

## **UC Merced**

# **Proceedings of the Annual Meeting of the Cognitive Science Society**

### **Title**

Interpreting implausible event descriptions under noise

### **Permalink**

<https://escholarship.org/uc/item/13n5660h>

### **Journal**

Proceedings of the Annual Meeting of the Cognitive Science Society, 46(0)

### **Authors**

Achimova, Asya  
van Os, Marjolein  
Demberg, Vera  
[et al.](#)

### **Publication Date**

2024

Peer reviewed

# Interpreting implausible event descriptions under noise

**Asya Achimova (asya.achimova@uni-tuebingen.de)**

Department of Linguistics, University of Tübingen  
Keplerstr. 2, 72074 Tübingen, Germany

**Marjolein van Os (vanos@lst.uni-saarland.de)**

Department of Language Science and Technology, Saarland University  
Campus C7.2, 66123, Saarbrücken, Germany

**Vera Demberg (vera@lst.uni-saarland.de)**

Department of Computer Science and Department of Language Science and Technology, Saarland University  
Campus C7.2, 66123 Saarbrücken, Germany

**Martin V. Butz (martin.butz@uni-tuebingen.de)**

Cognitive Modeling, Department of Computer Science and Department of Psychology, University of Tübingen  
Sand 14, 72076 Tübingen, Germany

## Abstract

Gricean maxims prescribe cooperative speakers to make their utterances maximally informative so that listeners have the highest chance of understanding the utterances as intended. At the same time, speakers are expected to save articulatory effort and not produce descriptions that are more explicit than necessary. In this work, we first ask how predictability of the described events affects the choice of anaphoric referring expressions. We show that speakers prefer phonologically overt descriptions, such as definite NPs, when they refer to agents that behave in an unexpected way. We further test how the interpretation of referring expressions changes depending on the listening conditions (low vs. high noise) and prior expectations about the plausibility of an event. Our work shows that the speaker's extra effort in choosing a more phonologically overt referring expression is justified by listeners' behavior: they report having heard an utterance which is more plausible than the originally spoken utterance and which contains additional phonological material.

**Keywords:** pronominalization, anaphoric references, events, predictability, noisy channel hypothesis

## Introduction

Within the paradigm of cooperative communication (Grice, 1975), speakers and listeners are expected to make rational choices that maximize their chances of efficient information transmission (Frank & Goodman, 2012; Goodman & Frank, 2016). For the speaker, being rational means choosing utterances that signal the intended message unambiguously. Game-theoretic models of speaking further include balancing utterance costs and their capacity to signal the intended meaning (Parikh, 2001, 2010). Avoiding ambiguity, either lexical or structural, often means having a more costly utterance if we think of costs in terms of the utterance length or articulatory effort.

Complicating matters further, both speakers and listeners typically operate in a noisy communication channel (Benz, 2012; Bergen & Goodman, 2015). Broadly speaking, a noisy channel hypothesis (Shannon, 1948) applied to pragmatics means that speakers and listeners sometimes make imperfect choices in either selecting messages (the intended meaning)

or the way of their encoding (utterances) (Benz, 2012). Inspired by theories of information processing in a noisy channel, current accounts of speech understanding view speech perception as a Bayesian process where a listener evaluates the incoming signal taking into account her prior beliefs and the possibility of noise (Gibson, Bergen, & Piantadosi, 2013; Jurafsky, 1996; Levy, 2008; Levy, Bicknell, Slattery, & Rayner, 2009). Expecting a noisy channel, in turn, also affects the speaker's choice of utterances.

The noisy channel hypothesis can also receive a narrow interpretation when we interpret noise as auditory noise. Speech perception rarely happens in perfect silence: environmental noise and speech of other people can interfere with utterance perception. Experimental work on degraded speech perception reveals both behavioral (Boothroyd & Nittrouer, 1988; Miller, Heise, & Lichten, 1951) and neurophysiological evidence (Hannemann, Obleser, & Eulitz, 2007; Obleser & Kotz, 2010; Sohoglu, Peelle, Carlyon, & Davis, 2012) supporting the idea of integration of prior expectations in the form of lexical or world knowledge and the imperfect auditory signal as two critical components of successful speech comprehension.

In this work, we focus on the production and interpretation of referring expressions, such as *the red monster* or *it*, under noise. Experiments 1 and 2 investigate (a) whether the speakers' choice of referring expressions depends on the plausibility of the events they describe and (b) how plausibility affects the interpretation of referring expressions under noisy conditions. Experiment 2a examines how the type of reference produced interacts with the priors, and whether certain combinations are more prone to misinterpretation than others. Experiment 2b assesses priors regarding the plausibility of different events.

