Every participant read 60 fillers that were all grammatical and plausible, e.g., “The professor conducted an inquiry into the status of the file.” In Experiment 2.1 to Experiment 2.5, half of the fillers were ungrammatical made by scrambling a few adjacent words, or inserting or deleting a function word in fillers used in Experiment 1, e.g., “The professor conducted an inquiry the status of the file”. This way, overall Experiment 2 had more ungrammatical sentences than Experiment 1 did – how Gibson and colleagues operationalized noise rate. In Experiment 3, each participant read all experimental sentences spanning all five construction pairs. This way, participants in Experiment 3 read more implausible sentences than participants

⁸The comma is most likely written off as a mistake in production rather than perception given the slow and strictly incremental nature of self-paced reading, which makes this process actually Gibson et al. (2013)’s noisy-channel inference rather than Levy (2008b)’s noisy-channel inference. More discussion on evidence for Gibson et al. (2013)’s noisy-channel inference in Section 4.4

⁹This pattern cannot be explained by saying that participants did not notice the comma and therefore 10a was read as a true garden-path sentence rather than a hallucinated garden-path sentence. Participants clearly noticed the comma since response times to the subordinate clause verb (and its comma in -PP conditions) were longer in both -PP conditions than in both +PP conditions, which is expected since verbs like “march” are usually followed by a PP.

Table 1: Example stimuli from Gibson et al., 2013, reproduced from their Table 1.

construction	perceived (implausible)	inferred production error	inferred intended (plausible)
active/passive	11a The ball kicked the girl.	two deletions	11c The ball was kicked by the girl.
	11b The girl was kicked by the ball.	two insertions	11d The girl kicked the ball.
non-inverted/inverted	11e The table jumped onto a cat.	one deletion one insertion	11g Onto the table jumped a cat.
	11f Onto the cat jumped a table.	one deletion one insertion	11h The cat jumped onto a table.
DO/PO goal	11i The mother gave the candle the daughter.	one deletion	11k The mother gave the candle to the daughter.
	11j The mother gave the daughter to the candle.	one insertion	11l The mother gave the daughter the candle.
transitive/intransitive	11m The businessman benefited the tax law.	one deletion	11o The businessman benefited from the tax law.
	11n The tax law benefited from the businessman.	one insertion	11p The tax law benefited the businessman.
DO/PO benefactive	11q The cook baked a cake Lucy.	deletion	11s The cook baked a cake for Lucy.
	11r The cook baked Lucy for a cake.	insertion	11t The cook baked Lucy a cake.

in Experiment 1 did.

All four of their predictions were completely confirmed by the results, except for prediction 2. Subjects interpreted implausible sentences non-literally more often (i) For DO/PO goal, transitive/intransitive, and DO/PO benefactive than for active/passive and uninverted/inverted, (ii) when the nearest plausible neighbor is deletion(s) than insertion(s) away for DO/PO goal, transitive/intransitive, and DO/PO benefactive, (iii) in Experiment 2 than in Experiment 1, and (iv) in Experiment 1 than in Experiment 3 (see their Fig. 2. For the full results). Results on the relative difference in proportion of non-literal interpretation between active and passive and between uninverted and inverted are inconsistent with their prediction 2. Since the implausible 11b is two insertions away from 11d, whereas the implausible 11a is two insertions away from 11c, it is predicted that participants are more likely to interpret 11a (active) than non-literally 11b (passive). Contrary to their prediction, participants more often interpreted passive non-literally than active. Since both implausible 11f and 11g are 1 deletion plus 1 insertion away from their nearest plausible neighbors, 11h and 11f respectively, it is predicted that participants are equally likely to interpret 11f (inverted) and 11e (uninverted) non-literally. Contrary to their prediction, participants more often interpreted inverted non-literally than uninverted. In response to the unexpected result on the interpretation of active versus passive, Gibson et al. (2013) briefly alluded to Ferreira (2003), which reported the same pattern using a similar task paradigm in listening. Ferreira (2003) showed that participants more often misinterpret non-garden-path sentences that simply “require thematic roles to be assigned in an atypical order” – passive/object cleft (e.g., “it was the ball that the girl kicked”) compared to active/subject cleft (e.g., “it was the girl who kicked the ball”) – “especially when they express implausible ideas.” Gibson et al. (2013)’s finding of more misinterpretation of inverted than of uninverted provides another instance of more “misinterpretation of non-canonical sentences”. Coming up next is another instance demonstrated by Poppels & Levy (2016). Why do participants make more noisy-channel inference for non-canonical sentences? Gibson et al. (2013) did not answer.

Table 2: Example stimuli from Poppels and Levy, 2016

construction	perceived (implausible)	inferred production error	inferred intended (plausible)
PP adjuncts	12a The package fell from the floor to the table. (canonical)	one exchange	12c The package fell to the floor from the table. (non-canonical)
	12b The package fell to the table from the floor. (non-canonical)	one exchange	12d The package fell from the table to the floor. (canonical)

4.1.4 Poppels & Levy (2016)

Assuming participants consider only two types of string edits, insertion and deletion, Gibson et al. (2013)’s finding that participants interpreted implausible sentences non-literally more often for DO/PO goal, transitive/intransitive, and DO/PO benefactive, which requires 1 edit to transform from the plausible to the implausible, than for active/passive and uninverted/inverted, which requires 2 edits, confirms their prediction 1. A commonality shared by implausible sentences across all five construction pairs is that they can be transformed from a plausible neighbor by a single exchange of content words. So it seems that indeed comprehenders do not consider exchange errors. To investigate whether participants in Gibson et al. (2013) refrained from inferring exchange errors because they refrained from inferring exchange errors in general or because they refrained from inferring errors of exchanging content words specifically, Poppels & Levy (2016) adopted the paradigm of Gibson et al. (2013) and retained three out of their five construction pairs: active/passive, transitive/intransitive, and DO/PO benefactive. In addition, they included prepositional phrase (PP) adjuncts, where an implausible version could result from exchanging the two function words in a plausible version. One example item of PP adjuncts is shown in 2.

Replicating Gibson et al. (2013)’s results regarding the relative difference in proportion of non-literal interpretation between the two construction in each pair, Poppels & Levy (2016) found that, when reading implausible sentences, participants were more-likely to make non-literal interpretation for passive than for active, and were more-likely to infer deletion than insertion errors for the other two constructions. Confirming their new hypothesis, participants were more-likely to make non-literal interpretations for PP adjuncts than for active/passive. In addition, they found that within PP adjuncts, participants were more-likely to make non-literal interpretations for non-canonical (12b) than for canonical (12a), but they could not say for sure that participants inferred exchanges of prepositions (12b transformed from 12d, 12a transformed from 12c) instead of exchanges of nouns (12b transformed from 12c, 12a transformed from 12d). Either way, the between-construction result that participants made non-literal interpretations for PP adjuncts (4 edits of function words to transform from plausible to implausible) as often as for transitive/intransitive (1 function word edit), examined in conjunction with the between-construction result from Gibson et al. (2013) that participants made non-literal interpretations for transitive/intransitive than for uninverted/inverted (2 edits of function words), is inconsistent with Gibson et al. (2013)’s noise model.

4.1.5 Ryskin et al. (2018)

More evidence disconfirming Gibson et al. (2013)’s noise model surfaced when Ryskin et al. (2018) attempted to replicate Gibson et al. (2013)’s results by employing a paradigm that is more communicative and explicitly encourages inference. In Ryskin et al. (2018), participants were told that they would be reading “transcriptions of someone’s speech and that these transcriptions might contain errors” and were asked to “retype each sentence in a text box and edit it if they thought the speaker had intended something different.” In addition, most participants were exposed to implausible sentences before they started the editing task. There were five exposure conditions: mixed errors, deletions, insertions, exchanges, and no errors. In the deletion condition, participants were exposed to sentences like 11i; in the insertion condition, participants were exposed to sentences like 11j; in the exchange condition, participants were exposed to sentences like 11a and 11b. This manipulation allowed Ryskin and colleagues to test if they could replicate Gibson et al. (2013), but also if participants would adapt to the language-production-error idiosyncrasy of the posed producer. Subjects’

editing responses were categorized into whether they inferred deletion error, insertion error, exchange error, no error, or other error. All experimental sentences and most fillers were taken from Gibson et al. (2013). All fillers were grammatical. As expected, participants adapted: across all inferred error types, participants most often inferred an error type when they had been exposed to that error type compared to when they had been exposed to other error types. Inconsistent with Gibson et al. (2013)’s noise model, across all exposure conditions, participants much more often inferred exchange than deletion/insertion errors, while deletion was inferred slightly more often than insertion was. Both findings persisted in their Experiment 2, which was almost of the same design as Experiment 1.¹⁰

Ryskin and colleagues briefly addressed the inconsistency, saying that “idiosyncrasies of the stimulus set may contribute to these different baselines.” This is implausible, since almost all sentences were from Gibson et al. (2013). I suspect that the overwhelming preference for exchange errors is at least in part due to the process underlying structural priming in production (and comprehension), that after exposure to a prime sentence, one is more likely to produce a target sentence that is of the same structure as the prime sentence (Bock, 1986; Bock & Griffin, 2000; Pickering & Branigan, 1998). 11b “The girl was kicked by the ball” primes 11c “The ball was kicked by the girl” more than it primes 11d “The girl kicked the ball.” However, structural priming would be present in both the editing task of Ryskin et al. (2018) and the sentence comprehension task of Gibson et al. (2013). After updating the noise model of Gibson et al. (2013)’s noisy-channel inference theory based on Ryskin et al. (2018)’s data, the updated theory predicts that non-literal interpretation be made equally likely for all constructions in Gibson et al. (2013), since all implausible sentences across all six constructions tested there can be transformed from a plausible neighbor through one exchange of two content words. This new prediction is inconsistent with the between-construction difference found by Gibson et al. (2013) and Poppels & Levy (2016) – the patterns observed in these studies would have to be caused by some other process.

4.2 A misperception explanation

The goal of language comprehension during communication is not to reconstruct what the interlocutor said but what the interlocutor meant. Consequently, human listeners sometimes infer that the producer intended or produced differently from what they perceived, especially if the communication system is noisy, whether it is because the listener has some perceptual deficit, the communication channel is noisy, or the speaker is still learning the language. However, the language comprehension task used by Gibson et al. (2013) and Poppels & Levy (2016) is not communicative – participants were simply filling out a survey and there was no communicative partner involved, not even a posed one. Every sentence was presented in isolation with no context. I suspect that the non-literal interpretation results shown by Ferreira (2003), Gibson et al. (2013), and Poppels & Levy (2016) actually reflect noisy reading (misperception) rather than intentional post-perceptual non-literal interpretation as in Gibson et al. (2013)’s noisy-channel inference.

The behavior of misperceiving an improbable sentence as its more probable neighbor was briefly discussed in Levy (2008b):

... the comprehender should in principle be able to override the linguistic input actually presented, so that a sentence is interpreted as meaning — and perhaps even being — something other than it actually meant or was. At one level, it is totally clear that comprehenders do this on a regular basis: the ability to do this is required for someone to act as a copy editor — that is, to notice and (crucially) correct mistakes on the printed page. In many cases, these types of

¹⁰Except for two modifications they made, both of which would motivate participants to “infer” more insertion/deletion errors than exchange errors. First, “four test sentences were changed between Experiment 1 and Experiment 2; because of the high rate of Inferred Exchanges, sentences that could only plausibly be interpreted as the result of exchanges (e.g., The paper wrote the student.) were switched for sentences from Gibson et al. (2013) that were rated most likely to be the result of exchanges but could also be interpreted as resulting from deletion (e.g., The bat swung the player).” Second, “The instructions for Experiment 2 additionally stated that participants were allowed to copy and paste sentences if they did not think they contained any errors.” Although this instruction explicitly encouraged participants to copy and paste when they infer no error only, it could also prompt participants to do so when they inferred errors. When participants were producing the sentence they thought the produced had intended, if they copied, pasted, and then edited the sentence given to them rather than typing from scratch, they would make fewer finger movements (fewer edits) if they “inferred” exchange instead of insertion/deletion errors.

correction happen at a level that may be below consciousness — thus we sometimes miss a typo but interpret the sentences as it was intended. . .

For expository convenience, I define autocorrecting as misperceiving A as B, when the reader (whether human or machine) has a higher prior for B than for A. When A is autocorrected into B, I define the transformation from B to A as an error.

Although autocorrecting inter-word errors has been demonstrated only in autocorrecting transposition, duplication, and omission of “the”, I suspect that autocorrecting transposition, omission, and insertion (including duplication) of function words happen in daily sentence processing and happened during Gibson et al. (2013)’s task¹¹. Under Levy (2008b)’s noisy-channel inference theory, comprehenders maintain uncertainty about what they have perceived knowing their own perception may be noisy. Why would comprehenders do that if perception is not actually noisy at least sometimes? Since Levy (2008b)’s noisy-channel inference implies noisy reading, all evidence for Levy (2008b)’s noisy-channel inference are indirect evidence for noisy reading.

Conducting any of two variations of Gibson et al. (2013)’s experiments can answer whether the disputed non-literal interpretation results reflect misperception or Gibson et al. (2013)’s noisy-channel inference. First, re-do Gibson et al. (2013) but record response time. Performing Gibson et al. (2013)’s noisy-channel inference takes time. Therefore, if participants respond to implausible sentences which they interpreted as plausible as fast as they do to plausible sentences, it should be inferred that in those cases participants misperceived rather than misinterpreted. Second, re-do Gibson et al. (2013) with eye-tracking. Reading an implausible sentence, for participants to make Gibson et al. (2013)’s noisy-channel inference they need to first notice that the sentence reads odd, which should cause disruption to eye-movement as in Levy et al. (2009) and Huang & Staub (2021a). Given eye-tracking results from Huang & Staub (2021a) and Staub et al. (2019), the misperception hypothesis predicts that the part of an implausible sentence where it differs from its plausible neighbor will be fixated less when question-answering suggests a plausible interpretation than question-answering suggests a implausible interpretation.

