

**\*\*MANUSCRIPT IN PREPARATION\*\***

PLEASE DO NOT CITE OR SHARE WITHOUT  
PERMISSION OF THE AUTHORS

---

# PRAGMATIC PREDICTION IN THE PROCESSING OF REFERRING EXPRESSIONS CONTAINING SCALAR QUANTIFIERS

---

A PREPRINT

**Vinicius Macuch Silva\***  
Institute of Cognitive Science  
Osnabrück University  
Germany  
vini.macuch@gmail.com

**Michael Franke**  
Institute of Cognitive Science  
Osnabrück University  
Germany  
mchfranke@gmail.com

November 27, 2020

## ABSTRACT

Previous research in cognitive science and psycholinguistics has shown that language users are able to predict upcoming linguistic input probabilistically, pre-activating material on the basis of cues emerging from different levels of linguistic abstraction, from phonology to semantics. Current evidence suggests that linguistic prediction also operates at the level of pragmatics, where processing is strongly constrained by interpersonal and supra-linguistic factors. In the current study, we investigated whether people generate predictions about upcoming lexical material on the basis of pragmatic considerations about how the non-linguistic context might be referred to. For that, we used a self-paced reading task where German natives were asked to view visual scenes and then read descriptions of those same scenes. The visual stimuli depicted two arrays of geometric shapes which varied consistently in regard to their color composition, their size, and their location relative to one another. The descriptions contained scalar quantifiers and referred to one and only one of the depicted shape arrays. Crucially, we manipulated whether the visual context biased readers into specific expectations about how the description might unfold word by word. We measured participants' reading times at sentence regions where critical information was revealed. Contrary to what was expected, we found that participants took longer reading the main critical term in scenarios where they were biased by context and pragmatic constraints to expect a given shape array as the target of the referential expression, as opposed to scenarios where there was no such bias for a specific shape array, in which case the main critical term was read faster. Our results provide new insights about how language users might predict upcoming linguistic input on the basis of pragmatic considerations. We discuss these results in light of possible processing constraints that may interact with people's pragmatic expectations during online incremental language processing.

**Keywords** Pragmatics · Predictive processing · Pragmatic prediction · Scalar quantifiers · Self-paced reading · German

---

\*Use footnote for providing further information about author (webpage, alternative address)—*not* for acknowledging funding agencies.

# 1 Introduction

Prediction in online processing has been a central theme in recent cognitive scientific and psycholinguistic research (see, *inter alia*, Bubic, Von Cramon, & Schubotz, 2010; Clark, 2013; Kuperberg & Jaeger, 2016). Numerous studies have shown that humans are able to predict upcoming input on the basis of perceptual, social, and linguistic cues (e.g. Kamide, Scheepers, & Altmann, 2003; Ridderinkhof, 2017; Winkler, Denham, & Nelken, 2009; see also Litwin & Miłkowski, 2020). As far as linguistic prediction is concerned, studies have shown that material is pre-activated at all levels of linguistic representation, from phonologically- and lexically-driven pre-activation to pre-activation derived from syntactic and semantic cues (e.g. Boudewyn, Long, & Swaab, 2015; Kuperberg & Jaeger, 2016; Mani & Huettig, 2012; Urbach, DeLong, & Kutas, 2015). Pragmatics is no different: many studies have provided evidence that high-level semantic and pragmatic prediction occurs while people process language, from the processing of negation (e.g. Haase, Spychalska, & Werning, 2019; Nieuwland, 2016; see also Scappini, Delfitto, Marzi, Vespignani, & Savazzi, 2015) to the processing of sentences containing pragmatic cues such as the scalar quantifier *some* (e.g. Augurzky, Franke, & Ulrich, 2019; Nieuwland, Ditman, & Kuperberg, 2010). Although the bulk of the evidence stems from neurolinguistic studies, other studies have drawn on behavioral methods such as eye-tracking (e.g. Degen & Tanenhaus, 2016; Huang & Snedeker, 2018; Scholman, Rohde, & Demberg, 2017) and self-paced reading (e.g. Bergen & Grodner, 2012; Bicknell & Rohde, 2009). Put together, the evidence suggests that people are capable of predicting linguistic as well as non-linguistic material on the basis of pragmatic and discourse-based expectations. For example, studies relying on the manipulation of the pictorial context which the linguistic stimuli refer to show that people are sensitive to contextually-driven changes in expectation (Augurzky et al., 2019; Degen & Tanenhaus, 2016). In other words, language users modulate their expectations about the unfolding linguistic signal depending on how a referring expression might map onto a visually-presented referent. In the present study, we are interested in understanding the extent to which visually-anchored language processing might be said to rely on pragmatic expectations about the perceived likelihood of a linguistic utterance referring to a particular visual referent. In order to address this question, we introduce a self-paced reading task where participants are asked to read quantified descriptions of visual scenes depicting arrays of geometric shapes. Crucially, we vary whether the descriptions are locally biased in terms of their potential referential target.

In order to investigate whether comprehenders modulate their expectations about upcoming linguistic material on the basis of pragmatic considerations, we probe their online behavior in a reading task. The referring expressions that participants are asked to read contain scalar quantifiers, which have been widely studied in both theoretical and experimental pragmatic research (see Huang and Snedeker (2009) and Geurts (2010) for an overview), and which are known to act as pragmatic cues in the online processing of language. In previous research, predicability has been linked to the relative probability of a word serving as a continuation of an unfolding linguistic signal (Kuperberg & Jaeger, 2016; Levy, 2008). In written language comprehension, the surprisal of a given word as a sentence continuation has been argued to be proportional to the cognitive effort of reading a word (Smith & Levy, 2013). Reading times (RTs) are assumed to be proportional to the effort involved in processing words in a written stream of language. In order to generate predictions about the reading patterns we expect to find in our study, we depart from an assumption that, at the level of utterance and discourse processing, language users attend to different magnitudes of pragmatic felicity. We posit a coarse-grained, Gricean-inspired account of utterance processing as our set of working assumptions, assuming, at the very least, that comprehenders are sensitive to the contrast between semantically congruent and semantically incongruent utterances, as well as to the contrast between pragmatically felicitous and pragmatically infelicitous utterances.

In the pragmatic tradition initiated by philosopher Paul Grice (Grice, 1975, 1989), linguistic utterances are usually modeled, among other dimensions, in terms of their degree of informativeness, where particular terms and expressions render an utterance either more or less informative depending on other semantically-related lexical alternatives which together with the uttered term constitute a so-called linguistic scale (Horn, 1989). For our current purposes, we operationalize pragmatic felicity as informativity in the Gricean sense. Therefore, we consider a pragmatically infelicitous linguistic utterance to be a description which is underinformative, i.e. a true description for which a salient alternative exists which is also true and logically stronger, so that the latter entails the former by semantic meaning but not the other way around. While there are other aspects of

pragmatic felicity, we restrict our analyses to these clearly defined benchmarks of informativity. To illustrate how pragmatic felicity in this sense can lead to interesting predictions about expectation-based processing difficulty, consider Figure 1 (which shows example items from the experiments presented in section 2 and 3 below).



Figure 1: Examples of visually presented context information which raises expectations of particular lexical material based on considerations of pragmatic felicity.

Under an informativity account, an unfolding linguistic signal such as *Some of the ...* may be said to be an informative description with reference to scene (a). Indeed, even though the quantified referring expression is in principle semantically congruent with either shape array depicted in the scene, from a pragmatic perspective one can expect such a description to eventually refer to the circle array and not the triangle array, given that a salient alternative expression, namely *All of the ...*, would have been semantically stronger and thus pragmatically more informative in case the triangle array was the actual intended target of the referring expression. This inferential jump from a lower-bounded interpretation of *some* to an upper-bounded one is regarded as a pragmatic enrichment, more specifically a scalar implicature, whereby an interpreter is able to derive the enriched, upper-bounded meaning of the referring expression on the basis of the assumed informativity relation between the scalar alternatives *some* and *all*, which, couched in a more general expectation of pragmatic felicity, gives rise to probabilistic expectations about the to-be-mentioned shape array. Importantly, notice that a similar inference would not be possible with reference to scene (b), where neither of the shape arrays can be truthfully described using the quantifier *all*, hence there being no stronger alternatives to *some*.