To refer to a previously mentioned entity, a speaker may choose several forms of reference, such as a definite noun phrase (NP) (1a), a pronoun (1b), or a conjoined verb phrase

with no overt subject repetition (1c).

- (1) The yellow monster attacked the red monster and...
  - a. the yellow monster fell down.
  - b. it fell down.
  - c. fell down.

Neo-Gricean models of pragmatic reasoning, and particularly Rational Speech Act (RSA) models (Frank & Goodman, 2012; Franke & Jäger, 2016; Goodman & Frank, 2016), assume that the speaker chooses utterances by simulating the potential listener’s interpretation of these utterances. Thus, speakers are expected to avoid ambiguous references (1b), following the Gricean maxim of manner (Grice, 1975). RSA models conveniently include a term for costs associated with utterance production. If we equate the costs with articulatory effort, we would predict that speakers will choose shorter descriptions, such as conjoined verb phrases (1c) or pronouns (1b) to refer to a previously mentioned agent.

However, if speakers are concerned with signal reliability they might avoid conjoined verb phrases (1c), since the absence of an overt subject reference may allow the listeners to “repair” the gap by inserting phonological material. This repair is more likely if it makes the description more predictable (Gibson et al., 2013). Predictability has been previously shown to affect the acoustic prominence of words (Lam & Watson, 2010), the presence of an overt complementizer (Jaeger, 2010), and case markings (Kurumada & Jaeger, 2015).

Interestingly, the effect of predictability on the choice of referring expressions, especially noun phrases vs. pronouns, has been subject to substantial debate: While some studies report effects of predictability on pronominalization (e.g., Arnold (2008); Tily and Piantadosi (2009); Rosa and Arnold (2017); Weatherford and Arnold (2021), others have failed to find support for such effects (Rohde & Kehler, 2014; Fukumura & van Gompel, 2010; Modi, Titov, Demberg, Sayeed, & Pinkal, 2017). In a recent systematic study of factors impacting the effect of predictability on pronominalization, Demberg, Kravtchenko, and Loy (2023) found that this effect crucially depended on whether stimuli were contextualized as part of larger stories with recurrent characters, thereby engaging participants better than previous decontextualized designs.

This observation suggests that predictability as a cognitive category might have a stronger effect on language production than the predictability at the level of linguistic form alone. The concept of predictability is closely related to a more cognitive view on predictability, or plausibility, of events. Events are central to human cognition (Baldwin & Kosie, 2021; Butz, Achimova, Bilkey, & Knott, 2021; Kuperberg, 2021; Zacks & Tversky, 2001; Zacks, Speer, Swallow, Braver, & Reynolds, 2007) and the structure of events is intricately connected to their linguistic encoding (Butz, 2017; Ünal, Ji, & Papafragou, 2021). Predictability of events has been shown to affect the likelihood of speakers preferring

fragments (subsential units (Morgan, 1973)) to full sentences (Lemke, Reich, Schäfer, & Drenhaus, 2021). Recent psycholinguistic evidence suggests that speakers opt for more phonologically overt descriptions when they describe implausible events (Stegemann-Philipps, Butz, Winkler, & Achimova, 2021). Achimova, Stegemann-Philipps, Winkler, and Butz (2022) further propose a formal model that leverages the reliability of the auditory signal and the strength of prior beliefs in determining speaker utterance choices. However, the evidence considered in both of these papers is limited and bound to a single type of transitive event. Moreover, the preference for overt noun phrases as a form of reference appears rather subtle, possibly due to confounds in the experimental design. In this paper, we aim to expand this paradigm to a number of different interactions and investigate whether event plausibility affects production and disambiguation of referring expressions.

## Experiment 1:

### Production of anaphoric referring expressions

In Experiment 1 ( $n = 44$ , Prolific), we assessed whether world knowledge, and more specifically, the fact that an event is implausible, affects the types of references speakers produce. The perception of event plausibility might be very person-specific: events that are plausible to one person might appear less plausible to another. To avoid the challenge of estimating plausibility of real-life events, we constructed an artificial world with a restricted number of characters and interactions (Stegemann-Philipps et al., 2021).

### Experiment design

The experiment consisted of three main phases. In phase 1, we trained the participants to recognize sequences of events as plausible and implausible in a given artificial world. All interactions involved two monsters that took part in two sub-events: first, one of the monsters performed an action towards the other monster, and then one of the monsters fell on the ground. We tested three types of actions: attack, jump over, and throw a rock. These actions differ in the number of agents and entities involved and the character of their interactions, thus possibly leading to a variety of syntactic structures that underlie their description.