I propose a misperception explanation of all findings on misinterpretation of ordinary sentences discussed so far:

- the within-construction plausibility effect demonstrated by Ferreira (2003), Gibson et al. (2013), and Poppels & Levy (2016), and the grammaticality effect demonstrated by “noisy reading” studies
- the transposed-word effect demonstrated by “noisy reading” studies and a similar effect demonstrated by Poppels & Levy (2016)
- the within-construction canonicity effect demonstrated by Ferreira (2003), Gibson et al. (2013), and Poppels & Levy (2016)
- and when there is not a canonical form, e.g., DO/PO goal, the within-construction asymmetry between “inferring” an insertion error and “inferring” a deletion error demonstrated by Gibson et al. (2013) and Poppels & Levy (2016)
- the between-construction difference demonstrated by Gibson et al. (2013) and Poppels & Levy (2016)
- the effect of frequency of implausible sentences demonstrated by Gibson et al. (2013)
- the effect of frequency of ungrammatical sentences demonstrated by Gibson et al. (2013)

In particular, I attribute the latter three findings from Gibson et al. (2013) to the way the experimental manipulations alter the likelihood of misperception for a rational reader. A rational reader balances between speed and accuracy and is adaptive to the linguistic environment. For most of these behaviors, I will also discuss how they can potentially be implemented by extending Levy (2008b)’s noisy-channel inference model and existing rational reading models (Bicknell & Levy, 2012, 2010a, 2010b).

¹¹Staub et al. (2019) introduced the paradigm of mixing grammaticality judgment questions with comprehension questions so that participants would read more naturally (less carefully). Since Gibson et al. (2013) had comprehension questions only and no grammaticality judgment, participants likely read even more naturally (less carefully).

the within-construction plausibility effect and the grammaticality effect. Both effects are intuitive and can be simulated when modeling perception as bayesian inference and calculating prior based on plausibility and grammaticality.

Recall that Levy (2008b)’s noisy-channel inference model calculates prior using a PCFG. Although it is possible to build PCFGs that learn lexical-structural co-occurrence statistics, such as that in double-object sentences an animate noun is more likely to be the indirect object than to be the direct object, and the reverse for inanimate nouns, standard annotated corpora that PCFGs are trained on have syntactic info only, and thus PCFGs learn structure occurrence statistic (e.g., that reduced relative clauses are rare) only. Consequently, Levy (2008b)’s noisy-channel inference model would autocorrect ungrammatical sentences with 0 prior probability to their grammatical neighbors but would not assign a higher prior for “The mother gave the daughter the candle” than for “The mother gave the candle the daughter”, since the two sentences have the same structure. This gap between the model and human data can be bridged by updating Levy (2008b)’s noisy-channel inference model such that the prior is calculated using a language model that learns lexical-structural co-occurrence statistics and (structure-dependent) lexical co-occurrence statistics, such as n-gram models or large language models.

the transposed-word effect. When misperceiving “The package fell to the table from the floor” as “The package fell from the table to the floor”, the reader autocorrects an outer-transposition error, which happens as demonstrated by Snell & Grainger (2017). The current Levy (2008b)’s noisy-channel inference model, however, is unlikely to exhibit such behaviors since it estimates the likelihood of inferring sentence B given sentence A based on Levenshtein distance. Under Levenshtein distance, a transposition counts as four edits, while an insertion/deletion just one edit. Also, this noise model considers inter-transposition and outer-transposition to be equally likely, which is inconsistent with findings from Snell & Grainger (2017).

the within-construction canonicity effect. If we think sentence processing as a kind of perception (more discussion in Section 6), there would be a readily and simple explanation of the within-construction canonicity effect. Because perception is shaped by statistical learning, perception is ubiquitously harder when the object is in a non-canonical form, e.g., when the object is in an unusual pose. When bottom-up evidence is less reliable, the perceptual system leans more on expectation. The language-specific mechanism could be that when the sentence is in a non-canonical form, individual word processing is sometimes more prolonged and thus words are processed more parallelly (recall the relaxed parallel word processing hypothesis described in Section 3). Consequently, sentence processing becomes more noisy and error-prone.

the asymmetry between insertion and deletion. I propose that duplication error and deletion error are more likely to be autocorrected than insertion. While duplication may not be noticed because of skipping, and not seeing a word could be attributed to skipping by Levy (2008b)’s noisy-channel inference (provided that the reader does not backtrack to confirm that they indeed have skipped the missing word rather than that the word was not there at all and they did not skip anything), seeing a word that does not belong to the sentence perceived has to be the result of hallucination. The asymmetry between autocorrecting insertion and autocorrecting deletion could be implemented by modifying wFSA in Levy (2008b)’s noisy-channel inference model such that deletion has a higher weight/cost than insertion.

the between-construction difference. I propose that the between-construction difference is the result of rational eye-movement. While listening is passive perception, reading is active perception. Reading is carried out by a feedback loop, the received visual signal influences sentence processing, which influences eye-movement, which influences what visual signal is received. When reading, a reader typically receives visual input that spans several words at any given moment, which allows multiple words being processed in one fixation. However, in human visual systems, visual acuity tends to decrease with retinal eccentricity. To maximize sentence perception accuracy, one could fixate every word and repeatedly refixate longer words, so that every part of a sentence is processed in fovea, and even further slow down eye-movement so that there is no parallel word processing at all. But then they would be reading very slowly. In reality, a typical reader rationally balances between accuracy and speed and processes certain words in parafoveal only. Readers tend to skip short words like function words (for a review, see Rayner, 1998). As mentioned in Section 1, more predictable words are more likely to be skipped as well. Using the gaze contingent boundary paradigm, Balota et al. (1985) found that when a word is highly predicted by the context, it is more likely to get skipped when the reader has got a parafoveal “preview” of it from a string that is its neighbor than from a string that

is not its neighbor (for a review, see Staub, 2015). This suggests that strings sufficiently resemble what is predicted can be misperceived as what is predicted. Meanwhile, in a sentence, the latter words are generally more predictable than the earlier ones. This is intuitive and is supported by the fact that as a sentence unfolds, its words tend to elicit progressively smaller N400 responses as shown by Kutas & Hillyard (1980). Everything taken together, one concludes that latter words are more likely to be misperceived than earlier words. This could explain the between-construction difference. The latter the discrepancy between the implausible sentence and its nearest plausible neighbor (in terms of Levenshtein distance) is located, the more often it was misinterpreted in Gibson et al. (2013) and Poppels & Levy (2016). Recall that across all six construction pairs tested: participants made more non-literal interpretations for DO/PO goal and PO/DO benefactive than for transitive/intransitive and PP adjuncts, and more for transitive/intransitive and PP adjuncts than for active/passive and uninverted/inverted. Now, examine their sentence stimuli. In implausible DO/PO goal and DO/PO benefactive, comparing 11i/j/q/r to 11k/l/s/t, the deletion/insertion of “to”/“for” occurs after three content words. In implausible transitive/intransitive, comparing 11m/n to 11o/p, the deletion/insertion of “from” occurs after two content words. In implausible PP adjuncts, comparing 12a/b to 12c/d, the transposition “from” or “to” occurs after two content words. In implausible active/passive, comparing 11a/b to 11c/d, the deletion/insertion of “was” occurs after one content word. In implausible non-inverted/inverted, comparing 11e/f to 11g/h, “onto” is deleted/inserted as early as at the beginning of the sentence. Existing rational reading models (Bicknell & Levy, 2012, 2010a, 2010b) already exhibit many human-like reading behaviors: they take in noisy character input, the noisiness of which depends on eccentricity, implement predictive processing using bigram model, and implement the effect of predictive processing on eye-movement. For example, if replacing the bigram model with a large language model whose predictive processing is sensitive to “deeper” left context, reading “From the closet, she pulled out a racket for the upcoming game”, the renewed model may misperceive “racket” as “jacket” in first-pass reading when fixating to the left of “racket”, skip “racket”, and then when it perceives “game” regresses back to “racket” to correct its previous misperception. However, existing rational reading models will not exhibit any of the misinterpretation behavior discussed. They can potentially correct intra-word errors but never inter-word errors because they are certain about word boundaries. This gap could be bridged by incorporating how Levy (2008b)’s noisy-channel inference model calculates the likelihood of the visual input given a hypothesized sentence identity, which permits both intra-word errors and inter-word errors.

the effect of frequency of implausible sentences. I propose that the mechanism behind the effect of predictive processing on eye-movement is that the language comprehension system decides the length of “progression” (forward saccade or saccade to the right) based on the posterior of the hypothesis about the sentence identity with the highest posterior: the higher this posterior is, the bigger a progression it will make. When the prior of a sentence is already high, it needs less likelihood for its posterior to reach the confidence threshold that would trigger a big progression that may skip a word. When the reading content has more implausible sentences as in Gibson et al. (2013)’s Experiment 3 compared to their Experiment 1, prior for plausible sentences decreases. Having processed “The mother gave the daughter”, while the likelihood of perceiving “. . . to the candle” given “. . . the candle” does not differ between these two experiments, but the prior and the posterior of “The mother gave the daughter the candle” is higher in Experiment 1 than in Experiment 3, which means that participants will make more fixations toward the end of the sentence and are less likely to misperceive the sentence in Experiment 3 than in Experiment 1.

the effect of frequency of ungrammatical sentences. The effect of frequency of ungrammatical sentences could be the result of similarly adaptive behaviors in reading. When the reading material has more ungrammatical sentences with inter-word errors (deletion, insertion, and transposition) as in Gibson et al. (2013)’s Experiment 2 compared to their Experiment 1, participants would read faster if they ramp up their autocorrecting skill by reading less incrementally (more parallelly), which would result in more misperception, specifically inter-word errors, for even grammatical sentences. This hypothesis predicts that when ungrammatical fillers contain intra-word errors as opposed to inter-word errors, participants are less likely to misperceive grammatical implausible experimental sentences. This hypothesis assumes that participants would try to mentally correct ungrammatical fillers in Experiment 2 into their grammatical neighbors and answer comprehension questions based on the grammatical version. Gibson et al. (2013) did not publish results on question-answering to fillers. Unlike other effects, I think this particular effect may actually be caused by Gibson et al. (2013)’s noisy-channel inference: participants seemed to make more non-literal interpretations in Experiment 3

compared to Experiment 1 because on top of (potentially more) misperception they indeed sometimes engaged in Gibson et al. (2013)'s noisy-channel inference when they perceived the implausible to be implausible. Perceiving ungrammatical fillers to be ungrammatical may cause participants to suspect that the experimenter may sometimes make production errors (deletion, insertion, and transposition) and potentially engage in Gibson et al. (2013)'s noisy-channel inference when reading grammatical implausible sentences. I will resume this discussion in Subsection 4.4.

4.3 (Which) noisy-channel inference during speech comprehension?

Despite the lack of compelling evidence, Gibson et al. (2013)'s noisy-channel inference as well as the four properties of it that Gibson et al. (2013) speculated are all intuitive and reasonable. Their predictions 3 and 4 say that Gibson et al. (2013)'s noisy-channel inference should increase with the perceived noise rate of the communicative situation and decrease if semantically anomalous meanings are more likely to be communicated. I completely agree. However, in reality, these two imagined scenarios tend to co-occur: someone who is not fluent in the language they are speaking is more likely not only to make production errors but also to intend information that you find implausible, as they likely come from a different culture. Therefore, when listening to such a person speaking, it is not obvious whether the comprehender should infer more or infer less. Meanwhile, deciphering speech in an unfamiliar accent is more difficult. When speech comprehension becomes challenging, individual word processing is sometimes more prolonged and thus words are processed more parallelly (recall the relaxed parallel word processing hypothesis described in Section 3). Consequently, sentence processing becomes more noisy and error-prone.

Gibson et al. (2017) had American participants listen to sentences from Gibson et al. (2013) in an American or foreign accent and answered comprehension questions. They found that participants made more non-literal interpretations when the speech was in a foreign accent than when in an American accent, and concluded that comprehenders assumed a higher error rate in foreign-accented speech and thus engaged in Gibson et al. (2013)'s noisy-channel inference more. My main criticism of the conclusion drawn in Gibson et al. (2013) applies here too: since the task was to comprehend a list of unconnected sentences not embedded in any communicative context, it is odd that participants would not comprehend these sentences literally but attempt to infer the intended message of some producer that was nowhere to be seen.

Again, I suspect that these results were the result of misperception. Recall from Section 3 Dufour et al. (2022) demonstrated that transposition errors can be autocorrected during speech comprehension. So may deletion and insertion errors. Although Gibson et al. (2017) preventively addressed this misperception explanation by conducting a norming study on the auditory stimuli, what they found actually somewhat supports the misperception hypothesis. In the norming study, participants transcribed experimental sentences presented in the main study. The transcription error rate was higher in DO foreign-accent than DO American-accent than the rest conditions of structure accent combinations, matching the trend in the rate of non-literal interpretation in the main study (in both rounds of data collection, see their Fig. 1. And Fig. 2.). Gibson et al. (2017) claimed to have eliminated the misperception hypothesis by pointing out that the error rate in the norming study is much lower than the non-literal interpretation rate in the main study for all constructions. However, idiosyncrasies of the task demand may contribute to these different baselines. Furthermore, in the norming study, "participants were asked to transcribe what the speakers said, even if it was implausible". While an instruction like this would discourage participants from engaging in Gibson et al. (2013)'s noisy-channel inference (which is likely what the researchers intended), it would also signal to participants that the sentences they are going to transcribe may be tricky, and thus they should pay extra attention. Gibson et al. (2017) did not say whether participants in the main study or in the norming study were allowed to re-play each sentence.

The other potential alternative explanation, which is not mutually exclusive with but reinforces the misperception hypothesis, is Levy (2008b)'s noisy-channel inference. When listening to speech of an unfamiliar accent, listeners face an additional layer of noise on top of the baseline noise in their perception. Therefore, they may be more uncertain about what they have perceived, and since they cannot backtrack as in reading, they may rely even less on bottom-up evidence and more on prior. If participants were not allowed to re-play the spoken sentences, on trials where they interpreted the sentence non-literally, they may have actually perceived the spoken sentences correctly but engaged in Levy (2008b)'s noisy-channel inference: they interpreted an

implausible sentence non-literally not because they thought the producer intended the plausible but produced the implausible, but because they thought the sentence presented to them was plausible.