Now, considering that people generate online linguistic predictions based on pragmatic expectations, their predictions should vary as a function of both the amount of linguistic information available at any given moment, i.e., how much of the unfolding linguistic signal has been processed, as well as one’s overall expectations regarding what is being communicated, in this case what is being referred to. Crucially, the two are interrelated, such that expectations might shift as new material is processed and integrated into competing discourse models. In practical terms, we need to establish a link hypothesis between the predictions derived from our theory and the empirically measured processing signatures which are to reflect the effects of the postulated underlying pragmatic causes. We do so in our case via surprisal, such that we expect that comprehenders will read a given description more slowly whenever their expectations fail to be met by whatever fragment of the description they are reading. That is, if a participant expects to read a specific word at a specific moment but instead reads a different word, she is expected to experience processing difficulties as a result of a mismatch between her expectations and the actual linguistic material she encounters. This is in line with previous work which has established surprisal as a possible link hypothesis in language processing (Smith & Levy, 2013), including processing as measured by means of online reading times.

Bringing together the theoretically-motivated considerations of pragmatic interpretation with the insights about the integration of linguistic cues in online processing via surprisal, we arrive at the following expected effects on RTs, referred henceforth as *pragmatic surprisal theory*: after having read a sentence initial fragment of words  $w_1, \dots, w_n$ , comprehenders read word  $w_{n+1}$  faster in (visual) context  $C$  than in context  $C'$  if they expect  $w_{n+1}$  with a higher probability to occur in  $C$  than in  $C'$ . Given a quantified expression like *Some of the triangles are green*, we hypothesize that, after having read the quantifier, a comprehender will read the next critical term, i.e., the shape term, more slowly in context (b) than in context (a). This is because after processing the quantifier, more than one true and pragmatically felicitous continuation is available in (b), whereas in (a) only one true and pragmatically felicitous continuation is available. Therefore, in a scenario like (a) comprehenders are expected to generate strong predictions about a specific sentence continuation, which implies strong expectations about the to-be-read shape term in the unfolding sentence. In a scenario like (b), however, comprehenders are expected to generate weaker predictions about one or the other possible sentence continuations, which implies weaker relative expectations about the to-be-read shape term in the

unfolding sentence. Whether a given prediction A or B is confirmed, participants are expected to be surprised by what they read, given that they did not have strong expectations about a given term X or Y in the first place. In short, given the cases illustrated in Fig. 1, a comprehender confronted with (a) should expect to read *circles* after reading *Some of the*, as that is the most informative continuation at that point in the sentence. Crucial for our argumentation, however, this prediction should only hold when the scalar term *some* is enriched pragmatically, giving rise to a so-called scalar implicature, as discussed above.

Aside from any underlying considerations of a pragmatic nature, we expect that participants in our experiment will also bring along to the task of reading descriptions in context general biases associated with efficient problem-solving, such that they will attempt to perform the task as quickly and as optimally as possible. We will consider the effects of task-specific optimization later in the discussion section.

Importantly, in addition to the reading task proper, we also had participants perform a separate task in which they were asked to produce descriptions of the visual stimuli used in the processing task. While in the present study we are primarily interested in investigating online language processing, this complementary task allows us to collect production probabilities for the critical linguistic expressions measured in the reading task. With that in hand, we are then able to probe whether the measured reading patterns, which we assume reflect underlying processes of pragmatic prediction, find any support in independently collected data which reveals the biases that people may have when producing descriptions using the same linguistic expressions they process in our main task. Although links between production and comprehension are not straightforward to establish, our aim here is simply to ground the processing patterns in empirically verified production preferences. We discuss these possible links in the general discussion at the end of the paper. In the next section, we describe the production task.

## 2 Task 1 - Production of referring expressions containing scalar quantifiers

In our main task, described below in section 3, we test the extent to which comprehenders process referring expressions containing scalar terms predictively on the basis of pragmatic considerations. Our working assumptions are derived from general, Gricean-inspired considerations of pragmatic felicity and informativity. More specifically, we assume that comprehenders prefer true and pragmatically felicitous utterances over true but pragmatically infelicitous utterances. However, in practice, comprehenders might have somewhat different expectations, expectations that perhaps diverge from or fuse together these different theoretically-motivated assumptions. In order to systematically tease apart pragmatic surprisal from other possible auxiliary assumptions, we first had our participants perform a production task in order to determine whether their observed production behavior gives rise to expectations that support a surprisal-based account of the reading data reported in section 3. Our aim here is to test the extent to which comprehenders' empirically verified reading patterns, assumed to be linked to their underlying pragmatic expectations, match those derived from a normative, theoretically-grounded account of pragmatic processing.

Before asking our participants to perform the reading task, we asked them to describe the same stimuli used in that task by completing sentences which had gaps in them. As explained earlier in the introduction, the aim of such task was to collect descriptions of the stimuli so as to know what the likelihood of producing specific descriptions might be in the first place. We are thus particularly interested in knowing what possible sentence configurations are more likely given the specific types of pictures found in our stimulus set.

### 2.1 Method

#### 2.1.1 Participants

Fifty-eight native speakers of German were recruited among the cognitive science and psychology student population of the University of Osnabrück. Participants were given course credit in exchange for their participation in the experiment. Data collection was conducted at the Institute of Cognitive Science of the same university, in a computer laboratory designed for the execution of behavioral experiments.

### 2.1.2 Materials and design

Participants were asked to complete 21 German sentences, each referring to a visual scene composed of two arrays of eight geometric shapes. Figure 2 exemplifies the different types of pictures contained in the study, which varied along three semantic dimensions: the color of the shapes (Fig. 2, first row), their size (Fig. 2, second row), and their position relative to a box also depicted in a subset of the scenes (Fig. 2, third row).

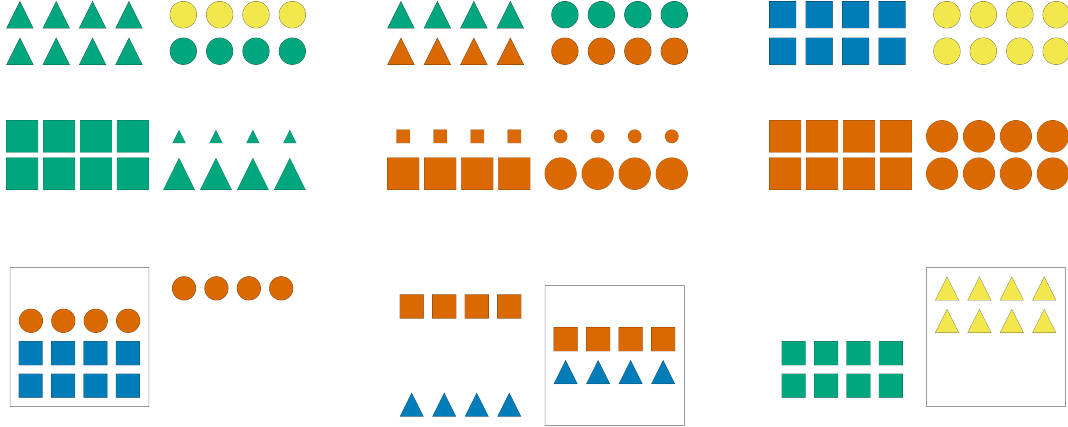


Figure 2: Sample visual scenes. Scenes in the first column contain one homogeneous shape array and one heterogeneous shape array; scenes in the second column contain two heterogeneous shape arrays; scenes in the third column contain two homogeneous shape arrays.