The monsters in our stories differed in strength. Red monsters were stronger than the yellow ones; yellow monsters, in turn, were stronger than the blue monsters. This strength pattern allowed us to construct events that were identical in their first part but differed in their final outcome. The setup yielded scenarios where the agent performs an action and falls down as a result, as well as scenarios where the patient experiences an action and falls down in the end. Thus, we maximally abstracted away from any prior expectations that experiment participants might have had about the outcome of monster interactions. We further gain an opportunity to study how event plausibility interacts with syntactic constraints on the interpretation of ambiguous references, such as pronouns (Kehler,

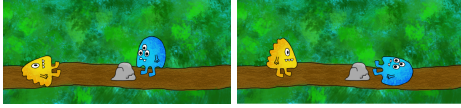


Figure 1: Event outcomes in Experiment 1: surprising (left panel) and familiar (right panel).

Kertz, Rohde, & Elman, 2008; Rohde & Kehler, 2014).

In the second phase of the experiment, we evaluated whether participants correctly learned the power dynamics between the monsters and were able to predict how the events unfolded. If the participants failed the test, they were invited to do the training once again. If they failed the test the second time, we excluded their data from the analysis. Experiments were carried out using the Prolific crowd-sourcing platform. The next paragraph details the third part of the experiment—the production task.

**Production task.** The main part of the experiment constituted a spoken free production task, where the participants were asked to describe the scenes they saw on the screen. We were interested in the types of anaphoric references they produce in cases like (1). The scenes that the participants watched in this phase differed in their plausibility: they either conformed to the course of the events participants witnessed before during the training phase or they violated the previously learned interaction patterns (e.g., the blue monster turned out to be stronger than the yellow one, so the yellow ended up on the ground). Figure 1 shows an example of a surprising (left panel) and familiar event (right panel). Each participant watched both familiar and surprising scenes, the type of action was treated as a between-subject variable.

We transcribed participants’ responses and then coded the type of referring expression produced for the second sub-event. The reference for the second sub-event was anaphoric: the participants mentioned one of the monsters once describing the first sub-event (attack, jump over, throw a rock) and the second sub-event (falling). Table 1 shows a sample of responses and their coding.

1	The yellow monster is approaching the blue monster and then <b>the blue monster</b> falls on the ground.	NP
2	The yellow monster bumped on the red monster and <b>he</b> fell to the ground.	pro
3	A yellow monster walking towards a red monster, the yellow monster jumps over the red monster and <b>then rests</b> on the floor.	zero

Table 1: Experiment 1. Sample responses and coding

## Results

The right panel of Figure 2 demonstrates that speakers overwhelmingly preferred a definite NP, as in (2a), when the char-

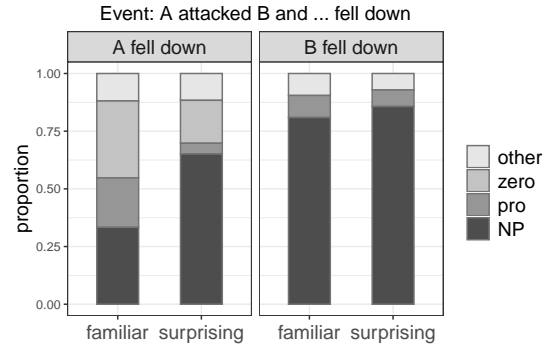


Figure 2: Experiment 1. Types of referring expression produced depending on syntactic structure and type of event.

acter switched thematic roles between the events (2, right panel): in the first sub-event the red monster was the patient and it became the agent in the second sub-event (falling). These results are in line with the previously reported effect of antecedent syntactic position: subject antecedents were found to be more likely to be pronominalized than object antecedents (Demberg et al., 2023). Furthermore, participants never produced a conjoined verb phrase (e.g., *attacked and fell down*) since the subject of the verb *fall* cannot be co-referenced with the patient (the red monster).

- (2) The yellow monster attacked the red monster and...
- a. the red monster fell down.
  - b. it fell down.
  - c. \*fell down.

**No change of thematic role cases.** The critical case for evaluating the effect of predictability lies in the trials where it was the agent of the first event who fell over in the second event (Figure 2, left panel). This scenario permitted the use of a definite NP (1a), a personal pronoun (1b), or a zero overt subject (1c). To examine the effect of plausibility on reference production, we excluded the responses categorized as “other” (11.8% of trials) and grouped pronouns and zero cases together to form a “reduced reference” category.

