COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 50 (2026) e70217

ISSN: 1551-6709 online
DOI: 10.1111/cogs.70217

"Just-Right" Is (Not) Always Right: Joint Effects of Typicality and Informativity During Real-Time Sentence Processing

Muxuan He, Elsi Kaiser

Department of Linguistics, University of Southern California

Received 17 March 2025; received in revised form 25 March 2026; accepted 20 April 2026

Abstract

A growing body of empirical research shows that, for both affirmative and negative sentences, language processing is not only shaped by the ease of integrating linguistic input into world knowledge, but also by expectations of informative communication. Yet, it remains unclear how comprehenders coordinate pressures from both aspects during real-time sentence processing. We propose that language comprehension reflects a dynamic balance between world-knowledge typicality and communication informativeness, which is subject to contextual modulation. Two self-paced reading studies test this proposal by examining affirmative and negative sentences describing part−whole relations that vary in typicality and informativity in unmarked and unexpectedness-signaling contexts. Our results provide evidence for joint effects of typicality and informativity, with distinct patterns across sentence polarities and contexts.

Keywords: Sentence processing; World knowledge; Informativity; Typicality; Negation; Trade-off; Pragmatics; Self-paced reading

1. Introduction

Our **real-world knowledge**, given some situations/events are more likely to occur than the others, affects language processing: comprehenders find sentences describing more typical situations/events easier to process than those describing atypical situations/events (e.g., Hagoort, Hald, Bastiaansen, & Petersson, 2004; Kutas & Hillyard, 1980; Matsuki et al., 2011; Rayner, Carlson, & Frazier, 1983). In addition, our **linguistic experience**, given that some things are more likely to be communicated than the others due to general communicative

Correspondence should be sent to Muxuan He, Department of Linguistics, University of Southern California, 3601 Watt Way, Los Angeles, CA 90089-1693, USA. E-mail: muxuanhe@usc.edu

norms (e.g., Grice, 1975), also guides language processing: comprehenders assume that more informative expressions are more likely to be communicated (Rohde & Rubio-Fernandez, 2022; Sedivy, 2003; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). Importantly, real-world knowledge and linguistic experience have an inherent negative correlation. On the one hand, people rarely mention highly typical situations/events (e.g., Brown & Dell, 1987; Bergey et al., 2020; Kravtchenko & Demberg, 2022), perhaps due to their low informativity. On the other hand, people find highly atypical situations/events informative and worth noting. This raises the question of how comprehenders coordinate real-world knowledge and linguistic experience in sentence processing, especially when they misalign.

Let us imagine the following statements about a house in the U.S. context:

(1) a. The house has a *bathroom*. (high typicality, low informativity)
b. The house has a *garage*. (medium typicality, medium informativity)
c. The house has a *ballroom*. (low typicality, high informativity)

Sentence (1a) describes a highly typical situation for a house, but at the same time, this situation seems too common to be informative or worth mentioning. In other words, it has low communicative informativeness. The situation described in (1c) is informative and worth noting, but highly atypical. It has high communicative informativeness. In a sentence like (1b), world knowledge typicality and communicative informativeness are both "middling": the sentence describes a moderately typical situation, which is moderately informative and somewhat worth mentioning.

If comprehenders prioritize either world knowledge typicality or communicative informativeness, sentences like (1a) or sentences like (1c) should presumably be easiest to process. But what if comprehenders incorporate both factors? We would expect that descriptions of highly typical situations (which violate expectations for informative communication) or highly atypical situations (which violate world knowledge expectations) are both hard to process, which suggests that sentences like (1b) would be the easiest to process among (1a−c). We refer to this later case as the *Just Right Hypothesis*: sentences should be easiest to process when they balance ease of integration into our world knowledge alongside expectations of informative communication.

The linguistic phenomenon of negation offers a unique tool for probing how comprehenders coordinate real-world knowledge and linguistic experience. Negation is usually used to deny expectations (e.g., Horn, 1989; Wason, 1965). For example, it is natural to say "The elevator is not working" when you see someone heading for the elevator doors, because that person is expecting it to work as usual. The stronger the expectation, the easier negative sentences are to process (e.g., Fischler et al., 1983; Lüdtke & Kaup, 2006; Nordmeyer & Frank, 2014; Tian, Breheny, & Ferguson, 2010).

Let us now consider the negative-polarity counterparts of (1):

(2) a. The house doesn't have a bathroom. (low typicality, high informativity)
b. The house doesn't have a garage. (medium typicality, medium informativity)
c. The house doesn't have a ballroom. (highly typicality; low informativity)

If we are considering only linguistic experience in which negation is often used to communicate expectation violation, we might expect negative sentences like (2c) to be more difficult to process than (2b) and (2a), since we do not normally expect a house to have a ballroom. However, again, what if comprehenders consider both linguistic experience and real-world knowledge, even in negative sentences? The situation described in a negative sentence like (2a) is informative and worth noting (a house *not* having a bathroom), but it is highly atypical. The negative sentence in (2c) describes a highly typical situation for a house (a house *not* having a ballroom), but at the same time, this situation seems too common to be informative or worth mentioning. Compared to negative sentences like (2a) and (2c), negative sentences like (2b) describe a moderately typical situation, which is moderately informative and somewhat worth mentioning. Thus, these patterns in negative sentences (2a−c) turn out to be quite similar to those in the affirmative sentences (1a−c). However, the question of how linguistic experience (i.e., the fact that some things are more likely to be communicated than the others due to general communicative norms) and world knowledge interact in the processing of negative sentences is not yet well-understood. Indeed, although sentence polarities are famously asymmetrical, for example, negation presupposes its affirmative counterpart and negation takes longer to process, it is not yet clear to what extent the processing of affirmative and negative sentences is shaped by similar pressures. In the present study, we propose and test the *Just Right Hypothesis* for both affirmative and negative sentences in contexts like (1) and (2).

As shown in (1) and (2), the kind of real-world knowledge we test in this paper is "situation typicality," that is, based on our prior knowledge of the world, how probable is a particular situation? We focus on part−whole relations (ex. 1), and the negation of part−whole relations (ex. 2). We refer to linguistic experience (defined as experience about some things being more likely to be communicated than the others) as communication likelihood and operationalize it via informativity: what is more informative is more likely to be communicated. Ex. (1) and (2) illustrate how we define the informativity of a sentence from an information-theoretic perspective (e.g., Shannon, 1948), in terms of how much it updates our prior world knowledge. This follows prior work (e.g., Frank & Goodman, 2012; Xiang et al., 2020).

It is worth noting that language also provides us with lexical devices that allow a speaker to signal that they are saying something not normally communicated. Prior work shows that comprehenders can use these kinds of lexical cues to anticipate events that are uncommon or very common based on real-world knowledge (e.g., see work using "even so" by Xiang & Kuperberg, 2015; "My cousin is a surprising person" vs. "My cousin is a boring person" by Rohde, Futrell, & Lucas, 2021; the adverb "obviously" by Albu et al., 2024). These kinds of expressions create a context where the comprehender knows to expect something typical or atypical.

To test processing in the presence of such expectations, we explore how lexical cues—specifically those signaling atypical situations—impact the processing of affirmative and negative sentences. Specifically, we use sentential adverbials like "surprisingly" in front of affirmative and negative sentences like (1) and (2). For example, "the house has a ballroom" and "the house *doesn't* have a bathroom" are highly atypical situations, but the addition of "surprisingly," as in (3) and (4), provides a signal to expect the unexpected.

(3) Surprisingly, the house has a ballroom.
(4) Surprisingly, the house doesn't have a bathroom.

This allows us to test whether, in the presence of adverbs like "surprisingly," sentences describing such atypical situations might become the easiest to process, despite (or specifically because) they describe atypical situations. This possibility gives rise to what we call the *Surprise Me Hypothesis*: when lexical cues signaling the unexpected are present, comprehenders should expect atypical situations.

In the present paper, we report two experiments that test how typicality and informativity affect the processing of affirmative and negative sentences, with and without lexical cues (e.g., "surprisingly") pointing to atypical situations. For ease of discussion, we refer to contexts without such lexical cues as unmarked (ex. 1 and 2) and contexts with such lexical cues as "contexts of unexpectedness" (ex. 3 and 4). We test the *Just Right Hypothesis* for affirmative and negative sentences in unmarked contexts in Experiment 1, and we test the *Surprise Me Hypothesis* for affirmative and negative sentences using contexts of unexpectedness in Experiment 2.

Before outlining the more specific aims and hypotheses of our study, we review relevant work on affirmative and negative sentences in the following sections.

1.1. Affirmative sentences: Typicality effects and context-dependent informativity effects

We use the term "typicality effects" in reference to the influence of real-world knowledge in sentence processing, linking higher probabilities in prior knowledge to processing ease. In affirmative sentences, typicality facilitates processing and atypicality increases processing cost (e.g., Ferretti, Kutas, & McRae, 2007; Kutas & Hillyard, 1980; Matsuki et al., 2011; McRae & Matsuki, 2009; Morris, 1994; Walker, 1975; see also Lee & Kaiser, 2026 for related work). For example, in self-paced reading and eye-tracking, comprehenders read the noun "hair" faster after a typical instrument−action pair than after an atypical pair (Matsuki et al., 2011: "use the shampoo to wash hair" vs. "use the hose to wash hair"). Similarly, Rayner, Warren, Juhasz, and Liversedge (2004) found that the more atypical an instrument−action pair was, the earlier it disrupted eye movements. Hagoort et al. (2004) tested Dutch speakers using sentences about culturally specific situations (e.g., "The Dutch trains are yellow/white/sour") and found N400 effects for atypical properties ("white" and "sour"), but not for a typical property ("yellow"). The strength of typicality effects even extends beyond the "real world": comprehenders can adapt to new worlds, for example, the magical world of Harry Potter (Troyer & Kutas, 2018) and the fictional world of a peanut protagonist (Nieuwland & Van Berkum, 2006), such that they process typical situations in these new worlds more easily than atypical situations.