Three choices were recorded per sentence, each of these mapping onto the critical regions from the reading task, as in the sentence structure below:

Original: QUANT | der | SHAPE | auf | dem | Bild | sind | PROPER | [der | Box].

English gloss: QUANT | of the | SHAPE | in | the | picture | are | PROPER | [the | box].

Each trial, participants' task was to select one scalar quantifier (QUANT), one shape term (SHAPE), and one adjective or preposition, which depending on the picture type related to a different visuo-semantic property (PROPER). There were two quantifiers to choose from, three shape terms, two prepositions, and six adjectives, as shown below.

- QUANT – *einige* (some), *alle* (all);
- SHAPE – *Kreise* (circles), *Quadrate* (squares), *Dreiecke* (triangles);
- PROPER – *orange* (orange), *blau* (blue), *gelb* (yellow), *grün* (green), *klein* (small), *groß* (big), *in* (inside), *neben* (next to).

### 2.1.3 Procedure

Written as well as oral instructions were provided prior to the actual task, followed by three practice trials which mimicked the exact procedure of the test trials. Each trial, participants were presented with a visual scene as well as with sentence with gaps in it. Participants were instructed to fill the gaps in the sentence by choosing between different words in drop-down menus available on screen. After making their choices and filling all gaps, a 'next' button appeared on screen, allowing participants to proceed to the next trial.

## 2.2 Results and discussion

For every test sentence, participants chose a quantifier, a shape term, as well as a third term which, depending on the type of picture, could be an adjective denoting color or size or a preposition. Figure 3 shows the raw frequencies of the produced combinations of quantifier + shape term + property term. The terms other than the quantifier are shown in terms of their congruency with the chosen quantifier: congruent terms are those that, when combined with either *alle* or *einige*, result in a semantically congruent picture description, such as *Alle Dreiecke grün* or *Einige Kreise gelb* in the case of the scene on the top row of the figure. Note that in the case of scenes with only one heterogeneous array [top row], when the quantifier under consideration is *einige*, the label *Incongruent* actually refers to shape terms which in conjunction with the quantifier render the description underinformative, such as triangle. Similarly, when the shape term itself is congruent, choosing either of the options available as a property term – in the example provided, green and yellow – results in a congruent description. As such, in such cases the label *Incongruent* in the legend actually refers to the property shared by both the heterogeneous and homogeneous arrays – in the example provided, the color green.

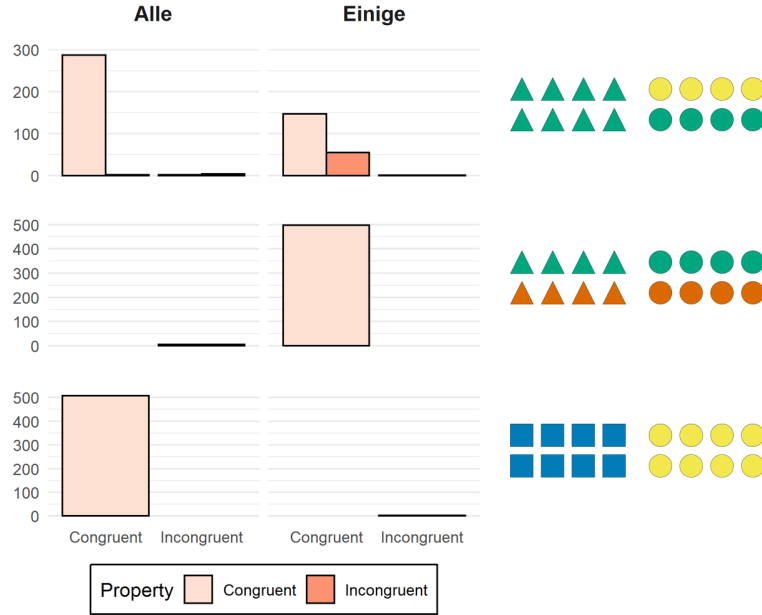


Figure 3: Production frequencies. The labels below each quantifier column indicate whether the selected shape term renders the description of the visual scene shown on the x-axis congruent or not. The color indicates whether the selected property term renders the description congruent or not.

Unsurprisingly, participants produced descriptions of pictures with two homogeneous arrays [bottom row] virtually always using the quantifier *alle*, much the same way that they produced descriptions of pictures with two heterogeneous arrays [middle row] virtually always using the quantifier *einige*. This is in line with the naive assumption that people prefer semantically congruent utterances over semantically incongruent ones. When describing pictures with only one heterogeneous array [top row], participants show a preference for producing descriptions containing the quantifier *alle*, that is, they show a bias for referring to the homogeneous array in the scene. This can be interpreted as a bias for producing referring expressions using quantifiers which are semantically stronger, and thus pragmatically more informative, than their scalar alternatives.

Another referential bias participants show when describing pictures with only one heterogeneous array is to prefer referring to the semantic dimension which is congruent exclusively with the heterogeneous array. In other words, when people use *einige* to refer to a scene like the one depicted on the top row of Fig. 3, they tend to couple the quantifier with a property term that renders the description maximally distinct from other possible descriptions of that same picture, such as *Einige der Kreise auf dem Bild sind gelb*. Both these

preferences are in line with the naive assumption that, all things being equal, people prefer pragmatically felicitous utterances over pragmatically infelicitous ones.

### 3 Task 2 - Processing of referring expressions containing scalar quantifiers

#### 3.1 Method

##### 3.1.1 Materials and design

Participants read 84 German sentences, each referring to a varying visual scene, as per the stimuli in Task 1. Participants saw four instances of each picture type for each condition, for a total of 84 trials. Reading times were measured at eight to ten sentence regions, each of these consisting of one word of the sentence structure introduced above:

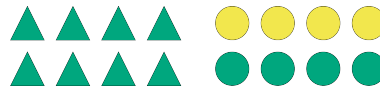
QUANT | der | SHAPE | auf | dem | Bild | sind | PROPER | [der | Box].

In the case of the reading task, the terms represented here by uppercase words varied each trial. Much like in Task 1, each sentence contained a scalar quantifier (QUANT), a shape term (SHAPE), and either an adjective denoting color or size or a preposition (PROPER):

- QUANT – *einige* (some), *alle* (all);
- SHAPE – *Kreise* (circles), *Quadrate* (squares), *Dreiecke* (triangles);
- PROPER – *orange* (orange), *blau* (blue), *gelb* (yellow), *grün* (green), *klein* (small), *groß* (big), *in* (inside), *neben* (next to).

Different sentences and pictures were paired so as to yield four critical experimental conditions. These conditions differ in terms of the quantifier which was processed as well as in terms of whether the matching visual scene induced specific linguistic expectations at the SHAPE region. In other words, the experimental manipulation involved modulating the predictability of the shape terms by means of varying the visual context which participants encountered immediately before reading the test sentence. In our experiment, we take predictability to be primarily dependent on pragmatic considerations of a Gricean nature, as described above, such that readers are expected to find a given critical term more or less predictable given the assumption that a referring expression produced by a cooperative describer is informative. As such, the resulting experimental conditions are as follows:

- Alle (Biased) - Sentences containing the quantifier *alle* were paired with visual scenes which were meant to increase the predictability of specific critical terms (i.e. one of the arrays is homogeneous while the other is heterogeneous - the contrast should bias the expectation of the critical terms which match the homogeneous array). A sample pairing is as follows:



Original: "Alle der Dreiecke auf dem Bild sind grün."

English gloss: "All of the triangles in the picture are green."

Upon reading *alle*, an interpreter who is expecting a semantically congruent utterance should predict the description to refer to the triangle array, and not the circle array, as only in the case of the former is the description congruent with the scene.