If plausibility affects the choice of referring expressions, we expect to see more reduced forms for familiar events. The same type of interaction, in which a yellow monster attacks another monster and then falls down, was either familiar (yellow monster interacting with the red one) or surprising (yellow monster interacting with the blue monster). We can therefore compare the choice of referring expressions for the events that follow the same event structure but differ in plausibility. A generalized linear mixed model analysis with response as the dependent variable and the event type (plausible or implausible) as the independent variable, participants and types of actions as random intercepts<sup>1</sup>, revealed that speak-

<sup>1</sup>Here and in experiments below we report maximally-converging random effect structure.

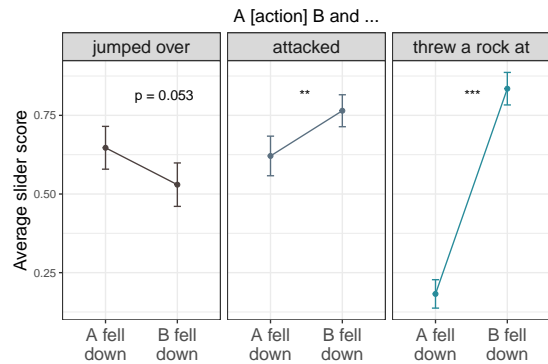


Figure 3: Slider ratings for both event outcomes. A positive slope indicates the patient bias. Error bars correspond to 95% confidence intervals.

ers were more likely to use a NP for an implausible event ( $\beta = 18.978, SE = 3.771, z = 5.033, p < 0.01$ ). We can witness this pattern in Figure 2: in the left panel the proportion of NPs for implausible events is almost double the number of NPs for plausible events.

**Discussion.** If speakers choose referring expressions to avoid ambiguity, we would expect NPs and conjoined verb phrases (zero overt subject) to be approximately equally likely when referring to the agent of the previous event since both of these forms are not ambiguous. However, Figure 2 reveals an asymmetry in these forms of reference for surprising events. Following the noisy channel hypothesis (Gibson et al., 2013; Levy, 2008), we expect speakers to avoid structures with zero overt subjects, since such utterances might be misinterpreted: a listener could “repair” such prompts by inserting a pronoun, thus making the prompt compatible with a reading where the patient of the previous action fell down. To evaluate this hypothesis, we tested how speakers interpret different types of referring expressions when speech is masked by noise. According to the noisy channel models of repair, some repairs are more costly than others (Gibson et al., 2013). Thus, filling in an element would be a less costly operation than first removing an element from the utterance and then filling it with another element.

## Experiments 2a and 2b:

### Perception of referring expressions under noise

#### Experiment 2a: Speech under noise

Experiment 2a assessed how participants ( $n = 49$ , Prolific) interpret utterances that contain anaphoric references. None of the participants who completed Experiment 1 took part in Experiments 2a or 2b.

**Experiment design.** In Experiment 2a, they watched two animated scenes that featured monster interactions where either the agent (Figure 4, left panel) or the patient (right panel)

of the previous event fell down. The participants then heard an audio description and were asked to pick the corresponding scene, as well as to type what they heard.

The audio prompts differed in two dimensions. First, they were either ambiguous (1b) or unambiguous (1c) (we did not use prompts with NPs in this phase). The unambiguous prompts contained conjoined verb phrases (e.g. *attacked and fell down*), which grammatically can only mean that it was the same character who attacked and fell down. Second, the prompts contained either low or high amounts of babble noise. We used cocktail party noise to mask the descriptions: the low noise condition was set at 0 dB SNR and thus had equal levels of speech and noise, while the high noise condition was set at -5 dB SNR (the noise was 5 dB louder than the speech).

Our predictions differ for ambiguous and unambiguous prompts. For the ambiguous prompts, like (1b), we expected the prior beliefs to favor the choice of the scene. For unambiguous prompts, such as (1c) technically, only the scene in the left panel of Figure 4 matches the description.

#### Experiment 2b: Priors

We independently assessed the prior expectations about event outcomes ( $n = 51$ , Prolific) in Experiment 2b. Here, participants watched two animations that featured an interaction between two monsters. The interactions followed a similar schema like in Experiment 1: monster A interacted with the monster B, and then one of the monsters fell down. Participants assessed how plausible each of the event outcomes was using a slider: thus, for each trial we obtained two slider ratings that assessed the plausibility of each outcome (either monster A or monster B falling down). In this experiment, participants did not receive any audio prompts: they evaluated only the plausibility of the event outcomes themselves and not the match between the prompt and the outcome.

