How do linguistic illusions arise? Rational inference and good-enough processing as competing latent processes within individuals

Dario Paape
University of Potsdam

Abstract

Non-literal interpretations of implausible sentences such as *The mother gave the candle the daughter* have been taken as evidence for a rational error-correction mechanism that reconstructs the intended utterance from the ill-formed input (*... gave the daughter the candle*). However, the good-enough processing framework offers an alternative explanation: readers sometimes miss problematic aspects of sentences because they are only processing them superficially, which leads to acceptability illusions. As a synthesis of these accounts, I propose that rational inferences about errors on the one hand and good-enough processing on the other are competing latent processes that simultaneously occur within the same comprehender. In support of this view, I present data from a two-dimensional grammaticality/interpretability judgment task with different types of subtly ill-formed sentences. Both rational inference and good-enough processing predict positive interpretability judgments for such sentences, but only good-enough processing also predicts positive grammaticality judgments. By fitting a lognormal race model jointly to judgments and response latencies, I show that rational inference and good-enough processing actively trade off with each other during reading. Furthermore, individual differences measures reveal that participant traits such as linguistic pedantry, interpretational charity, and analytic/intuitive cognitive styles contribute to variability in the processing patterns.

Keywords: linguistic illusions, good enough processing, rational inference, individual differences

Introduction

Consider the sentence in (1):

(1) The mother gave the candle the daughter.

Please send correspondence concerning this article to: Dario Paape, Department Linguistik, Universität Potsdam, Karl-Liebknecht-Straße 24–25, D-14476, Germany. Email: paape@uni-potsdam.de.

Depending on one's point of view, this sentence is either implausible or ungrammatical. If the sentence is interpreted literally, it means that the daughter was given to the candle. If the sentence is not interpreted literally, and we assume that the utterer meant to say that the candle was given to the daughter, they must have either swapped the order of the arguments by mistake, or forgotten the word *to* before *the daughter*. Furthermore, in a spoken conversation, it is also possible that the *comprehender* may have missed or misheard a word. It has been suggested that language users make these kinds of inferences about possible speech errors and other communicative slips all the time, in order to cope with the noisiness inherent in everyday communication (Gibson et al., 2013; Levy, 2008b). This behavior is *rational* in the sense that it is purposive (Chater & Oaksford, 1999): the language user's goal is to reconstruct the intended message by making use of prior knowledge about sensible things that people might say. By combining this knowledge with the (subjective) probability of different kinds of errors, they can recover the most likely interpretation. In the case of (1), the inference would be that it was probably the candle that was given to the daughter, given the plausibility of the scenario and the plausibility of an argument swap as a possible speech error (Poppels & Levy, 2016).

The pure "rationalist" perspective of language processing assumes an idealized comprehender with perfect knowledge of language statistics (Gibson et al., 2013) who is not bounded by cognitive resource constraints. Levy (2008b) assumes that the error-correction process in cases like (1), where mentally editing the input yields an a-priori more plausible analysis, causes additional mental effort compared to cases in which error correction is not needed or not possible (see also Chen et al., 2023; Gibson et al., 2013; Levy et al., 2009; Ryskin et al., 2021).[1] It is thus assumed that comprehenders may go to great lengths to find out what a given sentence is most likely supposed to mean: they may invest more cognitive resources into mental error correction (Levy, 2008b), and/or reread parts of the sentence (Levy et al., 2009). This assumption may not always be realistic.

There exists a longstanding and influential line of work in decision-making research that has highlighted the fact that human rationality is bounded by constraints such as time pressure, incomplete information, and fluctuating motivation (e.g., Gigerenzer and Selten, 2002; Selten, 1990; Simon, 1972). Drawing from this literature, the "good enough" processing framework (e.g., Ferreira and Patson, 2007) in psycholinguistics has highlighted the goal of saving cognitive resources by occasionally omitting effortful processing steps. In the case of (1), this could mean computing only a "bags of words" style parse of the sentence and otherwise mostly relying on prior event plausibility (Kuperberg, 2016; Paape et al., 2020), failing to register the absence of the word *to*, and/or failing to actively monitor one's comprehension (Glenberg et al., 1982).

Due to the shared prediction that readers sometimes adopt non-literal interpretations and partly rely on plausibility, it has historically proven difficult to disentangle the predictions of the good-enough processing framework from those of the rational inference framework (Brehm et al., 2021). For this reason, rational inference has recently been treated as a subtype of good-enough processing by some authors (e.g., Dempsey et al., 2023; Goldberg and Ferreira, 2022). However, a closer look at the assumptions and predictions of the two frameworks reveals important differences. In contrast with the view that examples like (1) activate a potentially costly error correction mechanism, good-enough processing predicts that such sentences can often be processed with relatively

[1] Starting with Levy (2011), a different line of work has integrated a noisy context representation with surprisal (Hale, 2001; Levy, 2008a) as the main determinant of processing difficulty. The resulting model makes very different predictions from the original noisy-channel model, which I present in more detail in the general discussion.

little effort, assuming that the actual structure is simply ignored or at least heavily downweighted.

Good-enough processing incorporates Simon's (1955, 1956) concept of satisficing: it is assumed that readers do not fully optimize their interpretation processes in the sense that they take all available information into account, but may instead settle for an imperfect representation once their current *aspiration level* is reached (Christianson, 2016; Ferreira et al., 2009). When proper incentives to process utterances deeply and attentively are lacking, readers may save cognitive resources by partly ignoring the syntactic structure of a sentence and adopting a superficially plausible reading (Christianson et al., 2010; Ferreira, 2003), failing to revise initial misinterpretations in the face of disambiguating material (e.g., Christianson et al., 2001), or underspecifying their syntactic analysis of a sentence so that its semantic representation remains vague (Dwivedi, 2013; Swets et al., 2008; von der Malsburg & Vasishth, 2013).

It is clear that rational inference and good-enough processing serve different, potentially conflicting goals: to reconstruct the intended meaning of a sentence by invoking additional processes beyond the "normal" parsing of the literal string, or to save cognitive resources by omitting some parts of "normal" parsing. Given that both goals are plausible drivers of human reading (and listening) behavior, how can they be reconciled? One possibility is to move away from the focus on average behavior that dominates most of psycholinguistics (Yadav et al., 2022), and which may obscure a more complex reality: reconstructive and effort-saving mechanisms may be active to different degrees across different individuals, or within the same individual at different times, or even concurrently (Brehm et al., 2021). Some speakers may set relatively low aspiration levels and often go with their "gut feeling" when interpreting utterances, while others may expend mental effort to try and reconstruct what the other person probably wanted to say.

Importantly, there likely exists a third type of individual or processing mode that is rarely discussed in the literature: people who are pedantic — which I use neutrally here — in the sense that they are completely faithful to the linguistic stimulus, and who take sentences like (1) literally, responding with "That's nonsense!", "I don't know what you're trying to say!", or "What an unusual thing to happen!". Literal interpretations of implausible sentences are robustly attested (e.g., Ferreira, 2003; Gibson et al., 2013) and have been linked to high verbal working memory (Bader & Meng, 2018; Meng & Bader, 2021; Stella & Engelhardt, 2022). A completely input-faithful speaker would arguably be an embodiment of pure grammatical competence in the Chomskyan sense (Wray, 1998), which makes literal responses theoretically highly interesting. Any realistic model of sentence comprehension should thus take into account that different people or even one and the same person may, depending on the situation, be "inferencers", "slackers", or "pedants".

How can each of these processing "modes" be identified? One promising avenue is to use error awareness as an indicator. In the "slacker" mode of processing, ill-formed sentences should be very likely to pass unnoticed. Such cases are known as linguistic illusions, where an ungrammatical sentence passes as grammatical and/or an implausible sentence passes as plausible (e.g., Muller, 2022; Phillips et al., 2011; Sanford and Sturt, 2002). The rational inference account is underspecified with regard to error awareness: Levy (2008b, p. 237) states that a copy editor needs to "notice and (crucially) correct mistakes on the printed page" but also that "in many cases, these types of correction happen at a level that may be below consciousness — thus we sometimes miss a typo but interpret the sentences as it was intended" (see also Huang and Staub, 2021b).

That rational inference *can* be conscious is implied by studies that have used highly explicit tasks such as retyping of sentences (Ryskin et al., 2018) or judging how likely one sentence is to be changed into a different one due to a speech error (Zhang et al., 2023). There is also evidence

that error correction via rational inference leads to increased P600 amplitudes (Li & Ettinger, 2023; Ryskin et al., 2021), which has been linked to conscious detection of anomalies during reading (Coulson et al., 1998; Rohaut & Naccache, 2017; Sanford et al., 2011). Crucially, while the additional processing steps involved in rational inference may not *always* be conscious, they should be comparatively more likely to rise to consciousness than good-enough processing, which implies the *absence* of one or more processing steps. Finally, the "pedantic" processing mode naturally predicts error awareness, in the sense that the utterance is identified as being ungrammatical or nonsensical.

In addition to differences between people, there are likely to be differences between error types that create variability in how a speaker responds to a sentence (Frazier & Clifton, 2015). Some sentences, such as the "depth charge" sentence *No head injury is too trivial to be ignored* (Wason & Reich, 1979) cause an illusion of acceptability almost invariably across individuals (Paape et al., 2020), while other sentence types may show large amounts of variability both across and within speakers (Christianson et al., 2022; Frank et al., 2021; Goldshtein, 2021; Hannon & Daneman, 2004; Leivada, 2020). The aim of the present study is to quantify this variability between speakers and sentences across six different constructions that are known to cause linguistic illusions, and to identify traits that correlate with speakers' dispositions towards rational inference or good-enough processing. Table 1 lists the six constructions under investigation.

An additional contribution of the present work is the use of a computational modeling approach to investigate how responses to sentences are generated in real time. In the model presented below, rational inference, good-enough processing and outright rejection of a sentence are treated as latent processes that compete and trade off with each other to produce a response. In combination with the empirical breadth of the experimental design, this allows for the comparison of the three processes across different linguistic constructions. Furthermore, the model allows for a systematic investigation of participant-level traits that affect each of the latent processes.

I will now introduce the logic of the experiment, followed by the description of the procedure, and finally the implementation of the computational model.

Experimental study

The sentence judgment study presented below had three main aims:

1. To quantify the relative contributions of rational inference and good-enough processing to different linguistic illusions.
2. To investigate individual-level trade-offs between rational inference and good-enough processing across different illusions.
3. To investigate the effect of individual-level traits such as linguistic pedantry on rational inference, good-enough processing, and outright rejection of illusion sentences.