4.4 Evidence for Gibson et al. (2013)’s noisy-channel inference from language comprehension in “communicative” settings

It is of everyday experience that we infer an interlocutor to intend something other than what we perceived they had said, especially when communication is noisy. The noisy-channel inference posited by Gibson et al. (2013) is definitely real though the theory needs to be updated in accordance with human participants results. Is there any less disputable evidence for Gibson et al. (2013)’s noisy-channel inference from existing studies? Any result mentioned so far that is still unexplained was likely caused by Gibson et al. (2013)’s noisy-channel inference. Recall studies demonstrating that participants seemed to misinterpret garden-path sentences (Christianson et al., 2001) and garden-path-like sentences (Levy, 2011; Levy et al., 2009; Tabor et al., 2004). Reading a garden-path sentence, which is of some rare structure but initially overlaps with a sentence of some frequent structure, one tends to make the wrong syntactic analysis and later correct the previous analysis. Reading a garden-path-like sentence, which is of some rare structure but initially resembles a sentence of some more frequent structure, one tends to suspect that they have misperceived the difference part and re-read those parts, or even mentally correct the difference part and then later correct that previous correction. The surprisal theory and Levy (2008b)’s noisy-channel inference theory together explain why participants would have difficulty processing these sentences, but they do not explain why after potential backtracking, which was possible in Christianson et al. (2001) and Levy et al. (2009), participants still misinterpreted some of these sentences more than they did control sentences. Given results that garden-path sentences are sometimes rejected as ungrammatical in grammaticality judgment tasks (Ferreira & Henderson, 1991; Warner & Glass, 1987), it is likely that the seeming misinterpretation of garden-path sentences is the result of engaging in Gibson et al. (2013)’s noisy-channel inference. Given that the comprehension questions had only two possible responses (i.e., “yes” or “no”), which implies that the sentence was comprehensible, participants may have felt compelled to engage in Gibson et al. (2013)’s noisy channel inference and infer that the experimenter had intended a similarly looking and grammatical sentence but made an error when typing it.

Although it may seem trivial to demonstrate that participants tend to infer a sentence they find ungrammatical to have been corrupted by a production error from a grammatical sentence that the producer intended, since no producer would ever intend an ungrammatical sentence¹², revealing among the infinite grammatical sentences the producer could have intended which one participants think is the most probable still sheds light on how humans reason about linguistic signals. Reading 1a “While the man hunted the deer ran into the woods”, participants are probably more likely to infer that the producer intended 3a “While the man hunted the deer it ran into the woods” (a function word deleted) than “While the man hunted the deer the pheasant ran into the woods” (a content word deleted) or “The man hunted the deer into the woods” (a content word inserted), consistent with prediction 1 and 2 of Gibson et al. (2013)’s noisy-channel inference theory. I will discuss even more potential evidence for Gibson et al. (2013)’s noisy-channel inference at the end of this paper.

5 The depth-charge illusion

The depth-charge illusion is a verbal illusion first described in Wason & Reich (1979). 12a is an example sentence. It reads/sounds like it means the plausible 12b while it literally means the implausible 12c:

12a No head injury is too trivial to be ignored.

12b No matter how trivial head injuries are, they should not be ignored.

12c All head injuries should be ignored however trivial.

¹²Unless the producer is still learning the language or has some language-related deficit, which is also an interesting thing to study

“As yet, there has not been a satisfactory explanation of how depth-charge illusions are understood, but researchers have investigated several potentially relevant factors” (Zhang et al., 2023). For a review of these factors and the origin of the name “depth-charge”, refer to the Introduction section of Zhang et al. (2023).

While the few studies to date that investigated the depth-charge illusion test sentences that are of the form “No X is too Y to Z” exclusively, here is a depth-charge illusion sentence I thought of that does not strictly conform to this form: “No student thinks too slow to be given up upon.”

5.1 Zhang et al. (2023)’s noisy-channel “explanation”

Zhang et al. (2023) came up with a noisy-channel inference explanation of the depth-charge illusion. They hypothesized that reading 12a participants inferred that the producer had actually intended the plausible 13a but made a production error by replacing *so...as to* with *too...to* (an error type they called structural substitution), or intended the plausible 13b but mistakenly replaced *treated* with *ignored* (an error type they called antonym substitution). This hypothesis is obviously wrong since an illusion is a perceptual phenomenon, not a communicative phenomenon (more discussion on perception versus reasoning in language comprehension in Section 6). People who experience this illusion misperceive the implausible sentence as its plausible neighbor rather than perceive it correctly and then infer that the producer has intended its plausible neighbor. Plus, the noisy-channel inference account predicts that reading the implausible 13c one would infer that the producer intended the plausible 13a but mistakenly replaced “ignored” with “treated”, or intended the plausible 13b but mistakenly replaced “too...to” with “so...as to”. 13c easily reads implausible.

13a No head injury is *so* trivial *as to* be ignored.

13b No head injury is too trivial to be *treated*.

13c No head injury is so trivial as to be treated.

5.2 A rule deactivation theory

So what causes this illusion? Clearly, this illusion has something to do with *too...to* since while 12a causes the depth-charge illusion 14a easily reads implausible. It also has something to do with the sentential negation, since 14b easily reads implausible.

14a No head injury is so severe as to be ignored.

14b Some head injuries are too trivial to be ignored.

Borrowing the concept of semantic transparency from morphology that the compound word “blueberry” is more semantically transparent than the compound word “strawberry”. Imagine you do not know the words “blueberry” or “strawberry”, but “blue”, “straw”, and “berry”. Presented with a basket of a variety of berries, you are more likely to correctly identify which is “blueberry” than which is “strawberry”. The construction *too...to* is more semantically opaque compared to *so...as to*. Without having learned the construction *so...as to*, one would decipher the meaning of *A is so B as to be C* as *A is so B and is thus to be C*, consistent with the construction’s real meaning. In contrast, without having learned the construction *too...to*, one would interpret *A is too B to be C* as that *A is too B and A is to be C*, contradicting the construction’s real meaning that *A is too B and thus is not to be C*. Applying this more naive parsing to 12a yields that no head injury is too trivial and is thus to be ignored, consistent with 12b. When parsing *too...to*, computing the rule-based interpretation following the rule of *too...to* is more difficult than computing the more transparent interpretation.

Processing sentential negation also incurs extra computational cost. No/not is no regular adjective/adverb. A typical reader would not have trouble processing the sentence “you can have a lovely little old rectangular green French silver whittling knife” even if they has not ever read a sentence with so many adjectives. In contrast, there is a limit to how many negations (e.g., never, not, no) someone’s language comprehension system can process in a sentence. Beyond that limit, comprehenders resort to logical reasoning to convert an unparsable sentence into a parsable equivalent. For example, by canceling out negatives, one converts a

double negatives sentence like “no student does not believe the class is useful” into “students believe the class is useful”.

So why would readers apply the more transparent interpretation for 12a whereas the rule-based interpretation for 13b and 14b? Here I propose a rule deactivation theory. I propose it exceeds the capacity of a typical human language comprehension system to follow the rule of *too...to* while incrementally processing *No A too B to C*. Reading *No A too B to C* and not aware of its trickiness, readers process the sentence incrementally and go with the more transparent interpretation of *too...to*. If the interpretation turns out to be plausible as in 12a, that was it. Otherwise, readers may try to break down the sentence and process it non-incrementally. For example, reading 13b, one could process “too trivial to be treated” first and then plug in the resulted representation back into the original sentence, thereby finding the sentence plausible.

According to Zhang et al. (2023), my rule deactivation theory is similar to that of Fortuin (2014), which Zhang and colleagues referred to as “a construction-based non-illusory account”.

My rule deactivation theory makes a wealth of predictions. 15a-p in Table 3 are 16 sentences created by crossing four binary variables: whether sentence initial is “No” or “Most” X whether construction is *so...as to* or *too...to* X whether the adjective is “trivial” or “severe” X whether the verb is “treat” or “ignore”. 15a-d are *All A so B as to C*, 15e-h are *All A too B to C*, 15i-l are *No A so B as to C*, 15m-p are *No A too B to C*. Assuming processing negation takes less time than applying the rule of *too...to*, my theory predicts that *All A so B as to C* is processed faster than *No A so B as to C*, which is processed faster than *All A too B to C*, which is processed faster than *No A too B to C*, with the exception that depth-charge illusion sentences are processed as fast as *No A so B as to C*. These processing time predictions are shown in column *task 1 predicted response time (if plausibility judgment as predicted)*.

Table 3: Predictions by my rule deactivation theory

sentence	transparent interpretation: intelligibility	transparent interpretation: plausibility	rule-based interpretation: intelligibility	rule-based interpretation: plausibility	task 1 predicted intelligibility judgment	task 1 predicted plausibility judgment	task 2 predicted plausibility judgment	task 1 predicted response time (if plausibility judgment as predicted)	task 2 predicted correlation between the main plausibility rating and the reference plausibility rating	task 2 predicted response time (if plausibility judgment as predicted)
15a All head injuries are so trivial as to be treated.	0	NA	NA	NA	0	NA	0	very fast	0	slower than very fast

15b All head injuries are so trivial as to be ignored.	1	0	NA	NA	1	0	0	very fast	-	very fast
15c All head injuries are so severe as to be treated.	1	1	NA	NA	1	1	1	very fast	+	very fast
15d All head injuries are so severe as to be ignored.	0	NA	NA	NA	0	NA	0	very fast	0	slower than very fast
15e All head injuries are too trivial to be treated.	NA	NA	1	0	1	0	0	medium	-	medium
15f All head injuries are too trivial to be ignored.	NA	NA	0	NA	0	NA	0	medium	0	slower than medium
15g All head injuries are too severe to be treated.	NA	NA	0	NA	0	NA	0	medium	0	slower than medium
15h All head injuries are too severe to be ignored.	NA	NA	1	1	1	1	1	medium	+	medium
15i No head injury is so trivial as to be treated.	0	NA	NA	NA	0	NA	0	fast	0	slower than fast
15j No head injury is so trivial as to be ignored.	1	1	NA	NA	1	1	1	fast	+	fast
15k No head injury is so severe as to be treated.	1	0	NA	NA	1	0	0	fast	-	fast
15l No head injury is so severe as to be ignored.	0	NA	NA	NA	0	NA	0	fast	0	slower than fast
15m No head injury is too trivial to be treated.	0	NA	1	1	1	1	1	slow	+	slow

15n No head injury is too trivial to be ignored.	1	1	0	NA	1	1	1	fast	+	fast
15o No head injury is too severe to be treated.	1	0	0	NA	0	NA	0	slow	0	slower than slow
15p No head injury is too severe to be ignored.	0	NA	1	0	1	0	0	slow	-	slow

Every column in Table 3 is about either plausibility or “intelligibility”. So far in this section, I have been labeling a sentence as either plausible or implausible as Zhang et al. (2023) did. Technically, while all implausible sentences are not plausible, not all sentences that are not plausible are implausible. To form a plausibility opinion about any sentence, the sentence needs to have some meaning or make some sense (i.e., be intelligible) first.

Some people argue that all sentences, including “Colorless green ideas sleep furiously”, make some sense however little it may be. For expository purpose, I adopt a relatively conservative definition of intelligibility here, under which “Colorless green ideas sleep furiously” is unintelligible and is neither plausible nor implausible. The sentence “My new coworker is a block of ice” would be unintelligible without knowledge of the metaphor. “a good desk has a flat surface”, “a good desk has a rugged surface”, and “a good desk eats a lot” are plausible, implausible, and unintelligible, respectively. Among the sentences I have labeled as implausible thus far, while 12c is truly implausible, while 13c, 14a, and 14b are unintelligible. For a quantifier *too...to* or quantifier *so...as to* sentence to be plausible, they need to meet three requirements that are progressively stringent, which means meeting the first requirement is implied by meeting the second requirement, and meeting the second requirement is implied by meeting the third requirement. First, both the adjective and the verb have to be semantically relevant to the noun, and through transitivity the adjective and the verb will be semantically relevant too. Second, the adjective and the verb have to be semantically congruent given the construction they are embedded in. For examples, embedded in *too...to*, “trivial” and “treated” are congruent with each other, “severe” and “ignored” are congruent with each other; embedded in *so...as to*, “trivial” and “ignored” are congruent with each other, “severe” and “treated” are congruent with each other. Third, the relationship between the noun and the predicate as defined by the quantifier has to be plausible. Sentences that do not meet the first two requirements are semantically anomalous and are neither plausible nor implausible. Whether a sentence is semantically anomalous can be assessed by asking “does it make any sense?” and is pretty much objective; whether a sentence is plausible can be assessed by asking “how much you agree with this sentence” and is more subjective. The first four columns of Table 3 reflect this analysis. Sentences that have 0 for *intelligibility* have NA for *plausibility*. All *A so B as to C* and *No A so B as to C* do not have a rule-based interpretation; therefore, they have NA in column *rule-based interpretation: intelligibility*. All *A too B to C* do not have a transparent interpretation as readers should activate the rule of *too...to* just fine. Therefore, they have NA in column *rule-based interpretation: intelligibility*. *No A too B to C* have both transparent interpretation and rule-based interpretation. Reading *No A too B to C* and not aware of its trickiness, readers process the sentence incrementally and go with the more transparent interpretation of *too...to*. If the interpretation turns out to be plausible as in 12a now reproduced as 15n, that was it. For the other three sentences of *No A too B to C*, readers may then attempt the rule-based interpretation, if they are sufficiently motivated.