Nevertheless, real-world knowledge expectations can be altered by contextual cues. Xiang and Kuperberg (2015) tested whether a concessive connective, *even so*, elicited expectations contradictory to one's world knowledge. They found that while comprehenders experience difficulty processing what is unexpected given their world knowledge, this effect was reversed when a concessive connective was present (e.g., "Elizabeth took the test and aced/failed it. (Even so,) she went home and celebrated wildly."). This line of research (e.g., Heller, Arnold,

Klein, & Tanenhaus, 2015; Xiang & Kuperberg, 2015; Köhne-Fuetterer et al., 2020) suggests that contextual cues incrementally integrate with the flexible use of world knowledge during processing. More recent studies have also found that typicality effects are reversed when the context shows a clear expectation for newsworthy information (e.g., Reksnes, Rees, Cummins, & Rohde, 2024; Rohde & Rubio-Fernandez, 2022; Rohde et al., 2021), which Rohde et al. (2021) refer to as "informativity effects." Rohde et al. (2021) showed that, when the context introduces a surprising protagonist, sentences like (5b) yielded shorter reading times (RTs) at the critical noun *shovel* than sentences like (5a), although the former describes situations that are atypical based on world knowledge.

(5) a. In order to dig some holes, s/he was using a shovel yesterday in the afternoon. [surprising persona] [typical instrument]
b. In order to chop some carrots, s/he was using a shovel yesterday in the afternoon. [surprising persona] [atypical instrument]

In addition to processing-based effects, affirmative sentences describing trivial, expected events (e.g., "John paid the cashier when he went grocery shopping tonight") have been shown to elicit pragmatic inferences. For example, when comprehenders are presented with sentences describing highly expected events and asked to rate the typicality of these events in an inference elicitation task (e.g., Rate how often you think it is that John pays when he gets groceries), they draw the so-called "atypicality inferences," and infer that the event is actually *atypical* (e.g., that John doesn't typically pay when he gets groceries, see, e.g., Kravtchenko & Demberg, 2015, 2022; Ryzhova, Mayn, & Demberg, 2023). While Ryzhova and Demberg (2023) found that generating such atypicality inferences in affirmative sentences is not cognitively demanding per se, evidence on the real-time processing of low-informativity sentences without an inference elicitation task is limited. Our study provides new evidence relevant to this question.

1.2. Negative sentences: Effects of negation are modulated by context

Turning to negative sentences, the first point to note is that prior work is inconclusive regarding whether people can access and use real-world knowledge immediately and successfully when processing negative sentences. Studies have shown that, for stand-alone negative sentences, negation does not seem to be integrated rapidly enough to influence incremental sentence processing (e.g., Fischler et al., 1983; Kounios & Holcomb, 1992; Nieuwland & Kuperberg, 2008). While some studies reported greater N400 responses for true statements compared to false statements (e.g., "The robin is not a tree" > "The robin is not a bird" in Fischler et al., 1983), others found no difference between true and false statements when they are stand-alone negative sentences (e.g., "A baby bunny's fur isn't very hard/soft" in Nieuwland & Kuperberg, 2008). However, when testing negative sentences in contexts that establish expectations for what might be negated, Nieuwland and Kuperberg (2008) found that false negative statements yielded greater N400 responses than true negative statements (e.g., "In moderation, drinking red wine isn't good" > "In moderation, drinking red wine isn't bad"), similar to the pattern observed in affirmative statements. This suggests that when

negation is informative (i.e., when it has a clear expectation to negate), people can access and use real-world knowledge immediately and successfully in evaluating negative sentences.

Focusing on the role of informativity, Nordmeyer and Frank (2014) found that a visual context providing probabilistic expectations (e.g., a picture of multiple boys having apples) facilitates truth-value judgments of negative sentences (e.g., "The boy has no apples"). Nordmeyer and Frank also manipulated the strength of these expectations by varying the number of people possessing a certain property (e.g., how many boys have apples) in the visual context. Intuitively, as the probability of a boy having apples increases, the informativity of a situation where a boy does not have apples increases as well. Results from Nordmeyer and Frank (2014; see also Xiang, Kramer, & Nordmeyer, 2020) revealed graded effects of informativity: negative sentences that are more informative tended to elicit faster verification times. More broadly, however, in most prior work on informativity and negative sentences, the processing of negation has been bundled with truth-value judgments. This leaves an open question: do the observed effects of expectations reflect mechanisms involved in verifying a sentence, or the processing of negation per se? Our study sheds light on this question.

Like affirmative sentences, negative sentences are sensitive to contextual cues that modulate expectations for what kinds of situations or events will be communicated. Albu et al. (2024) provided evidence for such contextual effects in negative sentences using sentence completion. Results from their study showed that in continuing negative sentences like (6a), people predominantly chose *yogurt*.

(6) a. The child will not eat the _____.
Four choices: **yogurt**, shellfish, branch, minivan
b. Obviously, the child will not eat the _____.
Four choices: **yogurt, shellfish, branch, minivan**

Crucially, when adverbs (e.g., "obviously," "certainly," "definitely") that signaled expectations for what is very likely to happen were added (e.g., (6b)), people started considering other choices than "yogurt" (e.g., "The child will not eat the minivan."). Albu et al. (2024)'s findings suggest that while people prefer to use negation to deny expectations in general, this preference is subject to contextual modulation, like in (6b). Parallel to the "obviously" study (Albu et al., 2024), we test sentential adverbials like "surprisingly" that signal unexpected, atypical situations in the real-time processing of negative sentences (and affirmative sentences).

1.3. Previous proposals regarding the joint effects of world knowledge and communicative informativeness

Proposals that language comprehension is jointly affected by world-knowledge typicality and communicative informativeness have come up in prior work on both affirmative and negative sentences.

Rohde et al. (2021) propose that comprehension can be approximated as a process of reverse engineering a speaker's utterance choice, while a speaker's utterance choice can be characterized as a generative architecture that combines the prior over meanings and the

likelihood of producing a particular surface form. Following this view, comprehender's expectations about what they are going to read (or hear) can be shaped by their prior knowledge of how likely a situation is and their expectations for how likely it is that an utterance (describing that situation) is produced given that situation. Thus, the comprehender's expectations can be expressed as follows, adapted from Rohde et al.'s (2021) Eq. 4:

$$(7) \quad P(utterance) = \sum_{situation} P(utterance \mid situation)\, P(situation)$$

While Rohde et al.'s generative architecture includes two terms—situation prior and utterance likelihood—their study focused on examining whether comprehenders can be guided to prefer newsworthy information (i.e., high utterance likelihood) even though the information is about atypical events (i.e., low situation prior). As will become clear below, we build on and go beyond their work by testing world knowledge and communicative informativeness at the same time, in both affirmative and negative sentences.

As mentioned above, Nordmeyer and Frank (2014) provide empirical evidence for graded effects of informativity in the speed of truth-value judgments of negative statements. As part of this work, Nordmeyer and Frank modeled reaction times as proportional to the surprisal of the utterance and a certain property (e.g., whether a boy has apples), and found that this parametric relationship applies to both affirmative and negative sentences. The idea behind the model fitting in Nordmeyer and Frank (2014) is similar to what we sketched out in the *Just Right Hypothesis*: language comprehension is a process that coordinates utterance-level possibility (in our case, communication likelihood) and situation-level possibility (in our case, situation typicality).

Considering the nature of a visual context, however, Nordmeyer and Frank's (2014) work leaves open the question of whether similar effects arise with typicality manipulations rooted in our real-world knowledge. Given that core aspects of real-world knowledge are based on a lifetime's worth of experience and not constrained to specific visual contexts, we regard them as a strong, robust test case for typicality effects. Furthermore, Nordmeyer and Frank focused on end-of-sentence reaction times data, and thus their results do not yield information about the time-course of these effects, that is, how rapidly they occur in real-time, incremental processing. Thus, our study extends Nordmeyer and Frank's work to situation typicality based on real-world knowledge and adds real-time sentence processing data to the joint effects of communication likelihood and situation typicality.

1.4. Aims of the present work

We seek to directly and systematically test the joint influence of situation typicality and informativity: how they are coordinated in language comprehension. To manipulate these two factors, we use situations involving real-world part−whole relations. It is well-known that humans have robust knowledge of the real-world properties of such relationships and use this knowledge during language processing (e.g., Chaffin, Herrmann, & Winston, 1988; Poesio, 2003; Sailor & Brooks, 2014).

Our first aim is to systematically investigate how different levels of situation typicality and informativity influence the ease of processing affirmative and negative sentences. Our second

aim is to examine the effects of context. In particular, we are interested in how the absence (in Experiment 1) and presence (in Experiment 2) of contextual cues signaling unexpectedness affects the processing of affirmative and negative sentences.

Ex. (8 and 9) provide examples of affirmative and negative sentences in an unmarked context (Experiment 1), where—in the absence of other cues—we assume situation typicality is estimated based on world knowledge.

(8) Emma just visited a house. The house has a bathroom/garage/ballroom, Emma said.
(9) Emma just visited a house. The house does**n't** have a bathroom/garage/ballroom, Emma said.

Experiment 2 tests contexts that signal an explicit intention of communicating unexpected things, as illustrated (10a,b). (Regarding the quotation marks, see Section 3.3.)

(10) a. Emma just visited a house. "Surprisingly, the house has a ballroom/garage/bathroom," Emma said.
b. Emma just visited a house. "Surprisingly, the house doesn't have a ballroom/garage/bathroom," Emma said.

1.5. Hypotheses

For both affirmative and negative sentences in the unmarked context, we propose the *Just Right Hypothesis*. This builds on our general hypothesis that language comprehension is sensitive to typicality and informativity, and seeks a balance between (i) the ease of integrating linguistic input into world knowledge and (ii) an expectation that communication is informative. Therefore, we expect a "just-right" pattern: the medium level of situation typicality and informativity should be easiest to process (e.g., in the case of "the house has/doesn't have a garage"). This hypothesis is summarized below:

- *Just Right Hypothesis*: The optimal status for language comprehension is when the sentences are "informative enough" and describe "typical enough" world knowledge.

Furthermore, we propose that this status can be modulated by lexical cues. Specifically, we hypothesize that lexical cues such as "surprisingly" can facilitate the processing of atypical situation descriptions, for example, "the house has a ballroom" and "the house doesn't have a bathroom." We call this the *Surprise Me Hypothesis*:

- *Surprise Me Hypothesis*: When lexical cues for unexpectedness are present, the optimal status for language comprehension is when the sentences satisfy such contextual expectations for the unexpected. That is, in the presence of a cue signaling unexpectedness, sentences describing atypical situations are easier to process.