For the “jumping over” event we did not register a difference in plausibility of the outcomes (mean ratings 0.65 vs. 0.53,  $p = 0.0536$ ): participants found both outcomes plausible. On the other hand, for “attack” and “throw rock” events, participants rated higher the animation that showed the patient (monster B) of the first sub-event falling down (attack, mean ratings 0.62 vs. 0.77 on a scale between 0 and 1,  $p < 0.01$ ; throw rock, mean ratings 0.17 vs. 0.90,  $p < 0.001$ , Figure 3). This prior assessment suggests the following order of biases: “jump over” (no bias), “attack” (weak patient bias), and “throw rock at” (strong patient bias). We use the term “patient bias” to refer to a situation where the patient of the first event is likely to suffer the consequences of the event and fall down. These results provide a baseline for experiment 2a, where participants heard a prompt after they watched both animations. Their task was to select the animation that was described in the prompt (Figure 5).

## Results

**Experiment 2a: Choice of event outcome.** We analyze the results of Experiment 2a separately for ambiguous and non-

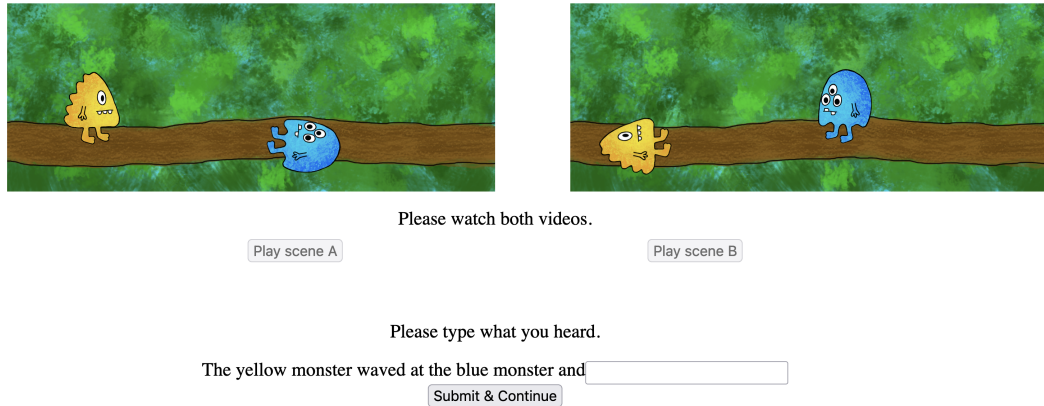


Figure 4: Experiment 2a, sample trial. After watching both videos participants heard the prompt: *The yellow monster threw a rock at the blue monster and (it) fell down*. Prompts with *it* are ambiguous, while prompts without *it* are not ambiguous and are only compatible with the event shown in the left panel.

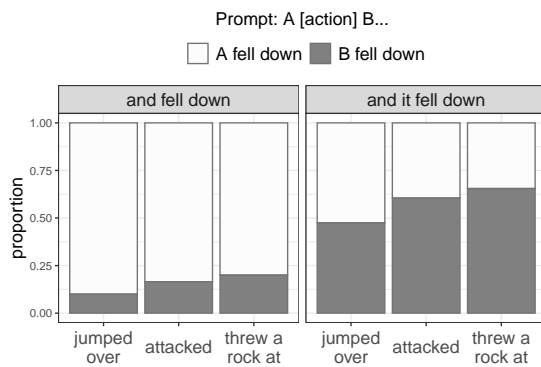


Figure 5: Experiment 2. Type of scene that the participants chose after hearing the prompt.

ambiguous prompts. Recall that ambiguous prompts (those containing the pronoun *it*) were grammatically compatible with both outcomes of the event. We expected the distribution of choices to align with the priors we obtained in Experiment 2b. The results of Experiment 2a demonstrate that indeed the prior expectations drove the choice of the scene for ambiguous prompts (Figure 5, right panel). We set “attack” as a reference level and compared how often participants chose a video where it was the agent vs. patient of the first sub-event who fell down in the end. The analysis shows that participants were less likely to select the videos where the patient fell down for “jump over” events ( $\beta = -0.7230, SE = 0.2359, z = -3.065, p < 0.01$ ). We know independently from Experiment 2b that the bias towards the patient is weak or even reversed here. There was no difference in the choice of outcome for “throw a rock” events compared to “attack” ( $\beta = -0.2913, SE = 0.2383, z = 1.223, p = 0.222$ ). Both of these actions showed a patient bias in experiment 2b.

Unexpectedly, predictability also affected the unambiguous prompts: 13.5% of the time participants chose scenes that appeared more plausible although the prompt did not li-

cense such a reading (Figure 5, left panel). For example, the prompt *The yellow monster attacked the red monster and fell down* can only be understood to mean that the yellow monster (monster A) fell down. Yet, we see that some participants selected the scene where the red monster (monster B) fell down. These responses are strictly speaking errors. In order to understand the nature of these errors, we now turn to the analysis of textual responses participants provided.