How to operationalize intelligibility, plausibility, and processing time? A sentence rating task can be conducted by instructing participants that they will be reading sentences. For each sentence, they will be asked if the sentence makes any sense, and if they answer yes, they will be asked a second question “on a scale from 1-5, 1 being strongly disagree, 3 being neither agree or disagree, 5 being strongly agree, how much do you agree with what the sentence expresses?” Processing times are operationalized as response times to the

first question. Predictions of answers to the first and second questions are shown in column *task 1 predicted intelligibility judgment* and column *task 1 predicted plausibility judgment* respectively. Predictions of response times to the first question are shown in column *task 1 predicted response time (if plausibility judgment as predicted)*. This task is a little unnatural as it encourages participants to partition language comprehension into two processes. I do not think humans can actually separate language comprehension that way, which is why I operationalize sentence processing time as response time to the first question. A more natural task is to ask participants the plausibility question only. This introduces two problems, both of which could be addressed. First, it implies that all sentences are intelligible and thus participants may dwell longer on intelligible sentences before responding. Column *task 2 predicted response time (if plausibility judgment as predicted)* reflects this consideration. For any sentence that is unintelligible, their *task 2 predicted response time (if plausibility judgment as predicted)* is slower than their *task 1 predicted response time (if plausibility judgment as predicted)*. Second, it is not clear how participants will rate intelligible sentences – will they strongly disagree, or neither agree or disagree? How can we tell if participants found a sentence unintelligible or implausible? This problem can be addressed by conducting a reference sentence rating task after the main task, where participants are asked to rate the plausibility of sentences such as “All head injuries should be treated, no matter how trivial they are”. This reference plausibility rating should correlate positively with plausible test sentences, correlate negatively with implausible sentences, and have no correlation with unintelligible sentences, as shown in column *predicted correlation between the main plausibility rating and the reference plausibility rating*.

Besides intelligibility, plausibility, and processing time, my rule deactivation theory also makes predictions about what words comprehenders predict to come up next through predictive processing. Recall my theory says that reading *No A too B to C* and not aware of its trickiness, readers process the sentence incrementally and go with the more transparent interpretation of *too...to*. This predicts that reading “No head injury is too trivial to be ...”, participants should be more likely to predict the next word to be “ignored” than “treated”, and are more likely to complete “All head injuries are so severe as to be ...”, “All head injuries are too severe to be ...”, and “No head injury is so trivial as to be ...” with “treated”, “ignored”, and “ignored” respectively. These predictions can be tested in a sentence completion task, which is the standard norming task employed by studies investigating predictive processing (see Federmeier et al., 2007 for their cloze norming procedure).

5.3 Relevant results from previous studies

My theory’s predictions about the correlation between the main plausibility rating and the reference plausibility rating are consistent with data from Zhang et al. (2023). Zhang et al. (2023) had participants rate the plausibility of *too...to* sentences in their Experiment 1 and another group of participants rate the plausibility of sentences like “All head injuries should be treated, no matter how trivial they are” in their two versions of Experiment 2. In Experiment 1, Zhang et al. (2023) presented participants sentences from four conditions and had them rate “whether the sentence made sense and whether it was written well using a 7-point Likert scale (1 = ‘makes no sense/written poorly’, 4 = ‘intermediate’, 7 = ‘makes perfect sense/written well’).” 16a-d are example sentences. At the end of each sentence, I note which color Zhang and colleagues plotted them in throughout their paper and how each condition would be labeled under my theory, which I will be using when referring to them.

16a *Some* head injuries are too *severe* to be ignored. (purple; plausible)

16b *Some* head injuries are too *trivial* to be ignored. (green; unintelligible)

16c *No* head injury is too *severe* to be ignored. (blue; implausible)

16d *No* head injury is too *trivial* to be ignored. (orange; illusion)

Subjects rated 16a as more plausible than 16d than 16c than 16b.

In the preliminary version of their Experiment 2, Zhang and colleagues presented participants with sentences like “Head injuries are in general too severe to be ignored” and for each sentence has them answer the question ‘According to what you believe about the world, how much do you agree with the sentence?’ on a fully labeled 7-point Likert scale (1 = ‘completely disagree’, 4 = ‘intermediate’, 7 = ‘completely agree’). The latter

version, which they put in the main body of their paper, was the same as the preliminary version except that the sentences had a different structure: “Head injuries should be treated, no matter how trivial they are”. Comparing rating results in their Experiment 1 against rating results in their Experiment 2, the plausibility ratings for the illusion correlated positively with the norming score in both versions of their Experiment 2, while the plausibility ratings for the implausible correlated negatively with the norming score in the first version of their Experiment 2. Recall that my theory predicted that the the reference plausibility rating should correlate positively with plausible test sentences, correlate negatively with implausible sentences, and have no correlation with unintelligible sentence. The reason that they did not find a positive correlation between the plausible and the norming score may be that the plausibility ratings for the plausible were already at the ceiling even for the less plausible items, which may be caused by the generosity of the sentence using “some” instead of “all” as the quantifier.

Their measurement was likely extra noisy because the norming score and the main rating were collected from different pools of participants. Furthermore, in their Experiment 1, instead of asking participants a question like “how much you agree with what the sentence”, they asked “whether the sentence made sense and whether it was written well”.

My theory’s predictions about sentence completion are confirmed by data from Paape et al. (2020), who tested German material on German participants. In their Experiment 4, they asked participants to complete sentences in two conditions. Here is an example item in English translation:

17a *Some* head injuries are too innocuous to ...

17b *No* head injury is too innocuous to ...

Pappe and colleagues had two groups of coders blind to the experiment design code participants’ responses following two slightly different coding schemes. They found that under both schemes while participants rarely complete 17a with something like “be ignored”, they more often completed 17b with responses like “be ignored” than “be treated”.

6 Proposition: the core process of language comprehension is a kind of perception

While language comprehension as a behavior, especially in real-life settings, calls upon cognition to apply world knowledge, pragmatic/social knowledge, knowledge of the communicative context and to model the interlocutor’s mental state including their world knowledge, pragmatic/social knowledge, and their knowledge of the communicative context in order to infer the message they intends by what they says, which may contain deceptions, irony, humor, and metaphor, I argue that the core process of language comprehension, which is indispensable to extracting the literal meaning of a sentence by decoding concepts/referents and the relationships among them (hereafter referred to as sentence processing), is a kind of perception, not cognition. Firestone & Scholl (2015) said “The challenge of understanding the natural world is to capture generalizations – to ‘carve nature at its joints’. Where are the joints of the mind? Easily, the most natural and robust distinction between types of mental processes is that between perception and cognition.”

Based on Firestone & Scholl (2015)’s characterization of perception as “encompassing both (typically unconscious) visual processing and the (conscious) percepts that result”, I define cognition as mental processes that operate on conscious input, perception as mental processes that operate on unconscious input, both of which yield at least some conscious outputs. Under this definition, sentence processing would be cognitive if its input was perceived individual words. However, this does not seem to be the case given the results discussed in 3: although sentence processing definitely requires word processing, unreliable (conscious) percepts of individual words seem to be mere byproducts of sentence processing, the goal of which is to produce percepts of meaning of sentences – even percepts of sentence identities may be byproducts as well. Results from studies using the masked word priming paradigm (Forster & Davis, 1984; for a review, see Forster, 1998) suggest that word processing can be subliminal, even though we are obviously able to perceive words.

Similarly in vision, object recognition does not require conscious experience of lower-level features, many of which we are nonetheless able to perceive. Some may counterargue that certain psycholinguistic processes

cannot be classified as either perception or cognition. While noisy reading and Gibson et al. (2013)’s noisy-channel inference clearly fall into the categories of perception and cognition, respectively, it is more ambiguous on which side of the line lies Levy (2008b)’s noisy-channel inference. However, similarly ambiguous cases also exist in the domain of vision. Are recognizing whether two people in a scene are interacting or not (Isik et al., 2019, 2017) and whether objects in a scene are stable or unstable (Pramod et al., 2022) the result of seeing or thinking?

Firestone & Scholl (2015) said that the difference between perception and cognition “is especially clear when perception and cognition deliver conflicting evidence about the world”, as in the case of illusions. While there may not be as many verbal illusions as there are visual illusions, the depth-charge illusion is one. The inability of participants to see how a garden-path sentence can be grammatical may count as a grammaticality illusion, a term that usually refers to the phenomenon that English sentences containing double center embedded relative clauses from which the middle verb is omitted tend to be processed more easily than their grammatical counterparts (Frank et al., 2016; Gibson & Thomas, 1999; Vasishth et al., 2010). For example, “The apartment that the maid who the service had sent over was well decorated” compared to “The apartment that the maid who the service had sent over was cleaning every week was well decorated”.

Similar to language comprehension, visual scene analysis can also sometimes use help from cognition, such as when inferring the nuanced social dynamics among humans from a picture, which may depict deceptions, when understanding a cartoon with deep meaning, which may contain irony, humor, or metaphor, and when solving visual relational reasoning problems like Raven’s Progressive Matrices (RPM), which are commonly used in IQ tests. In the meantime, we likely all agree that the core process of visual scene analysis, which is indispensable to decoding objects and the spatial relationships among them from a picture (hereafter referred to as basic visual scene analysis), is a kind of perception, not cognition.

For both sentence processing and basic visual scene analysis, the key building blocks are knowledge of feature mappings. As previously discussed, the goal of sentence processing is to decode concepts/referents and the relationships among them, while the goal of basic visual scene analysis is to decode objects and the spatial relationships among them. While decoding spatial relationships is a built-in feature of the visual system and is probably innate, decoding concepts from words, decoding the relationships among concepts/referents from syntactic signals, and decoding objects all require knowledge of feature mappings, which mostly need to be learned. Visual object recognition requires knowledge of how edges compose curvature, curvatures compose shapes, shapes compose parts, and parts compose objects. Auditory object recognition (recognizing a friend’s voice, an instrument, or a natural sound such as a dog barking) requires knowledge of spectro-temporal features of objects. Visual and auditory word recognition, which is really just a special case of object recognition, requires knowledge of visuo-orthographic mappings and acoustic-phonetic mappings respectively, and knowledge of word form in the respective modality. As to sentence processing, decoding concepts/referents from words requires knowledge of lexico-semantic mappings, while decoding relationships among concepts/referents requires knowledge of syntactic-thematic mappings. Humans absorb these kinds of knowledge from the environment since infancy, the representation of at least some of which seems to remain malleable into adulthood. For the remainder of this section, I will illustrate two properties shared by sentence processing and basic visual scene analysis: adaptation and predictive processing. I will also briefly discuss features of machine implementation shared by sentence processing and basic visual scene analysis.

6.1 Adaptation in object recognition

Statistical learning is obviously crucial for object recognition. Through mere observation, humans constantly learn the forms of novel objects and the new forms of familiar objects. Additionally, after sufficient exposure to speech in an unfamiliar accent, individuals can learn new acoustic-phonetic mappings and adapt to the unfamiliar accent (for reviews on adaptive speech perception, see Kleinschmidt & Jaeger, 2015; Xie et al., 2023). Many studies have tested participants’ adaptation to unfamiliar accents by training and testing them with natural speech (Bradlow & Bent, 2008; Clarke & Garrett, 2004; Sidaras et al., 2009; Tzeng et al., 2016; Xie et al., 2018, 2021). The key process underlying real-life accent adaptation is phonetic adaptation, specifically recalibration (Bertelson et al., 2003; Kleinschmidt, 2020; Kleinschmidt & Jaeger, 2015; Norris et al., 2003; Vroomen et al., 2007) as opposed to selective adaptation (Bertelson et al., 2003; Eimas &

Corbit, 1973; Kleinschmidt & Jaeger, 2015; Samuel, 1986; Vroomen et al., 2007), both of which have been demonstrated using parametrically manipulated consonants or syllables instead of natural speech as stimuli. Vroomen et al. (2007) created a continuum of sounds from a prototypical /aba/ to a prototypical /ada/. During exposure blocks, participants were presented with one sound on each trial accompanied by a video recording of a talker articulating either /aba/ or /ada/. This paradigm leverages the McGurk illusion and was first developed by Bertelson et al. (2003). During test blocks, participants categorized a sound on each trial. Exposure blocks and test blocks were interleaved. Replicating previous studies, Vroomen et al. (2007) found that participants are more likely to categorize ambiguous sounds as /aba/ as opposed to /ada/ after exposure to ambiguous sounds visually “labeled” as /aba/ (recalibration) and after repeated exposure to prototypical /ada/ “labeled” as /ada/ (selective adaptation).

Selective adaptation has been demonstrated in other subdomains of auditory object recognition: instruments (e.g., morphing between the sound of clarinet and the sound of oboe) and natural sounds (e.g., morphing between the sound of rain and the sound of wind) (Piazza et al., 2018), and voice identities (Zäske et al., 2010). As to the visual domain, a similar effect called identity aftereffect has been demonstrated in face recognition (Leopold et al., 2001; Rhodes & Jeffery, 2006). Leopold et al. (2001) created a space of faces bounded by four original faces and their anti-faces. For example, one original face has big cheeks, so its anti-face has small cheeks. Participants were first trained to recognize the four original faces. On baseline test trials, participants categorized faces from the face space into one of the four original faces. An adaptation test trial is the same as a baseline test trial except that before the onset of the test face, participants were exposed to an anti-face for 5 seconds, which biased participants’ identification of an average face toward the original face that the adapter is an anti-face to.

6.2 Adaptation in sentence processing

While it is obvious that we adapt our lexico-semantic mappings throughout the lifespan as words change their meaning (Hamilton et al., 2016), the syntactic-thematic mappings of a language tend to stay more stable. Still, adaptation in syntactic-thematic mappings can be tested in labs. Kaschak & Glenberg (2004) and Fraundorf & Jaeger (2016) demonstrated that participants who had never encountered the western Pennsylvania dialectal structure “needs + [past participle]”, such as using “The car needs washed” to mean “The car needs to be washed”, learn to comprehend this structure quickly through mere exposure to naturalistic passages that use it (Fraundorf & Jaeger, 2016; Kaschak & Glenberg, 2004). In Kaschak & Glenberg (2004), participants on each trial read a narrative of two interacting characters, where one character in one sentence used the standard “needs to be” or the idiomatic “needs + [past participle]”. Participants pressed a key to proceed to the next sentence and their per-sentence reading times were recorded. While reading times of both standard and idiomatic structures decreased over trials, the reading time of “need” sentences decreased drastically within the first few trials. Fraundorf & Jaeger (2016)’s leveraged crowdsourcing to achieve a between-subjects design. On each trial, participants read a paragraph constructed to resemble a work email that may or may not contain an idiomatic “need” structure. One group of participants was recruited from Ohio and western Pennsylvania, whereas the other Colorado. Participants pressed a key to proceed to the next word and their per-word reading times were recorded. A reader unfamiliar with the idiomatic “need” structure reading “The car needs washed” will likely interpret “washed” as an adjective and expect the sentence to continue with something like “seat covers”, making “The car needs washed before tomorrow” a garden-path sentence disambiguated at “before”, which predicts great precessing difficulty with “before” (like with “fell” in “The horse raced past the barn fell”). As expected, response times to the disambiguating word were initially longer for the group unfamiliar with the idiomatic “need” structure than for the familiar group. However, this difference diminished almost completely within the first few trials, replicating Kaschak & Glenberg (2004).