These two hypotheses complement each other and are not mutually exclusive. They predict a "just-right" pattern for affirmative and negative sentences in unmarked contexts (Experiment 1). However, in a context of unexpectedness (Experiment 2)—when expectations of low situation typicality are made explicit by an adverbial cue—they predict a "surprise me" pattern for both affirmative and negative sentences. Because comprehenders expect to be "surprised,"

sentences should now be the easiest to process at the low level of situation typicality and the most difficult to process at the high level of situation typicality.

2. Experiment 1

Experiment 1 used word-by-word self-paced reading to test whether typicality and informativity jointly affect RTs, and whether affirmative and negative sentences show a "just-right" pattern. Materials, data, and analysis scripts for all studies reported in this paper are available through the Open Science Framework (OSF) at: https://osf.io/un6aq.

2.1. Participants

The number of participants was determined by a power analysis through simulations based on pilot data ($n = 54$) using the simr R package (Green & MacLeod, 2016). The power analysis showed that a sample size of 240 participants would yield 85% power [95% CI: 82.63, 87.16] to observe the effect of interest. Supplementary Materials on OSF provide more details of the power analysis.

All participants in the studies reported here were recruited via Prolific (https://www.prolific.com). Each participant only participated in one of the experiments.

Two hundred seventy-three adult native speakers of American English participated in Experiment 1. Participants received \$2−4 dollars for their participation. We only included data from self-reported U.S.-born native English speakers with normal (or corrected-to-normal) vision and hearing. Six participants were excluded for not meeting the demographic requirements. Twenty-two additional participants were excluded for not meeting the prespecified comprehension question accuracy threshold (80%) across all the trials. Thus, 245 participants were included in the final analysis. All exclusion criteria were determined prior to data collection. The current study was reviewed and approved by the IRB at the authors' institution.

2.2. Typicality norming study

To ensure that target items have the intended typicality levels, we ran a norming study. This allows us to obtain a measure of situation typicality, namely, how likely a part−whole relation is. Each part−whole pair was presented individually, as exemplified in (11): The location (e.g., house) is shown in all capitals, and the part entity is shown in lower case below it. Participants rated how likely it is that a location contains a part entity. We report typicality ratings from 40 adult native speakers of American English. Two additional people were excluded for reporting hearing impairments (which can impact their language input) and one for not being a self-reported native speaker of American English. All participants received \$2.

We normed 94 part−whole pairs (24 location nouns, each with 3–5 critical noun candidates). For each location, we used the norming data to identify those part nouns that could be most clearly classified as high, medium, and low typicality. This yielded a total of 72 pairs, 24

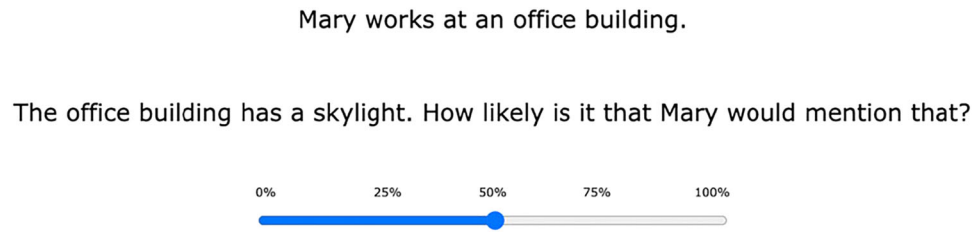

Fig. 1. Communication likelihood norming study: A sample affirmative trial.

for each typicality level. In the selected pairs, the mean likelihood ratings of three typicality levels were 6.43/7 (high typicality), 4.21/7 (medium typicality), and 1.93/7 (low typicality).

(11)

			HOUSE bathroom			
not likely at all					highly likely	
1	2	3	4	5	6	7

2.3. Communication likelihood norming study

Afterward, we also ran a second norming study to ensure that the selected 72 pairs have the intended relation between communication likelihood and situation typicality in both their negative and affirmative versions. On each trial, the participant saw a lead-in sentence and a sentence presenting a part−whole relationship (see Fig. 1). We tested both the affirmative sentences (e.g., "The library has a bookstore," intended to elicit high communication likelihood in a U.S. context) and the negative sentences (e.g., "The library doesn't have a bookstore," intended to elicit low communication likelihood in a U.S. context).

We obtained likelihood ratings given on a slider scale from 52 adult native speakers of American English. Beta regression analysis (Cribari-Neto & Zelleis, 2010, betareg package in R) shows that the communication likelihood of affirmative sentences is negatively correlated with situation typicality ($\beta = -1.89$, $SE = 0.24$, $z = -7.86$, $p < .001$), and the communication likelihood of negative sentences is positively correlated with the situation typicality of their affirmative counterparts ($\beta = 4.51$, $SE = 0.29$, $z = 15.68$, $p < .001$). The results confirm that communication likelihood and situation typicality are closely correlated, as expected.

2.4. Design and materials

In the main experiment, we used self-paced reading and manipulated (i) the typicality of the part−whole relationship described in the critical sentence (low, medium, high) and (ii) the sentence polarity of the critical sentence (affirmative, negative). This yields a 3×2 within-subject design, presented to participants using Latin-Square (six lists). Each target consisted of a lead-in sentence and a critical sentence (shown on one screen), followed by a comprehension question (on the next screen).

(12) a. [affirmative] [high situation typicality; low informativity]
Emma just visited a house. The house has a bathroom, Emma said.
b. [affirmative] [medium situation typicality; medium informativity]
Emma just visited a house. The house has a garage, Emma said.
c. [affirmative] [low situation typicality; high informativity]
Emma just visited a house. The house has a ballroom, Emma said.
d. [negative] [low situation typicality; high informativity]
Emma just visited a house. The house doesn't have a bathroom, Emma said.
e. [negative] [medium situation typicality; medium informativity]
Emma just visited a house. The house doesn't have a garage, Emma said.
f. [negative] [high situation typicality; low informativity]
Emma just visited a house. The house doesn't have a ballroom, Emma said.
***Sample comprehension question**: Did Emma visit a house?*

Example stimuli are given in (12): The lead-in sentence introduces a location (e.g., house), and the critical sentence makes a statement about that location containing something. Thus, target items describe a relation between a location (e.g., house) and a part of that location (e.g., bathroom). We refer to the part noun (e.g., bathroom, garage, ballroom) as the critical noun. Each critical sentence ends with two additional words, which provides a spill-over region: the name of the person from the lead-in sentence and the verb "say" in past tense (e.g., "Emma said"). The reporting clause (e.g., "Emma said") appears at the end of the sentence in half of the items, including targets and fillers.

The experiment included 24 target items, each followed by a comprehension question. Each target item had three versions, associated with three distinct levels of typicality (high, medium, low typicality). These levels vary based on the critical noun used (e.g., $\text{bathroom}_{\text{HIGH\_TYPI}}$, $\text{garage}_{\text{MEDIUM\_TYPI}}$, $\text{ballroom}_{\text{LOW\_TYPI}}$). Based on the norming study described above, we selected 72 nouns that best fit the three levels of typicality (24 nouns per level). In addition to 24 targets, the study included 36 fillers drawn from three published self-paced studies (Sturt, 2007; Patson, Darowski, Moon, & Ferreira, 2009; Roland, Mauner, O'Meara, & Yun, 2012).

All items are followed by comprehension questions. Overall, half of the comprehension questions that a participant sees over the course of the experiment require a "Yes" response and half a "No" response. The presentation order of targets and fillers was randomized for each participant.

2.5. Procedure

The experiment used self-paced reading and was conducted online using PCIbex (Zehr & Schwarz, 2018; https://www.pcibex.net/). Participants completed two practice trials before the main experiment. In each trial, the two-sentence sequence initially appeared as rows of dashes. Participants then pressed the space bar to read the sentences word by word. After the last word, participants saw a comprehension question in the next screen and answered yes or no (F for "Yes," J for "No"). No feedback on the correctness of question answers was

provided. We recorded RTs for each word and responses to the comprehension questions. The estimated completion time for Experiment 1 is approximately 15 min.

2.6. Predictions

If both affirmative and negative sentences show a "just-right" balance between communication likelihood and situation typicality—as predicted by the *Just Right Hypothesis*—we expect that the medium typicality level condition is read faster than the high and the low typicality level conditions (e.g., Affirmative sentences: $\text{garage}_{\text{MEDIUM\_TYPI}} < (\text{bathroom}_{\text{HIGH\_TYPI}}, \text{ballroom}_{\text{LOW\_TYPI}})$; Negative sentences: $\text{not-garage}_{\text{MEDIUM\_TYPI}} < (\text{not-ballroom}_{\text{HIGH\_TYPI}}, \text{not-bathroom}_{\text{LOW\_TYPI}})$).

2.7. Results

2.7.1. Data processing and analysis

We only analyzed target trials where participants answered the associated comprehension question correctly. Furthermore, out of these trials, we removed two trials with consecutive RTs under 50 ms (0.04% of the data), indicating that the participant was simply holding down a key during that trial. We also removed RTs shorter than 150 ms and longer than 3000 ms (0.56% of the data) and RTs more than 3 SDs from each participant's mean RTs (1.8% of the data).

For statistical analyses, we used linear mixed-effect regression models (*lmer*; Baayen, Davidson, & Bates, 2008), using the *lme4* package in R (Bates, Maechler, Bolker, & Walker, 2015). We analyzed affirmative and negative sentences together in a single model, with Polarity and Typicality as predictors. In this model, Typicality was nested within Polarity (written as Polarity/Typicality), allowing us to test whether the three levels of Typicality differed significantly within each sentence polarity. In other words, we examined the simple main effects of Typicality within each level of Polarity, rather than the overall interaction. This approach was chosen because our hypotheses concerned the within-Polarity effects of Typicality, not the interaction term itself.

Following common practice, we also included Trial Order, Word Length, and Word Frequency (log of frequency per million words from the SubtlexUS database, Brysbaert & New, 2009) as fixed effects to control for differences between critical nouns that vary between conditions. In addition, we included log-transformed RTs of the previous word (henceforth, preword RTs) as a covariate in the linear mixed-effects models to control for potential spillover effects and to reduce residual noise (e.g., Haeuser & Borovsky, 2025; Pimentel, Meister, Wilcox, Levy, & Cotterell, 2023; Demberg, 2013; Bartek, Lewis, Vasishth, & Smith, 2011).