**Experiment 2a: Analysis of textual responses.** Alongside with choosing the animation, participants were asked to type what they heard in a text box following each trial. The screen showed the first part of the prompt, as in (3).

- (3) The yellow monster [action] the red monster and...

We then analyzed what kind of responses participants provided. More specifically, we coded whether they entered any overt form of reference (*it, the red monster*) into the field. We then analyzed the cases when the entered referring expression matched the prompt and those, where participants either omitted or inserted phonological material. Table 2 shows the distribution of responses. When the prompt contained no pronoun (left column) participants typed the responses without a pronoun 84% of the time. For the prompts with pronoun, participants included the pronoun in their responses 62.8% of the time, and omitted it 28.5% of the time.

Table 2: Referring expressions typed by the participants

	and fell down	and it fell down
NP	3%	4.8%
pro	10.2%	<b>62.8%</b>
zero	<b>84%</b>	28.5%
missing data	2.8%	3.8%

We will now focus on the critical cases—these include

unambiguous prompts (*The yellow monster attacked the red monster and fell down*), where participants selected an event outcome that was grammatically incompatible with the prompt (where the red monster fell down). Out of those cases, participants inserted phonological material 63% of the time, as opposed to only 3% of the time when they chose the correct scene. We analyzed the insertions by fitting the data with a binomial mixed effect logistic regression model with random intercepts for participants. Setting “attack” as a reference level, we evaluated whether the type of action, and hence the type of bias, affected the insertions. Overall, participants were less likely to insert a pronoun or a NP for the action with no bias, that is, “jump over” ( $\beta - 0.709, SE = 0.351, z = -2.018, p < 0.05$ ). The insertion rate did not differ for “attack” and “throw rock”—the two actions with a patient bias ( $\beta - 0.266, SE = 0.327, z = 0.813, p = 0.42$ ). Insertions were less likely to occur in the low noise condition ( $\beta - 1.095, SE = 0.303, z = -3.617, p < 0.001$ ).

## Discussion

These data suggest that under noisy conditions, participants are more likely to misinterpret forms of reference if the described event does not conform to their prior expectations. Moreover, listeners actively repair the auditory signal that is incoherent with their prior beliefs. In our case, participants inserted a pronoun into the perceived prompt to make it ambiguous and, as a consequence, compatible with a scenario where the patient of the first sub-event fell down.

### General discussion

RSA models of utterance choice predict that in the context of anaphoric references, speakers should take into account signal reliability, as well as content plausibility when making a choice of a referring expression (Achimova et al., 2022). In this work, we assess empirically whether plausibility of an event affects the type of referring expressions speakers produce and how listeners interpret such references. We have created an artificial world setup that allowed flexibly manipulating the prior expectations about a normal cause of interactions between two monsters. We then assessed whether speakers describe events that conform to their priors differently than events that violate these priors. We have demonstrated that speakers produce more definite NPs to refer to agents of implausible events. We have expanded the empirical scope of considered event structures and registered this pattern for actions with different level of bias and syntactic frames.

This result agrees with a recently reported effect of predictability documented in Demberg et al. (2023). The authors have found that providing a richer context to the prompt containing the reference made the predictability effect much stronger compared to isolated prompts, where the effect of predictability was manipulated solely through the choice of implicit causality or transfer of possession verb. This contrast between poor and rich context experimental setups in Demberg et al. (2023) implicitly suggests that predictability

as a cognitive category might have a stronger effect on the choice of references than the linguistic form alone.

We have previously reported that NPs were preferred as a form of reference for patient antecedents in general, and for agent antecedents if the new event was implausible. The flip side of a preference for NPs was the avoidance of shorter forms of reference, such as pronouns. This avoidance is likely associated with the ambiguity that the pronouns introduce: they can refer to either the agent or the patient of the first sub-event when used anaphorically for the second sub-event. Interestingly, when speakers described implausible events, they avoided not only ambiguous pronouns, but also conjoined verb phrases where the second sub-event can only apply to the agent of the previous event, and not to the patient. Following predictions of the noisy channel hypothesis (Gibson et al., 2013) and models of speech production in the RSA framework (Achimova et al., 2022; Demberg et al., 2023), we have suggested that this behavior might be driven by the expectation of how such descriptions might be interpreted by the listener. A listener who encounters an implausible event description might choose to reinterpret the speech signal to make it more coherent with a normal cause of events. A rational speaker would then be expected to use a more phonologically reliable definite NP despite added articulatory costs.