In order to achieve the first aim, the study used a novel two-dimensional sentence judgment task in which participants simultaneously judge whether they feel that they understand the sentences ("get it"/"don't get it"), in addition to whether they think that the sentences are formally correct ("correct"/"incorrect"). Under good-enough processing, readers should occasionally miss formal grammatical errors, resulting in the impression that illusion sentences are both well-formed and interpretable ("get it, correct"). Under rational inference, by contrast, it is plausible to assume that readers notice the errors — especially when instructed to look out for them — but can nevertheless reconstruct the intended sentence ("get it, but incorrect"). The remaining two judgment

Inversion: Order of direct and indirect object is swapped (e.g. Gibson et al., 2013) *The mother gave the candle the daughter.*
Agreement attraction: Verb agrees with intervening noun phrase instead of subject noun phrase (e.g., Bock et al., 2001) *The waitress who sat the girls unsurprisingly were unhappy about all the noise.*
Depth charge: Incongruous degree phrase "saved" by negation (e.g., Wason and Reich, 1979) *In Maria's class, no test is too difficult to fail.*
Comparative illusion: Number of individuals compared to number of events (e.g., Wellwood et al., 2018) *More engineers relocated to San Francisco than our accountant did.*
Missing VP illusion: Three clauses with three subjects but only two verbs (e.g., Gibson and Thomas, 1999) *The manuscript that the student who the catalog had confused was missing a page.*
NPI illusion: Negative polarity item *ever* licensed by embedded negation (e.g., Drenhaus et al., 2005) *The authors that no critics recommended have ever received acknowledgment for a novel.*

Table 1

Constructions used in the experimental study.

options ("don't get it, incorrect"/"don't get it, but correct") are not covered in any depth by either theory, but may nevertheless show differences between illusions: readers may outright reject some ungrammatical constructions more readily than others.

To achieve the second aim, I analyze the correlations between the subject-level random effects in a hierarchical computational model. The model assumes that a given manipulation has an average effect across all participants, and that the individual effects for each participant are normally distributed around this average. Analyzing the correlations between individual effects allows for statements of the form "Participants who are more likely than average to do good-enough processing for illusion X are also more likely than average to do good-enough processing for illusion Y". Furthermore, and perhaps more interestingly, correlations can be analyzed not only within but across response types: "Participants who are more likely than average to do good-enough processing for illusion X are less likely than average to engage in rational inferences for illusion Y". Finding such negative correlations would strengthen the case for shared cognitive mechanisms across different

illusions, and crucially also yield insights into how these mechanisms compete within individuals.[2]

The third aim is to uncover the underlying factors that contribute to individual differences in the processing of illusion sentences by collecting additional measures outside of the sentence judgment task. The first measure I use is a simple questionnaire that covers linguistic pedantry, interpretational charity, motivation, and attention. The second measure comes from a syllogistic reasoning task with believable and unbelievable syllogisms, which is intended to uncover individual differences in cognitive style (Stupple et al., 2011; Trippas et al., 2015, 2018): a more analytic cognitive style "denotes a propensity to set aside highly salient intuitions when engaging in problem solving" (Pennycook et al., 2012, p. 335). Having an analytic cognitive style should increase an individual's tendency to favor analytical grammar rules and logic over "quick and dirty", good-enough processing of illusion sentences.

Participants

Participants were tested in three groups that were recruited over Prolific (https://www.prolific.co; Palan and Schitter, 2018). Group 1 completed only the sentence judgment task, Group 2 additionally completed the questionnaire, and Group 3 additionally completed the questionnaire and the syllogistic reasoning task. Group 1 consisted of 100 self-identified native speakers of English currently living in the US. The first 50 participants were initially paid £2.83 each[3], which was adjusted to £3.55 after review, as the completion time estimate had been too low. The remaining 50 participants were paid £3.55 each. The data of one participant were subsequently removed because response times were consistently too short to be realistic, leaving data from 99 participants. Groups 2 and 3 consisted of 157 participants and 100 participants, respectively, recruited from the same subject pool. Participants in Group 2 were paid £3.55 each while participants in Group 3 were paid £5.65 each due to the longer experiment duration.

Materials

12 inversion sentences were adapted from Cai et al. (2022). Inversion sentences appeared in two conditions, the normal condition and the inverted condition, as shown in (2). In 6 items, both the inverted sentence and the control sentence used the direct object construction, while in the other 6 items both sentences used the prepositional object construction with *to*.

(2) The mother gave { a. the daughter the candle / b. the candle the daughter } before bedtime.

12 agreement attraction sentences were adapted from Parker and An (2018). Agreement attraction sentences appeared in three conditions, as shown in (3). All participants saw the *waitress*/*girls* (attraction) condition. In Group 1, 50 participants saw only the *waitress*/*girl* (ungrammatical) condition as a control, while the other 50 participants saw only the *waitresses*/*girls* (grammatical) condition as a control. Group 2 saw only the grammatical control condition, while Group 3 saw no agreement attraction sentences at all.

[2] See Brown (2021) for a related discussion of how random-effects correlations in linear mixed models can be leveraged to answer specific research questions.

[3] Prolific is a UK-based company, so renumeration is calculated in British Pounds.

(3) The { waitresses / waitress } who sat the { a. girls / b. girl } unsurprisingly were unhappy about all the noise.

12 depth charge sentences were adapted from O'Connor (2015). Depth charge sentences appeared in three conditions, as shown in (4). All participants saw the *no*/*too* (illusion) condition. In Group 1, 50 participants saw only the *no*/*so* condition (sensible) as a control, while the other 50 participants saw only the *some*/*too* condition (not sensible) as a control. The remaining participants saw only the *so* condition as a control.

(4) In Maria's class, ...

a. ... no test is so difficult that she would fail it.

b. ... no test is too difficult to fail.

c. ... some tests are too difficult to fail.

The depth charge illusion is often difficult to spot even for trained linguists (Wason & Reich, 1979). The crucial insight is that the degree phrase *too difficult to fail* is pragmatically incongruous (compare *too easy to fail*). The illusory property of this construction is that having a negation at the beginning of the sentence (*no test*) can completely mask the incongruity. However, in logical terms, asserting that no test has the property of being *too difficult to fail* should not repair the incongruity (Paape et al., 2020).

12 comparative illusion sentences were adapted from O'Connor (2015). Comparative illusion sentences appeared in two conditions, the plural condition (grammatical), and the singular condition (ungrammatical), as shown in (5). All comparative illusion sentences contained a sentence-final subordinate clause (*because* ... ; *due to* ...) to make them sound more natural.

(5) Last fall, more engineers relocated to San Francisco than { a. accountants / b. our accountant } did, ...

12 missing VP sentences were taken from Langsford et al. (2019). Missing VP sentences appeared in two conditions, the VP2 condition (grammatical) and the no-VP2 condition (ungrammatical), as shown in (6).

(6) The ancient manuscript that the grad student who the new card catalog had confused a great deal { a. was studying in the library / b. —— } was missing a page.

12 NPI sentences were adapted from Muller (2022). NPI sentences appeared in three conditions, as shown in (7). All participants saw the *the authors*/*no critics* (embedded negation) condition. 50 participants saw only the *no authors*/*the critics* condition (grammatical) as a control, while the other 50 participants saw only the *the authors*/*the critics* condition (ungrammatical) as a control. The remaining participants saw only the grammatical control condition.

(7) { No authors / The authors } that { a. no critics / b. the critics } recommended have ever received acknowledgment for a best-selling novel.

In addition to the 72 illusion and control sentences, there were also 24 fillers. Of these, 12 were garden-path sentences (e.g., *The farm hand believed that while the fox stalked the geese continued to peck* . . .), 6 contained "malaphors" or "idiom blends" (e.g., *Bill wasn't the sharpest bulb in the box*), and 6 were relatively long but well-formed sentences of different types.

Group 2 completed the same judgment task as Group 1, followed by an additional questionnaire that appeared at the end of the experiment. Participants were asked to what extent they agreed or disagreed with the following four statements:

1. Doing the experiment was fun for me.
2. It bothers me a lot when people use incorrect grammar.
3. I usually assume that what people say makes sense.
4. In my everyday life, when I read a text, I always pay close attention to every sentence.

There were five levels on the scale: "completely disagree", "somewhat disagree", "undecided", "somewhat agree", "completely agree".

Group 3 completed a syllogistic reasoning task in addition to the judgment task and the questionnaire. The materials consisted of 64 syllogisms in four conditions resulting from crossing the factors validity (valid versus invalid) and believability (believable versus unbelievable). These factors were manipulated between items, that is, each syllogism appeared in only one condition. Examples for each condition are shown in (8). The structure of the syllogisms varied between items. Some materials were novel while others were adapted from previous studies (Goel & Vartanian, 2011; Hayes et al., 2022; Solcz, 2011).

(8) a. Either the sky is blue or it is green. The sky is not green. Therefore, the sky must be blue. (valid, believable)

b. All rabbits are fluffy. All fluffy creatures are tadpoles. Therefore, all rabbits are tadpoles. (valid, not believable)

c. If an animal is a feline, then it purrs. If an animal purrs, then it is a cat. Therefore, if an animal is a cat, then it is a feline. (invalid, believable)

d. Some sodas are beverages. All sodas are carbonated drinks. Therefore, some carbonated drinks are not beverages. (invalid, not believable)

Procedure

All subjects gave informed consent to participate in the study, which was run on the PCIbex farm (Schwarz & Zehr, 2021). Given the potential importance of task demands for the results, the experimental instructions are reproduced here in their entirety[4]:

[4]The speech error examples are taken from Frazier (2015).

> People often make mistakes when they speak or write. They will say things like "is sufficient enough for", "pales next to comparison of", or even "I'm going to get some bed". Such utterances can be considered incorrect or nonsensical, but it is nevertheless clear what the person in question was trying to say. In this study, you are supposed to judge both the **formal correctness** of the sentences you will read, as well as indicate whether you know what the other person **meant**.
>
> Some of the sentences will be very complex, and you may feel that you don't understand them, but still get the impression that they are formally correct. Sometimes you may be completely unsure. This is fine; you can just choose the appropriate answer option ("no idea").
>
> Don't "overanalyze" the sentences — try to read normally as much as possible. There will be no detailed comprehension test. We are mainly interested in whether you **feel** that you understood the sentence.

In each trial, the stimulus sentence was presented at the center of the screen, with five possible judgment options shown directly below:

1. 😊👍 Get it, correct
2. 😊👎 Get it, but incorrect
3. 😕👍 Don't get it, (probably) correct
4. 😕👎 Don't get it, incorrect
5. ❓❓ No idea

The time elapsed from presentation of the sentence to choosing a response was recorded for each trial. There was no time limit. Responses could be chosen by either clicking on them or by pressing the appropriate number key. Sentences were rotated through the conditions in a Latin squares fashion. There were no practice trials. The median duration of an experimental session in Group 1 was 20 minutes. The median duration of an experimental session for Group 2 was 21 minutes.

In Group 3, the order of the sentence judgment task and the syllogistic reasoning task was counterbalanced across participants, so that 50 participants completed the reasoning task first and the remaining 50 participants completed the judgment task first. In the syllogistic reasoning task, participants were instructed to judge the syllogisms purely based on logic ("logically valid" versus "not logically valid"), and to assume that the premises were true, even if they didn't make sense. As for the judgment task, the entire syllogism was presented on the screen, with the response options shown underneath. At the end of the experiment, the same questionnaire as in Group 2 was administered. The median duration of an experimental session in Group 3 was 33 minutes.

Data preparation

The data were analyzed in R (R Core Team, 2022). The reaction time and response data from all three participant groups were combined into one data set. Trials with response times below 2 seconds or above 60 seconds were dropped, which resulted in a loss of 5% of the data. Additionally, trials with "no idea" responses were dropped, which resulted in a loss of 0.5% of the remaining data.[5] The final data set contained 22,884 observations from 355 participants.

[5] This step resulted in the complete removal of data from one subject, who always responded with "no idea".