Incorporating random effects, we started with fully crossed and fully specified by-subject and by-item effects of Polarity and Typicality. The maximal random effects structure was then reduced (beginning with item-level effects) via model comparison (Baayen et al., 2008). The final statistical model we used is log(RTs) $\sim$ Polarity/Typicality + log(Pre_word_RTs) + Trial_Order + scale(Word_frequency) + scale(Word_length) + random effects.

To test the "just-right" pattern, we used *helmert* contrast coding for the three-level factor Typicality. The contrast coding includes the M versus H+L contrast, which compares the

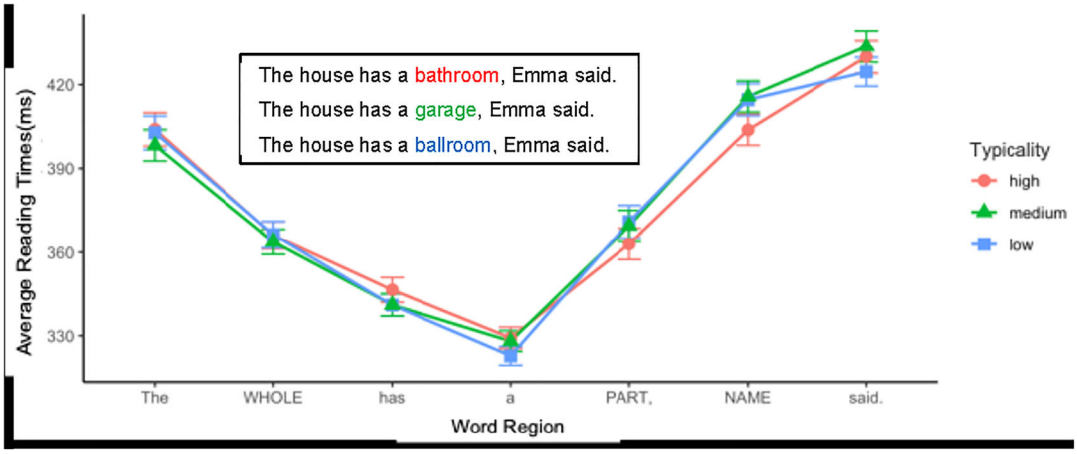

Fig. 2. Experiment 1: Mean reading times by word region in affirmative sentences (error bars represent +/−1 SE).

mean of the medium typicality level with the mean of the other two typicality levels. It also includes the contrast H versus L, which compares the high and low typicality levels with each other.

In all analyses, significance is determined using Satterthwaite's method in the *lmerTest* package (Kuznetsova, Brockhoff, & Christensen, 2017). The significance level was set at $p < .05$.

2.7.2. *Reading times*

We plot raw RTs for each sentence polarity in the figures for ease of interpretability and conducted statistical analyses on log-transformed RTs, in line with current standard practice (e.g., Orth, Sloggett, & Yoshida, 2025; Keshev & Meltzer-Asscher, 2024; Cartson & Yan, 2023). We analyzed RTs for all regions of the critical sentence.

In what follows, we first report the RT patterns for affirmative sentences, then for negative sentences, and finally compare the two sentence polarities. We focus here on the RTs starting at the critical noun. As expected, the RTs before the critical noun show no significant effects of Typicality; see Supplementary Materials on OSF for details.

2.7.2.1. RTs for affirmative sentences: Fig. 2 illustrates the RTs for the critical sentence in affirmative conditions. Visual inspection suggests that, at the critical noun (i.e., PART) and the first spillover region (i.e., NAME), the RTs are shorter in the high typicality level condition (e.g., "bathroom"), than the medium (e.g., "garage") and low typicality level (e.g., "ballroom") conditions, but this pattern disappears by the second spillover region (i.e., "said"). However, when we tested these patterns statistically, we found no significant effects of Typicality in affirmative sentences at any regions of interest (see Table 1 for statistical details). In other words, we did not find evidence that affirmative sentences in an unmarked context show a "just-right" pattern.

Table 1
Experiment 1: Results of the *lmer* model of log (RTs) in *affirmative* and *negative* sentences

Word region	Coefficient	β	*SE*	*t*	*p*
PART (critical noun)	(Intercept)	3.723	0.083	44.597	< 2e−16***
	Polarity	−0.004	0.006	−0.692	.490
	TrialOrder	−0.003	< 0.001	16.339	< 2e−16***
	log(pre_word_RT)	0.383	0.014	27.526	< 2e−16***
	Word_frequency	−0.009	0.004	−2.509	.012*
	Word_length	0.016	0.004	4.394	< .0001***
	Negative: Contrast (M vs. HL)	−0.018	0.008	−2.084	.037*
	Negative: Contrast (H vs. L)	−0.013	0.010	−1.262	.207
	Affirmative: Contrast (H vs. ML)	−0.002	0.009	0.226	.822
	Affirmative: Contrast (M vs. L)	0.002	0.010	0.197	.844
NAME (spillover1)	(Intercept)	4.895	0.078	63.041	< 2e−16***
	Polarity	0.014	0.007	2.134	.033*
	TrialOrder	−0.004	< 0.001	−19.286	< 2e−16***
	log(pre_word_RT)	0.199	0.013	15.716	< 2e−16***
	Word_frequency	−0.005	0.005	−0.999	.319
	Word_length	0.016	0.005	3.427	.001***
	Negative: Contrast (M vs. HL)	−0.017	0.010	−1.690	.091
	Negative: Contrast (H vs. L)	0.002	0.012	0.177	.860
	Affirmative: Contrast (H vs. ML)	−0.005	0.010	−0.489	.625
	Affirmative: Contrast (M vs. L)	0.012	0.012	−1.010	.313
Said (spillover2)	(Intercept)	5.449	0.073	74.799	< 2e−16***
	Polarity	<0.001	0.007	−0.135	.893
	TrialOrder	−0.004	<0.001	−19.580	< 2e−16***
	log(pre_word_RT)	0.115	0.012	9.840	< 2e−16***
	Word_frequency	0.001	0.004	0.274	.784
	Word_length	0.004	0.004	1.052	.293
	Negative: Contrast (M vs. HL)	0.004	0.010	0.005	.698
	Negative: Contrast (H vs. L)	0.001	0.012	0.128	.899
	Affirmative: Contrast (H vs. ML)	0.009	0.010	0.859	.390
	Affirmative: Contrast (M vs. L)	−0.010	0.012	−0.853	.394

Note. Let us now compare *reading times for affirmative* versus *negative sentences*. Among the regions of interest, there are no significant effects of Polarity except at the spillover1 region (i.e., NAME, e.g., "Emma"), where affirmative sentences were read slower than negative sentences ($\beta = 0.014$, $SE = 0.007$, $t = 2.134$, $p = .033$). This is not relevant to the claims we are making in this paper, and we do not discuss this further.[1]

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

2.7.2.2. RTs for negative sentences: Fig. 3 illustrates the RTs for the critical sentence in negative conditions. Visual inspection suggests that, at the critical noun (i.e., PART) and the first spillover region (i.e., NAME), the RTs are shorter in the medium typicality level condition (e.g., not-garage) than the low (e.g., not-bathroom) and high typicality level (e.g., not-ballroom) conditions, but this pattern disappears by the second spillover region (i.e., "said"). When we tested these patterns statistically, we found evidence for a "just-right" pattern at the critical noun: the medium typicality condition yielded significantly lower RTs

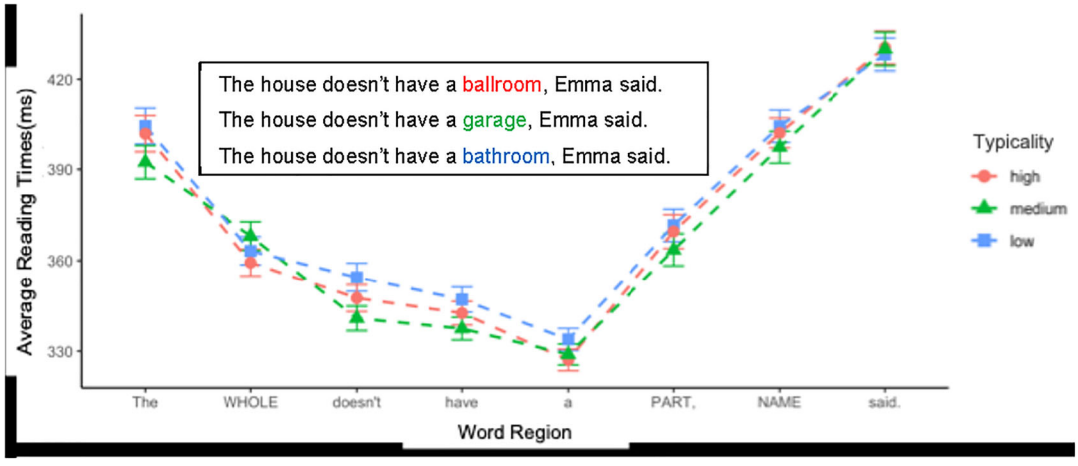

Fig. 3. Experiment 1: Mean reading times by word region in negative sentences (error bars represent +/−1 SE).

($\beta = -0.018$, $SE = 0.008$, $t = -2.084$, $p = .037$) than the other two levels of Typicality that did not differ from each other (see Table 1 for more statistical details). At the first spillover region, we found no significant effects of Typicality (see Table 1).

2.8. Discussion

Experiment 1 examined how situation typicality and informativity influence the processing of affirmative and negative sentences. The *Just Right Hypothesis* predicts that sentences—whether affirmative or negative—are the easiest to process when they describe moderately typical situations and are reasonably likely to be communicated.

The results for *affirmative* sentences did not show a "just right" pattern: we found no significant effects of typicality/informativity. Given the large sample size, this null result is unlikely to be due to a lack of power. We return to this in the General Discussion.

In contrast, the results for *negative* sentences do show the predicted "just-right" pattern at the critical noun: the medium typicality level is read fastest, suggesting that it induced the lowest processing effort. It is important to note that this "just-right" pattern cannot simply be explained by effects of situation typicality or informativity alone; rather, it indicates that situation typicality and informativity interact to affect the processing of negative sentences. We regard the results here as supporting the *Just Right Hypothesis* for negative sentences.