- Alle (Unbiased) - Sentences containing the quantifier *alle* were paired with visual scenes which were meant not to increase the predictability of any subsequent critical term (i.e. both arrays are homogeneous). A sample pairing is as follows:





Original: "Alle der Kreise auf dem Bild sind gelb."

English gloss: "All of the circles in the picture are yellow."

- Einige (Biased) - Sentences containing the quantifier *einige* were paired with visual scenes which were meant to increase the predictability of specific critical terms (i.e. one of the arrays is heterogeneous while the other is homogeneous - the contrast should bias the expectation of the critical terms which match the heterogeneous array, however, only if *einige* is read as *some-but-not-all*). A sample pairing is as follows:

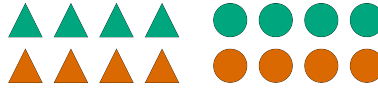


Original: "Einige der Kreise auf dem Bild sind gelb."

English gloss: "Some of the circles in the picture are yellow."

While in the case of the Alle (Biased) condition the sentence bias originates from an expectation of truthfulness, in the case of Einige (Biased) the bias originates from an expectation of informativeness, which, according to the inferential account we are entertaining, should give rise to an implicature.

- Einige (Unbiased) - Sentences containing the quantifier *einige* were paired with visual scenes which were meant not to increase the predictability of any subsequent critical term (i.e. both arrays are heterogeneous). A sample pairing is as follows:

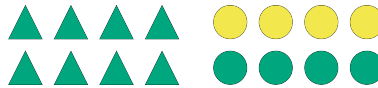


Original: "Einige der Kreise auf dem Bild sind orange."

English gloss: "Some of the circles in the picture are orange."

In addition to the four critical conditions, the experiment includes three conditions which serve as a baseline for the manipulations involving pragmatic expectations. These consist of pairings of sentences and pictures which resulted in semantically incongruent descriptions, as well as a condition in which descriptions containing the quantifier *einige* are semantically congruent but underspecified:

- Alle (False) - Sentences containing the quantifier *alle* were paired with visual scenes which are semantically incongruent with the referring expression (i.e. the referred array is not homogeneous). A sample pairing is as follows:



Original: "Alle der Kreise auf dem Bild sind gelb."

English gloss: "All of the circles in the picture are yellow."

- Einige (False) - Sentences containing the quantifier *einige* were paired with visual scenes which are semantically incongruent with the referring expression (i.e. both arrays are heterogeneous but the shape term renders the description semantically incongruent). A sample pairing is as follows:



"Einige der Quadrate auf dem Bild sind orange."

- Einige (Infelicitous) - Sentences containing the quantifier *einige* were paired with visual scenes which were meant to increase the predictability of specific critical terms (i.e. one of the arrays is heterogeneous while the other is homogeneous - the contrast should bias the expectation of the critical terms which match the heterogenous array). A sample pairing is as follows:



"Einige der Dreiecke auf dem Bild sind grün."

English gloss: "Some of the triangles in the picture are green."

Note that in this condition the description, which contains the quantifier *einige*, refers to the homogeneous array, which implies that if the utterance is ultimately interpreted as a congruent description of the scene no pragmatic enrichment must have taken place, suggesting a strictly semantic interpretation of the quantifier *einige*, meaning *some-and-maybe-all*.

In summary, the study contains seven experimental conditions - four critical conditions and three control conditions - as schematized in Table 1.

Quantifier	
	<i>Alle</i> <i>Einige</i>
Bias	Alle (Biased)                      Einige (Biased)
	Alle (Unbiased)                      Einige (Unbiased)
Control	Alle (False)                      Einige (False)
	Einige (Infelict)

Table 1: Study design.

### 3.1.2 Procedure

Written as well as oral instructions were provided to participants prior to the task, followed by three practice trials which mimicked the exact procedure of the test trials<sup>2</sup>. All trial elements flashed in and out of the screen in a pre-specified order. First, a fixation cross appeared at the center of the screen. After 500 milliseconds, the cross disappeared and the visual scene became visible at the cross location. The picture remained on screen for as long as participants wished, and it was dismissed by a press of the space bar. Immediately after the picture disappeared, underscores appeared below the picture location, along with the written indication "Press the SPACE bar to reveal the words", presented in German. As indicated by the cue, each press of the

<sup>2</sup>The experimental program is available for inspection under <https://spr-in-lab-june-2019-direct-folder.netlify.com/>.

space bar revealed one of the sentence chunks foreshadowed by the underscores. Once participants reached the last chunk in the sentence, their next key press triggered the question "How accurate was the sentence as a description of the picture?", presented in German. As indicated by the question, participants' task was to rate, on a 7-point scale, how appropriate the sentence was as a description of the visual display they saw on screen. After making their choice, participants were forwarded to the next trial.

There were 14 trials of each trial type and all participants saw all of these, for a total of 84 trials. Given the constraints imposed by the experimental manipulations, similar or even identical visual scenes were paired with different referring expressions. However, the matching between an image and a description always resulted in unique trial instantiations. The 84 trials were administered in four blocks of 21 trials. In between blocks, participants encountered a pause screen, and they were encouraged to take as much time as needed before proceeding to the next block.

### 3.2 Hypotheses

Given the design of the study and the considerations outlined in the introduction, we put forward the following general prediction: participants will read the critical terms more slowly if they are unexpected. Thus, we hypothesize that at the shape region, i.e., the next critical region after the quantifier, participants will read the critical term more slowly in the unbiased conditions than in the biased conditions, regardless of the quantifier they encounter. This is under the assumption that, after processing the quantifier, more than one informative continuation is available per scene in the unbiased scenarios, whereas in the biased scenarios only one informative continuation is available. Therefore, participants in the biased conditions are expected to generate strong predictions about a specific sentence continuation, which implies strong expectations about the to-be-read shape term in the unfolding sentence. In the unbiased conditions, however, participants are expected to generate weaker predictions about one or the other possible sentence continuations, which implies weaker relative expectations about the to-be-read shape term in the unfolding sentence. Whether a given prediction A or B is confirmed, participants are expected to be surprised by what they read, given that they did not have strong expectations about a given term X or Y in the first place. Fig. 4 shows what a biased and an unbiased scenario look like in relation to a description containing the quantifier *einige*.



Figure 4: Biased and unbiased scenarios for descriptions containing the quantifier *einige*.

Considering the cases illustrated in Fig. 4, a participant confronted with (a) should expect *Kreise* (English: circles) after reading *Einige der* (English: Some of the), as that is the most informative continuation at that point in the sentence. This prediction should only hold, however, when the quantifier *einige* is taken as a pragmatic, as opposed to a semantic, cue. Thus, we expect that the bias manipulation will yield different results depending on whether participants process the sentences as pragmatically enriched or strictly semantic descriptions. Namely, we expect that when *einige* is interpreted as *some-and-maybe-all* reading times will be similar in both the biased and unbiased conditions, as in both cases participants are expected not to generate strong predictions about one or the other possible sentence continuations.

#### Pragmatic interpretation

- (a) Einige | der | **Kreise** \ Dreiecke ...
- (b) Einige | der | **Kreise** \ **Dreiecke** ...

#### Semantic interpretation

- (a) Einige | der | **Kreise** \ Dreiecke ...
- (b) Einige | der | **Kreise** \ **Dreiecke** ...

We assume that, after processing the quantifier and the shape term, participants will then generate predictions about the last semantically relevant term in the sentence. Considering again the cases in Fig. 4, we may assume that in the unbiased scenario, even if participants already know the shape term, it is still unclear which

color might be referred to. Much the same way, in the biased scenario, even if participants already know that the sentence refers to circles, the color of the referred array cannot be easily predicted. In fact, even if *einige* is taken as a pragmatic cue (*some-but-not-all*), participants should not be able to generate strong predictions about a particular color term.