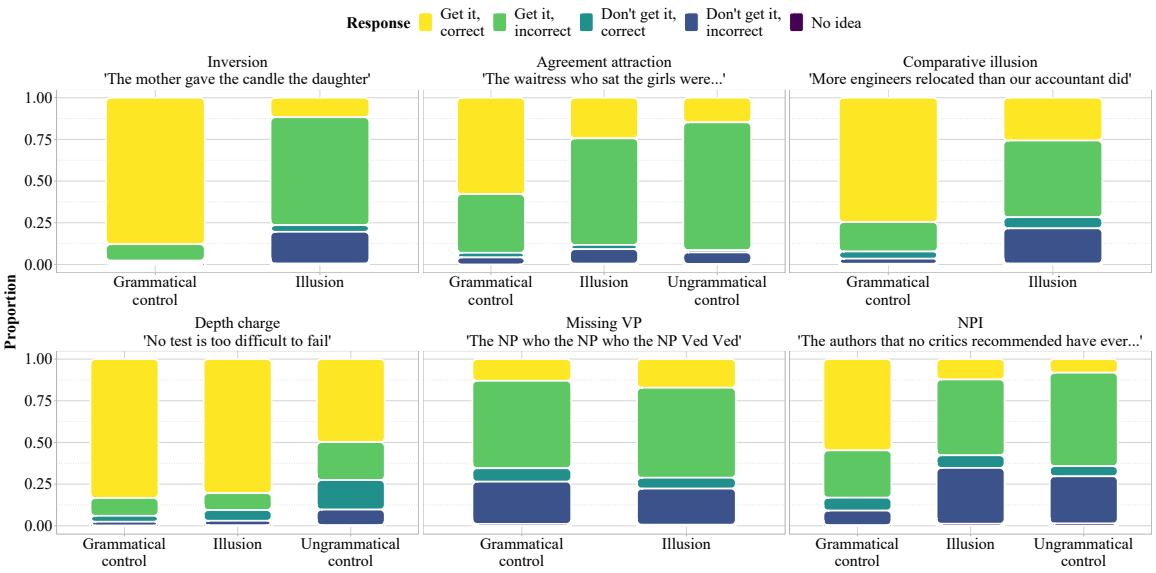

Figure 1

Response proportions across constructions and conditions.

Descriptive results

Figure 1 shows response proportions across constructions and conditions. Figure 2 shows judgment times (reading time + response selection time) across constructions, conditions, and response types.

A number of high-level observations can be made about the descriptive results:

- Across constructions, "get it, incorrect" judgments are more common in the illusion conditions than "get it, correct" judgments, suggesting that rational inference is more common under the current instructions than good-enough processing. The only exception is the depth charge construction, which shows more than 75% "get it, correct" judgments.
- Across constructions, even the fully grammatical control sentences show non-zero proportions of "incorrect" judgments. This tendency is especially pronounced for the agreement attraction sentences of Parker and An (2018)[6] and the NPI sentences of Muller (2022).
- It is informative to compare the illusion conditions to both grammatical and ungrammatical control conditions. For agreement attraction sentences, the judgment pattern in the illusion condition is close to midway between the grammatical and ungrammatical conditions in terms of "correct" judgments, while for NPI sentences, the illusion condition is much closer to the ungrammatical condition. This is in line with the qualitative pattern observed by Xiang et al. (2013) and Langsford et al. (2019), who also compared the two constructions.

[6] Speculatively, the high rejection rates may be due to the placement of the adverb between the attractor and the verb (*The waitress who sat the girls unsurprisingly were . . .*), which is anecdotally regarded as incorrect by some speakers.

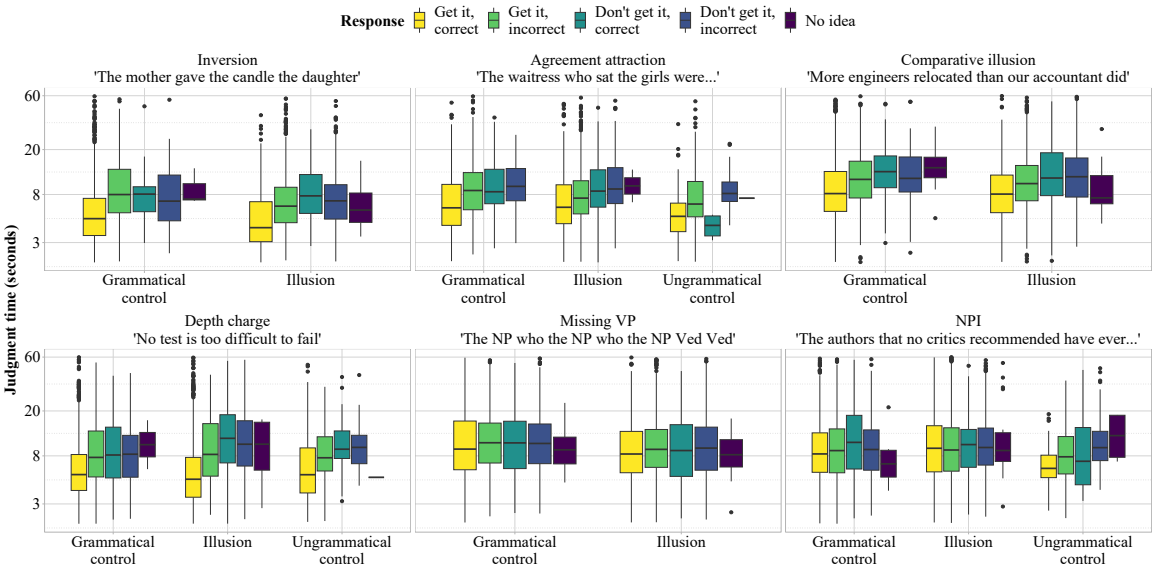

Figure 2

Judgment times (reading time + response selection time) across constructions, conditions, and response types. Note that the y-axis is log-scaled.

- For the depth charge illusion, the grammatical condition with *so* is almost indistinguishable from the illusion condition with *too*, which essentially patterns like a fully grammatical sentence. This result is unexpected if the depth charge illusion is the result of a rational inference mechanism, as recently argued by Zhang et al. (2023).
- Across constructions, latencies tend to be shorter for "get it, correct" judgments, as would be expected if the sentence is either grammatical or an ungrammatical sentence is processed superficially. However, this pattern does not seem to hold for missing VP and NPI sentences, which casts doubt on the assumption that positive grammaticality judgments for illusion sentences are always the result of effort-saving mechanisms.

Computational modeling

Preprocessing of individual differences predictors

Model-based preprocessing was applied to the individual differences measures from Groups 2 and 3 prior entering them into the computational model. The questionnaire responses were subjected to a principal components analysis with polychoric correlations using the psych package (Revelle, 2023). The number of factors was set to 2. The resulting factor loadings are shown in Table 1. As the table shows, Factor 1 loaded positively on having fun during the experiment, being bothered by grammatical errors, and paying attention during reading, but loaded negatively on assuming that utterances typically make sense. Factor 2, by contrast, loaded positively on making

this default assumption of sensibleness, but less positively on attention, and negatively on being bothered by grammatical errors.

	Factor 1 (PEDANTRY)	Factor 2 (CHARITY)
had fun	0.62	0.58
bothered by incorrect grammar	0.66	−0.46
assume that people make sense	−0.16	0.82
attentive reading	0.84	0.09

Table 2

Factor loadings from the principal components analysis of the questionnaire data.

Given these loadings, I will assume that Factor 1 captures motivation to rigorously apply grammatical rules and pay attention to linguistic detail, which I will call PEDANTRY, while Factor 2 captures the default assumption that sentences are sensible without much regard for correct grammar, which I will call CHARITY. If this interpretation is on the right track, individuals who score highly on PEDANTRY should show less good-enough processing for illusion sentences, meaning fewer "correct" responses, while individuals who score highly on CHARITY should give more "correct" responses and fewer "don't get it" responses. For both factors, the factor scores for each participant were entered as scaled, centered predictors into the main analysis.

The data from the syllogistic reasoning task completed by Group 3 were analyzed using a hierarchical logistic regression model in brms (Bürkner, 2017). The factors validity and believability were sum-coded for this analysis. Response time was also entered as a predictor to account for speed-accuracy trade-off. The subject-level random effects for validity, believability and their interaction were extracted from this model and entered into the main analysis as centered, scaled predictors. I will call these predictors LOGIC, BELIEF, and CONFLICT.

Model implementation

In psycholinguistics, so-called "online" measures such as reading times are usually analyzed separately from "offline" measures such as acceptability ratings. In cognitive psychology, by contrast, the standard is to look at the measured latencies and observed responses in a task as reflecting the same mental process, which is often conceptualized in terms of evidence accumulation or sequential sampling (Evans & Wagenmakers, 2019; Ratcliff et al., 2016). The core assumption of evidence accumulation models is that the time spent processing a given stimulus — modulo the time required to, say, press a keyboard key — reflects the time needed to extract enough information from the stimulus to be able to respond to it. In the context of a sentence judgment study, one can assume that while the participant reads a given sentence, they are extracting information that is relevant to the task of making the judgment.

A relatively simple type of of evidence accumulation model is the lognormal race model (Rouder et al., 2015). In the lognormal race model, the different response options independently accrue evidence, with the fastest option winning and determining the observed response in a given trial. Under this model assumption, each trial also yields information about the unobserved responses, given that they must have been slower than the observed response.

I implement the lognormal race model in Stan (Stan Development Team, 2023) and fit it to the judgments and their associated reading/judgment times. The two types of "don't get it" responses ("correct"/"incorrect") are coded as a single response category, so that there are three accumulators whose finishing times FT in a given trial i are each sampled from a lognormal distribution with mean μ and standard deviation σ:

$$\begin{aligned} FT\_\text{REJECT}_i &\sim Lognormal(\mu_1, \sigma_1) && \text{(Response "Don't get it")} \\ FT\_\text{INFER}_i &\sim Lognormal(\mu_2, \sigma_2) && \text{(Response "Get it, but incorrect")} \\ FT\_\text{GOOD}_i &\sim Lognormal(\mu_3, \sigma_3) && \text{(Response "Get it, correct")} \end{aligned} \tag{9}$$

The FT_REJECT accumulator represents complete failure to understand the sentence, the INFER accumulator represents rational inference processes, and the GOOD accumulator represents good-enough processing. The observed reaction time in a given trial i is the finishing time of the winning accumulator, plus a shift parameter estimated from the data that represents non-decision time, which can vary between subjects.

$$RT_i = min(FT\_\text{REJECT}_i, FT\_\text{INFER}_i, FT\_\text{GOOD}_i) + \text{SHIFT} \tag{10}$$

The μ parameter of each accumulator is assumed to be affected by the experimental manipulations as well as by individual differences between participants. The mathematical structure of μ_x for each accumulator x in trial i is that of a linear mixed-effects model with an intercept α, slopes β for the predictors $1...n$, and normally distributed by-subject and by-item adjustments u and w for both intercepts and slopes. Predictors include the factor-coded conditions within each construction type, the individual differences measures, and their interactions.

$$\mu_{x,i} = \alpha_x + u_i + w_i + \beta_{n,x,i} \cdot \text{predictors} \tag{11}$$

IF...ELSE constructions in Stan were used to exclude the individual differences predictors where no data was available, that is, for the syllogistic reasoning predictors in Groups 2 and 3, and for the questionnaire-based predictors in Group 1. To account for differences in sentence length between constructions, sentence length in characters was added as a predictor to each μ.