3. Experiment 2

In contrast to Experiment 1, which used unmarked contexts, in Experiment 2, the critical sentences are preceded by an adverb indicating unexpectedness (e.g., *surprisingly*) and are presented as direct speech (using quoted utterances). For both affirmative and negative sentences, these contextual cues provide an explicit signal for readers to expect low-situation-

typicality content. This set-up allows us to test the *Surprise Me Hypothesis*, which predicts that in the context of unexpectedness, both sentence polarities are expected to show a "surprise me" pattern where the low situation typicality level is the easiest to process.

3.1. Participants

The number of participants was determined by a power analysis through simulations based on pilot data ($n = 64$) following the same procedure in Experiment 1. The power analysis showed that a sample size of 138 participants would yield 83.4% power [95% CI: 80.95, 85.66] to observe the effect of interest. Supplementary Materials on the OSF repository provide more details.

All participants in the studies reported here were recruited via Prolific (https://www.prolific.com). Each participant only participated in one of the experiments.

One hundred eighty-three adult native speakers of American English participated in Experiment 2. Participants received \$2.6−\$4 for their participation. We only included data from self-reported U.S.-born native English speakers with normal (or corrected-to-normal) vision and hearing. Eight participants were excluded for not meeting the demographic requirements. Forty additional participants were excluded for not meeting the comprehension question accuracy threshold (80%) across all the trials. Thus, 138 participants were included in the final analysis. All exclusion criteria were determined prior to data collection. The current study was reviewed and approved by the IRB at the authors' institution.

3.2. Design and materials

As in Experiment 1, we used self-paced reading and manipulated (i) typicality of the situation described in the critical sentence (low, medium, high) and (ii) the sentence polarity of the critical sentences (affirmative, negative). The materials were adapted from Experiment 1, with changes to the critical sentence and its wrap-up region (see (13)).

First, a sentence-initial evaluative adverb that indicates unexpectedness ("surprisingly," "amazingly," "unexpectedly," or "strangely") was added. Each of the four adverbs appears in four target items per list. Second, we added quotation marks around the words that the named character (e.g., Emma) says. This presents the critical sentences as direct speech. Recent work (e.g., He & Kaiser, 2025) shows that direct speech (in the form of quoted utterances) introduces expectations for noteworthy information. Furthermore, using direct speech renders the unexpectedness adverbs more natural and clarifies what part of the critical sentence is unexpected (namely, the part inside quotations, not the final clause). This set-up, together with the adverbs, allows us to accomplish our aim of testing how contexts of unexpectedness facilitate the expectations for low-typicality, high-informativity situations. (Recall that the unmarked context examined in Experiment 1 yields different predictions.) Third, the post-critical region was lengthened by four words to make potential late spillover effects detectable.

(13) a. [affirmative] [high situation typicality]
Emma just visited a house. "Surprisingly, the house has a bathroom," Emma told her friend that night.

b. [affirmative] [medium situation typicality]
Emma just visited a house. "Surprisingly, the house has a garage," Emma told her friend that night.
c. [affirmative] [low situation typicality]
Emma just visited a house. "Surprisingly, the house has a ballroom," Emma told her friend that night.
d. [negative] [low situation typicality]
Emma just visited a house. "Surprisingly, the house doesn't have a bathroom," Emma told her friend that night.
e. [negative] [medium situation typicality]
Emma just visited a house. "Surprisingly, the house doesn't have a garage," Emma told her friend that night.
f. [negative] [high situation typicality]
Emma just visited a house. "Surprisingly, the house doesn't have a ballroom," Emma told her friend that night.
Sample comprehension question*: Did Emma visit a house?*

In light of the edits made to targets, we also slightly modified the fillers from Experiment 1 by adding sentence-initial adverbs or phrases and using quotation marks to help fillers and targets blend together. The same comprehension questions in Experiment 1 were used in all trials of Experiment 2.

3.3. Procedure

The procedure is the same as in Experiment 1. The estimated completion time for Experiment 2 is approximately 18 min, due to the lengthening of the materials.

3.4. Predictions

If both affirmative and negative sentences show a "surprise me" pattern, we expect that the low typicality level condition is read fastest, while the high typicality level condition is read slowest (e.g., Affirmative sentences: $\text{ballroom}_{\text{LOW\_TYPI}} < \text{garage}_{\text{MEDIUM\_TYPI}} < \text{bathroom}_{\text{HIGH\_TYPI}}$; Negative sentences: $\text{not-bathroom}_{\text{LOW\_TYPI}} < \text{not-garage}_{\text{MEDIUM\_TYPI}} < \text{not-ballroom}_{\text{HIGH\_TYPI}}$).

3.5. Results

3.5.1. Data processing and analysis

Data processing was done as in Experiment 1. The numerical cutoff of RTs excluded 0.4% of data, and the 3SD-based trimming of RTs excluded another 1.8% of data.

We also used the same statistical model as in Experiment 1. To test for the "surprise me" pattern in both affirmative and negative sentences, we used two pairs of contrast coding. The first pair tests whether the low typicality level condition is read fastest. In this pair, one contrast, L versus H+M, compares the mean of the low typicality level with the mean of the other two typicality levels, and the other contrast, H versus M, compares the high and medium typ-

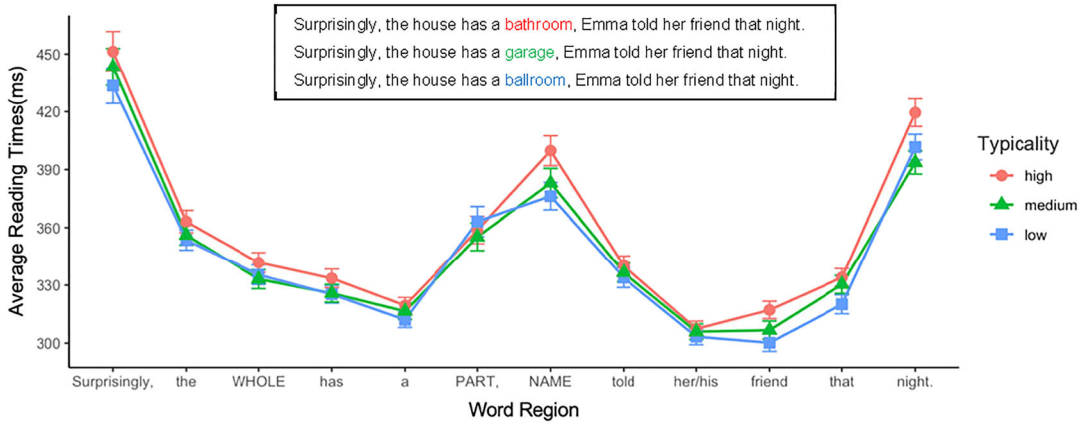

Fig. 4. Experiment 2: Mean reading times by word region in affirmative sentences (error bars represent +/−1 SE).

icality levels with each other. The second pair tests whether the high typicality level condition is read slowest. In this pair, H versus M+L compares the mean of the high typicality level with the mean of the other two typicality levels, and the other contrast, M versus L, compares the medium and low typicality levels with each other. Both pairs of contrasts were applied to the same dataset containing both sentence polarities and run individually using the same statistical model formula.

Additionally, while our statistical analyses were done in a similar way as in Experiment 1—focusing on the main effects of Typicality nested in each level of Polarity—we ran a separate linear mixed effects model with the full interaction between Typicality and Polarity. Supplementary Materials on the OSF repository provide more details and results of this model.

3.5.2. Reading times

We plot raw RTs for each sentence polarity in the figures and conducted statistical analyses on log-transformed RTs. For the ease of comparison, we use the same Typicality level labels as in Experiment 1, even though we introduced contextual factors in Experiment 2. We analyzed RTs for all regions of the critical sentence. In what follows, we first report the RT patterns for affirmative sentences, then for negative sentences, and finally compare the two sentence polarities. As in Experiment 1, we focus here on the RTs starting at the critical noun. The RTs before the critical noun show no significant effects of Typicality, apart from one region ("the") in negative sentences, where the effect is driven by one item; see Supplementary Materials on OSF for details.

3.5.2.1. RTs for affirmative sentences: Fig. 4 illustrates the RTs for the critical sentence in affirmative conditions. Visual inspection suggests (i) that the RTs show little variation at the critical noun (i.e., PART), and that (ii) at the first spillover region (i.e., NAME, e.g., "Emma"), the RTs are shorter in the low typicality level condition (e.g., "ballroom") than the medium

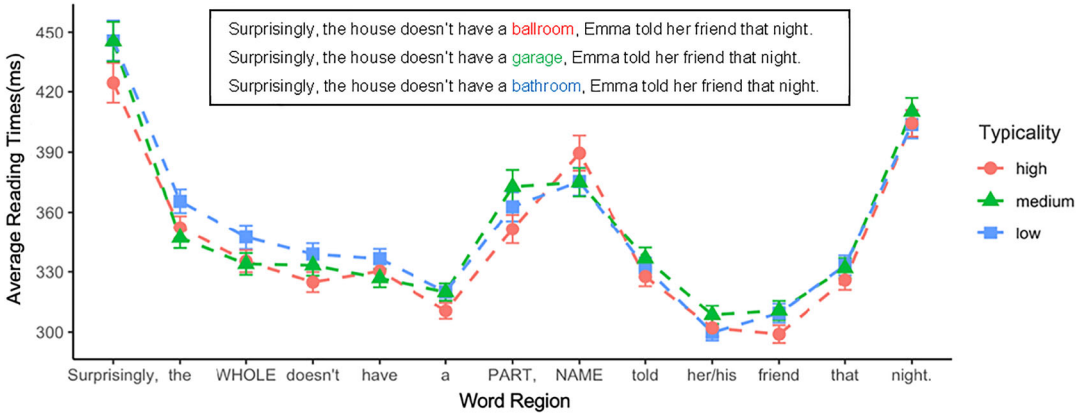

Fig. 5. Experiment 2: Mean reading times by word region in negative sentences (error bars represent +/−1 SE).

(e.g., "garage") and high typicality level (e.g., "bathroom") conditions, but (iii) this pattern disappears in the next two spillover regions (e.g., "told her/his").