Humans learn not only new syntactic-thematic mappings, but also the statistics of structures¹³. Garden-path sentences are difficult to comprehend because they are rare in the natural linguistic environment. When they become more frequent, participants quickly adapt (Fine et al., 2013; Fine & Jaeger, 2016; Jaeger et al., 2018; Yan & Jaeger, 2020). In Fine et al. (2013), participants performed a word-by-word self-paced reading task on sentences like 20a-d. 20c is a garden-path sentence with a similar structure to “The horse raced past the barn

¹³Learning new syntactic-thematic mappings can be viewed as a special case of statistical learning: comprehenders learn that the frequency of this new structure goes from 0 to positive.

fell”. While *spoke* in 20b and *told* in 20d are unambiguously a main verb and part of a reduced relative clause respectively, it is ambiguous whether *warned* in 20a/c was a main verb or part of a reduced relative clause. Obviously, *warned* is more likely to be a main verb than to be part of a reduced relative clause. Response times to the disambiguating word in RC+ambiguous (**conducted** in 20c) and its counterparts in the other conditions were measured.

20a The experienced soldiers *warned* about the dangers **before** the midnight raid. (MV+ambiguous)

20b The experienced soldiers *spoke* about the dangers **before** the midnight raid. (MV+unambiguous)

20c The experienced soldiers *warned* about the dangers **conducted** the midnight raid. (RC+ambiguous)

20d The experienced soldiers who were *told* about the dangers **conducted** the midnight raid. (RC+unambiguous)

The results passed both sanity checks: 1) response times to the disambiguating word were longer for RC+ambiguous than for the other conditions, demonstrating the effect of garden-path; 2) response times to the disambiguating word decreased for all conditions over the course of the experiment, demonstrating the effect of task adaptation. As predicted by the authors’ sentence processing adaptation hypothesis, response times to the disambiguating word in RC+ambiguous decreased more drastically over the course of the experiment than response times to the other conditions, demonstrating the effect of adaptation to reduced relative clauses.

6.3 Predictive processing in sentence processing

All of the studies discussed in the last section as evidence for adaptation in sentence processing are also evidence for predictive processing in sentence processing. By measuring reading time as an index of learned linguistic probability, researchers assume predictive processing. Only when researchers have assumed that the more predicted a word is by the comprehender, the faster it is processed by the comprehender, when they observe that the structure that becomes more frequent is getting processed faster by the participants, can they conclude that such occurrence statistics are getting absorbed by the comprehender. Because predictive processing requires and manifests learning, all of the studies I will be discussing in this section as evidence for predictive processing are also evidence for adaptation.

While the studies from the last section manifest learning of structure occurrence statistics, most studies investigating predictive processing demonstrate learning of (structure-dependent) lexical co-occurrence statistics. For example, “He spread the warm bread with. . .” is more likely to be completed with “butter” than with “socks” (Kutas & Hillyard, 1980). In contrast, “He spread the warm butter with. . .” is more likely to be completed with “a knife” than with “bread”. The lack of predictive reciprocity from “butter” to “bread” highlights how humans learn lexical co-occurrence statistics that are dependent on the structure of the language.

Humans also learn lexical-structural co-occurrence statistics such as verb bias, the likelihood of the co-occurrence of a particular verb and a syntactic alternative that can be associated with it (e.g., that “raced” is more frequently intransitive than transitive). While both “The farmer loaded apples into the cart” and “The farmer loaded the cart with apples” sound fine, “The farmer dumped apples into the cart” sounds fine but “The farmer dumped the cart with apples” does not (Levin, 1993). Recall the first sentence example used in this paper “The women discussed the dogs on the beach”. According to Jurafsky (1996), readers tend to attach the prepositional phrase “on the beach” to the noun “dogs” than to the verb “discussed”. On the contrary, reading “The women kept the dogs on the beach”, readers tend to attach the prepositional phrase to the verb “kept” than to the noun “dogs”. Similarly, some verbs, like *admit*, tend to be followed by a clause as the object, while others, like *confirm*, tend to be followed by a noun phrase. In their Experiment 1, Wilson & Garnsey (2009) had participants perform self-paced reading on potentially temporarily ambiguous sentences like 21a-f and measured reading times to disambiguating regions. Reading 21a/c, it is ambiguous whether “the mistake” is the object of *admitted* or the participant of a clause until they reach the disambiguating **might/because**. Similarly, reading 21d/f, it is ambiguous whether “the rumor” is the object of *confirmed* or the participant of the clause until they reach the disambiguating **could/when**. All verbs were categorized as clause biased versus noun phrase biased based on the results from their norming study, where a different

group of participants were asked to complete prompts like “Susan admitted...”, following Garnsey et al. (1997).

21a The ticket agent *admitted* the mistake **might** not have been caught. (clause biased, clause without “that”)

21b The ticket agent *admitted* that the mistake **might** not have been caught. (clause biased, clause with “that”)

21c The ticket agent *admitted* the mistake **because** she had been caught. (clause biased, noun phrase)

21d The CIA director *confirmed* the rumor **could** mean a security leak. (noun phrase biased, clause without “that”)

21e The CIA director *confirmed* that the rumor **could** mean a security leak. (noun phrase biased, clause with “that”)

21f The CIA director *confirmed* the rumor **when** he testified before Congress. (noun phrase biased, noun phrase)

There were three key findings. First, participants processed the disambiguating region more slowly for clauses without “that” than for clauses with “that” for both verb types, which is already evidence for predictive processing. Second, the aforementioned difference was bigger when the verb was noun phrase biased than when the verb was clause biased.

Third, participants processed the disambiguating region more slowly in the clauses without “that” condition than in the noun phrase condition when the verb was noun phrase biased, a pattern that was reversed when the verb was clause biased. The latter two findings demonstrate the effect of verb bias on predictive processing. All three findings were replicated in Experiment 2, which was an eye-tracking version of Experiment 1.

6.4 Predictive processing and adaptation in basic visual scene analysis

Humans learn statistical regularities of not only the linguistic environment but also the visual environment, such as the co-occurrence between objects and their spatial relationship. Predictive processing in vision has been manifested as context-dependent object search and object recognition (for reviews see Oliva & Torralba, 2007; Vö, 2021). Seeing a scene, participants automatically predict what kind of objects will be there (behavioral: Zhang et al., 2020; Bomatter et al., 2021; EEG: Ganis & Kutas, 2003; Vö & Wolfe, 2013), where they will be (Castelhano & Henderson, 2007; Eckstein et al., 2006; Torralba et al., 2006), and what sizes they will be in (Eckstein et al., 2017). Looking for a toothbrush in a picture of some bathroom is faster when the toothbrush is normally sized than oversized (Eckstein et al., 2017). Searching for objects in pictures of indoor home environments, participants tend to look higher when they search for a painting than when they search for a mug (Torralba et al., 2006). When searching for a chimney, participants tend to look around the roof of buildings before they look elsewhere (Eckstein et al., 2006). Castelhano & Henderson (2007) developed what they call a flash-preview moving-window paradigm. On each trial, participants were presented briefly a preview picture followed by a mask, then the name of the target object, then the search scene shown only through a circular window centered at fixation that moved as participants’ eyes moved. Subjects were instructed to press a button when they found the target. A trial was scored as correct when the participant both fixated on the target object and pressed the response button. The target was found faster when the preview picture was the search scene itself, the search scene in a different size, or the search scene without the target, than when the preview picture was a different scene (similar or not) or a meaningless mask. In Ganis & Kutas (2003)’s EEG study, on each trial, participants were presented with a blank picture with a cue at the target location, followed by a scene (e.g., football players), followed by a congruent target object (e.g., a football) or an incongruent target object (e.g., a roll of toilet paper) appearing on the scene at the cued location. Participants were instructed to identify the target. The target elicited a bigger N400 and a bigger P600 when it was incongruent with the scene than when congruent. This result seems to be not fully replicated by Vö & Wolfe (2013). In Vö & Wolfe (2013)’s EEG study, on each trial, participant were presented with a scene (e.g., an opened laptop on a desk), followed by a cue appearing on the scene at the target location (e.g., to the right side of the laptop or on the laptop screen), and then by a semantically congruent target object (e.g., a mouse) or incongruent target object (e.g., a bar soap) replacing the cue.

Participants were instructed to press a key when the target on the current trial was the same as the preview trial (a 1-back task).

Compared to targets that were fully congruent (e.g., a mouse to the right side of the laptop), semantically incongruent targets (e.g., a soap bar to the right side of the laptop) elicited larger N400, whereas “syntactically” incongruent targets (e.g., a mouse leaning on the laptop screen) elicited larger P600. Using a similar trial procedure as the two EEG studies, Zhang et al. (2020) and Bomatter et al. (2021) found that participants identified objects more accurately when the object was in a congruent context than when the object was in an incongruent context or when there was no context at all. Replicating Eckstein et al. (2017), Bomatter et al. (2021) found impaired performance when targets are enlarged.

Since predictive processing requires and manifests learning, all the studies discussed in the last section as evidence for predictive processing in basic visual scene analysis are also evidence for adaptation in basic visual scene analysis. As far as I could find, more direct evidence for adaptation in basic visual scene analysis has only been shown using extremely unnatural “scenes” composed of oriented bars or rotated Ts amongst upright Ls (Chun & Jiang, 1998) or novel geometric shapes (Chun & Jiang, 1999) by a research program called contextual cueing.

6.5 From humans to machines

Because object recognition and sentence processing are essentially processes of detection, mapping, and composition of features, at least most of which are learned from the environment, they are better modeled neurally than symbolically (vision: Yamins et al., 2014; audition: Kell et al., 2018; language: Schrimpf et al., 2021). The evaluation of how well neural networks match brain internally is usually done by regressing measures of neural activity (fMRI or EEG) on model activation.

In contrast, reasoning is better modeled symbolically than neurally (Falkenhainer et al., 1989; Hofstadter & Mitchell, 1994; Newell et al., 1959). In the subdomain of visual relational reasoning, currently, the state of the art neuro-symbolic model that implements the reasoning part symbolically (Hersche et al., 2023) outperforms state of the art purely neural models (Wu et al., 2020; Zhang et al., 2021) on RPM benchmarks (Barrett et al., 2018; Hu et al., 2021; Zhang et al., 2019).

Neural networks have made great success in modeling sentence processing, yet they are lacking in “cognitive abilities required for language understanding and use in the real world” (Mahowald et al., 2023).

Because both basic visual scene analysis and sentence processing require representations of long-range dependencies, they are better modeled by neural networks that implement not just feedforward connections but also recurrent connections, such as RNN (Elman, 1990) and LSTM (Hochreiter & Schmidhuber, 1997), or attention, such as transformer (Vaswani et al., 2017), which is a key component of GPT-3 (Brown et al., 2020). Recurrent but not purely feedforward networks can solve spatial relation problems such as whether two dots are on the same object (Kim et al., 2020; Linsley et al., 2019) and whether a dot is inside an object (Villalobos et al., 2021). Zhang et al. (2020) and Bomatter et al. (2021) implemented recurrent or transformer models of object recognition in scenes that are sensitive to the visual context. By experimenting on both humans and models, the researchers found that their models’ recognition patterns resemble those of human participants, as discussed in the last section. Specifically, models performed better in congruent context, especially when the target was smaller; when the context was similarly congruent but the target was enlarged, performance worsened. In contrast, off-the-shelf feedforward object detection neural networks do not share this human “weakness” in finding oversized targets (Eckstein et al., 2017).

7 Final discussion

In this paper, I contend that the noisy-channel inference theory proposed by Gibson et al. (2013) does not explain the data it has been claimed to explain, namely, the misinterpretation of ordinary grammatical sentences (Gibson et al., 2013) and the depth-charge illusion (Zhang et al., 2023). I reviewed the original noisy-channel inference theory by Levy (2008b) and then Gibson et al. (2013)’s noisy-channel inference theory, and studies conducted to test these theories. I proposed a rational reading theory to explain the misinterpretation of ordinary grammatical sentences during reading, and a rule deactivation theory to explain the depth-charge illusion. To support my rational reading theory, I reviewed recent studies on misperception

during reading, and briefly reviewed data on eye-movement during reading and previous rational reading models. I also suggested that the seeming misinterpretation of garden-path sentences demonstrated by studies that employed the same kind of task that Gibson et al. (2013) did may be the result of Gibson et al. (2013)'s noisy-channel inference, which I will continue to discuss at the end of this section. Last, I argued that the core process of language comprehension is a kind of perception, not cognition, mostly through analogy between language comprehension and visual scene analysis.

I brought up many properties that the two behaviors share and discussed extensively the statistical learning property they share. Through the experience of sentence processing, humans learn and apply implicit knowledge of structure occurrence statistics, structure-dependent lexical co-occurrence statistics, and lexical-structural co-occurrence statistics.¹⁴ Through the experience of viewing the surroundings, humans learn and apply implicit knowledge of the spatial-relationship-dependent object co-occurrence statistics, which is analogous to structure-dependent lexical co-occurrence statistics. Statistical learning is a kind of adaptation, and the implicit knowledge gained is applied in predictive processing.