Given the assumptions of the lognormal race model, if one accumulator becomes faster in a given condition, there will be more responses of the associated type, and fewer responses in the other response categories. This is true even if the finishing times of the other accumulators are unaffected by the manipulation, because, like in a real-life race, it is the *relative* speed of the competitors that determines the winner. Thus, if a given manipulation speeds up the INFER accumulator, it does not follow that the GOOD and REJECT accumulators are slowed down by the same amount, or even slowed down at all.

I will call situations in which one accumulator is affected by a manipulation but the other accumulators remain unaffected *passive trade-offs*. Perhaps of more theoretical interest are what I will call *active trade-offs*: cases in which a speedup or slowdown on one accumulator is accompanied by the reverse effect on another accumulator. For instance, it might be that rational inference as represented by the INFER accumulator trades off in this way with good-enough processing as

represented by the GOOD accumulator, so that faster evidence accumulation for one latent process results in slower evidence accumulation for the other one. Whether such active trade-offs exist can be determined by jointly fitting the model to the response and latency data.

Modeling results and discussion

Population-level effects

Figure 3 shows the predicted finishing times of the three accumulators in the lognormal race model across constructions and conditions. The figure allows for several types of visual comparisons: each panel shows one construction, while the outline colors show the different conditions (illusion versus control). For each accumulator (GOOD, INFER, REJECT), the conditions can be compared, with *shorter* finishing times meaning *more* responses of that type. In addition, finishing times can be compared *across* constructions by visually comparing two panels. For instance, finishing times for the GOOD accumulator in the grammatical control condition (green) are faster for inversion sentences than for comparative illusion sentences, so that more and faster "get it, correct" judgments are predicted for that condition, in line with the descriptive results.

Figure 4 shows highest density intervals for the slope parameters (differences between conditions) across constructions.[7] Across all constructions, the estimated mean finishing times of the accumulators are slower than the empirically observed judgment times, which follows from the model assumptions: in trials where a given response is *not* observed, it is assumed that the associated accumulator was slower than the observed response time.

In general, the parameter estimates and predicted finishing times align well with the descriptive results: the inversion construction shows the largest difference in judgments between the illusion and control conditions, followed by the comparative illusion and the NPI illusion. The agreement attraction effect is somewhat smaller, while for depth charge sentences and missing VP sentences the illusion and grammatical control conditions are almost indistinguishable, though the REJECT accumulator does show some graded sensitivity to the depth charge manipulation.

Most constructions show active trade-offs between the accumulators: the inversion manipulation slows down GOOD while simultaneously speeding up INFER and REJECT, as do the comparative illusion manipulation, the agreement attraction manipulation, and the NPI illusion manipulation. The presence of these active trade-offs is informed by the latency data: based only on the responses, it would be unclear if one process dominates the response pattern because it is strengthened by a manipulation, or because the competing processes are weakened.

REJECT is the slowest accumulator on average, being [11 s, 26 s] (95% highest density interval) slower than GOOD across all constructions. Across illusion and control sentences, INFER accumulates evidence more slowly than GOOD by [0.2 s, 9 s] on average. In terms of variability, the opposite picture emerges: REJECT shows the lowest amount of variability in finishing times at [11 s, 12 s], followed by INFER [14 s, 15 s], and finally GOOD [20 s, 23 s] with the highest amount of variability. The GOOD accumulator being fastest but less consistent in its speed of evidence accumulation than the other accumulators is in line with the basic assumptions of the good-enough processing framework, which assumes that readers create imperfect, "quick and dirty" sentence representations in some proportion of trials.

[7] The slope estimates for ungrammatical agreement attraction sentences, ungrammatical depth charge sentences, and ungrammatical NPI sentences are notably wider than for the other constructions, given that these sentence types were only used for one half of Group 1.

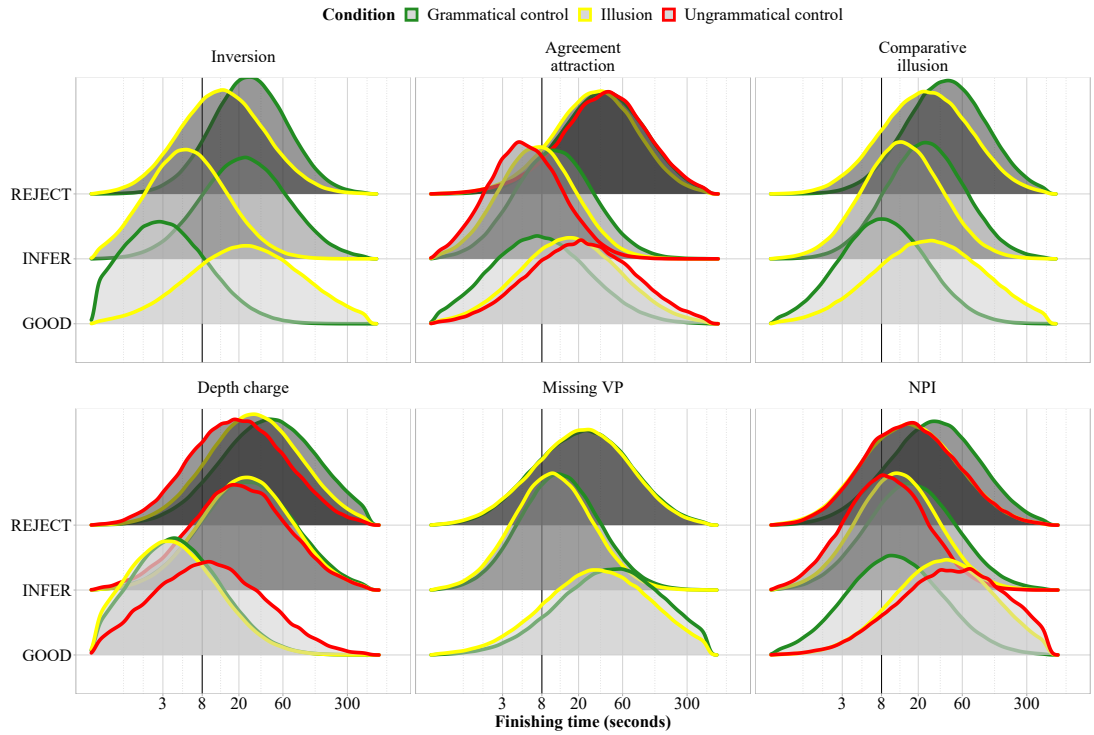

Figure 3

Posterior predictive distributions of finishing times (250 samples) of the three accumulators across constructions and conditions. Faster finishing times correspond to a higher expected number of responses of the respective type. Finishing times more than ± 2.5 SD away from the log mean have been removed. Reference line added at 8 seconds. Note that the x-axis is log-scaled.

Increasing sentence length in characters by one standard deviation slows down REJECT by [0.2 s, 4.5 s], and GOOD by [3.8 s, 10.1 s], while the effect on INFER crosses zero: [−0.9 s, 2.4 s]. This pattern is somewhat unexpected, given that longer sentences should, in principle, offer more opportunities for errors and "repairs", but recall that there is a passive trade-off between the accumulators: REJECT and GOOD being slowed down in longer sentences results in more trials in which INFER wins, thus predicting more "get it, incorrect" judgments compared to shorter sentences.

In Experiment 3, receiving the logic task prior to the sentence judgment task as opposed to the other way around also affected the speed of the accumulators: doing the logic task first speeds up REJECT by [−24.2 s, −3.2 s] and GOOD by [−7.2 s, −3.9 s]. The order effect on INFER crosses zero: [−6.5 s, 2.8 s]. Due to trade-off between the accumulators, this means that there were fewer "get it, incorrect" answers but more "don't get it" and "get it, correct" answers when the logic task was completed first. The observed pattern is in line with participants being mentally exhausted after completing the logic task, and preferring to either completely reject sentences or to uncritically accept them rather than engaging in effortful rational inference processes.

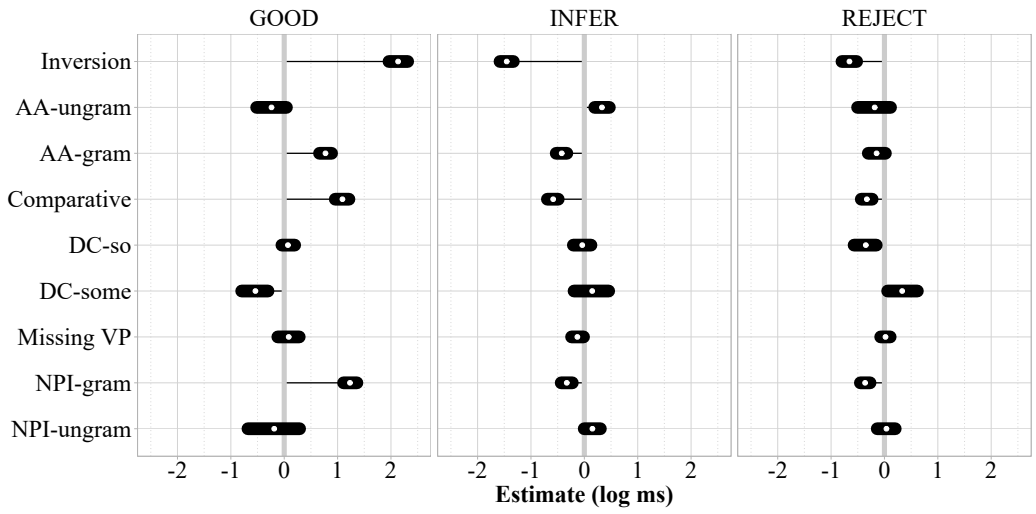

Figure 4

95% highest density intervals of slope parameters across all constructions on the log ms scale. Across all constructions, the slope is the difference between the illusion condition and the indicated control condition(s). Positive estimates correspond to a slowdown of the accumulator, that is, fewer responses of the associated category in the illusion condition. Reference line added at 0 log ms.

Individual-level trade-offs

Figure 5 shows the correlation estimates between subject-level random effects for which the probability of direction is above 0.95, that is, for which the correlation estimate is mostly positive or negative.[8] The correlations of subject-level random effects show some active trade-offs at the individual level within constructions, but crucially also across constructions.

As the figure shows, participants who show larger-than-average effects of the inversion illusion on the GOOD accumulator tend to show smaller-than-average effects on the INFER accumulator for the same manipulation. This negative correlation suggests that participants who tend to do good-enough processing in the inversion construction ("slackers") are less likely to engage in rational inference. Such a trade-off is expected under the assumption that good-enough processing and rational inference serve different, conflicting goals: to conserve mental energy or to expend additional energy to infer meaning. The pattern of trade-offs holds across all correlations shown in Figure 5: larger participant-level effects on the GOOD accumulator correlate with smaller participant-level effects on the INFER accumulator. Importantly, all correlation estimates that reached the 0.95 probability-of-direction cutoff include the inversion construction, likely because the average effects in this construction are the largest. The inversion construction is thus a good candidate for a stable indicator of individual differences in illusion processing: an individual's processing preferences for inversion sentences can be used to predict the same individual's processing preferences for compar-

[8]Note that this criterion does not correspond to a frequentist test of significance, but merely singles out effects for which the parameter estimates mostly point in a given direction.

