RESEARCH PAPER

Rational over-specification in visually-situated comprehension and production

Elli N. Tourtouri [ID] · **Francesca Delogu** [ID] · **Les Sikos** [ID] · **Matthew W. Crocker** [ID]

Abstract Contrary to the Gricean maxims of quantity (Grice, in: Cole, Morgan (eds) Syntax and semantics: speech acts, vol III, pp 41–58, Academic Press, New York, 1975), it has been repeatedly shown that speakers often include redundant information in their utterances (over-specifications). Previous research on referential communication has long debated whether this redundancy is the result of speaker-internal or addressee-oriented processes, while it is also unclear whether referential redundancy hinders or facilitates comprehension. We present an information-theoretic explanation for the use of over-specification in visually-situated communication, which quantifies the amount of uncertainty regarding the referent as entropy (Shannon in Bell Syst Tech J 5:10, https://doi.org/10.1002/j.1538-7305.1948.tb01338.x, 1948). Examining both the comprehension and production of over-specifications, we present evidence that (a) listeners' processing is facilitated by the use of redundancy as well as by a greater reduction of uncertainty early on in the utterance, and (b) that at least for some speakers, listeners' processing concerns influence their encoding of over-specifications: Speakers were more likely to use redundant adjectives when these adjectives reduced entropy to a higher degree than adjectives necessary for target identification.

Keywords Maxim of quantity · Over-specification · Entropy reduction · Redundancy · Audience-design · Index of cognitive activity

Introduction

In situations that require interlocutors to collaborate in order to manipulate objects around them, the visual environment plays a crucial part in establishing reference and creating meaning. Thus, in order to successfully identify a target object in a visually-situated communication task, speakers need to mention precisely those properties that distinguish an intended referent from the other objects in the environment. For example, in a setting that contains a blue and a green ball, the expression 'the ball' fails to select a referent (it is under-specified), because the mentioned property (shape) is shared between the two objects. By contrast, the modified expression 'the blue ball' (which is minimally-specified) successfully establishes reference, by exploiting a contrast between the two objects. According to Grice's maxims of quantity (Grice 1975, 1989), in order for communication to be successful speakers' expressions should

E. N. Tourtouri (✉) · F. Delogu · L. Sikos · M. W. Crocker
Department of Language Science and Technology (FR4.7), Saarland University, 66123 Saarbrücken, Germany
e-mail: elli@coli.uni-saarland.de

convey the minimal amount of information that is necessary—no less (first maxim), and crucially no more (second maxim)—unless speakers intend their listeners to infer some implicit meaning (implicature). Returning to the previous example, if the visual context contained a blue ball and a green mug, the expression 'the blue ball' would no longer be appropriate (it is over-specified) since the target referent could be disambiguated by mention of its shape alone. In this case, the adjective 'blue' would be redundant.

Research on reference production has nevertheless shown that speakers frequently use such over-specified expressions to refer to singleton objects (see Engelhardt et al. 2006 for a 10–60% estimation), thereby violating the second maxim of Quantity. Even though the Gricean theory does not make any predictions regarding the online processing of utterances that violate the maxims, it does have implications for the addressees (Grice 1989), in that they should expect speakers to observe the conversational principle and the maxims that follow from it. Redundant information may therefore engage addressees in unintended pragmatic inferencing (e.g., in the previous example, that a second ball is relevant but not visible to them), which might lead to comprehension difficulties (cf. Sedivy et al. 1999). This point raises an important question: What motivates speakers' tendency to include redundant information in their utterances, if doing so may impede comprehension for their listeners? As we will see below, however, it is still under debate whether over-specification hinders comprehension (e.g., Davies and Katsos 2013; Engelhardt et al. 2006; Engelhardt et al. 2011) or not (e.g., Arts et al. 2011a; Tourtouri et al. 2015).

Whatever the effect of over-specification on comprehension, the frequent use of redundancy poses a challenge to traditional *Gricean* pragmatic accounts of communication, which argue that rational speakers should not over-specify, when their goal is to merely establish reference to an object. More recent *bounded-rational* approaches to communication (cf. Hale 2003, 2006; Frank and Jaeger 2008; Jaeger 2010; Levy and Jaeger 2007) may be better suited to explain this behaviour. Hale (2006) argues that processing effort is proportional to the reduction of uncertainty (entropy; Shannon 1948) about upcoming material in a sentence. In addition, Levy and Jaeger (2007) propose that due to cognitive resource limitations, peaks in the amount of information conveyed by words can increase processing effort for the addressees, and consequently speakers' production choices are motivated by an intent to distribute this information (and thereby processing effort) more evenly across their utterances. Under these accounts, redundant expressions may be preferred to minimal descriptions, because they distribute the same content across more linguistic units, which strengthens the signal and provides addressees with additional cues to guide visual search, thereby making target identification faster and less effortful.

In this article, we crucially consider both the comprehension and production sides of visually-situated communication. In a first experiment, we seek to determine the impact of over-specification on situated comprehension, in order to gain a better understanding of what factors may contribute to speakers' use of redundancy in a subsequent referential communication experiment. In both experiments, we manipulate the contribution of a word to the reduction of uncertainty regarding the target referent, which we quantify as entropy (Shannon 1948). We then examine how the rate of entropy reduction in a referring expression may influence listeners' comprehension above and beyond any effects of specificity, and whether an adjective's potential to reduce uncertainty can explain its redundant use by speakers.