While our experiment did not directly target the interactions of two communication partners, our perception experiment demonstrates that the speaker’s preference for a more phonologically prominent phrase in the case of implausible events is warranted. Under noisy conditions, prior expectations made listeners choose events that are strictly speaking incompatible with the grammatical form of the utterance. They tended to fix this mismatch by inserting a pronominal or nominal reference. Critically, these insertions were more frequent when participants had strong prior expectations about the course of the events and the prompt conflicted those priors. In other words, under noisy conditions, participants were more likely to trust their prior expectations rather than the unreliable signal, and altered the signal post-hoc to make the description coherent with their priors.

The phenomenon of phonological repair is often discussed in the literature on phoneme restoration (Warren, 1970). There, the effect of filling in missing auditory information is discussed in the context of prediction: participants expect a particular sound to occur in a word, and they fill in the missing sound if it is missing (Leonard, Baud, Sjerps, & Chang, 2016). Computational models of speech perception further show that prediction and integration of top-down information of lexical information is critical for phoneme identification (McClelland & Elman, 1986). Experiment 2b reported in this paper takes prediction to a different level: it is the expectation about a course of events that drives the phonological repair. Thus, in addition to theoretical insights, our paper offers a paradigm for investigating the role of prior beliefs on speech perception by offering a controlled environment where both beliefs and the signal can be flexibly manipulated.

## Acknowledgements

We would like to thank the anonymous reviewers for sharing their insights and pointers to relevant literature. This project was funded by the German Research Foundation (DFG) project number 519109539. Martin Butz is a member of the Machine Learning Cluster of Excellence, EXC number 2064/1 – Project number 390727645.

## Data availability

Data and analysis code are available at <https://tinyurl.com/8dzw6e24>.

## References

- Achimova, A., Stegemann-Philipps, C., Winkler, S., & Butz, M. V. (2022). Anaphoric reference in descriptions of surprising events. In D. Gutzmann & S. Repp (Eds.), *Proceedings of Sinn und Bedeutung 26* (pp. 20–34). doi: 10.31234/osf.io/d5krz
- Arnold, J. E. (2008). Reference production: Production-internal and addressee-oriented processes. *Language and Cognitive Processes*, 23(4), 495–527.
- Baldwin, D. A., & Kosie, J. E. (2021). How does the mind render streaming experience as events? *Topics in Cognitive Science*, 13(1), 79–105.
- Benz, A. (2012). Errors in pragmatics. *Journal of Logic, Language and Information*, 21(1), 97–116.
- Bergen, L., & Goodman, N. D. (2015). The strategic use of noise in pragmatic reasoning. *Topics in Cognitive Science*, 7(2), 336–350.
- Boothroyd, A., & Nittrouer, S. (1988). Mathematical treatment of context effects in phoneme and word recognition. *The Journal of the Acoustical Society of America*, 84(1), 101–114.
- Butz, M. V. (2017). Which structures are out there? Learning predictive compositional concepts based on social sensorimotor explorations. In T. K. Metzinger & W. Wiese (Eds.), *Philosophy and Predictive Processing*. Frankfurt am Main: MIND Group. doi: 10.15502/9783958573093
- Butz, M. V., Achimova, A., Bilkey, D., & Knott, A. (2021). Event-predictive cognition: A root for conceptual human thought. *Topics in Cognitive Science*, 13, 10–24. doi: 10.1111/tops.12522
- Demberg, V., Kravtchenko, E., & Loy, J. E. (2023). A systematic evaluation of factors affecting referring expression choice in passage completion tasks. *Journal of Memory and Language*, 130, 104413.
- Frank, M. C., & Goodman, N. D. (2012). Predicting Pragmatic Reasoning in Language Games. *Science*, 336, 998–998.
- Franke, M., & Jäger, G. (2016). Probabilistic pragmatics, or why Bayes’ rule is probably important for pragmatics. *Zeitschrift für Sprachwissenschaft*, 35(1), 3–44.
- Fukumura, K., & van Gompel, P. G. (2010). Choosing anaphoric expressions: Do people take into account likelihood of reference? *Journal of Memory and Language*(62), 52–66.
- 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.
- Goodman, N. D., & Frank, M. C. (2016). Pragmatic Language Interpretation as Probabilistic Inference. *Trends in Cognitive Sciences*, 20(11), 818–829.
- Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and Semantics 3: Speech Acts* (p. 26–40). New York: Academic Press.
- Hannemann, R., Obleser, J., & Eulitz, C. (2007). Top-down knowledge supports the retrieval of lexical information from degraded speech. *Brain research*, 1153, 134–143.
- Jaeger, T. F. (2010). Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology*, 61(1), 23–62.
- Jurafsky, D. (1996). A probabilistic model of lexical and syntactic access and disambiguation. *Cognitive Science*, 20(2), 137–194.
- Kehler, A., Kertz, L., Rohde, H., & Elman, J. L. (2008). Coherence and coreference revisited. *Journal of Semantics*, 25(1), 1–44.
- Kuperberg, G. R. (2021). Tea with milk? a hierarchical generative framework of sequential event comprehension. *Topics in Cognitive Science*, 13(1), 256–298.
- Kurumada, C., & Jaeger, T. F. (2015). Communicative efficiency in language production: Optional case-marking in Japanese. *Journal of Memory and Language*, 83, 152–178.
- Lam, T. Q., & Watson, D. G. (2010). Repetition is easy: Why repeated referents have reduced prominence. *Memory & cognition*, 38(8), 1137–1146.
- Lemke, R., Reich, I., Schäfer, L., & Drenhaus, H. (2021). Predictable words are more likely to be omitted in fragments—evidence from production data. *Frontiers in Psychology*, 12, 662125.
- Leonard, M. K., Baud, M. O., Sjerps, M. J., & Chang, E. F. (2016). Perceptual restoration of masked speech in human cortex. *Nature Communications*, 7(1), 13619.
- Levy, R. (2008, October). A noisy-channel model of human sentence comprehension under uncertain input. In M. Lapata & H. T. Ng (Eds.), *Proceedings of the 2008 conference on empirical methods in natural language processing* (pp. 234–243). Honolulu, Hawaii: Association for Computational Linguistics. Retrieved from <https://aclanthology.org/D08-1025>
- 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.
- McClelland, J. L., & Elman, J. L. (1986). The trace model of speech perception. *Cognitive Psychology*, 18(1), 1–86.
- Miller, G. A., Heise, G. A., & Lichten, W. (1951). The in-