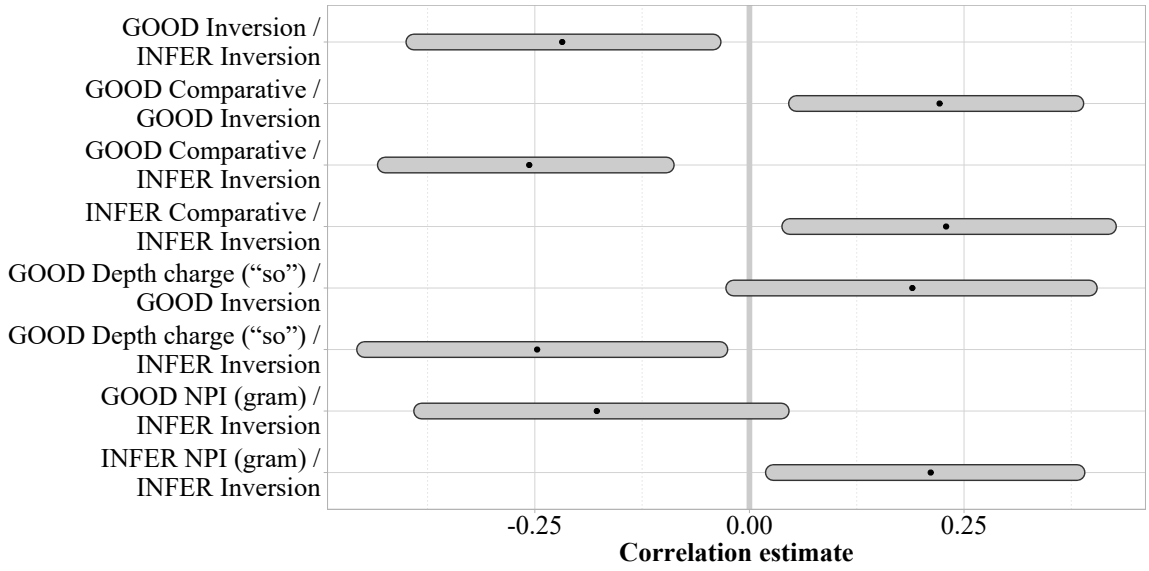

Figure 5

Correlation estimates and associated 95% highest density intervals of subject-level random effects. Only correlations with probability of direction > 0.95 for slope parameters (differences between conditions) are shown.

ative illusion sentences, depth charge sentences, and NPI illusion sentences.

Individual differences measures

In terms of interactions with the individual differences measures, the main question of interest is whether the difference between the illusion and control conditions for a particular constructions varies according to a participant's self-reported pedantry and interpretational charity, and/or according to their logical reasoning ability. Due to the large number of possible interactions across constructions and accumulators, I will report only the results for constructions in which at least one interaction parameter reached probability of direction > 0.95. Across all predictors, I plot the finishing time distributions of the three evidence accumulators for the 20 highest- and lowest-scoring participants against each other.

Pedantry. Figure 6 shows interactions with the PEDANTRY predictor, which is thought to reflect a participant's motivation to rigorously apply grammatical rules. PEDANTRY effects are seen in the agreement attraction, depth charge, and missing VP constructions. Contrary to my prediction, the interactions mainly affect the REJECT accumulator rather than the GOOD accumulator. For agreement attraction and missing VP sentences, high-PEDANTRY participants show faster finishing times for the REJECT accumulator compared to the control condition, that is, more "don't get it" responses. This pattern is still broadly in line with the assumption that PEDANTRY captures the strict application of grammatical rules in ungrammatical illusion sentences: pedantic individuals tend to give "don't get it" responses if the grammar does not license an interpretation, as opposed to

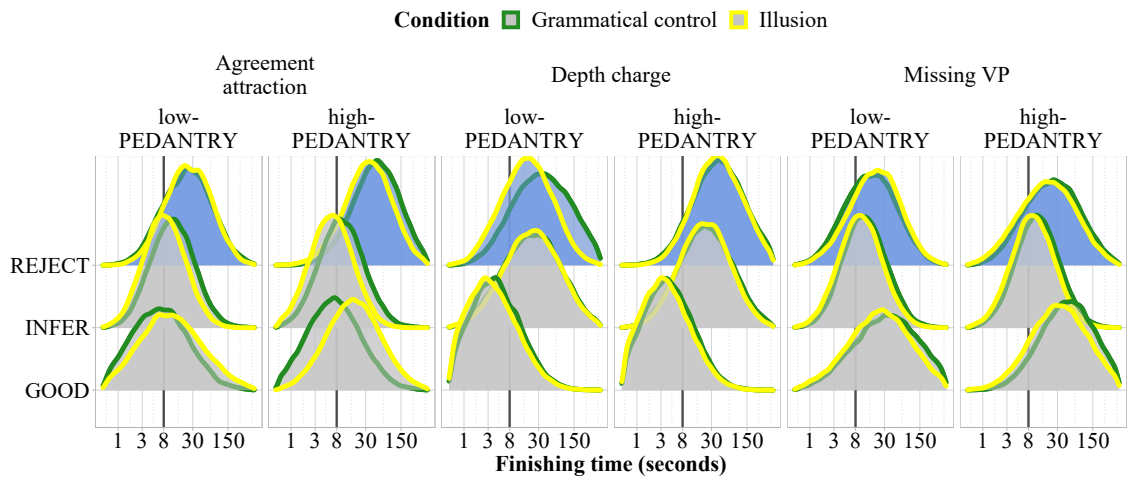

Figure 6

Posterior predictive distributions of finishing times (250 samples) of the three accumulators across constructions and conditions, by PEDANTRY *score. Interactions with probability of direction > 0.95 are highlighted in blue.*

ignoring errors or drawing inferences beyond the literal input. For depth charge sentences, however, the pattern is reversed: low-PEDANTRY participants tend to respond "I don't get it" more often in the illusion condition. I discuss this finding below.

Charity. Figure 7 shows interactions with the CHARITY predictor, which is thought to reflect a participant's default assumption that sentences are formally correct and sensible. Interactions with CHARITY are seen for inversion, agreement attraction, and missing VP sentences. As a general pattern, high-CHARITY participants tend to distinguish less between illusion and control sentences for these constructions, which is plausible given the interpretation of the predictor. However, there are differences between the constructions with regard to which accumulator is affected: for inversion illusion and agreement attraction sentences, high CHARITY leads to more "get it, correct" judgments (good-enough processing), while for missing VP sentences, *low* CHARITY leads to more "get it, incorrect" judgments (rational inference) for illusion compared to control sentences; this suggests that conscious rational inference in missing VP sentences may require *less* charitable assumptions about sentences usually being correct and sensible.

Moving on to the predictors derived from the syllogistic reasoning task, recall that there are three individual-differences measures: the main effect of logical validity (LOGIC), the main effect of believability (BELIEF), and the validity $\times$ believability interaction, which is commonly interpreted to signal a conflict between rule-based reasoning and intuition (CONFLICT). No effects reached probability of direction > 0.95 for BELIEF, so I will focus on the other two predictors.

Logic. Figure 8 shows effects of the LOGIC predictor for inversion sentences, depth charge sentences, missing VP sentences, and NPI sentences. Similarly to the PEDANTRY predictor, the effect of LOGIC is such that high-LOGIC participants distinguish more strongly between grammatical and illusion sentences for these constructions, which mainly affects the GOOD accumulator. Depth

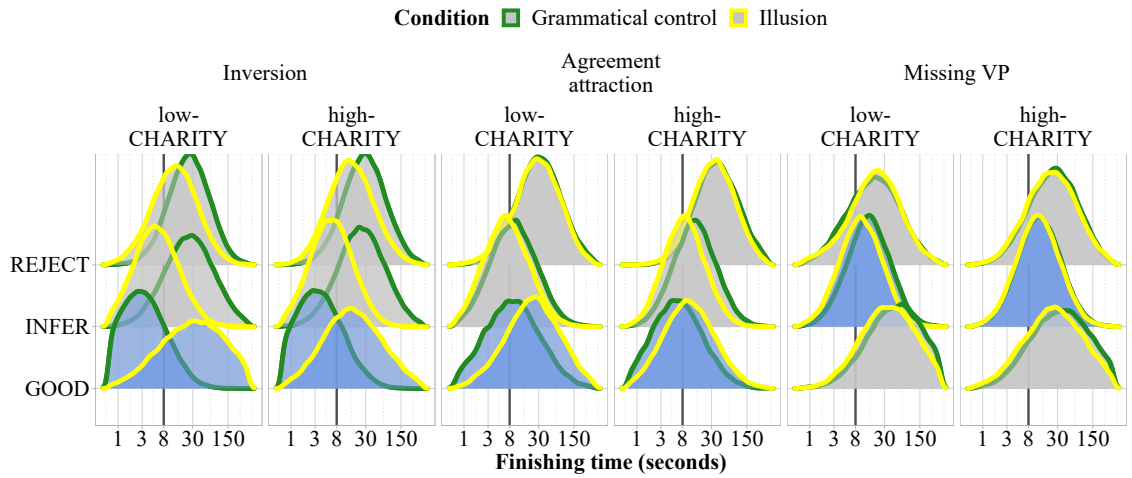

Figure 7

Posterior predictive distributions of finishing times (250 samples) of the three accumulators across constructions and conditions, by CHARITY *score. Interactions with probability of direction > 0.95 are highlighted in blue*

charge sentences are unique in that LOGIC mainly affects "don't get it" responses: high-LOGIC participants tend to reject illusion sentences of this type more often than control sentences. Overall, the results are in line with the assumption that individuals with strong logical abilities are less likely to process sentences superficially, and are thus more likely to spot grammatical errors, and/or, in the case of depth charge sentences, fail to understand the sentence when the compositional meaning is nonsensical.

Logic-belief conflict. The CONFLICT predictor affects comparative illusion, depth charge, and NPI sentences, as shown in Figure 9. For comparative illusion and depth charge sentences, low-CONFLICT participants distinguish more between illusion and control sentences; this difference is visible in rational inferences ("get it, incorrect") for comparative illusion sentences, but in rejections ("don't get it") for depth charge sentences. This pattern is broadly in line with the assumption that CONFLICT measures the amount to which intuition interferes with rule-based processing. The pattern in NPI sentences, however, is unexpected: high-CONFLICT participants show a slightly larger difference between conditions on the INFER accumulator than low-CONFLICT participants.

To summarize, pedantic individuals and individuals who consistently apply logical rules in the face of distracting believability information tend to experience fewer acceptability illusions for some constructions, presumably because they tend to stick to closely to prescriptive grammar when evaluating meaning. By contrast, individuals who make charitable assumptions about sentence interpretability and formal correctness appear to tend more towards good-enough processing, at least for some illusions.