Statistical learning is a key premise of the surprisal theory, which explains reading patterns of not only garden-path sentences but also double-center-embeddings (recall “The apartment that the maid who the service had sent over was cleaning every week was well decorated” from the last section). One obvious factor behind why double-center-embeddings are difficult to process is their nested/recursive structure. This, however, cannot be the only factor, since double-edge-embeddings (e.g., “Joe believes Mary thinks John took the trash out”) are much easier to process. Contrasting double-center-embeddings with double-edge-embeddings reveals that the other factor may be the verb-final structure of center embedded relative clauses, as in “the apartment that the maid was cleaning every week” and “the maid who the service had sent over”. Given English word order is strictly participant-verb-object (SVO) and thus verb-final is rare, whereas German and Dutch are SOV though sometimes SVO and thus verb-final is much more frequent, the surprisal theory predicts the grammaticality illusion to be stronger in English than in German or Dutch. Indeed, the grammaticality illusion has been found in English native speakers reading English, in German or Dutch native speakers reading English, but not in German or Dutch native speakers reading German or Dutch (Frank et al., 2016; Vasishth et al., 2010). An RNN trained in Dutch and English to predict the surprisal of each word exhibits the cross-linguistic difference in the grammaticality effect: for English, surprisal is higher for grammatical than ungrammatical sentences; for Dutch, surprisal is higher for ungrammatical than grammatical sentences.

My main purpose to propose that the core process of language comprehension is a kind of perception in this paper is to preemptively address any future study that like Zhang et al. (2021) attempts to explain some phenomenon of incremental “first-pass” sentence processing as the result of Gibson et al. (2013)'s noisy-channel inference, a post-perceptual process as Huang & Staub (2021a) labeled it. The implications of this proposition, however, go beyond that. It provides a simple answer to some perennial questions in the field of psycholinguistics: why cannot participants find the correct parse/construction of a garden-path sentence? Why do participants have the grammaticality illusion when reading English sentences containing double-center-embedded relative clauses? Why do participants' analyses of syntactically unusual sentences seem to be “shallow and incomplete” (Ferreira et al., 2002)? Well, maybe, participants were not analyzing at all. Maybe, sentence processing is not a process of reasoning about the sentence with an explicit set of grammatical rules, which would be infallible so long as one's linguistic knowledge is perfect (infallible in the sense that a grammatical sentence is always found grammatical and that grammatical sentences are found more acceptable than ungrammatical sentences). However, sentence processing is imperfect in labs and in real life. Given “My friend said he knew a man with a wooden leg named Smith” (Stevenson, 1964), a linguistic analysis would reveal that this sentence is ambiguous. It is unclear whether the man is named Smith or the man's wooden leg is named Smith. However, an average person would not even think of the second interpretation, why the response “What was the name of his other leg?” implying the second interpretation, is funny. Even chatGPT, who seems to have superb language skills, does not analyze sentences the way linguists do.

I asked chatGPT if 1a and 2a are grammatically correct. ChatGPT correctly explained how 1a can be grammatically correct but was adamant that 2a is grammatically incorrect. Here is its analysis of 2a:

¹⁴I did not discuss learning and the effect of learning lexical occurrence statistics. According to Rayner (1998), “high-frequency words are skipped more than low-frequency words when words are 6 letters or less.”

The sentence “The coach smiled at the player tossed the frisbee by the opposing team.” is not grammatically correct.

Here are a few possible ways to correct it depending on the intended meaning:

1. “The coach smiled at the player who tossed the frisbee by the opposing team.” - This suggests that the coach smiled at the player who threw the frisbee.
2. “The coach smiled as the player tossed the frisbee by the opposing team.” - This implies that the coach was smiling while the player threw the frisbee.
3. “The coach smiled, despite the player being tossed by the opposing team’s frisbee.” - This suggests that the coach smiled even though the player was hit by the opposing team’s frisbee.

It’s important to clarify the intended meaning to convey the message accurately.

Sentence processing is incremental and probabilistic/parallel with limits. It is not so incremental given the transposed-word effect and that comprehenders maintain uncertainty about their perception of linguistic input in the past¹⁵ (Levy et al., 2009). It is not so probabilistic/parallel given that the correct parse of a garden-path sentence is typically not considered at all. Otherwise, participants would not need to backtrack upon encountering the disambiguating word (e.g., “fell” in “The horse raced past the barn fell”). Calling the word in a garden-path sentence that eliminates the more probable parse the disambiguating word is linguistically appropriate but psycholinguistically inappropriate. If participants indeed maintained uncertainty about the structure of the sentence, they should have an easier time, not a harder time, processing the disambiguating word, as the ambiguity is finally resolved.

7.1 Evidence for a generalized Gibson et al. (2013)’s noisy-channel inference from language comprehension in challenging situations

At the beginning of the last section, I briefly described how realife language comprehension and visual scene analysis often tap into cognition. Cognition can get tapped even when performing “contextless” sentence processing or basic visual scene analysis if perception alone does not do the job, which tends to happen when comprehension becomes challenging. Comprehension in challenging situations is accompanied by pupil dilation (Winn & Teece, 2021, 2022) and activation of brain areas involved in cognitive control (Blanco-Elorrieta et al., 2021; Mollica et al., 2020), resulting in lower accuracy and longer processing time (Tang et al., 2018). Language comprehension becomes challenging when the comprehender has trouble making sense of the linguistic signal they received, which may result from the comprehender not being proficient in the language and thus having inaccurate or incomplete lexico-semantic or syntactic-thematic mappings, the comprehender possessing some sensory (Winn & Teece, 2021, 2022) or language deficit, the comprehender being unfamiliar with the accent (Blanco-Elorrieta et al., 2021; Gibson et al., 2017) or the handwriting style of the producer, presence of background noise (Winn & Teece, 2021, 2022), or the producer producing sentences containing egregious errors (Mollica et al., 2020), which may be the result of the producer not being proficient in the language or possessing some motor or language deficit. Basic visual scene analysis becomes challenging when objects are heavily occluded (Tang et al., 2018), “presented under different viewpoints, at extreme scales, or under poor illumination” (Kreiman & Serre, 2020). On the fly, some challenges may be ameliorated by predictive processing (perception) and compensated for by postdictive reasoning such as Gibson et al. (2013)’s noisy-channel inference (cognition). Over time, some challenges may be overcome through adaptation. As discussed earlier in Section 4.1, Gibson et al. (2013)’s noisy-channel inference can be a double-edged sword when communicating with a non-native speaker: you may think that the producer did not mean what they said but they actually did. Predictive processing can be a double-edged sword too: because of predictive processing, misperceiving a word early on in the sentence could mislead the perception of subsequent words. When the downside of predictive processing is more pronounced as when the communication system is extra noisy, it would be rational to do less predictive processing. Indeed, compared to participants with normal hearing (NHs), who process sentences that are more semantically constraining with less effort (Winn,

¹⁵Considering that Levy (2008b)’s noisy-channel inference may be unconscious, a more accurate description would be that the comprehender’s language comprehension system maintains uncertainty about its “perception” of linguistic input in the past.

2016), the effect of word predictability on effort is much smaller for participants with cochlear implants (CIs) (Experiment 2 of Winn & Teece, 2021).

Here I generalize Gibson et al. (2013)'s noisy-channel inference to encompass any post-perceptual reasoning about the identity of the sentence produced or intended¹⁶ that accounts for not just production noise but also perception noise and noise in the communication channel. I hypothesize that when perception fails, that is when no posterior of any "hypothesis" or posteriors of more than one hypothesis reaches some confidence threshold within some very short time window¹⁷, the generalized Gibson et al. (2013)'s noisy-channel inference is activated and sustained until one and only one posterior of any "hypothesis" reaches some confidence threshold or until the comprehender gives up. My hypothesis predicts that it should generally take the comprehender more effort to end up inferring an intended sentence that makes little sense (incoherent) – they likely tried to find a different "hypothesis" with a higher posterior but could not. Results from Winn & Teece (2022) and Winn & Teece (2021) are consistent with this prediction. In Winn & Teece (2022), participants with cochlear implants (CIs) were asked to listen to and repeat sentences that were either fully intact or had a target word early on in the sentence mispronounced (one consonant replaced) or masked by noise. The target word can be inferred from the later sentential context, e.g. "picture" in "they say a picture is worth a thousand words". Measuring pupil dilation as the index of effort, the researchers found that when participants inferred the sentence to be different from the actual sentence, comprehension was more effortful when what participants ended up inferring was incoherent rather than coherent, not necessarily when what participants ended up inferring was more deviant from the actual sentence, replicating findings from Winn & Teece (2021)' Experiment 2, where CIs listened to intact sentences. For example, inferring "The dealership sold the cars" from "The dealer shuffled the cards" is less effortful than inferring "We played a game as cough and mouse" from "We played a game of cat and mouse".

I suspect that the generalized Gibson et al. (2013)'s noisy-channel inference (mental correction/repair) is ubiquitous in language comprehension but does not apply when language comprehension is neither challenging nor communicative, e.g., for an average English speaker to read a list of unconnected ordinary grammatical sentences on their own computer screen at the pace they decides as in Gibson et al. (2013).

References

- Altarriba, J., Kroll, J. F., Sholl, A., & Rayner, K. (1996). The influence of lexical and conceptual constraints on reading mixed-language sentences: Evidence from eye fixations and naming times. *Memory & Cognition*, 24(4), 477–492. <https://doi.org/10.3758/BF03200936>
- Balota, D. A., Pollatsek, A., & Rayner, K. (1985). The interaction of contextual constraints and parafoveal visual information in reading. *Cognitive Psychology*, 17(3), 364–390. [https://doi.org/10.1016/0010-0285\(85\)90013-1](https://doi.org/10.1016/0010-0285(85)90013-1)
- Barrett, D., Hill, F., Santoro, A., Morcos, A., & Lillicrap, T. (2018). Measuring abstract reasoning in neural networks. *Proceedings of the 35th International Conference on Machine Learning*, 511–520.
- Bertelson, P., Vroomen, J., & de Gelder, B. (2003). Visual Recalibration of Auditory Speech Identification: A McGurk Aftereffect. *Psychological Science*, 14(6), 592–597. https://doi.org/10.1046/j.0956-7976.2003.psci_1470.x
- Bicknell, K., & Levy, R. (2012). Why long words take longer to read: The role of uncertainty about word length. *Proceedings of the 3rd Workshop on Cognitive Modeling and Computational Linguistics (CMCL 2012)*, 21–30.
- Bicknell, K., & Levy, R. (2010a). A Rational Model of Eye Movement Control in Reading. *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, 1168–1178.

¹⁶If the comprehender is still learning the language, they may be unable to infer the sentence produced or intended and can only attempt to infer the message intended

¹⁷I hypothesize that postdictions within this very short time window, e.g., Levy (2008b)'s noisy-channel inference as demonstrated in Levy et al. (2009), are achieved through predictive processing. I do not think participants thought to themselves "hm, I wonder if that 'at' was actually 'at' not 'and'/'as'" before they made a regression. Beyond this very short time window, postdictions are achieved through the generalized Gibson et al. (2013)'s noisy-channel inference.

- Bicknell, K., & Levy, R. (2010b). Rational eye movements in reading combining uncertainty about previous words with contextual probability. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 32(32).
- Blanco-Elorrieta, E., Gwilliams, L., Marantz, A., & Pykkänen, L. (2021). Adaptation to mis-pronounced speech: Evidence for a prefrontal-cortex repair mechanism. *Scientific Reports*, 11(1), 97. <https://doi.org/10.1038/s41598-020-79640-0>
- Bock, J. K. (1986). Syntactic persistence in language production. *Cognitive Psychology*, 18(3), 355–387. [https://doi.org/10.1016/0010-0285\(86\)90004-6](https://doi.org/10.1016/0010-0285(86)90004-6)
- Bock, K., & Griffin, Z. M. (2000). The persistence of structural priming: Transient activation or implicit learning? *Journal of Experimental Psychology. General*, 129(2), 177–192. <https://doi.org/10.1037//0096-3445.129.2.177>
- Bomatter, P., Zhang, M., Karev, D., Madan, S., Tseng, C., & Kreiman, G. (2021). When Pigs Fly: Contextual Reasoning in Synthetic and Natural Scenes. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 255–264.
- Boston, M. F., Hale, J., Kliegl, R., Patil, U., & Vasishth, S. (2008). Parsing costs as predictors of reading difficulty: An evaluation using the Potsdam Sentence Corpus. *Journal of Eye Movement Research*, 2(1). <https://doi.org/10.16910/jemr.2.1.1>
- Bradlow, A. R., & Bent, T. (2008). *Perceptual Adaptation to Non-Native Speech*. 22.
- Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D., Wu, J., Winter, C., ... Amodei, D. (2020). Language Models are Few-Shot Learners. *Advances in Neural Information Processing Systems*, 33, 1877–1901.
- Bruner, J. S., & O'Dowd, D. (1958). A Note on the Informativeness of Parts of Words. *Language and Speech*, 1(2), 98–101. <https://doi.org/10.1177/002383095800100203>
- Castelhano, M. S., & Henderson, J. M. (2007). Initial scene representations facilitate eye movement guidance in visual search. *Journal of Experimental Psychology: Human Perception and Performance*, 33, 753–763. <https://doi.org/10.1037/0096-1523.33.4.753>
- Chambers, S. M. (1979). Letter and order information in lexical access. *Journal of Verbal Learning and Verbal Behavior*, 18(2), 225–241. [https://doi.org/10.1016/S0022-5371\(79\)90136-1](https://doi.org/10.1016/S0022-5371(79)90136-1)
- Christianson, K., Hollingworth, A., Halliwell, J. F., & Ferreira, F. (2001). Thematic Roles Assigned along the Garden Path Linger. *Cognitive Psychology*, 42(4), 368–407. <https://doi.org/10.1006/cogp.2001.0752>
- Chun, M. M., & Jiang, Y. (1998). Contextual Cueing: Implicit Learning and Memory of Visual Context Guides Spatial Attention. *Cognitive Psychology*, 36(1), 28–71. <https://doi.org/10.1006/cogp.1998.0681>
- Chun, M. M., & Jiang, Y. (1999). Top-Down Attentional Guidance Based on Implicit Learning of Visual Covariation. *Psychological Science*, 10(4), 360–365. <https://doi.org/10.1111/1467-9280.00168>
- Clarke, C. M., & Garrett, M. F. (2004). Rapid adaptation to foreign-accented English. *The Journal of the Acoustical Society of America*, 116(6), 3647–3658. <https://doi.org/10.1121/1.1815131>
- Demberg, V., & Keller, F. (2008). Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition*, 109(2), 193–210. <https://doi.org/10.1016/j.cognition.2008.07.008>
- Dufour, S., Mirault, J., & Grainger, J. (2022). Transposed-word effects in speeded grammatical decisions to sequences of spoken words. *Scientific Reports*, 12(1), 22035. <https://doi.org/10.1038/s41598-022-26584-2>
- Eckstein, M. P., Drescher, B. A., & Shimozaki, S. S. (2006). Attentional Cues in Real Scenes, Saccadic Targeting, and Bayesian Priors. *Psychological Science*, 17(11), 973–980. <https://doi.org/10.1111/j.1467-9280.2006.01815.x>