When we tested these patterns statistically, we found no significant effects of Typicality at the critical noun (see Table 2 for statistical details). Importantly, we found significant evidence for a "surprise me" pattern at the first spillover region: the low typicality level condition yielded significantly lower RTs than the other two levels ($\beta = -0.037$, $SE = 0.013$, $t = -2.789$, $p = .005$), and the medium typicality level condition yielded significantly lower RTs than the high typicality level condition ($\beta = 0.036$, $SE = 0.015$, $t = 2.300$, $p = .022$). In other words, at the first spillover region in affirmative sentences, the low typicality level was read fastest, and the high typicality level was read slowest—as predicted by the *Surprise Me Hypothesis*.

We also observed significant effects of Typicality at the fourth and the last spillover regions (e.g., "friend," "night"). However, given that the critical region and the first spillover region already showed reliable effects, while intermediate spillover regions did not (see Table 2 for more details), effects observed at later regions are likely too far downstream to be meaningfully attributed to the critical manipulation. We approach these observations with caution and do not discuss this further.

3.5.2.2. RTs for negative sentences: Fig. 5 illustrates the RTs for the critical sentence in negative conditions. Visual inspection suggests (i) that the RTs are fastest in the high typicality level condition (e.g., "ballroom") at the critical noun (i.e., PART), and that (ii) at the first spillover region (i.e., NAME, e.g., "Emma"), the RTs are slowest in the high typicality level condition (e.g., "ballroom"), but (iii) this pattern disappears in the next two spillover regions (e.g., "told her/his").

Statistical analyses revealed effects of Typicality at the critical noun: the high typicality level condition (e.g., not-ballroom) yielded significantly shorter RTs than the other two conditions ($\beta = -0.036$, $SE = 0.014$, $t = 2.882$, $p = .012$), while the medium typicality level

Table 2
Experiment 2: Results of the *lmer* model of log (RTs) in *affirmative* and *negative* sentences

Word region	Coefficient	β	*SE*	*t*	*p*
PART (critical noun)	(Intercept)	5.688	0.260	21.866	< 2e−16***
	Polarity	−0.004	0.012	0.003	.769
	TrialOrder	−0.005	<0.001	−17.065	< 2e−16***
	log(pre_word_RT)	0.107	0.106	1.013	.311
	Word_frequency	−0.008	0.006	−1.331	.183
	Word_length	0.015	0.006	2.684	.007**
	Negative: Contrast (H vs. ML)	−0.036	0.014	−2.524	.012*
	Negative: Contrast (M vs. L)	0.010	0.016	0.653	.514
	Affirmative: Contrast (H vs. ML)	0.010	0.014	0.735	.463
	Affirmative: Contrast (M vs. L)	−0.006	0.016	−0.426	.670
	Negative: Contrast (L vs. HM)	0.010	0.014	0.714	.475
	Negative: Contrast (H vs. M)	−0.041	0.016	−2.515	.012*
	Affirmative: Contrast (L vs. HM)	−0.001	0.013	−0.005	.996
	Affirmative: Contrast (H vs. M)	0.014	0.016	0.854	.393
NAME (spillover1)	(Intercept)	5.121	0.092	55.479	< 2e−16***
	Polarity	0.015	0.009	1.631	.177
	TrialOrder	−0.005	<0.001	−16.886	< 2e−16***
	log(pre_word_RT)	0.151	0.015	10.164	< 2e−16***
	Word_frequency	−0.009	0.005	−1.722	.085
	Word_length	0.003	0.005	0.477	.633
	Negative: Contrast (H vs. ML)	0.029	0.014	2.092	.037*
	Negative: Contrast (M vs. L)	−0.008	0.016	−0.524	.601
	Affirmative Contrast (H vs. ML)	0.045	0.013	3.369	<.001***
	Affirmative: Contrast (M vs. L)	0.020	0.015	1.273	.203
	Negative: Contrast (L vs. HM)	−0.008	0.014	−0.608	.543
	Negative: Contrast (H vs. M)	0.033	0.016	2.077	.038*
	Affirmative Contrast (L vs. HM)	−0.037	0.013	−2.789	.005**
	Affirmative: Contrast (H vs. M)	0.036	0.015	2.300	.022*
Told (spillover2)	(Intercept)	4.850	0.075	64.723	<2e−16***
	Polarity	0.011	0.007	1.464	.143
	TrialOrder	−0.004	<0.001	−19.922	< 2e−16***
	log(pre_word_RT)	0.178	0.012	14.726	< 2e−16***
	Negative: Contrast (H vs. ML)	−0.012	0.011	−1.085	.278
	Negative: Contrast (M vs. L)	0.004	0.013	0.323	.746
	Affirmative: Contrast (H vs. ML)	−0.003	0.011	−0.290	.772
	Affirmative: Contrast (M vs. L)	0.011	0.013	0.894	.371
	Negative: Contrast (L vs. HM)	0.003	0.013	0.270	.787
	Negative: Contrast (H vs. M)	−0.015	0.013	−1.104	.270
	Affirmative Contrast (L vs. HM)	−0.007	0.011	−0.623	.533
	Affirmative: Contrast (H vs. M)	−0.009	0.013	−0.699	.485
her/his (spillover3)	(Intercept)	4.032	0.079	50.969	<2e−16***
	Polarity	0.004	0.006	0.598	.550
	TrialOrder	−0.004	<0.001	−18.203	< 2e−16***
	log(pre_word_RT)	0.304	0.013	23.155	< 2e−16***
	Negative: Contrast (H vs. ML)	−0.002	0.010	−0.162	.871
	Negative: Contrast (M vs. L)	0.017	0.011	1.516	.130
	Affirmative: Contrast (H vs. ML)	0.009	0.010	0.901	.368
	Affirmative: Contrast (M vs. L)	0.010	0.011	0.950	.342
	Negative: Contrast (L vs. HM)	−0.012	0.010	−1.226	.220
	Negative: Contrast (H vs. M)	−0.010	0.011	−0.892	.373
	Affirmative: Contrast (L vs. HM)	−0.012	0.010	−1.272	.203
	Affirmative: Contrast (H vs. M)	0.003	0.011	0.313	.754

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

did not differ from the low typicality level ($\beta = 0.010$, $SE = 0.016$, $t = 0.653$, $p = .514$). This does not align with the "surprise me" pattern. We return to this effect near the end of this section.

Crucially, we found evidence for the predicted "surprise me" pattern at the first spillover region: the high typicality level condition yielded longer RTs than the other two conditions ($\beta = 0.029$, $SE = 0.014$, $t = 2.092$, $p = .037$), while the medium typicality level did not differ from the high typicality level (see Table 2). Thus, we observe the "surprise me" pattern at the same point in affirmative and negative sentences; the finding that the contextual effect emerges also in negative sentences supports the *Surprise Me Hypothesis*.

We also observed marginal significant effects of Typicality at the fourth spillover region (e.g., *friend*). However, for the same reasons we mentioned above, we approach these observations with caution and do not discuss this further.

Taking the observations at the critical noun and the spillover region together, it appears that negative sentences briefly showed what seems to be an inverted "surprise me" pattern before showing the predicted "surprise me" pattern. We do not want to overinterpret the effect at the critical noun, but we take it to suggest that an underspecified context of unexpectedness—one that signals surprise without clarifying/implying what was expected—may have caused temporary confusion in the processing of negative sentences. *Importantly*, however, the emergence of the predicted "surprise me" predictions at the spillover region reinforces our main findings and provides strong support for the *Surprise Me Hypothesis*.

3.6. Discussion

Experiment 2 tests the *Surprise Me Hypothesis*, which predicts that in a context of unexpectedness, both sentence polarities are expected to show a "surprise me" pattern, such that the low situation typicality level is the easiest to process. In other words, in contexts of unexpectedness, sentences should be easiest to process when they describe low-situation typicality content, and most effortful to process when they describe high-situation typicality content.

The results for *affirmative* sentences provide strong evidence for a "surprise me" pattern at the first spillover region: the low typicality level induced the smallest processing effort, and the high typicality level induced the greatest processing effort. We regard the results here as supporting our hypothesis that context can influence expectations for what kinds of situations are being communicated.

The results for *negative* sentences also provide evidence for a "surprise me" pattern at the first spillover region: the high typicality level induced the greatest processing effort.

As a whole, we regard these results as providing support for the *Surprise Me Hypothesis*.

4. General discussion

Do comprehenders take into account both world knowledge and linguistic experience in language processing? How do situation typicality and informativity work together in guiding the processing of both affirmative and negative sentences? Prior work has observed individual effects of situation typicality and informativity; nevertheless, how these two jointly contribute

to language comprehension has received little attention, despite their conceptual correlation. Building on and going beyond prior work, we propose the *Just Right Hypothesis*, which posits that language comprehension is the least effortful when the sentences are "informative enough" and describe situations that are "typical enough" in terms of our world knowledge. Furthermore, to further understand how contextual cues for unexpectedness guide comprehenders' expectations for what kinds of situations are being communicated, we propose the *Surprise Me Hypothesis*, which posits that language comprehension is the least effortful when sentences satisfy the contextual expectations for atypical situations being communicated.

To test these two hypotheses, we examined how varying levels of situation typicality and informativity affect the processing of sentences describing part−whole relationships in different contexts. We conducted two self-paced reading experiments: Experiment 1 tested the *Just Right Hypothesis* and Experiment 2 tested the *Surprise Me Hypothesis*.

4.1. "Just Right" is not always right

In Experiment 1, we used an unmarked context that provides minimum contextual information. For both affirmative and negative sentences, in this kind of context, the *Just Right Hypothesis* predicts a "just-right" pattern whereby the medium level of situation typicality and communication likelihood is the easiest to process. Indeed, we find evidence for the "just-right" pattern from negative sentences in Experiment 1: at the critical noun, the medium typicality level condition has the fastest RTs, suggesting that it elicited the lowest processing effort. The "just-right" pattern we observed for negative sentences in an unmarked context is likely among the first empirical demonstrations of a Goldilocks effect in language comprehension. More generally, our findings contribute evidence from the comprehension domain to broader proposals in pragmatics and language production that there exists an ideal level of informativity that is both cognitively and communicatively desirable (e.g., Grice, 1975; Brown & Dell, 1987; Frank & Goodman, 2012; Levy & Jaeger, 2007; Jaeger, 2010; see also Goldilocks effects in attention allocation: Kidd, Piantadosi, & Aslin, 2012, 2014).