#### Pragmatic interpretation

- (a) Einige | der | Kreise | auf | dem | Bild | sind |  
gelb \ grün
- (b) Einige | der | Kreise | auf | dem | Bild | sind |  
grün \ orange

#### Semantic interpretation

- (a) Einige | der | Kreise | auf | dem | Bild | sind |  
gelb \ grün
- (b) Einige | der | Kreise | auf | dem | Bild | sind |  
grün \ orange

In summary, then, the following predictions are generated from our surprisal-based account of pragmatic prediction:

1. At the shape term, descriptions will be read more slowly in the unbiased conditions compared to the biased conditions. This will be the case for descriptions containing *einige* only if comprehenders interpret the quantifier as a pragmatic cue;
2. At the shape term, descriptions containing *einige* will be read equally fast in both conditions, if comprehenders interpret *einige* as a semantic cue.

A visual representation of the predictions is provided in Figure 5.

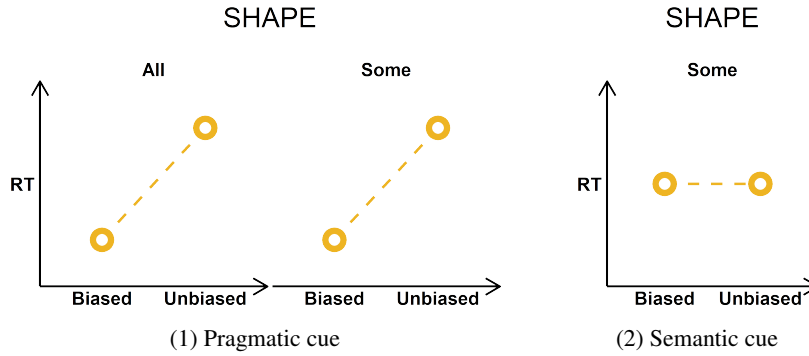


Figure 5: Study predictions.

### 3.3 Results

#### 3.3.1 Data cleaning

We recorded participants' reading times (RTs) as well as their sentence ratings. Two participants were excluded from the original sample due to faulty data recording which resulted in incomplete data sets, meaning that prior to data cleaning the sample consisted of 56 participants. For each test sentence, RTs were measured across all words, for a total of 8 to 10 measurement regions per sentence depending on the sentence type (descriptions referring to the position of the shapes contained two additional regions compared to the other two types of description). The data was cleaned according to two criteria: first, for any given trial, if the total RT differed by 2.5 positive or negative standard deviations from the mean total reading time for the respective condition, then the trial was excluded from any subsequent analysis; then, for any given participant, if their number of excluded trials was superior to 30% of the total number of trials, then the participant was excluded from any subsequent analysis. While no participant was excluded on the basis of these criteria, 776 individual trials were excluded from any subsequent analysis. Moreover, inspection of the trial data inputted to the experimental program showed that in 1/3 of the *Einige* (False) items were coded erroneously, such that

the critical term which rendered the descriptions false was the property term, when it should have been, in all cases, the shape term. In practice, what this particular sentence configuration did was to shift the incongruent critical term one region downstream. We therefore excluded the faulty trials from any subsequent analysis.

### 3.3.2 Confirmatory analyses

Recall that on each trial participants were asked to rate the accuracy of the description they read using a 7-point scale. In the case of the biased and unbiased conditions, all trials consisted of semantically congruent and thus, a priori, accurate image-description pairs. In the case of the false conditions, all trials consisted of semantically incongruent and thus, a priori, inaccurate image-description pairs. The distribution of participants' ratings can be seen in Figure 6 below:

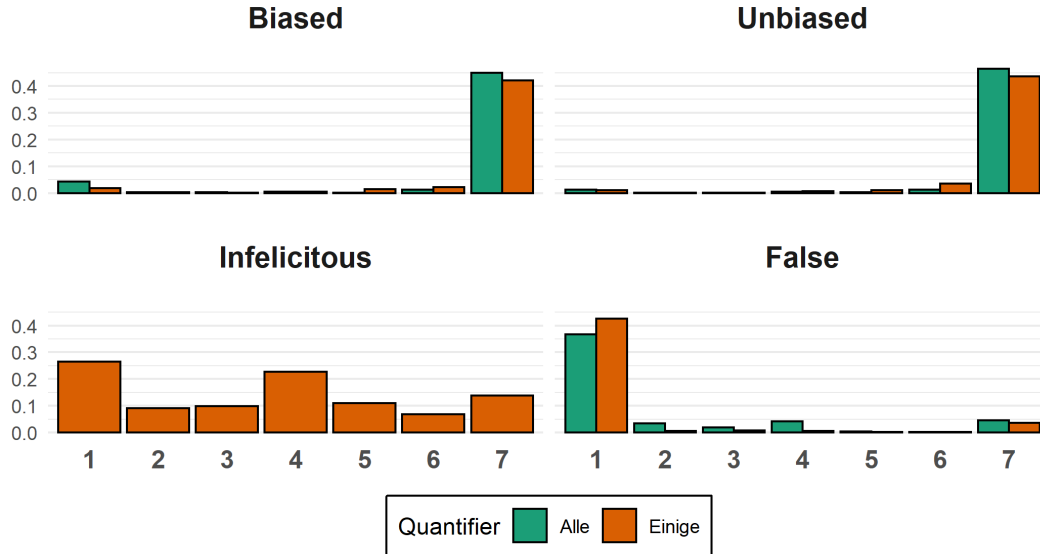


Figure 6: Sentence ratings. Each panel shows the proportion of ratings for a given critical condition. Ratings for sentences containing the quantifier *alle* are shown in green, while ratings for sentences containing the quantifier *einige* are shown in orange.

Visual inspection of the graph suggests that only rarely did participants rate the semantically congruent expressions as inaccurate descriptions of the pictures (rating 1, 2, and 3 on the 7-point scale). Similarly, only rarely did they rate the semantically incongruent expressions as accurate descriptions of the pictures (rating 5, 6, and 7 on the 7-point scale). Note that the ratings for the *Einige* (Infelict) condition are somewhat evenly distributed over the whole scale, which means that there is wide variability in how participants interpreted descriptions containing the quantifier *einige*.

Below, in Figure 7, we plot the RT data. In order to amass quantitative evidence in favor of the reported results, we fitted Bayesian hierarchical models predicting RTs at the critical region, the SHAPE region, as a function of the experimental conditions, which themselves reflect different combinations of the quantifiers and the sentence bias. The models included, if possible, the maximal random effect structure justified by the design, which in our case is random intercepts for items – the actual pictures seen by participants, which varied systematically according to the experimental condition – and random slopes and intercepts for participants. Our models, fitted using the R package *brms* (Bürkner, 2017) and described in detail in the supporting material, had the following general form, shown in *brms* syntax:

$$\log(\text{RT}) \sim \text{condition} + \\ (1 + \text{condition} \mid \text{participant}) + \\ (1 \mid \text{item})$$

For each highlighted result, we report whether or not the respective statistical model provides strong evidence in favor of the empirically attested differences (or lack thereof). In a Bayesian statistical framework, one is interested in the joint posterior distribution of the parameters of the model, which indicates a plausible range of values for the parameters given the model and the data at hand. We report a 95% credibility interval (CI) and the posterior probability that the parameter of interest,  $\beta$ , is larger than zero ( $P(\beta > 0)$ ). We speak of strong evidence for an effect when zero is not included in the CI and  $P(\beta > 0)$  is close to either zero or one. Concretely, we are interested in the difference between estimated values for cell means of conditions *Einige* (Biased) and *Einige* (Unbiased), as well as that between *Alle* (Biased) and *Alle* (Unbiased). Pragmatic surprisal theory predicts that, in the posterior distribution of the Bayesian regression model, the difference in cell means  $\beta_{\text{einige}} = [\text{estimates for cell mean of } \textit{Einige} \text{ (Biased)}] - [\text{estimates for cell mean of } \textit{Einige} \text{ (Unbiased)}]$  should be credibly bigger than zero, so that  $P(\beta_{\text{einige}} > 0)$  should be large, i.e., very close to 1; similarly for the conditions with the quantifier *alle*.