- Eckstein, M. P., Koehler, K., Welbourne, L. E., & Akbas, E. (2017). Humans, but Not Deep Neural Networks, Often Miss Giant Targets in Scenes. *Current Biology*, *27*(18), 2827–2832.e3. <https://doi.org/10.1016/j.cub.2017.07.068>
- Ehrlich, S. F., & Rayner, K. (1981). Contextual effects on word perception and eye movements during reading. *Journal of Verbal Learning and Verbal Behavior*, *20*(6), 641–655. [https://doi.org/10.1016/S0022-5371\(81\)90220-6](https://doi.org/10.1016/S0022-5371(81)90220-6)
- Eimas, P. D., & Corbit, J. D. (1973). Selective adaptation of linguistic feature detectors. *Cognitive Psychology*, *4*(1), 99–109. [https://doi.org/10.1016/0010-0285\(73\)90006-6](https://doi.org/10.1016/0010-0285(73)90006-6)
- Elman, J. L. (1990). Finding Structure in Time. *Cognitive Science*, *14*(2), 179–211. https://doi.org/10.1207/s15516709cog1402_1
- Falkenhainer, B., Forbus, K. D., & Gentner, D. (1989). The structure-mapping engine: Algorithm and examples. *Artificial Intelligence*, *41*(1), 1–63. [https://doi.org/10.1016/0004-3702\(89\)90077-5](https://doi.org/10.1016/0004-3702(89)90077-5)
- Federmeier, K. D., Wlotko, E. W., De Ochoa-Dewald, E., & Kutas, M. (2007). Multiple effects of sentential constraint on word processing. *Brain Research*, *1146*, 75–84. <https://doi.org/10.1016/j.brainres.2006.06.101>
- Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive Psychology*, *47*(2), 164–203. [https://doi.org/10.1016/S0010-0285\(03\)00005-7](https://doi.org/10.1016/S0010-0285(03)00005-7)
- Ferreira, F., Bailey, K. G., & Ferraro, V. (2002). Good-Enough Representations in Language Comprehension. *Current Directions in Psychological Science*, *11*(1), 11–15. <https://doi.org/10.1111/1467-8721.00158>
- Ferreira, F., & Henderson, J. M. (1991). Recovery from misanalyses of garden-path sentences. *Journal of Memory and Language*, *30*(6), 725–745. [https://doi.org/10.1016/0749-596X\(91\)90034-H](https://doi.org/10.1016/0749-596X(91)90034-H)
- Fine, A. B., & Jaeger, T. F. (2016). The role of verb repetition in cumulative structural priming in comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*, 1362–1376. <https://doi.org/10.1037/xlm0000236>
- Fine, A. B., Jaeger, T. F., Farmer, T. A., & Qian, T. (2013). Rapid Expectation Adaptation during Syntactic Comprehension. *PLOS ONE*, *8*(10), e77661. <https://doi.org/10.1371/journal.pone.0077661>
- Firestone, C., & Scholl, B. J. (2015). Cognition does not affect perception: Evaluating the evidence for “top-down” effects. *Behavioral and Brain Sciences*, *39*. <https://doi.org/10.1017/S0140525X15000965>
- Fodor, J. D. (2017). Ambiguity, parsing, and the evaluation measure. *Language Acquisition*, *24*(2), 85–99. <https://doi.org/10.1080/10489223.2016.1270948>
- Forster, K. (1998). The Pros and Cons of Masked Priming. *Journal of Psycholinguistic Research*, *27*, 203–233. <https://doi.org/10.1023/A:1023202116609>
- Forster, K. I., & Davis, C. (1984). Repetition priming and frequency attenuation in lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*(4), 680–698. <https://doi.org/10.1037/0278-7393.10.4.680>
- Forster, K. I., Davis, C., Schoknecht, C., & Carter, R. (1987). Masked priming with graphemically related forms: Repetition or partial activation? *The Quarterly Journal of Experimental Psychology Section A*, *39*(2), 211–251. <https://doi.org/10.1080/14640748708401785>
- Fortuin, E. (2014). Deconstructing a verbal illusion: The “No X is too Y to Z” construction and the rhetoric of negation. *Cognitive Linguistics*, *25*(2), 249–292. <https://doi.org/10.1515/cog-2014-0014>
- Frank, S. L., Trompenaars, T., & Vasishth, S. (2016). Cross-Linguistic Differences in Processing Double-Embedded Relative Clauses: Working-Memory Constraints or Language Statistics? *Cognitive Science*, *40*(3), 554–578. <https://doi.org/10.1111/cogs.12247>
- Fraundorf, S. H., & Jaeger, T. F. (2016). Readers generalize adaptation to newly-encountered dialectal structures to other unfamiliar structures. *Journal of Memory and Language*, *91*, 28–58. <https://doi.org/10.1016/j.jml.2016.05.006>

- Frazier, L. (1979). On comprehending sentences: Syntactic parsing strategies (doctoral dissertation). *Ph. D. Thesis, UMass at Amherst*.
- Ganis, G., & Kutas, M. (2003). An electrophysiological study of scene effects on object identification. *Cognitive Brain Research, 16*(2), 123–144. [https://doi.org/10.1016/S0926-6410\(02\)00244-6](https://doi.org/10.1016/S0926-6410(02)00244-6)
- Garnsey, S. M., Pearlmutter, N. J., Myers, E., & Lotocky, M. A. (1997). The Contributions of Verb Bias and Plausibility to the Comprehension of Temporarily Ambiguous Sentences. *Journal of Memory and Language, 37*(1), 58–93. <https://doi.org/10.1006/jmla.1997.2512>
- Gibson, E., Bergen, L., & Piantadosi, S. T. (2013). Rational integration of noisy evidence and prior semantic expectations in sentence interpretation. *Proceedings of the National Academy of Sciences, 110*(20), 8051–8056. <https://doi.org/10.1073/pnas.1216438110>
- 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. <https://doi.org/10.1177/0956797617690277>
- Gibson, E., & Thomas, J. (1999). Memory Limitations and Structural Forgetting: The Perception of Complex Ungrammatical Sentences as Grammatical. *Language and Cognitive Processes, 14*(3), 225–248. <https://doi.org/10.1080/016909699386293>
- Hale, J. (2001). A probabilistic earley parser as a psycholinguistic model. *Second Meeting of the North American Chapter of the Association for Computational Linguistics on Language Technologies 2001 - NAACL '01*, 1–8. <https://doi.org/10.3115/1073336.1073357>
- Hamilton, W. L., Leskovec, J., & Jurafsky, D. (2016). *Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change* (arXiv:1605.09096). arXiv. <https://doi.org/10.48550/arXiv.1605.09096>
- Hersche, M., Zeqiri, M., Benini, L., Sebastian, A., & Rahimi, A. (2023). A neuro-vector-symbolic architecture for solving Raven's progressive matrices. *Nature Machine Intelligence, 1*–13. <https://doi.org/10.1038/s42256-023-00630-8>
- Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation, 9*(8), 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- Hofstadter, D. R., & Mitchell, M. (1994). The Copycat project: A model of mental fluidity and analogy-making. In *Analogical connections* (pp. 31–112). Ablex Publishing.
- Hu, S., Ma, Y., Liu, X., Wei, Y., & Bai, S. (2021). Stratified Rule-Aware Network for Abstract Visual Reasoning. *Proceedings of the AAAI Conference on Artificial Intelligence, 35*(2), 1567–1574. <https://doi.org/10.1609/aaai.v35i2.16248>
- Huang, K.-J., & Staub, A. (2022). The transposed-word effect does not require parallel word processing: Failure to notice transpositions with serial presentation of words. *Psychonomic Bulletin & Review, 29*(1), 1–13. <https://doi.org/10.3758/s13423-022-02150-9>
- Huang, K.-J., & Staub, A. (2021a). Using eye tracking to investigate failure to notice word transpositions in reading. *Cognition, 216*, 104846. <https://doi.org/10.1016/j.cognition.2021.104846>
- Huang, K.-J., & Staub, A. (2021b). Why do readers fail to notice word transpositions, omissions, and repetitions? A review of recent evidence and theory. *Language and Linguistics Compass, 15*(7), e12434. <https://doi.org/10.1111/lnc3.12434>
- Inhoff, A. W. (1984). Two stages of word processing during eye fixations in the reading of prose. *Journal of Verbal Learning and Verbal Behavior, 23*(5), 612–624. [https://doi.org/10.1016/S0022-5371\(84\)90382-7](https://doi.org/10.1016/S0022-5371(84)90382-7)
- Isik, L., Koldewyn, K., Beeler, D., & Kanwisher, N. (2017). Perceiving social interactions in the posterior superior temporal sulcus. *Proceedings of the National Academy of Sciences, 114*(43), E9145–E9152. <https://doi.org/10.1073/pnas.1714471114>

- Isik, L., Mynick, A., Pantazis, D., & Kanwisher, N. (2019). The speed of human social interaction perception. *bioRxiv*, 579375. <https://doi.org/10.1101/579375>
- Jaeger, T. F., Zachary, B., & Wednesday, B. (2018). *Strong evidence for expectation adaptation during language understanding, not a replication failure. A reply to Harrington Stack, James, and Watson (2018)*.
- Jurafsky, D. (1996). A Probabilistic Model of Lexical and Syntactic Access and Disambiguation. *Cognitive Science*, 20, 137–194.
- Kaschak, M., & Glenberg, A. (2004). This Construction Needs Learned. *Journal of Experimental Psychology. General*, 133, 450–467. <https://doi.org/10.1037/0096-3445.133.3.450>
- Kell, A. J. E., Yamins, D. L. K., Shook, E. N., Norman-Haignere, S. V., & McDermott, J. H. (2018). A Task-Optimized Neural Network Replicates Human Auditory Behavior, Predicts Brain Responses, and Reveals a Cortical Processing Hierarchy. *Neuron*, 98(3), 630–644.e16. <https://doi.org/10.1016/j.neuron.2018.03.044>
- Kernighan, M. D., Church, K. W., & Gale, W. A. (1990). A Spelling Correction Program Based on a Noisy Channel Model. *COLING 1990 Volume 2: Papers Presented to the 13th International Conference on Computational Linguistics*.
- Kim, J., Linsley*, D., Thakkar, K., & Serre, T. (2020, March). Disentangling neural mechanisms for perceptual grouping. *International Conference on Learning Representations*.
- Kleinschmidt, D. (2020). *What constrains distributional learning in adults?* PsyArXiv. <https://doi.org/10.31234/osf.io/6yhbe>
- Kleinschmidt, D. F., & Jaeger, T. F. (2015). Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review*, 122(2), 148–203. <https://doi.org/10.1037/a0038695>
- Kreiman, G., & Serre, T. (2020). Beyond the feedforward sweep: Feedback computations in the visual cortex. *Annals of the New York Academy of Sciences*, 1464(1), 222–241. <https://doi.org/10.1111/nyas.14320>
- Kutas, M., & Federmeier, K. D. (2011). Thirty Years and Counting: Finding Meaning in the N400 Component of the Event-Related Brain Potential (ERP). *Annual Review of Psychology*, 62(1), 621–647. <https://doi.org/10.1146/annurev.psych.093008.131123>
- Kutas, M., & Hillyard, S. A. (1980). Reading Senseless Sentences: Brain Potentials Reflect Semantic Incongruity. *Science*, 207(4427), 203–205. <https://doi.org/10.1126/science.7350657>
- Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, 307(5947), 161–163. <https://doi.org/10.1038/307161a0>
- Leopold, D. A., O’Toole, A. J., Vetter, T., & Blanz, V. (2001). Prototype-referenced shape encoding revealed by high-level aftereffects. *Nature Neuroscience*, 4(1), 89–94. <https://doi.org/10.1038/82947>
- Levin, B. (1993). *English Verb Classes and Alternations: A Preliminary Investigation*. University of Chicago Press.
- Levy, R. (2008a). Expectation-based syntactic comprehension. *Cognition*, 106(3), 1126–1177. <https://doi.org/10.1016/j.cognition.2007.05.006>
- Levy, R. (2011). Integrating surprisal and uncertain-input models in online sentence comprehension: Formal techniques and empirical results. *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, 1055–1065.
- Levy, R. (2008b). A noisy-channel model of rational human sentence comprehension under uncertain input. *Proceedings of the Conference on Empirical Methods in Natural Language Processing - EMNLP ’08*, 234. <https://doi.org/10.3115/1613715.1613749>
- 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(50), 21086–21090. <https://doi.org/10.1073/pnas.0907664106>