In affirmative sentences, however, we found no significant effects of situation typicality. Given our large sample size and adequate power, we take this null result as meaningful. This absence of an effect is somewhat surprising, considering the large prior literature on affirmative sentences that took a binary (low/high) approach and reported typicality effects. Our findings suggest that the real-time processing of low-informativity sentences is not as effortless as the long-standing effects of world knowledge might lead us to expect. Importantly, we would like to emphasize that our null result for affirmative sentences does not directly contradict the view that situation typicality and communicative informativeness jointly affect the processing of affirmative sentences; neither typicality nor informativity effects alone can account for the null difference across levels of typicality. Rather, it indicates that a "just-right" balance between typicality and informativity may be incomplete for affirmative sentences. In particular, instead of weighing typicality and informativity equally, comprehenders may give different weights to these two factors in processing affirmative sentences.

Taken together, the Experiment 1 results suggest that although it is likely that both affirmative and negative sentences are subject to similar pressures, that is, from both world knowl-

edge and expectations for communicative informativeness, how exactly these two factors work together differs across sentence polarities. In this sense, Experiment 1 provides new evidence as well as a new perspective for sentence polarity asymmetry at the processing level.

In addition to addressing our primary research questions, the present findings also speak to two open issues in the literature (see Sections 1.1 and 1.2). First, regarding the processing of low-informativity sentences, results of both sentence polarities in Experiment 1 suggest that comprehenders are sensitive to low informativity in the real-time processing, even without a task that directs their attention toward sentence informativeness. Second, regarding the processing of negation, our results for negative sentences in Experiment 1 suggest that prior findings about the role of informativity in negation processing may reflect the combined efforts of processing negation per se and making truth-value judgments of negative sentences. Our present results thus provide initial evidence that these questions about the real-time processing of negation and low-informativity sentences are tractable and offer a methodological foundation for investigating them directly.

4.2. "Surprise Me" and how comprehenders follow speakers' lead

In addition to testing the *Just Right Hypothesis* in Experiment 1, in Experiment 2, we investigated how contextual cues can guide comprehenders' expectations toward what people do not normally communicate in a context of unexpectedness. We used a context that signals an explicit intention of communicating atypical situations by means of adverbs like "surprisingly." This allows us to test the *Surprise Me Hypothesis*, which predicts a "surprise me" pattern in contexts of unexpectedness, regardless of sentence polarity: affirmative and negative sentences describing low situation typicality should be the easiest to process, and those describing high situation typicality should be the hardest to process. Overall, the RT patterns from Experiment 2 provide evidence for this kind of "surprise me" pattern in both affirmative and negative sentences. For affirmative sentences, at the first spillover region, the low typicality level condition has the fastest RTs, suggesting that it elicited the lowest processing effort, and the high typicality level condition has the slowest RTs, suggesting that it elicited the greatest processing effort. For negative sentences, at the first spillover region, the high typicality level condition has the slowest RTs, suggesting that it elicited the greatest processing effort.

Broadly speaking, the "surprise me" pattern observed in Experiment 2 illustrates that comprehenders are able to detect and use contextual cues that speakers embed in their utterances to guide interpretation, thereby enabling more efficient communication and reducing processing effort. This contrast is particularly clear when comparing the findings in our two experiments: in Experiment 1, such contextual cues were absent, whereas in Experiment 2, they were explicitly provided, and people showed a different processing pattern than in Experiment 1. Prior research shows that these cues are not limited to overt lexical markers such as "surprisingly." They can also take the form of discourse connectives like "even so," which signal an unexpected transition (Xiang & Kuperberg, 2015), speech disfluencies such as "um" (Heller et al., 2015), or a protagonist's tendency to conduct unconventional behaviors (a surprising person vs. a boring person, Rohde et al., 2021). Moreover, such cues do not only

facilitate the communication of atypical or unexpected information. The communication of obvious information can likewise be signaled through lexical markers such as "obviously" (Albu et al., 2024).

Alongside prior work, our findings across both experiments highlight that communication can be seen as a joint effort between speakers and comprehenders, with speakers providing cues that shape expectations and comprehenders actively using those cues to guide comprehension.

4.3. Preliminary attempts to formalize the joint effects of world knowledge and communicative informativeness across sentence polarities

While it is not one of the aims of the current study to formalize the joint effects of world knowledge and communicative informativeness across sentence polarities, in this section, we present some initial speculative explorations built on prior work.

As mentioned above, comprehenders' expectations for language input (i.e., utterances) can be seen as jointly shaped by their prior knowledge of how likely a situation is and their expectations for how likely it is that an utterance (describing that situation) is produced given that situation. This can be expressed as shown in (7) and repeated below (adapted from Eq. 4 in Rohde et al., 2021):

$$(14) \quad P(utterance) = \sum_{situation} P(utterance \mid situation)\, P(situation)$$

In the unmarked context (Experiment 1), our findings suggest that situation typicality and communicative informativeness jointly affect language comprehension, supporting the conceptualization in (14). However, in the context of unexpectedness (Experiment 2), the conceptualization in (14) may have the potential to capture our findings of the "surprise me" pattern, but that is hard to verify without explicit modeling. As utterances and situations become more complex, explicitly computing or estimating *P(utterance)* following (14) quickly becomes intractable. Consider a complex situation like "The house has a garage AND someone finds it surprising." Here, *P(situation)* refers to the joint probability of these two subsituations occurring together, denoted as P(The house has a garage, Someone finds it surprising). To further complicate matters, these subsituations are probably not independent; estimating their joint probability requires accounting for their dependency, which is more difficult than estimating them individually. Therefore, for current preliminary attempts at formalization, we focus on the findings in the unmarked context and leave further conceptualization of more complex cases to future work.

To formalize findings in the unmarked context across both sentence polarities, we suggest the possible post-hoc modification shown in (15). Here, communicative informativeness and situation typicality have individual weights, indicated by the superscripts α and β. These weights determine how much each factor contributes to a comprehender's expectations of what they are going to read (or hear), and allow the two factors to contribute differently to different sentence polarities as well.

$$(15) \quad P(utterance) = \sum_{situation} P(utterance \mid situation)^{\alpha}\, P(situation)^{\beta}$$

We present this idea of having weight parameters simply at the conceptual level. Nonetheless, a version of this idea has been implemented in Nordmeyer and Frank (2014). Nordmeyer and Frank explored the graded effects of expectations in the truth-value judgments of affirmative and negative sentences using visual contexts. They modeled reaction times as proportional to utterance surprisal and property surprisal (e.g., regarding whether a boy has apples in a visual context) by weighing these two factors differently. Their model, however, used the same weighting for both affirmative and negative sentences and was able to account for much of the variance in their truth-value judgment reaction time data.

Considering that we obtained null results for affirmative sentences and a "just-right" pattern for negative sentences in Experiment 1, it is possible that *affirmative sentences* weigh situation typicality more than communicative informativeness. This is supported by the numerical trend we observed showing shorter RTs in the high typicality level condition than the other two conditions in affirmative sentences, and by the stronger representation of typicality effects in affirmative sentences overall. It is also possible that *negative sentences* weigh communicative informativeness more than situation typicality. This is supported by the stronger representation of informativity effects in negative sentences. We leave this possibility of polarity-dependent weighing as a promising direction for future modeling work.

5. Conclusion

The experiments reported in this paper provide empirical evidence for how comprehenders coordinate situation typicality and communicative informativeness, revealing differences regarding this coordination across sentence polarities and different contexts. These findings have novel implications for understanding the joint effects of world knowledge and linguistic experience in language comprehension, as well as the (a)symmetry between affirmative and negative sentences in real-time sentence processing.

Acknowledgments

Thanks to Deniz Rudin and Zuzanna Fuchs for their valuable feedback at the early stages of this work. We thank the anonymous reviewers for their generosity in sharing their expertise, insights, and time. An earlier version of some of the results from both experiments appeared in the *Proceedings of the Chicago Linguistic Society 60*.

Ethics approval statement

The current study was reviewed and approved by the IRB at the University of Southern California (UP-22-00845).

Conflicts of interest

The authors report there are no conflicts of interest to declare.

Data availability statement

Materials, data, and analysis scripts for all studies reported in this paper are available through Open Science Framework at: https://osf.io/un6aq.

Note

1 Further analysis showed that affirmative sentences were numerically read slower than negatives in only one of the six stimulus lists, and numerically slightly shorter in the remaining lists. We, therefore, treated the significant Polarity effect at the spillover1 region as an artifact. Overall, these results suggest that, in an unmarked context like ours, negative sentences are not necessarily read slower than affirmative sentences.

References

Albu, E., Dudschig, C., Warren, T., & Kaup, B. (2024). Does negation influence the choice of sentence continuations? Evidence from a four-choice cloze task. *Quarterly Journal of Experimental Psychology*, *77*(1), 90–110. https://doi.org/10.1177/17470218231158109

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Bartek, B., Lewis, R. L., Vasishth, S., & Smith, M. R. (2011). In search of on-line locality effects in sentence comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(5), 1178.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bergey, C., Morris, B. C., & Yurovsky, D. (2020). Children hear more about what is atypical than what is typical. In *Proceedings of the 42th Annual Meeting of the Cognitive Science Society* (pp. 501–507). Cognitive Science Society.

Brown, P. M., & Dell, G. S. (1987). Adapting production to comprehension: The explicit mention of instruments. *Cognitive Psychology*, *19*(4), 441–472. https://doi.org/10.1016/0010-0285(87)90015-6

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*(4), 977–990. https://doi.org/10.3758/BRM.41.4.977

Carston, R., & Yan, X. (2023). Metaphor processing: Referring and predicating. *Cognition*, *238*, 105534. https://doi.org/10.1016/j.cognition.2023.105534

Chaffin, R., Herrmann, D. J., & Winston, M. (1988). An empirical taxonomy of part−whole relations: Effects of part−whole relation type on relation identification. *Language and Cognitive Processes*, *3*(1), 17–48. https://doi.org/10.1080/01690968808402080

Cribari-Neto, F., & Zeileis, A. (2010). Beta regression in R. *Journal of Statistical Software*, *34*(2), 1–24. https://doi.org/10.18637/jss.v034.i02

Demberg, V. (2013). Integration costs on auxiliaries?–A self-paced reading study using WebExp. In *Proceedings of the 35th Annual Meeting of the Cognitive Science Society* (pp. 2169–2165). Cognitive Science Society.