Figure 7 shows the mean RTs at each sentence region up until the last critical term, across all seven experimental conditions. QUANT, SHAPE, and PROPER are the critical regions where task-relevant information was read, namely, the quantifier, the shape term, and the property term, respectively. Visual inspection of the graph suggests that there are differences between the critical conditions and the control conditions, which serve as a diagnostic of participants’ online sensibility to the semantic congruency of the descriptions. At the SHAPE region, participants read the critical term faster in the critical conditions [orange and yellow] compared to their respective controls [red and purple] (for the regression coefficients, see the respective table in the supporting material), as is to be expected.

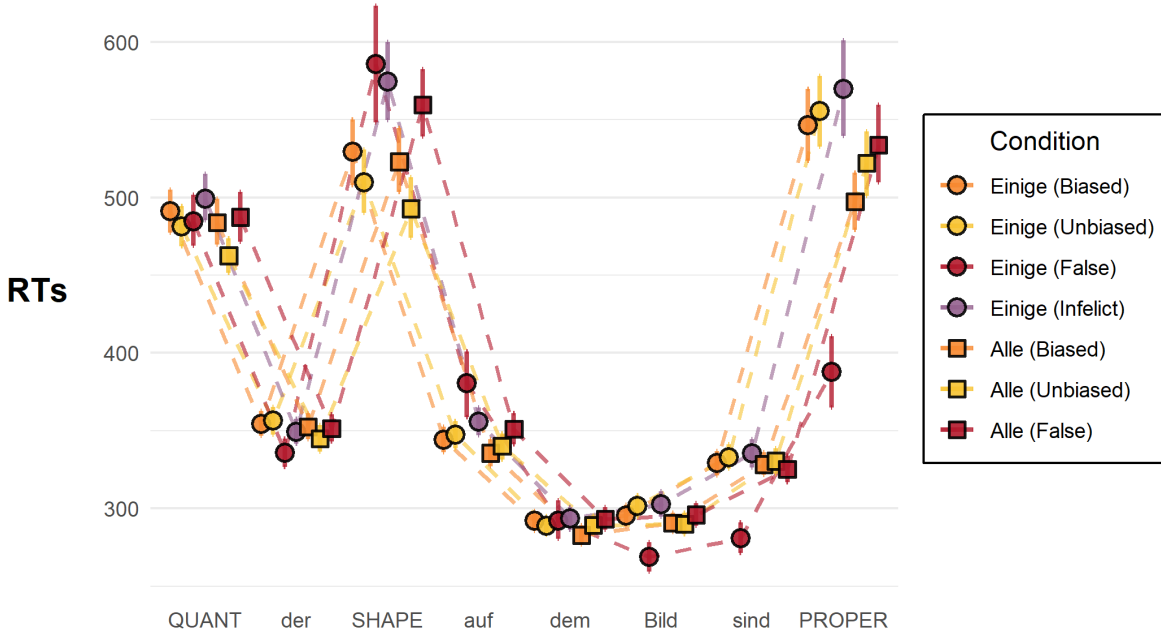


Figure 7: Mean reading times across all sentence regions up until the last critical term. The error bars represent 95% bootstrapped confidence intervals.

As for the critical condition-quantifier pairs, visual inspection of the graph suggests that there are differences at the SHAPE region. Interestingly, however, the results go in the opposite direction of the predictions of pragmatic surprisal theory, contradicting our hypothesis: we find strong evidence that in the case of both *einige* and *alle* participants took longer reading the shape term in the biased condition compared to the unbiased condition. Indeed, our statistical models show that there is practically no reason to believe that the RTs in the unbiased conditions are larger than those in the biased conditions ( $\beta_{\text{alle}} = -0.06$ , 95% CI [-0.12, 0.00],  $P(\beta_{\text{alle}} > 0) = 0.02$ ;  $\beta_{\text{einige}} = -0.07$ , 95% CI [-0.13, 0.00],  $P(\beta_{\text{einige}} > 0) = 0.02$ ).

### 3.3.3 Exploratory analysis: Processing by pragmatic profile

In order to investigate whether there are further processing differences depending on the pragmatic profile exhibited by participants, we plot the RT data according to the different types of respondents, namely those who entirely consistently rated underspecified expressions as accurate scene descriptions (semantic respondents,  $n = 6/56$ ), those who entirely consistently rated those same expressions as inaccurate scene descriptions (pragmatic respondents,  $n = 10/56$ ), and those who show a mixed response pattern (mixed respondents,  $n = 40/56$ ). For these comparisons, we fitted hierarchical models predicting RTs at the different critical regions as a function of both the experimental conditions and the pragmatic profiles, again including, if possible, the maximal random effects structure justified by the design, which in this case is random intercepts for both items and participants. Since each participant is of exactly one pragmatic profile and since each item belongs to exactly one experimental condition, our models had the following general form:

$$\log(\text{RT}) \sim \text{condition} * \text{pragmatic profile} + \\ (1 + \text{condition} \mid \text{participant}) + \\ (1 + \text{pragmatic profile} \mid \text{item})$$

Figure 8 shows the mean RTs at the SHAPE region according to each respondent type. Visual inspection of the graph suggests that there is variation between the critical condition-quantifier pairs across the different profiles. While we find strong evidence that pragmatic respondents read the critical term more slowly in the Alle (Biased) condition [orange square] compared to Alle (Unbiased) [yellow square] ( $\beta_{\text{alle}} = -0.15$ , 95% CI [-0.26, -0.04],  $P(\beta_{\text{alle}} > 0) = 0.00$ ), we only find weak evidence for the same effect in the case of semantic respondents ( $\beta_{\text{alle}} = -0.11$ , 95% CI [-0.25, 0.03],  $P(\beta_{\text{alle}} > 0) = 0.07$ ), which matches the visibly less pronounced descriptive difference. As for the *einige* conditions [circles], we find weak evidence that both pragmatic and semantic respondents read the shape term more slowly in the biased condition [orange circle] (semantic:  $\beta_{\text{einige}} = -0.06$ , 95% CI [-0.23, 0.11],  $P(\beta_{\text{einige}} > 0) = 0.25$ ; pragmatic:  $\beta_{\text{einige}} = -0.06$ , 95% CI [-0.19, 0.07],  $P(\beta_{\text{einige}} > 0) = 0.19$ ). As for mixed respondents, we find again only weak evidence of an effect for both critical condition-quantifier pairs, which matches the lack of any visible descriptive differences (see supporting material for regression coefficients). All in all, these results seem to conflict with the prediction that  $\text{RT}_{\text{Einige (Biased)}} < \text{RT}_{\text{Einige (Unbiased)}}$ , when considering pragmatic respondents, as well as with the prediction that  $\text{RT}_{\text{Einige (Biased)}} = \text{RT}_{\text{Einige (Unbiased)}}$ , when considering semantic respondents.