In addition to this general plausible pattern of results, there were also two unexpected findings: low-PEDANTRY participants rejected depth charge illusion sentences (*No test is too difficult to fail*) more often than high-PEDANTRY participants, and high-CONFLICT participants made more

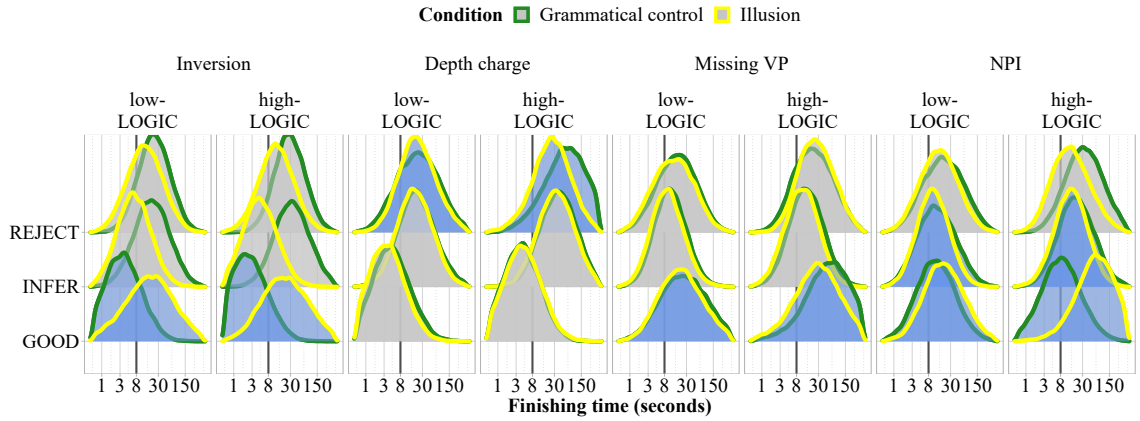

Figure 8

Posterior predictive distributions of finishing times (250 samples) of the three accumulators across constructions and conditions, by LOGIC *score. Interactions with probability of direction > 0.95 are highlighted in blue.*

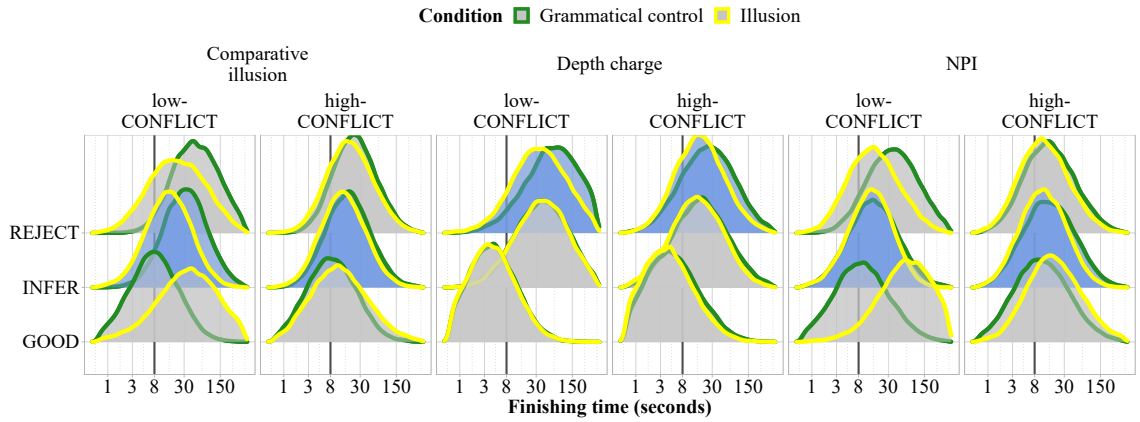

Figure 9

Posterior predictive distributions of finishing times (250 samples) of the three accumulators across constructions and conditions, by CONFLICT *score. Interactions with probability of direction > 0.95 are highlighted in blue.*

rational inferences for NPI illusion sentences than low-CONFLICT participants. In this context, recall that PEDANTRY not only includes disposition but also motivation: pedantic individuals invest energy into finding errors. Speculatively, low-PEDANTRY participants, who are less motivated, may reject depth charge sentences in the illusion condition because they are semantically more complex than the sentences in the control condition (*No test is so difficult that she would fail it*). It has been suggested that depth charge sentences are so complex that it may be impossible to parse them correctly even for highly motivated participants (Wason & Reich, 1979), which would explain why low-PEDANTRY participants reject them while high-PEDANTRY participants show no difference between the illusion and control conditions. Regarding the unexpected interaction between CONFLICT and rational inference in NPI sentences, it may be related to the fact that both NPI licensing and syllogistic reasoning have been linked to extralinguistic pragmatic processing (Tessler et al., 2022; Xiang et al., 2013). Speculatively, high-CONFLICT individuals may be more sensitive to the pragmatic licensing conditions on NPIs, and thus better able to draw inferences beyond the literal input. Alternatively, due to the large number of parameters in the model, the unexpected results may simply be false positives. This possibility can be addressed in future replication studies.

General discussion

The main aim of the current study was to investigate trade-offs between rational inference mechanisms and good-enough processing. To study these trade-offs, I collected data from a large participant sample on different sentence types that are known to cause linguistic illusions: sentences with argument inversions (*The mother gave the candle the daughter*), agreement attraction sentences, depth charge sentences (*No test is too difficult to fail*), comparative illusion sentences, missing VP sentences, and NPI illusion sentences. The current study is, to my knowledge, the only one to date that tested all of these constructions with the same sample of participants, that used combined judgments of grammaticality and interpretability, and that additionally collected individual differences measures. The study of Langsford et al. (2019) was comparable in terms of sample size and empirical breadth, but didn't cover individual differences. In addition, the data from the current study are complemented by a computational cognitive model that jointly accounts for judgments and judgment latencies, and that can dissociate the underlying latent processes.

There is already a wealth of work that has explored specific illusions in depth by using targeted manipulations of sentence structure and lexical content, as well as grammatical differences between languages (e.g., Bhatia and Dillon, 2022; Frank and Ernst, 2019; O'Connor, 2015; Orth et al., 2021; Wellwood et al., 2018). This kind of work is highly valuable, but introduces the risk of overfitting theories to particular constructions instead of aiming to develop a "theory of everything". Studies that compare different illusions and try to identify possible shared mechanisms, on the other hand, are comparatively rare (e.g., Brehm et al., 2021; Langsford et al., 2019; Parker and Phillips, 2016). The current study aimed to further broaden the empirical picture by presenting six illusion types to the same participants, and by testing the predictions of two accounts that are general enough to cover a wide range of illusions.

As discussed in the introduction, the rational inference approach and the good-enough processing approach are different both in spirit and in their predictions. The rational inference approach in the version originally proposed by Levy (2008b) posits that readers can mentally correct errors in sentences they read, which can result in non-literal interpretations that mainly rely on prior expectations about plausible meanings (e.g., Cai et al., 2022; Gibson et al., 2013). By contrast, good-enough processing posits that readers sometimes process sentences superficially, which can

result in input information being ignored (e.g., Ferreira, 2003). Crucially, under both rational inference and good-enough processing, readers should be under the impression that they *understood* the meaning of the sentence. The predictions of the two accounts only start to diverge when the precise nature of the rational error-correction mechanism is taken into account: given an appropriate task, such as judging the formal correctness of sentences, it is likely that error corrections will rise to consciousness (Levy, 2008b; Ryskin et al., 2018). By contrast, an obvious corollary of the good-enough processing assumption is that sentence anomalies are sometimes missed when processing is shallow (Christianson, 2008; Frazier & Clifton, 2015; Karimi & Ferreira, 2016; Sanford & Sturt, 2002), even when readers/listeners are specifically instructed to monitor for errors or give acceptability ratings (Erickson & Mattson, 1981; Paape et al., 2020; Sanford et al., 2011).

Two-dimensional judgments, competing latent processes, and individual differences

In the current experiments, readers were asked to simultaneously judge the formal correctness and interpretability of linguistic illusion sentences and control sentences. The relevant linking assumption is that if processing is shallow, readers should not even notice that an "inversion" sentence like *The mother gave the candle the daughter* is implausible, and thus give "get it, correct" judgments. By contrast, if readers engage in rational error correction, and if error correction is conscious given the explicit task demands, they should give "get it, incorrect" judgments. Finally, if readers are pedantic about grammar, and assume that grammar is the only mediator between input and meaning, they may notice the error but refrain from trying to infer the (presumably) intended interpretation, responding with "I don't get it".

Across the six illusions tested, results showed a mixture of the three response types, with "get it, incorrect" judgments dominating in the illusion conditions, consistent with rational inference under task demands that favor conscious error detection. The notable exception to this pattern were depth charge sentences such as *No test is too difficult to fail*, which were judged as formally correct in about 75% of trials. This finding suggests that the depth charge illusion is likely not due to rational error correction, as recently proposed by Zhang et al. (2023), but either caused by superficial processing (Paape et al., 2020; Wason & Reich, 1979) or even by (partial) grammaticalization of the anomalous structure (Cook & Stevenson, 2010; Fortuin, 2014).

In order to further dissociate the latent processes involved in the processing of linguistic illusions, I fitted the lognormal race model of Rouder et al. (2015) to the judgment and latency data.[9] The lognormal race model assumes that while the stimulus — in this case, the sentence — is processed, the different response options accumulate evidence in parallel. The first option that accumulates enough evidence to pass a threshold determines the observed response. Under this model, rational inference, good-enough processing, and outright rejection of the sentence are seen as parallel, competing processes within the same individual in each trial. Given this perspective, it is also important to ask to what extent underlying individual differences between participants contribute to different response patterns. To this end, two of the three participant groups completed a questionnaire in which participants indicated how motivated they had been throughout the experiment, how strongly grammatical mistakes bothered them (pedantry), and to what extent they tended to assume by default that utterances make sense (charity). The third group of participants additionally included a syllogistic reasoning task with believable and unbelievable conclusions, which

[9]For previous applications of race models to psycholinguistic data, see e.g. Nicenboim and Vasishth (2018), Paape and Zimmermann (2020), Lissón et al. (2021), Logačev and Vasishth (2016).

was intended to measure the logical abilities of participants as well as the amount of conflict they experienced between analytical, rule-based reasoning and intuition.

The modeling results showed that there are active trade-offs between rational inference and good-enough processing within individuals, both within and across illusion constructions: participants who compute fast rational inferences tended to show less good-enough processing, especially for inversion sentences, and vice versa. There was no indication that rejection ("don't get it") traded off with the other response options at the individual level, but this null result may be due to insufficient data. There was, however, some indication that individuals who scored high on pedantry were more likely to reject agreement attraction and missing VP sentences, and that participants who scored high on charity were more likely to do good-enough processing of inversion and agreement attraction sentences. By contrast, individuals who scored high on logic and individuals who scored low on logic-belief conflict tended to distinguish *more* strongly between illusion and control conditions for inversion sentences, depth charge sentences, and NPI illusion sentences.

Integrating breadth- and depth-focused approaches, and the possible role of surprisal

Overall, the current work has highlighted the value of adopting a breadth-focused approach when investigating linguistic illusions. However, it is clear that the more widely used "depth"-focused approach, in which a single sentence type is manipulated in different ways, continues to yield crucial insights into the cognitive mechanisms involved in acceptability illusions, especially with regard to the role of fine-grained grammatical constraints. The current work was not intended to uncover the precise processing steps that give rise to illusory acceptability, but to highlight the fact that broad frameworks exists into which more specific mechanisms can be integrated. For instance, it has been suggested that agreement attraction, missing VP, and NPI illusions can be explained by a cue-based retrieval mechanism that sometimes accesses incorrect elements in working memory (Häussler & Bader, 2015; Vasishth et al., 2008; Wagers et al., 2009). Cue-based retrieval can potentially be integrated with good-enough processing, either in the sense that memory representations can be faulty because not all aspects of the stimulus were processed (Yadav et al., 2023), or in the sense that there is an additional error monitoring step after retrieval that may be omitted when processing is superficial.