Ferretti, T. R., Kutas, M., & McRae, K. (2007). Verb aspect and the activation of event knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(1), 182–196. https://doi.org/10.1037/0278-7393.33.1.182

Fischler, I., Bloom, P. A., Childers, D. G., Roucos, S. E., & Perry Jr, N. W. (1983). Brain potentials related to stages of sentence verification. *Psychophysiology*, *20*(4), 400–409. https://doi.org/10.1111/j.1469-8986.1983.tb00920.x

Frank, M. C., & Goodman, N. D. (2012). Predicting pragmatic reasoning in language games. *Science*, *336*(6084), 998–998. https://doi.org/10.1126/science.1218633

Green, P., & MacLeod, C. J. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, *7*(4), 493–498.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), *Syntax and semantics: Speech acts* (pp. 41–58). Academic Press.

Haeuser, K. I., & Borovsky, A. (2025). Got it right up front? Further evidence for parallel graded prediction during prenominal article processing in a self-paced reading study. *Glossa Psycholinguistics*, *4*(1): 7, 1–40. http://doi.org/10.5070/G6011.1636

Hale, J. (2001). A probabilistic Earley parser as a psycholinguistic model. In *Second Meeting of the North American Chapter of the Association for Computational Linguistics*.

Hagoort, P., Hald, L., Bastiaansen, M., & Petersson, K. M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science*, *304*, 438–441. https://doi.org/10.1126/science.1095455

He, M., & Kaiser, E. (2025). Communication form modulates sentence interpretation: (A) typicality inferences from descriptions vs. direct speech. In *Proceedings of the 47th Annual Meeting of the Cognitive Science Society* (pp. 885–891). Cognitive Science Society.

Heller, D., Arnold, J. E., Klein, N., & Tanenhaus, M. K. (2015). Inferring difficulty: Flexibility in the real-time processing of disfluency. *Language and Speech*, *58*, 190–203.

Horn, L. (1989). *A natural history of negation*. University of Chicago Press.

Jaeger, T. F. (2010). Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology*, *61*(1), 23–62. https://doi.org/10.1016/j.cogpsych.2010.02.002

Levy, R., & Jaeger, T. (2007). Speakers optimize information density through syntactic reduction. *Advances in Neural Information Processing Systems*, *19*, 849–856.

Keshev, M., & Meltzer-Asscher, A. (2024). The representation of agreement features in memory is updated during sentence processing: Evidence from verb−reflexive interactions. *Journal of Memory and Language*, *135*, 104495. https://doi.org/10.1016/j.jml.2023.104495

Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2012). The Goldilocks effect: Human infants allocate attention to visual sequences that are neither too simple nor too complex. *PLOS ONE*, *7*(5), e36399. https://doi.org/10.1371/journal.pone.0036399

Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2014). The Goldilocks effect in infant auditory attention. *Child Development*, *85*(5), 1795–1804. https://doi.org/10.1111/cdev.12263

Kounios, J., & Holcomb, P. J. (1992). Structure and process in semantic memory: Evidence from event-related brain potentials and reaction times. *Journal of Experimental Psychology: General*, *121*(4), 459–479.

Kravtchenko, E., & Demberg, V. (2015). Semantically underinformative utterances trigger pragmatic inferences. In R. Dale, C. Jennings, P. Maglio, T. Matlock, D. Noelle, A. Warlaumont, & J. Yoshimi (Eds.), *Proceedings of the 37th Annual Meeting of the Cognitive Science Society* (pp. 1207–1212). Cognitive Science Society.

Kravtchenko, E., & Demberg, V. (2022). Informationally redundant utterances elicit pragmatic inferences. *Cognition*, *225*, 105159. https://doi.org/10.1016/j.cognition.2022.105159

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, *207*, 203–205. https://doi.org/10.1126/science.7350657

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

Lee, S. H. Y., & Kaiser, E. (2026). Event knowledge modulates real-time mental representations of object state-change. *Cognitive Science*, *50*(1), e70165.

Lüdtke, J., & Kaup, B. (2006). Context effects when reading negative and affirmative sentences. In *Proceedings of the 28th Annual Conference of the Cognitive Science Society* (pp. 1735–1740). Cognitive Science Society.

Matsuki, K., Chow, T., Hare, M., Elman, J. L., Scheepers, C., & McRae, K. (2011). Event-based plausibility immediately influences on-line language comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(4): 913–914. doi: 10.1037/a0022964

McRae, K., & Matsuki, K. (2009). People use their knowledge of common events to understand language, and do so as quickly as possible. *Language and Linguistics Compass*, *3*(6), 1417–1429. https://doi.org/10.1111/j.1749-818X.2009.00174.x

Morris, R. K. (1994). Lexical and message-level sentence context effects on fixation times in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(1), 92. https://doi.org/10.1037/0278-7393.20.1.92

Nieuwland, M. S., & Kuperberg, G. R. (2008). When the truth is not too hard to handle: An event-related potential study on the pragmatics of negation. *Psychological Science*, *19*(12), 1213–1218. https://doi.org/10.1111/j.1467-9280.2008.02226.x

Nieuwland, M. S., & Van Berkum, J. J. A. (2006). When peanuts fall in love: N400 evidence for the power of discourse. *Journal of Cognitive Neuroscience*, *18*, 1098–1111. https://doi.org/10.1162/jocn.2006.18.7.1098

Nordmeyer, A. E., & Frank, M. C. (2014). The role of context in young children's comprehension of negation. *Journal of Memory and Language*, *77*, 25–39. https://doi.org/10.1016/j.jml.2014.08.002

Orth, W., Sloggett, S., & Yoshida, M. (2025). Positive polarity items: An illusion of ungrammaticality. *Language, Cognition and Neuroscience*, *40*(4): 547–571. https://doi.org/10.1080/23273798.2025.2453183

Patson, N. D., Darowski, E. S., Moon, N., & Ferreira, F. (2009). Lingering misinterpretations in garden-path sentences: Evidence from a paraphrasing task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(1), 280.

Pimentel, T., Meister, C., Wilcox, E. G., Levy, R. P., & Cotterell, R. (2023). On the effect of anticipation on reading times. *Transactions of the Association for Computational Linguistics*, *11*, 1624–1642.

Poesio, M. (2003). Associative descriptions and salience: A preliminary investigation. In *Proceedings of the 2003 EACL Workshop on the Computational Treatment of Anaphora*.

Rayner, K., Carlson, M., & Frazier, L. (1983). The interaction of syntax and semantics during sentence processing: Eye movements in the analysis of semantically biased sentences. *Journal of Verbal Learning and Verbal Behavior*, *22*(3), 358–374. https://doi.org/10.1016/S0022-5371(83)90236-0

Rayner, K., Warren, T., Juhasz, B. J., & Liversedge, S. P. (2004). The effect of plausibility on eye movements in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(6), 1290–1301. https://doi.org/10.1037/0278-7393.30.6.1290

Reksnes, V. R., Rees, A., Cummins, C., & Rohde, H. (2024). Tell me something I don't know: Speaker salience and style affect comprehenders' expectations for informativity. *Information Structure and Information Theory*, *10*, 177–198.

Rohde, H., Futrell, R., & Lucas, C. G. (2021). What's new? A comprehension bias in favor of informativity. *Cognition*, *209*, 104491. https://doi.org/10.1016/j.cognition.2020.104491

Rohde, H., & Rubio-Fernandez, P. (2022). Color interpretation is guided by informativity expectations, not by world knowledge about colors. *Journal of Memory and Language*, *127*, 104371. https://doi.org/10.1016/j.jml.2022.104371

Roland, D., Mauner, G., O'Meara, C., & Yun, H. (2012). Discourse expectations and relative clause processing. *Journal of Memory and Language*, *66*(3), 479–508.

Ryzhova, M., & Demberg, V. (2023). Processing cost effects of atypicality inferences in a dual-task setup. *Journal of Pragmatics*, *211*, 47–80.

Ryzhova, M., Mayn, A., & Demberg, V. (2023). What inferences do people actually make upon encountering informationally redundant utterances? An individual differences study. In *Proceedings of the 45th Annual Meeting of the Cognitive Science Society* (pp. 2631–2638). Cognitive Science Society.

Sailor, K., & Brooks, P. J. (2014). Do part–whole relations produce facilitation in the picture–word interference task? *Quarterly Journal of Experimental Psychology*, *67*(9), 1768–1785. https://doi.org/10.1080/17470218.2013.870589

Sedivy, J. C. (2003). Pragmatic versus form-based accounts of referential contrast: Evidence for effects of informativity expectations. *Journal of Psycholinguistic Research*, *32*, 3–23. https://doi.org/10.1023/A:1021928914454

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

Sturt, P. (2007). Semantic re-interpretation and garden path recovery. *Cognition*, *105*(2), 477–488.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*(5217), 1632–1634.

Tian, Y., Breheny, R., & Ferguson, H. J. (2010). Why we simulate negated information: A dynamic pragmatic account. *Quarterly Journal of Experimental Psychology*, *63*(12), 2305–2312. https://doi.org/10.1080/17470218.2010.525712

Troyer, M., & Kutas, M. (2018). Harry Potter and the chamber of what?: The impact of what individuals know on word processing during reading. *Language, Cognition, and Neuroscience*, *35*(5): 641–657. https://doi.org/10.1080/23273798.2018.1503309

Walker, J. H. (1975). Real-world variability, reasonableness judgments, and memory representations for concepts. *Journal of Verbal Learning and Verbal Behavior*, *14*, 241–252. https://doi.org/10.1016/S0022-5371(75)80067-3

Wason, P. C. (1965). The contexts of plausible denial. *Journal of Verbal Learning and Verbal Behavior*, *4*(1), 7–11. https://doi.org/10.1016/S0022-5371(65)80060-3

Xiang, M., Kramer, A., & Nordmeyer, A. E. (2020). An informativity-based account of negation complexity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*(10), 1857–1873. https://doi.org/10.1037/xlm0000851

Xiang, M., & Kuperberg, G. (2015). Reversing expectations during discourse comprehension. *Language, Cognition and Neuroscience*, *30*(6), 648–672.

Zehr, J., & Schwarz, F. (2018). PennController for Internet Based Experiments: IBEX. *Open Science Framework*, https://doi.org/10.17605/OSF.IO/MD832