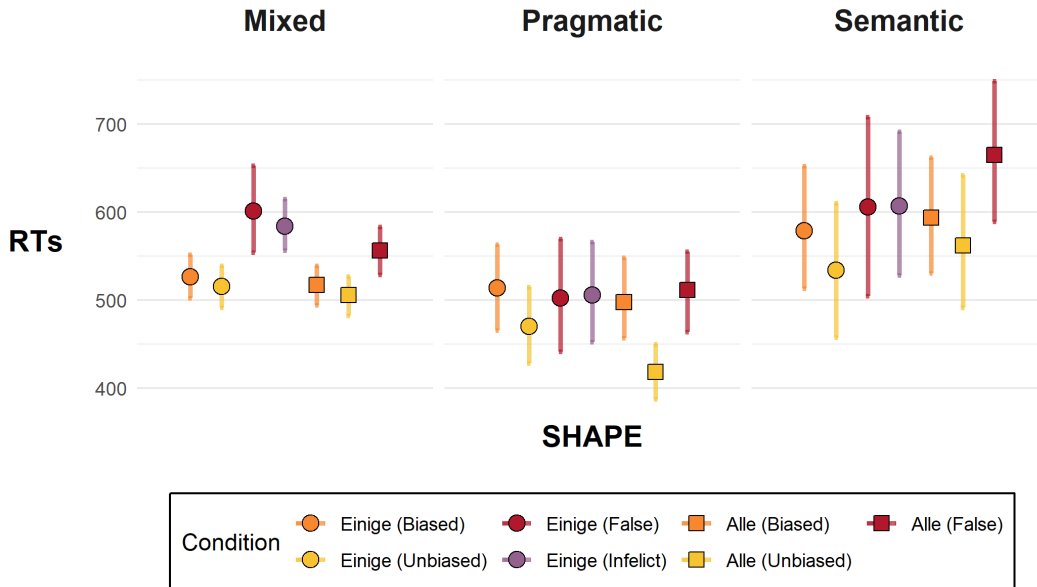


Figure 8: Mean reading times at the SHAPE regions by type of respondent. The error bars represent 95% bootstrapped confidence intervals.

### 3.3.4 Summary

We found reliable differences in RTs at the SHAPE region, but these differences were in the opposite direction of what was predicted by pragmatic surprisal theory, which explains reading times as a monotone decreasing function of pragmatic expectability of the shape term in the context of the picture and the initial sentence segment. This theory predicts that RTs for *Einige* (Biased) should be lower than for *Einige* (Unbiased) and similarly for *Alle* (Biased) and *Alle* (Unbiased). However, in the case of both quantifiers, participants read the shape term more slowly in the biased conditions compared to the unbiased conditions. When exploring the data in terms of the different pragmatic profiles exhibited by participants, only in the case of pragmatic respondents did we find strong evidence of a difference, namely that these participants read the shape term more slowly in the *Alle* (Biased) condition compared to *Alle* (Unbiased).

Table 2, below, summarizes the main results against the original predictions.

Prediction		Result	
		<i>Alle</i>	<i>Einige</i>
SHAPE	Biased < Unbiased		
		Sem: Biased = Unbiased	Biased > Unbiased
		Prag: Biased < Unbiased	Sem: Biased = Unbiased
			Prag: Biased = Unbiased

Table 2: Summary of the results.

## 3.4 Discussion

All in all, the reported results warrant careful conclusions. At face value, the observed pattern directly contradicts the predictions of pragmatic surprisal theory. But since pragmatic surprisal theory consists of two components, the problem could lie with either component or both. Remember that pragmatic surprisal theory assumes that (i) the reading times on a word are lower for more predictable words (the link function), and (ii) that probability of the next word is in turn influenced by contextual and pragmatic factors, in particular a preference for semantically true and pragmatically informative descriptions of the presented picture. Evidence against this conjunction of assumptions could be evidence against any one, or both, of these ideas.

Previous related work on pragmatic processing in context (e.g. Augurzky et al., 2019; Spychalska, Kontinen, & Werning, 2016) provides evidence for the idea that violations of pragmatic expectations, in the sense we are after here, do correspond with another assumed marker of next-word surprisal, namely the amplitude of an N400 component in ERPs (e.g. Kuperberg & Jaeger, 2016; Kutas & Federmeier, 2011). This suggests that context-induced pragmatic expectations, as conceived here, do seem to inform next-word expectations and may lead to surprisal-related processing difficulties.

There is also evidence that scalar implicature inferences (from *some* to *some but not all*) affect reading speed in self-paced reading studies. For example, Breheny, Katsos, and Williams (2006) found increased reading times on a continuation with a phrase like *the rest* in contexts where a scalar implicature of a preceding occurrence of *some* was more expectable given a textual manipulation of expectation based on general world knowledge (similar results are presented by Bergen & Grodner, 2012). This suggests that self-paced reading times are, in principle, susceptible to pragmatic expectations of a sort.



Given that there is some evidence for the ideas that (i) comprehenders do, at least sometimes, entertain pragmatic expectations of the kind of relevance here, induced by a visually presented context, and also that (ii) the self-paced reading method is susceptible pragmatic factors, one possible explanation for the our results could be that the specific combination of the kind of pragmatic expectations (induced by a context picture), on the one hand, and self-paced reading, on the other hand, does, for some reason or other, not work. It could, for example, be that since self-paced reading is a less natural way of reading than on-screen reading in rapid visual serial presentation (as used, e.g., in comparable EEG studies), the burden on working memory of remembering a complex picture, forming (pragmatic expectations) and reading text in a self-paced manner is too onerous a task. Indeed, to address this general issue more would need to be known about how exactly participants decide to advance to the next chunk of textual information in self-paced reading.

But even if limited processing resources are an issue, this does not straightforwardly reconcile our findings with pragmatic surprisal theory. The most natural effect of limited processing resources, on the assumption that the predictions of pragmatic surprisal theory are basically correct, would arguably be that the predicted difference would be deflated, perhaps to the extent that they are completely unattested in the data, i.e., we should expect lesser or no differences where pragmatic surprisal theory predicts differences. However, we see the exact opposite pattern in the data than what pragmatic surprisal theory predicts. This is unexpected even if we make room for limitations of processing resources due to the complexity of the task.

Based on the data at hand, we are therefore invited to speculate about alternative *post hoc* explanations. One conceivable alternative explanation for the faster reading of the unbiased conditions, compared to the biased conditions, revolves around *strategic allocation of attentional resources*. In this picture, what matters to self-paced reading speed is the immediate relevance of a chunk or word to the assessment of the pragmatic felicity of the unfolding descriptions. In other words, participants may be said to have read the shape terms more carefully, and thus more slowly, if the information at a given sentence region was relevant for assessing the pragmatic felicity of the description. Consider, for instance, a scenario composed of a homogeneous triangle array and a heterogeneous circle array (Fig. 1(a)): having read *Einige der* (English: Some of the), even though a participant might be biased, as per the design, to expect *circles* next, she needs to know exactly whether *triangles* or *circles* are actually referred to – if *triangles*, the description is very likely to be either underspecified or downright false; if *circles*, then the load of determining pragmatic felicity is shifted to the the subsequent critical region. Similarly, in the same scenario composed of a homogeneous triangle array and a heterogeneous circle array, having read *Alle der* (English: All of the), even though a participant might be biased, by design, to expect *triangles*, the shape term is key in determining whether the unfolding description is congruent or not: if *circles* are referred to, the description is rendered false at that very sentence region; if, instead, *triangles* are referred to, then the load of determining the congruency of the description is shifted to the the subsequent critical region. In essence, a shallower semantic parsing seems to be, on the face of it, a more reasonable explanation for the patterns found in our data.

Strategic allocation of attentional resources might explain the observed difference between unbiased conditions (read fast because they are irrelevant to the truth-judgement of the sentence) and biased conditions (potentially relevant information at the shape position to the truth-judgement of the sentence). This idea also explains why *Einige (False)* is read faster than *Einige (Infelicitous)*. Notice that the context picture associated with *Einige (False)* is the same as that for *Einige (Unbiased)*. Of course, to explain the increased reading times for conditions where the shape term makes the sentence (most likely) false, possibly by implicature, this alternative explanation must also stipulate a reading time increment for falsity.