- telligibility of speech as a function of the context of the test materials. *Journal of Experimental Psychology*, 41(5), 329.
- Modi, A., Titov, I., Demberg, V., Sayeed, A., & Pinkal, M. (2017). Modelling semantic expectation: Using script knowledge for referent prediction. *Transactions of the Association for Computational Linguistics*, 5, 31-44.
- Morgan, J. (1973). Sentence fragments and the notion 'sentence'. In B. B. Kachru, R. Lees, Y. Malkiel, A. Pietrangeli, & S. Saporta (Eds.), *Issues in linguistics. papers in honor of henry and renée kahane* (pp. 719–751). Urbana, IL: University of Illinois Press.
- Obleser, J., & Kotz, S. A. (2010). Expectancy constraints in degraded speech modulate the language comprehension network. *Cerebral Cortex*, 20(3), 633–640.
- Parikh, P. (2001). *The use of language*. Stanford: CSLI Publications.
- Parikh, P. (2010). *Language and equilibrium*. Cambridge: MIT Press.
- Rohde, H., & Kehler, A. (2014). Grammatical and information-structural influences on pronoun production. *Language, Cognition and Neuroscience*, 29(8), 912–927.
- Rosa, E. C., & Arnold, J. E. (2017). Predictability affects production: Thematic roles can affect reference form selection. *Journal of Memory and Language*, 94, 43–60.
- Shannon, C. E. (1948). A mathematical theory of communication. *The Bell system technical journal*, 27(3), 379–423.
- Sohoglu, E., Peelle, J. E., Carlyon, R. P., & Davis, M. H. (2012). Predictive top-down integration of prior knowledge during speech perception. *Journal of Neuroscience*, 32(25), 8443–8453.
- Stegemann-Philipps, C., Butz, M. V., Winkler, S., & Achimova, A. (2021). Speakers use more informative referring expressions to describe surprising events. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 43, pp. 2897–2903).
- Tily, H., & Piantadosi, S. (2009). Refer efficiently: Use less informative expressions for more predictable meanings. In *Proceedings of the workshop on the production of referring expressions: Bridging the gap between computational and empirical approaches to reference*.
- Ünal, E., Ji, Y., & Papafragou, A. (2021). From event representation to linguistic meaning. *Topics in Cognitive Science*, 13(1), 224–242.
- Warren, R. M. (1970). Perceptual restoration of missing speech sounds. *Science*, 167(3917), 392–393.
- Weatherford, K. C., & Arnold, J. E. (2021). Semantic predictability of implicit causality can affect referential form choice. *Cognition*, 214, 104759.
- Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S., & Reynolds, J. R. (2007). Event perception: A mind/brain perspective. *Psychological Bulletin*, 133(2), 273–293.
- Zacks, J. M., & Tversky, B. (2001). Event structure in perception and conception. *Psychological Bulletin*, 127(1), 3–21.