In a similar vein, the fundamental idea of the rational inference framework that mental edits can be made to preceding context can be integrated with predictive processing approaches. The concept of surprisal (Hale, 2001; Levy, 2008a) posits that a word's processing difficulty is proportional to its predictability in a given context. Levy (2011) presented a model in which surprisal can be affected by mental edits to the sentence representation *before* the critical word arrives. For instance, in an agreement attraction sentence like *The key to the cabinets were rusty*, if the preamble is mentally edited into *The keys to the cabinets . . .* , the ungrammatical word *were* is now much less surprising than in a setting where the preamble is always represented veridically (Yadav et al., 2023). Similarly, in a missing VP sentence like *The apartment that the maid who the cleaning service sent over was well-decorated*, if the preamble is edited to contain only two as opposed to three subject nouns, *not* encountering a third verb would not be surprising to the reader (Futrell et al., 2020; Hahn et al., 2022).

The predictions of such a model for linguistic illusions are quite different from those of the error-correction/rational inference model of Levy (2008b) and Gibson et al. (2013). In the error-correction model, the reader encounters an unexpected, potentially ungrammatical word. The unexpectedness of the critical word given the true context makes the reader doubt their representa-

tion of that context, which can result in potentially costly rational inferences. By contrast, under the surprisal-plus-mental-edits model, the context has possibly already been distorted when the critical word arrives, so that the anomaly will be *less* difficult to process. Under the surprisal-based model, there is no reason to assume that errors/mental edits should rise to consciousness: if the context has been distorted and the critical word fits with the distorted context, the reader should be under the impression that everything is fine. This is especially true if it is assumed that context edits happen pre-perceptually (Huang & Staub, 2021a, 2021b), and would explain why illusion sentences can be perceived as being formally correct. The surprisal-plus-mental edits model thus offers an alternative explanation for the "get it, correct" judgments seen in the present study. Whether such a model can account for the full range of data is a question for future work.

In any case, a theory is needed that can explain the interplay between inferring likely sentence meanings and consciously noticing formal errors, including effects of task demands, linguistic pedantry, interpretational charity, and individual preferences for more intuitive versus more analytical processing. Such a theory will likely include aspects of both the rational inference framework and the good-enough processing framework. For instance, both frameworks assume that the prior plausibility of a given interpretation plays a role, and it has been suggested that there is a correlation between the "naturalness" of an error ("Could have happened to me!") and its noticeability and/or subjective probability (Frazier & Clifton, 2015; Poppels & Levy, 2016; Zhang et al., 2023). Furthermore, the rational inference account assumes that readers and listeners rationally adapt to the type and frequency of errors in their environment, and change their interpretation strategies accordingly (Ryskin et al., 2018). What is unclear, however, is how detailed the reader's mental model of the interlocutor needs to be: does the reader or listener engage in a full simulation of the speaker that reverse-engineers the mental processes behind an utterance (Pickering & Garrod, 2013), including likely errors (Frazier & Clifton, 2015; Poppels & Levy, 2016)? A full simulation would possibly be too effortful in realistic scenarios, where cognitive resources are finite (Pöppel, 2023), and requires a comprehender who is able to accurately simulate the errors of the interlocutor without introducing errors of their own in the process.

The data collected in the current study can serve as a benchmark for the predictions of yet-to-be-developed models that can make detailed predictions about error noticeability and rational inference, ideally based on fine-grained properties of sentences. In this context, it will likely be fruitful to take into account how state-of-the-art language and dialogue models like GPT and ChatGPT process illusion sentences, as they are purely data-driven: language models neither have an explicit error correction mechanism, nor do they have a reservoir of motivation and/or attention that can be depleted, unlike humans. There is already some data showing how language models differ (or not) from humans in the domain of linguistic illusions (Cai et al., 2023; Dentella et al., 2023; Paape, 2023; Shin et al., 2023), which can inform future investigations of the mechanisms that may be unique to human sentence processing.

Conclusion

The current work has shown that rational inference and good-enough processing can be understood and modeled as competing latent processes within one and the same individual, which in turn compete with a "pedantic" process that causes ill-formed sentences to be rejected as uninterpretable. Depending on individual traits, task demands, and the specific error type, processing may be dominated by the "inferencer", "slacker", or "pedant" modes. For instance, the more bothered a given subject is by grammatical mistakes, the quicker they accumulate evidence that subtly

ill-formed "depth charge" sentences such as *No test is too difficult to fail* are uninterpretable. Similarly, subjects with a more analytic cognitive style are less likely to experience a grammaticality illusion for ill-formed comparative constructions such as *More engineers relocated than our accountant did* than subjects with a more intuitive cognitive style, as measured by a syllogistic reasoning task. By contrast, subjects who tend to make charitable assumptions about sentence interpretability and correctness are more likely to experience illusions in sentences featuring agreement attraction (*The waitress who sat the girls were . . .*) or argument inversion (*The mother gave the candle the daughter*). Further investigating such individual differences in the context of implemented cognitive models will prove fruitful for uncovering the fine-grained properties of sentences that support or weaken the "inferencer", "slacker" and "pedant" modes within a given reader or listener, and ultimately result in a better understanding of the interaction between sentence processing and other aspects of cognition.

Acknowledgments

The author would like to thank Shravan Vasishth and Garrett Smith for helpful comments on the paper. Additional thanks go to Roger Levy, Ted Gibson, and the audiences at AMLaP 2022 and HSP 2023 for fruitful discussions. The experiment was funded by the University of Potsdam.

Conflict of interest

The author has no conflicts of interest to declare.

References

Bader, M., & Meng, M. (2018). The misinterpretation of noncanonical sentences revisited. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(8), 1286–1311.

Bhatia, S., & Dillon, B. (2022). Processing agreement in Hindi: When agreement feeds attraction. *Journal of Memory and Language*, *125*, 104322.

Bock, K., Eberhard, K. M., Cutting, J. C., Meyer, A. S., & Schriefers, H. (2001). Some attractions of verb agreement. *Cognitive Psychology*, *43*(2), 83–128.

Brehm, L., Jackson, C. N., & Miller, K. L. (2021). Probabilistic online processing of sentence anomalies. *Language, Cognition and Neuroscience*, *36*(8), 959–983.

Brown, V. A. (2021). An introduction to linear mixed-effects modeling in R. *Advances in Methods and Practices in Psychological Science*, *4*(1), 2515245920960351.

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28.

Cai, Z. G., Haslett, D. A., Duan, X., Wang, S., & Pickering, M. J. (2023). Does ChatGPT resemble humans in language use? *arXiv preprint*, *2303.08014*.

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

Chater, N., & Oaksford, M. (1999). Ten years of the rational analysis of cognition. *Trends in Cognitive Sciences*, *3*(2), 57–65.

Chen, S., Nathaniel, S., Ryskin, R., & Gibson, E. (2023). The effect of context on noisy-channel sentence comprehension. *Cognition*, *238*, 105503. https://doi.org/10.1016/j.cognition.2023.105503

Christianson, K. (2008). Sensitivity to syntactic changes in garden path sentences. *Journal of Psycholinguistic Research*, *37*, 391–403.

Christianson, K. (2016). When language comprehension goes wrong for the right reasons: Good-enough, underspecified, or shallow language processing. *Quarterly Journal of Experimental Psychology*, *69*(5), 817–828.

Christianson, K., Dempsey, J., Tsiola, A., & Goldshtein, M. (2022). What if they're just not that into you (or your experiment)? On motivation and psycholinguistics. In K. Federmeier (Ed.), *Psychology of learning and motivation – advances in research and theory* (pp. 51–88). Academic Press.

Christianson, K., Hollingworth, A., Halliwell, J. F., & Ferreira, F. (2001). Thematic roles assigned along the garden path linger. *Cognitive Psychology*, *42*(4), 368–407.

Christianson, K., Luke, S. G., & Ferreira, F. (2010). Effects of plausibility on structural priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(2), 538–544.

Cook, P., & Stevenson, S. (2010). No sentence is too confusing to ignore. *Proceedings of the 2010 Workshop on NLP and Linguistics: Finding the Common Ground*, 61–69.

Coulson, S., King, J. W., & Kutas, M. (1998). Expect the unexpected: Event-related brain response to morphosyntactic violations. *Language and Cognitive Processes*, *13*(1), 21–58.

Dempsey, J., Tsiola, A., Chantavarin, S., Ferreira, F., & Christianson, K. (2023). Nonce word evidence for the misinterpretation of implausible events. *Journal of Cognitive Psychology*, 1–19.

Dentella, V., Murphy, E., Marcus, G., & Leivada, E. (2023). Testing AI performance on less frequent aspects of language reveals insensitivity to underlying meaning. *arXiv preprint*, *2302.12313*.

Drenhaus, H., Saddy, D., & Frisch, S. (2005). Processing negative polarity items: When negation comes through the backdoor (S. Kepser & M. Reis, Eds.). *Linguistic evidence: Empirical, theoretical, and computational perspectives*, 145–165.

Dwivedi, V. D. (2013). Interpreting quantifier scope ambiguity: Evidence of heuristic first, algorithmic second processing. *PloS one*, *8*(11), e81461.

Erickson, T. D., & Mattson, M. E. (1981). From words to meaning: A semantic illusion. *Journal of Verbal Learning and Verbal Behavior*, *20*(5), 540–551.

Evans, N. J., & Wagenmakers, E.-J. (2019). Evidence accumulation models: Current limitations and future directions. *The Quantitative Methods for Psychology*, *16*(2), 73–90.

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

Ferreira, F., Engelhardt, P. E., & Jones, M. W. (2009). Good enough language processing: A satisficing approach. *Proceedings of the 31st Annual Conference of the Cognitive Science Society*, *1*, 413–418.

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

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.

Frank, S. L., & Ernst, P. (2019). Judgements about double-embedded relative clauses differ between languages. *Psychological Research*, *83*(7), 1581–1593.

Frank, S. L., Ernst, P., Thompson, R. L., & Cozijn, R. (2021). The missing-VP effect in readers of English as a second language. *Memory & Cognition*, *49*(6), 1204–1219.

Frazier, L. (2015). Two interpretive systems for natural language? *Journal of Psycholinguistic Research*, *44*, 7–25.

Frazier, L., & Clifton, C., Jr. (2015). Without his shirt off he saved the child from almost drowning: Interpreting an uncertain input. *Language, Cognition and Neuroscience*, *30*(6), 635–647.

Futrell, R., Gibson, E., & Levy, R. P. (2020). Lossy-context surprisal: An information-theoretic model of memory effects in sentence processing. *Cognitive Science*, *44*(3), e12814.

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.

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.

Gigerenzer, G., & Selten, R. (2002). *Bounded rationality: The adaptive toolbox*. MIT press.

Glenberg, A. M., Wilkinson, A. C., & Epstein, W. (1982). The illusion of knowing: Failure in the self-assessment of comprehension. *Memory & Cognition*, *10*(6), 597–602.