Ultimately, then, our results seem to suggest that predictability in online pragmatic processing is linked not only to purely pragmatic-based prediction but also to other processing constraints, such as, in our case, one derived from a pressure to integrate crucial semantic information incrementally during processing. This is in line with constraint-based accounts of language comprehension which assume not only that multiple sources of information need to be integrated online during processing but perhaps more importantly that the weight of different constraints varies depending on the specific processing demands as well as on the larger discourse and communicative context (Degen & Tanenhaus, 2015; Levy, 2008). In fact, other studies on online pragmatic processing have also shown that pragmatic inferencing, including scalar inferencing cued by quantifiers, shows variable time courses and strong context-dependence (Bergen & Grodner, 2012; Huang & Snedeker, 2018; Urbach et al., 2015). What these different studies seem to agree on is that there is immense

variability in how pragmatic interpretation works, from its contexts of occurrence, to its online signatures, and ultimately its underlying mechanistic processes.

In conclusion, our data seem to conflict with the predictions of pragmatic surprisal theory. While limitations of processing resources might be relevant for our particular experimental design, it is not obvious how taking these into account would reconcile our empirical findings with pragmatic surprisal theory. We tentatively considered an alternative explanation, close in spirit to pragmatic surprisal theory, according to which reading times are a function of the strategic allocation of attentional resources, a process which is itself informed by context-induced pragmatic considerations.

## 4 General discussion and conclusion

The results of the reading task suggest that there is no surprisal-based pragmatic prediction in the online processing of referring expressions containing scalar quantifiers, though it remains uncertain to what extent task-specific constraints might have modulated the specific processing signatures we had expected to find as a result of pragmatic surprisal theory. The results of the production task confirm the intuition that people have very clear preferences when it comes to the production of referring expressions containing scalar quantifiers. On the one hand, in the absence of clearly defined communicative goals, these preferences seem to reflect biases which may or may not be related to pragmatic reasoning, such as, in our case, the bias for referring to homogeneous arrays when presented with a scene composed of both a homogeneous and a heterogeneous array. On the other hand, however, the preferences we found seem to suggest that people do indeed rely on a pragmatic principle of informativeness when producing referring expressions, as evidenced by participants' preference to refer to heterogeneous arrays in mixed scenes using terms which are congruent exclusively with those arrays.

In summary, the production results are in line with the assumptions we have entertained at the onset of this paper, namely that semantically congruent and pragmatically felicitous utterances are preferred over incongruent and infelicitous alternatives, as evidenced by the higher production frequencies of terms which result in informative, as opposed to underinformative, referring expressions. However, these production frequencies do not seem to map straightforwardly onto the processing patterns found in the reading task.

Combining the insights gained from our analyses of the reported data, it remains uncertain to what extent language users bring along with them to the task of reading quantified referring expressions biases similar to those displayed when producing such expressions, although their processing might, in principle, be guided by considerations related to the strategic allocation of attentional resources, which itself might interact or perhaps even conflict with pragmatic considerations of a Gricean nature. Ultimately, future investigation should provide more conclusive evidence as to whether pragmatic surprisal theory provides a satisfactory explanation of online incremental pragmatic processing.

## References

- Augurzky, P., Franke, M., & Ulrich, R. (2019). Gricean expectations in online sentence comprehension: An erp study on the processing of scalar inferences. *Cognitive science*, 43(8), e12776.
- Bergen, L., & Grodner, D. J. (2012). Speaker knowledge influences the comprehension of pragmatic inferences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 38(5), 1450.
- Bicknell, K., & Rohde, H. (2009). Dynamic integration of pragmatic expectations and real-world event knowledge in syntactic ambiguity resolution. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 31).
- Boudewyn, M. A., Long, D. L., & Swaab, T. Y. (2015). Graded expectations: Predictive processing and the adjustment of expectations during spoken language comprehension. *Cognitive, Affective, & Behavioral Neuroscience*, 15(3), 607–624.
- Breheny, R., Katsos, N., & Williams, J. (2006). Are generalised scalar implicatures generated by default? An on-line investigation into the role of context in generating pragmatic inferences. *Cognition*, 100(3), 434–463.
- Bubic, A., Von Cramon, D. Y., & Schubotz, R. I. (2010). Prediction, cognition and the brain. *Frontiers in human neuroscience*, 4, 25.
- Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, 80(1), 1–28. doi: 10.18637/jss.v080.i01
- Clark, A. (2013). Whatever next? predictive brains, situated agents, and the future of cognitive science. *Behavioral and brain sciences*, 36(3), 181–204.
- Degen, J., & Tanenhaus, M. K. (2015). Processing scalar implicature: A constraint-based approach. *Cognitive science*, 39(4), 667–710.
- Degen, J., & Tanenhaus, M. K. (2016). Availability of alternatives and the processing of scalar implicatures: A visual world eye-tracking study. *Cognitive science*, 40(1), 172–201.
- Geurts, B. (2010). *Quantity implicatures*. Cambridge University Press.
- Grice, H. P. (1975). Logic and conversation. *Syntax and semantics*, 3, 41–58.
- Grice, H. P. (1989). *Studies in the way of words*. Harvard University Press.
- Haase, V., Spsychalska, M., & Werning, M. (2019). Investigating the comprehension of negated sentences employing world knowledge: an event-related potential study. *Frontiers in psychology*, 10, 2184.
- Horn, L. (1989). A natural history of negation.
- Huang, Y. T., & Snedeker, J. (2009). Online interpretation of scalar quantifiers: Insight into the semantics–pragmatics interface. *Cognitive psychology*, 58(3), 376–415.
- Huang, Y. T., & Snedeker, J. (2018). Some inferences still take time: Prosody, predictability, and the speed of scalar implicatures. *Cognitive psychology*, 102, 105–126.
- Kamide, Y., Scheepers, C., & Altmann, G. T. (2003). Integration of syntactic and semantic information in predictive processing: Cross-linguistic evidence from german and english. *Journal of psycholinguistic research*, 32(1), 37–55.
- Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, cognition and neuroscience*, 31(1), 32–59.
- Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the n400 component of the event-related brain potential (erp). *Annual Review of Psychology*, 62, 621–647.
- Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, 106(3), 1126–1177.
- Litwin, P., & Miłkowski, M. (2020). Unification by fiat: arrested development of predictive processing. *Cognitive Science*, 44(7), e12867.
- Mani, N., & Huettig, F. (2012). Prediction during language processing is a piece of cake—but only for skilled producers. *Journal of Experimental Psychology: Human Perception and Performance*, 38(4), 843.
- Nieuwland, M. S. (2016). Quantification, prediction, and the online impact of sentence truth-value: Evidence from event-related potentials. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 42(2), 316.
- Nieuwland, M. S., Ditman, T., & Kuperberg, G. R. (2010). On the incrementality of pragmatic processing: An erp investigation of informativeness and pragmatic abilities. *Journal of memory and language*, 63(3), 324–346.
- Ridderinkhof, K. R. (2017). Emotion in action: A predictive processing perspective and theoretical synthesis.

- Emotion Review*, 9(4), 319–325.
- Scappini, M., Delfitto, D., Marzi, C. A., Vespignani, F., & Savazzi, S. (2015). On the non-incremental processing of negation: A pragmatically licensed sentence-picture verification study with italian (dyslexic) adults. *Nouveaux cahiers de linguistique française*, 32, 45–58.
- Scholman, M. C., Rohde, H., & Demberg, V. (2017). “on the one hand” as a cue to anticipate upcoming discourse structure. *Journal of Memory and Language*, 97, 47–60.
- Smith, N. J., & Levy, R. (2013). The effect of word predictability on reading time is logarithmic. *Cognition*, 128(3), 302–319.
- Spychalska, M., Kontinen, J., & Werning, M. (2016). Investigating scalar implicatures in a truth-value judgement task: evidence from event-related brain potentials. *Language, Cognition and Neuroscience*, 31(6), 817–840.
- Urbach, T. P., DeLong, K. A., & Kutas, M. (2015). Quantifiers are incrementally interpreted in context, more than less. *Journal of memory and language*, 83, 79–96.
- Winkler, I., Denham, S. L., & Nelken, I. (2009). Modeling the auditory scene: predictive regularity representations and perceptual objects. *Trends in cognitive sciences*, 13(12), 532–540.