- Linsley, D., Shiebler, D., Eberhardt, S., & Serre, T. (2019). *Learning what and where to attend* (arXiv:1805.08819). arXiv.
- Liu, Z., Li, Y., Cutter, M. G., Paterson, K. B., & Wang, J. (2022). A transposed-word effect across space and time: Evidence from Chinese. *Cognition*, *218*, 104922. <https://doi.org/10.1016/j.cognition.2021.104922>
- Liu, Z., Li, Y., Paterson, K. B., & Wang, J. (2020). A transposed-word effect in Chinese reading. *Attention, Perception, & Psychophysics*, *82*(8), 3788–3794. <https://doi.org/10.3758/s13414-020-02114-y>
- Mahowald, K., Ivanova, A. A., Blank, I. A., Kanwisher, N., Tenenbaum, J. B., & Fedorenko, E. (2023). *Dissociating language and thought in large language models: A cognitive perspective* (arXiv:2301.06627). arXiv.
- Mays, E., Damerau, F. J., & Mercer, R. L. (1991). Context based spelling correction. *Information Processing & Management*, *27*(5), 517–522. [https://doi.org/10.1016/0306-4573\(91\)90066-U](https://doi.org/10.1016/0306-4573(91)90066-U)
- Mirault, J., Snell, J., & Grainger, J. (2018). You That Read Wrong Again! A Transposed-Word Effect in Grammaticality Judgments. *Psychological Science*, *29*(12), 1922–1929. <https://doi.org/10.1177/0956797618806296>
- Mirault, J., Vandendaele, A., Pegado, F., & Grainger, J. (2022). Transposed-word effects when reading serially. *PLOS ONE*, *17*(11), e0277116. <https://doi.org/10.1371/journal.pone.0277116>
- Mollica, F., Siegelman, M., Diachek, E., Piantadosi, S. T., Mineroff, Z., Futrell, R., Kean, H., Qian, P., & Fedorenko, E. (2020). Composition is the Core Driver of the Language-selective Network. *Neurobiology of Language*, *1*(1), 104–134. https://doi.org/10.1162/nol_a_00005
- Newell, A., Shaw, J. C., & Simon, H. A. (1959). Report on a general problem solving program. *IFIP Congress*, *256*, 64.
- Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology*, *47*(2), 204–238. [https://doi.org/10.1016/S0010-0285\(03\)00006-9](https://doi.org/10.1016/S0010-0285(03)00006-9)
- Oliva, A., & Torralba, A. (2007). The role of context in object recognition. *Trends in Cognitive Sciences*, *11*(12), 520–527. <https://doi.org/10.1016/j.tics.2007.09.009>
- Paape, D., Vasishth, S., & von der Malsburg, T. (2020). Quadruplex Negatio Invertit? The On-Line Processing of Depth Charge Sentences. *Journal of Semantics*, *37*(4), 509–555. <https://doi.org/10.1093/jos/ffaa009>
- Perea, M., & Lupker, S. J. (2003). Does jugde activate COURT? Transposed-letter similarity effects in masked associative priming. *Memory & Cognition*, *31*(6), 829–841. <https://doi.org/10.3758/BF03196438>
- Piazza, E. A., Theunissen, F. E., Wessel, D., & Whitney, D. (2018). Rapid Adaptation to the Timbre of Natural Sounds. *Scientific Reports*, *8*(1), 13826. <https://doi.org/10.1038/s41598-018-32018-9>
- Pickering, M. J., & Branigan, H. P. (1998). The representation of verbs: Evidence from syntactic priming in language production. *Journal of Memory and Language*, *39*, 633–651. <https://doi.org/10.1006/jmla.1998.2592>
- Poppels, T., & Levy, R. P. (2016). Structure-sensitive Noise Inference: Comprehenders Expect Exchange Errors. *Proceedings of the 38th Annual Conference of the Cognitive Science Society*, 378–383.
- Pramod, R., Cohen, M. A., Tenenbaum, J. B., & Kanwisher, N. (2022). Invariant representation of physical stability in the human brain. *eLife*, *11*, e71736. <https://doi.org/10.7554/eLife.71736>
- Raviv, J. (1967). Decision making in Markov chains applied to the problem of pattern recognition. *IEEE Transactions on Information Theory*, *13*(4), 536–551. <https://doi.org/10.1109/TIT.1967.1054060>
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*, 372–422. <https://doi.org/10.1037/0033-2909.124.3.372>
- Rayner, K., Slattery, T. J., Drieghe, D., & Liversedge, S. P. (2011). Eye movements and word skipping during reading: Effects of word length and predictability. *Journal of Experimental Psychology. Human Perception and Performance*, *37*(2), 514–528. <https://doi.org/10.1037/a0020990>

- Rayner, K., & Well, A. D. (1996). Effects of contextual constraint on eye movements in reading: A further examination. *Psychonomic Bulletin & Review*, 3(4), 504–509. <https://doi.org/10.3758/BF03214555>
- Rhodes, G., & Jeffery, L. (2006). Adaptive norm-based coding of facial identity. *Vision Research*, 46(18), 2977–2987. <https://doi.org/10.1016/j.visres.2006.03.002>
- Ryskin, R., Futrell, R., Kiran, S., & Gibson, E. (2018). Comprehenders model the nature of noise in the environment. *Cognition*, 181, 141–150. <https://doi.org/10.1016/j.cognition.2018.08.018>
- Samuel, A. G. (1986). Red herring detectors and speech perception: In defense of selective adaptation. *Cognitive Psychology*, 18(4), 452–499. [https://doi.org/10.1016/0010-0285\(86\)90007-1](https://doi.org/10.1016/0010-0285(86)90007-1)
- Schrimpf, M., Blank, I. A., Tuckute, G., Kauf, C., Hosseini, E. A., Kanwisher, N., Tenenbaum, J. B., & Fedorenko, E. (2021). The neural architecture of language: Integrative modeling converges on predictive processing. *Proceedings of the National Academy of Sciences*, 118(45), e2105646118. <https://doi.org/10.1073/pnas.2105646118>
- Schustack, M. W., Ehrlich, S. F., & Rayner, K. (1987). Local and global sources of contextual facilitation in reading. *Journal of Memory and Language*, 26(3), 322–340. [https://doi.org/10.1016/0749-596X\(87\)90117-3](https://doi.org/10.1016/0749-596X(87)90117-3)
- Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal*, 27(3), 379–423. <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>
- Sidaras, S. K., Alexander, J. E. D., & Nygaard, L. C. (2009). Perceptual learning of systematic variation in Spanish-accented speech. *The Journal of the Acoustical Society of America*, 125(5), 3306–3316. <https://doi.org/10.1121/1.3101452>
- Smith, N. J., & Levy, R. (2008). Optimal Processing Times in Reading: A Formal Model and Empirical Investigation. *Proceedings of the Annual Meeting of the Cognitive Science Society*.
- Snell, J., & Grainger, J. (2019). Word position coding in reading is noisy. *Psychonomic Bulletin & Review*, 26(2), 609–615. <https://doi.org/10.3758/s13423-019-01574-0>
- Snell, J., & Grainger, J. (2017). The sentence superiority effect revisited. *Cognition*, 168, 217–221. <https://doi.org/10.1016/j.cognition.2017.07.003>
- Staub, A. (2015). The Effect of Lexical Predictability on Eye Movements in Reading: Critical Review and Theoretical Interpretation. *Language and Linguistics Compass*, 9(8), 311–327. <https://doi.org/10.1111/lnc3.12151>
- Staub, A., & Clifton, C. (2006). Syntactic prediction in language comprehension: Evidence from either...or. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32(2), 425–436. <https://doi.org/10.1037/0278-7393.32.2.425>
- Staub, A., Dodge, S., & Cohen, A. L. (2019). Failure to detect function word repetitions and omissions in reading: Are eye movements to blame? *Psychonomic Bulletin & Review*, 26(1), 340–346. <https://doi.org/10.3758/s13423-018-1492-z>
- Stevenson, R. (1964). *Mary Poppins*. Buena Vista Distribution Company.
- Tabor, W., Galantucci, B., & Richardson, D. (2004). Effects of merely local syntactic coherence on sentence processing. *Journal of Memory and Language*, 50(4), 355–370. <https://doi.org/10.1016/j.jml.2004.01.001>
- Tang, H., Schrimpf, M., Lotter, W., Moerman, C., Paredes, A., Ortega Caro, J., Hardesty, W., Cox, D., & Kreiman, G. (2018). Recurrent computations for visual pattern completion. *Proceedings of the National Academy of Sciences*, 115(35), 8835–8840. <https://doi.org/10.1073/pnas.1719397115>
- Torralba, A., Oliva, A., Castelano, M. S., & Henderson, J. M. (2006). Contextual guidance of eye movements and attention in real-world scenes: The role of global features in object search. *Psychological Review*, 113(4), 766–786. <https://doi.org/10.1037/0033-295X.113.4.766>

- Tzeng, C. Y., Alexander, J. E. D., Sidaras, S. K., & Nygaard, L. C. (2016). The role of training structure in perceptual learning of accented speech. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(11), 1793–1805. <https://doi.org/10.1037/xhp0000260>
- Vasishth, S., Suckow, K., Lewis, R. L., & Kern, S. (2010). Short-term forgetting in sentence comprehension: Crosslinguistic evidence from verb-final structures. *Language and Cognitive Processes*, *25*(4), 533–567. <https://doi.org/10.1080/01690960903310587>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). *Attention Is All You Need* (arXiv:1706.03762). arXiv. <https://doi.org/10.48550/arXiv.1706.03762>
- Villalobos, K., Štih, V., Ahmadinejad, A., Sundaram, S., Dozier, J., Francl, A., Azevedo, F., Sasaki, T., & Boix, X. (2021). Do Neural Networks for Segmentation Understand Insiderness? *Neural Computation*, *33*(9), 2511–2549. https://doi.org/10.1162/neco_a_01413
- Vitu, F. (1991). The influence of parafoveal preprocessing and linguistic context on the optimal landing position effect. *Perception & Psychophysics*, *50*(1), 58–75. <https://doi.org/10.3758/BF03212205>
- Võ, M. L.-H. (2021). The meaning and structure of scenes. *Vision Research*, *181*, 10–20. <https://doi.org/10.1016/j.visres.2020.11.003>
- Võ, M. L.-H., & Wolfe, J. M. (2013). Differential Electrophysiological Signatures of Semantic and Syntactic Scene Processing. *Psychological Science*, *24*(9), 1816–1823. <https://doi.org/10.1177/0956797613476955>
- Vroomen, J., van Linden, S., de Gelder, B., & Bertelson, P. (2007). Visual recalibration and selective adaptation in auditory–visual speech perception: Contrasting build-up courses. *Neuropsychologia*, *45*(3), 572–577. <https://doi.org/10.1016/j.neuropsychologia.2006.01.031>
- Warner, J., & Glass, A. L. (1987). Context and distance-to-disambiguation effects in ambiguity resolution: Evidence from grammaticality judgments of garden path sentences. *Journal of Memory and Language*, *26*(6), 714–738. [https://doi.org/10.1016/0749-596X\(87\)90111-2](https://doi.org/10.1016/0749-596X(87)90111-2)
- Wason, P. C., & Reich, S. S. (1979). A Verbal Illusion. *Quarterly Journal of Experimental Psychology*, *31*(4), 591–597. <https://doi.org/10.1080/14640747908400750>
- Wen, Y., Mirault, J., & Grainger, J. (2021). The transposed-word effect revisited: The role of syntax in word position coding. *Language, Cognition and Neuroscience*, *36*(5), 668–673. <https://doi.org/10.1080/23273798.2021.1880608>
- Wilson, M. P., & Garnsey, S. M. (2009). Making simple sentences hard: Verb bias effects in simple direct object sentences. *Journal of Memory and Language*, *60*(3), 368–392. <https://doi.org/10.1016/j.jml.2008.09.005>
- Winn, M. B. (2016). Rapid Release From Listening Effort Resulting From Semantic Context, and Effects of Spectral Degradation and Cochlear Implants. *Trends in Hearing*, *20*, 2331216516669723. <https://doi.org/10.1177/2331216516669723>
- Winn, M. B., & Teece, K. H. (2021). Listening Effort Is Not the Same as Speech Intelligibility Score. *Trends in Hearing*, *25*, 23312165211027688. <https://doi.org/10.1177/23312165211027688>
- Winn, M. B., & Teece, K. H. (2022). Effortful Listening Despite Correct Responses: The Cost of Mental Repair in Sentence Recognition by Listeners With Cochlear Implants. *Journal of Speech, Language, and Hearing Research*, *65*(10), 3966–3980. https://doi.org/10.1044/2022_JSLHR-21-00631
- Wu, Y., Dong, H., Grosse, R., & Ba, J. (2020). *The Scattering Compositional Learner: Discovering Objects, Attributes, Relationships in Analogical Reasoning* (arXiv:2007.04212). arXiv. <https://doi.org/10.48550/arXiv.2007.04212>
- Xie, X., Kurumada, C., & Jaeger, T. F. (2023). *What we do (not) know about the mechanisms underlying adaptive speech perception: A computational framework and review*. PsyArXiv. <https://doi.org/10.31234/osf.io/hgdn4>

- Xie, X., Liu, L., & Jaeger, T. F. (2021). Cross-talker generalization in the perception of nonnative speech: A large-scale replication. *Journal of Experimental Psychology: General*, No Pagination Specified–No Pagination Specified. <https://doi.org/10.1037/xge0001039>
- Xie, X., Weatherholtz, K., Bainton, L., Rowe, E., Burchill, Z., Liu, L., & Jaeger, T. F. (2018). Rapid adaptation to foreign-accented speech and its transfer to an unfamiliar talker. *The Journal of the Acoustical Society of America*, *143*(4), 2013–2031. <https://doi.org/10.1121/1.5027410>
- Yamins, D. L. K., Hong, H., Cadieu, C. F., Solomon, E. A., Seibert, D., & DiCarlo, J. J. (2014). Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proceedings of the National Academy of Sciences*, *111*(23), 8619–8624. <https://doi.org/10.1073/pnas.1403112111>
- Yan, S., & Jaeger, T. F. (2020). Expectation adaptation during natural reading. *Language, Cognition and Neuroscience*, *35*(10), 1394–1422. <https://doi.org/10.1080/23273798.2020.1784447>
- Zäske, R., Schweinberger, S. R., & Kawahara, H. (2010). Voice aftereffects of adaptation to speaker identity. *Hearing Research*, *268*(1), 38–45. <https://doi.org/10.1016/j.heares.2010.04.011>
- Zhang, C., Gao, F., Jia, B., Zhu, Y., & Zhu, S.-C. (2019). RAVEN: A Dataset for Relational and Analogical Visual REasoning. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 5317–5327.
- Zhang, C., Jia, B., Zhu, S.-C., & Zhu, Y. (2021). Abstract Spatial-Temporal Reasoning via Probabilistic Abduction and Execution. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 9736–9746.
- Zhang, M., Tseng, C., & Kreiman, G. (2020). Putting Visual Object Recognition in Context. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 12985–12994.
- Zhang, Y., Ryskin, R., & Gibson, E. (2023). A noisy-channel approach to depth-charge illusions. *Cognition*, *232*, 105346. <https://doi.org/10.1016/j.cognition.2022.105346>
- Zola, D. (1984). Redundancy and word perception during reading. *Perception & Psychophysics*, *36*(3), 277–284. <https://doi.org/10.3758/BF03206369>