Goel, V., & Vartanian, O. (2011). Negative emotions can attenuate the influence of beliefs on logical reasoning. *Cognition and Emotion*, *25*(1), 121–131.

Goldberg, A. E., & Ferreira, F. (2022). Good-enough language production. *Trends in Cognitive Sciences*, *26*(4), 300–311. https://doi.org/10.1016/j.tics.2022.01.005

Goldshtein, M. (2021). *Going beyond our means: A proposal for improving psycholinguistic methods* [Doctoral dissertation, University of Illinois Urbana-Champaign].

Hahn, M., Futrell, R., Levy, R., & Gibson, E. (2022). A resource-rational model of human processing of recursive linguistic structure. *Proceedings of the National Academy of Sciences*, *119*(43), e2122602119.

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

Hannon, B., & Daneman, M. (2004). Shallow semantic processing of text: An individual-differences account. *Discourse Processes*, *37*(3), 187–204.

Häussler, J., & Bader, M. (2015). An interference account of the missing-VP effect. *Frontiers in Psychology*, *6*, 766.

Hayes, B. K., Stephens, R. G., Lee, M. D., Dunn, J. C., Kaluve, A., Choi-Christou, J., & Cruz, N. (2022). Always look on the bright side of logic? testing explanations of intuitive sensitivity to logic in perceptual tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *48*(11), 1598–1617.

Huang, K.-J., & Staub, A. (2021a). Using eye tracking to investigate failure to notice word transpositions in reading. *Cognition*, *216*, 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.

Karimi, H., & Ferreira, F. (2016). Good-enough linguistic representations and online cognitive equilibrium in language processing. *Quarterly Journal of Experimental Psychology*, *69*(5), 1013–1040.

Kuperberg, G. R. (2016). Separate streams or probabilistic inference? what the n400 can tell us about the comprehension of events. *Language, Cognition and Neuroscience*, *31*(5), 602–616.

Langsford, S., Stephens, R. G., Dunn, J. C., & Lewis, R. L. (2019). In search of the factors behind naive sentence judgments: A state trace analysis of grammaticality and acceptability ratings. *Frontiers in Psychology*, *10*, 2886.

Leivada, E. (2020). Language processing at its trickiest: Grammatical illusions and heuristics of judgment. *Languages*, *5*(3), 29.

Levy, R. (2008a). Expectation-based syntactic comprehension. *Cognition*, *106*(3), 1126–1177.

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

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., 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*, *106*(50), 21086–21090.

Li, J., & Ettinger, A. (2023). Heuristic interpretation as rational inference: A computational model of the n400 and p600 in language processing. *Cognition*, *233*, 105359. https://doi.org/10.1016/j.cognition.2022.105359

Lissón, P., Pregla, D., Nicenboim, B., Paape, D., Van het Nederend, M. L., Burchert, F., Stadie, N., Caplan, D., & Vasishth, S. (2021). A computational evaluation of two models of retrieval processes in sentence processing in aphasia. *Cognitive Science*, *45*(4), e12956.

Logačev, P., & Vasishth, S. (2016). A multiple-channel model of task-dependent ambiguity resolution in sentence comprehension. *Cognitive Science*, *40*(2), 266–298.

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

Muller, H. E. (2022). *What could go wrong? Linguistic illusions and incremental interpretation* [Doctoral dissertation, University of Maryland, College Park].

Nicenboim, B., & Vasishth, S. (2018). Models of retrieval in sentence comprehension: A computational evaluation using bayesian hierarchical modeling. *Journal of Memory and Language*, *99*, 1–34.

O'Connor, E. (2015). *Comparative illusions at the syntax-semantics interface* [Doctoral dissertation, University of Southern California].

Orth, W., Yoshida, M., & Sloggett, S. (2021). Negative polarity item (npi) illusion is a quantification phenomenon. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *47*(6), 906–947.

Paape, D. (2023). When Transformer models are more compositional than humans: The case of the depth charge illusion. *Experiments in Linguistic Meaning*, *2*, 202–218.

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.

Paape, D., & Zimmermann, M. (2020). Conditionals on crutches: Expanding the modal horizon. *Proceedings of Sinn und Bedeutung*, *24*(2), 108–126.

Palan, S., & Schitter, C. (2018). Prolific.ac – A subject pool for online experiments. *Journal of Behavioral and Experimental Finance*, *17*, 22–27. https://doi.org/10.1016/j.jbef.2017.12.004

Parker, D., & An, A. (2018). Not all phrases are equally attractive: Experimental evidence for selective agreement attraction effects. *Frontiers in Psychology*, *9*, 1566.

Parker, D., & Phillips, C. (2016). Negative polarity illusions and the format of hierarchical encodings in memory. *Cognition*, *157*, 321–339.

Pennycook, G., Cheyne, J. A., Seli, P., Koehler, D. J., & Fugelsang, J. A. (2012). Analytic cognitive style predicts religious and paranormal belief. *Cognition*, *123*(3), 335–346.

Phillips, C., Wagers, M. W., & Lau, E. F. (2011). Grammatical illusions and selective fallibility in real-time language comprehension. In J. Runner (Ed.), *Experiments at the interfaces* (pp. 147–180). Brill.

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, *36*(4), 329–347.

Pöppel, J. (2023). *Models for satisficing mentalizing* [Doctoral dissertation, University of Bielefeld].

Poppels, T., & Levy, R. (2016). Structure-sensitive noise inference: Comprehenders expect exchange errors. In A. Papafragou, D. Grodner, D. Mirman, & J. Trueswell (Eds.), *Proceedings of the 38th Annual Meeting of the Cognitive Science Society*.

R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org/

Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016). Diffusion decision model: Current issues and history. *Trends in Cognitive Sciences*, *20*(4), 260–281.

Revelle, W. (2023). *Psych: Procedures for psychological, psychometric, and personality research* [R package version 2.3.3]. Northwestern University. Evanston, Illinois. https://CRAN.R-project.org/package=psych

Rohaut, B., & Naccache, L. (2017). Disentangling conscious from unconscious cognitive processing with event-related eeg potentials. *Revue neurologique*, *173*(7-8), 521–528.

Rouder, J. N., Province, J. M., Morey, R. D., Gomez, P., & Heathcote, A. (2015). The lognormal race: A cognitive-process model of choice and latency with desirable psychometric properties. *Psychometrika*, *80*, 491–513.

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

Ryskin, R., Stearns, L., Bergen, L., Eddy, M., Fedorenko, E., & Gibson, E. (2021). An ERP index of real-time error correction within a noisy-channel framework of human communication. *Neuropsychologia*, *158*, 107855.

Sanford, A. J., Leuthold, H., Bohan, J., & Sanford, A. J. (2011). Anomalies at the borderline of awareness: An ERP study. *Journal of Cognitive Neuroscience*, *23*(3), 514–523.

Sanford, A. J., & Sturt, P. (2002). Depth of processing in language comprehension: Not noticing the evidence. *Trends in Cognitive Sciences*, *6*(9), 382–386.

Schwarz, F., & Zehr, J. (2021). Tutorial: Introduction to PCIbex – An Open-Science Platform for Online Experiments: Design, Data-Collection and Code-Sharing. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *43*(43).

Selten, R. (1990). Bounded rationality. *Journal of Institutional and Theoretical Economics (JITE)/Zeitschrift für die gesamte Staatswissenschaft*, *146*(4), 649–658.

Shin, U., Yi, E., & Song, S. (2023). Investigating a neural language model's replicability of psycholinguistic experiments: A case study of NPI licensing. *Frontiers in Psychology*, *14*.

Simon, H. A. (1955). A behavioral model of rational choice. *The Quarterly Journal of Economics*, *69*(1), 99–118.

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review*, *63*(2), 129–138.

Simon, H. A. (1972). Theories of bounded rationality. *Decision and Organization*, *1*, 161–176.

Solcz, S. (2011). *Not all syllogisms are created equal: Varying premise believability reveals differences between conditional and categorical syllogisms* [Doctoral dissertation, University of Waterloo].

Stan Development Team. (2023). RStan: The R interface to Stan [R package version 2.26.8]. https://mc-stan.org/

Stella, M., & Engelhardt, P. E. (2022). Use of parsing heuristics in the comprehension of passive sentences: Evidence from dyslexia and individual differences. *Brain Sciences*, *12*(2), 209.

Stupple, E. J., Ball, L. J., Evans, J. S. B., & Kamal-Smith, E. (2011). When logic and belief collide: Individual differences in reasoning times support a selective processing model. *Journal of Cognitive Psychology*, *23*(8), 931–941.

Swets, B., Desmet, T., Clifton, C., & Ferreira, F. (2008). Underspecification of syntactic ambiguities: Evidence from self-paced reading. *Memory & Cognition*, *36*, 201–216.

Tessler, M. H., Tenenbaum, J. B., & Goodman, N. D. (2022). Logic, probability, and pragmatics in syllogistic reasoning. *Topics in Cognitive Science*, *14*(3), 574–601.

Trippas, D., Kellen, D., Singmann, H., Pennycook, G., Koehler, D. J., Fugelsang, J. A., & Dubé, C. (2018). Characterizing belief bias in syllogistic reasoning: A hierarchical Bayesian meta-analysis of ROC data. *Psychonomic Bulletin & Review*, *25*, 2141–2174.

Trippas, D., Pennycook, G., Verde, M. F., & Handley, S. J. (2015). Better but still biased: Analytic cognitive style and belief bias. *Thinking & Reasoning*, *21*(4), 431–445.

Vasishth, S., Brüssow, S., Lewis, R. L., & Drenhaus, H. (2008). Processing polarity: How the ungrammatical intrudes on the grammatical. *Cognitive Science*, *32*(4), 685–712.

von der Malsburg, T., & Vasishth, S. (2013). Scanpaths reveal syntactic underspecification and reanalysis strategies. *Language and Cognitive Processes*, *28*(10), 1545–1578.

Wagers, M. W., Lau, E. F., & Phillips, C. (2009). Agreement attraction in comprehension: Representations and processes. *Journal of Memory and Language*, *61*(2), 206–237.

Wason, P. C., & Reich, S. S. (1979). A verbal illusion. *The Quarterly Journal of Experimental Psychology*, *31*(4), 591–597.

Wellwood, A., Pancheva, R., Hacquard, V., & Phillips, C. (2018). The anatomy of a comparative illusion. *Journal of Semantics*, *35*(3), 543–583.

Wray, A. (1998). Protolanguage as a holistic system for social interaction. *Language & Communication*, *18*, 47–67.

Xiang, M., Grove, J., & Giannakidou, A. (2013). Dependency-dependent interference: NPI interference, agreement attraction, and global pragmatic inferences. *Frontiers in Psychology*, *4*, 708.

Yadav, H., Paape, D., Smith, G., Dillon, B. W., & Vasishth, S. (2022). Individual Differences in Cue Weighting in Sentence Comprehension: An Evaluation Using Approximate Bayesian Computation. *Open Mind*, *6*, 1–24. https://doi.org/10.1162/opmi_a_00052

Yadav, H., Smith, G., Reich, S., & Vasishth, S. (2023). Number feature distortion modulates cue-based retrieval in reading. *Journal of Memory and Language*, *129*, 104400.

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