Production of over-specified referring expressions

Since Grice put forth his cooperative principle, much work has investigated whether speakers do in fact observe the conversational maxims in everyday language use (e.g., Arts et al. 2011b; Belke and Meyer 2002; Davies and Katsos 2013; Deutsch and Pechmann 1982; Engelhardt et al. 2006, 2011; Koolen et al. 2011; Koolen et al. 2013; Koolen et al. 2015; Maes et al. 2004; Pechmann 1989; Rubio-Fernández 2016; Tarenskeen et al. 2015; Vogels et al. 2019; Vonk et al. 1992, among others). Despite the different visual settings, tasks or languages employed, these studies share a common finding: that speakers frequently use redundant information in their referring expressions. This redundancy is not in line with (a strict interpretation of) the Gricean maxims (see however Bach 2006; Geurts and Rubio-Fernández 2015), especially when compared to the low proportion of under-specifications (violations of the first maxim of Quantity). The consistency with which over-specification

appears in referential communication gives rise to the question: Why do speakers over-specify?

Generally speaking, two kinds of explanations have been offered, namely, that over-specification is the result of production-internal processes (*egocentric view*) or that it is addressee-oriented (*audience-design view*) (cf. Arnold 2008). Under the egocentric view, in the presence of a visual display that contains referents differing in various attributes, speakers may start to speak before they have fully scanned the display for possible competitors to their intended referent; they may therefore include attributes that turn out to be unnecessary (cf. Pechmann 1989). It is also possible that in the interest of easing attribute selection and production processes, speakers use features that are visually salient and therefore preferred, such as colour (cf. Belke and Meyer 2002; Koolen et al. 2015; among others). By contrast, the audience-design account holds that speakers include redundant information in an effort to facilitate comprehension for their addressees, for instance by including properties that are visually salient or those which allow the addressees to create a mental image of the target to guide their visual search (cf. Arts et al. 2011b; Paraboni et al. 2007).

To determine the extent to which egocentric or audience-design concerns underlie referential over-specification, past research has tried to identify which factors contribute to the use of redundancy. That is, if speakers are found to over-specify more frequently when the experimentally manipulated factors are associated with the addressees' performance, this should constitute evidence for the audience-design view. Two studies that manipulated exactly this found a higher over-specification rate for speakers who thought that their addressees had to carry out a demanding task such as *performing* a surgical operation (Arts et al. 2011b), or *learning* instructions on how to set an alarm clock (Maes et al. 2004).

On the other hand, if the use of over-specification is influenced by factors that are mostly relevant to the speaker, this would provide evidence for the egocentric view. Previous work has also manipulated such factors. For instance, properties of the target object such as *cardinality* (Koolen et al. 2011), or perceptual features such as colour *salience* (Belke and Meyer 2002; Belke 2006; Tarenskeen et al. 2015) have been shown to affect the rate of over-specifications produced. Other research underlines the role of *availability* in the production of redundant adjectives, such

that properties that are conceptually more available to the speaker, such as colour or category, tend to be redundantly included in object descriptions more frequently (e.g., Schriefers and Pechmann 1988).

As speakers need to contrast the intended referent with the distractor objects in order to identify the properties in which they differ, the role of the visual context in the production of over-specified reference has also been investigated. For instance, some studies (Gatt et al. 2017; Rubio-Fernández 2016) have found that the rate of over-specification increased with context *size* (number of distractors). Furthermore, *scene variation* was also shown to play a role, as it was more likely for speakers to produce redundant colour adjectives in polychrome compared to monochrome displays (Rubio-Fernández 2016), and in displays where more distinguishing properties were relevant for the disambiguation of the target referent (Koolen et al. 2013). Finally, the presence of *visual clutter* (thematically related objects) was also shown to contribute to the production of redundant references (Koolen et al. 2015). All these factors are, however, related to perceptual characteristics of the referents, and it is possible that while they ease attribute selection for the speaker, they may also facilitate visual search and identification of the target for the listener.

In sum, despite Gricean expectations speakers frequently over-specify in referential production studies, but there is no agreement regarding whether the use of redundancy is driven by egocentric or audience-design considerations.

Comprehension of over-specified referring expressions

As mentioned above, existing research is divided over whether referential redundancy impedes comprehension or not. Some studies report that over-specification hinders listeners' online processing and results in slower and less accurate identification of the target referent (cf. Davies and Katsos 2013; Engelhardt et al. 2011), while other work suggests that over-specification may even facilitate comprehension (cf. Arts et al. 2011a; Brodbeck et al. 2015; Tourtouri et al. 2015).

For instance, in an event-related brain potential (ERP) study, Engelhardt et al. (2011) found that when visual scenes contained two objects of different shapes, redundant prenominal adjectives (colour and

size) yielded larger N400-like amplitudes time-locked to the onset of the adjective when compared to scenes with two objects of the same shape (i.e., where the adjective was required for identifying the target). The N400 component is generally thought to reflect the degree to which the context supports semantic processing, and larger N400 amplitudes are associated with increased processing difficulty (see Kutas and Federmeier 2011 for a review). Therefore, Engelhardt and colleagues took this N400-like effect to indicate that over-specification hampers comprehension. The observation of this effect may, however, hinge on the simplicity of the visual context. Namely, it is possible that extra information was strikingly redundant with visual contexts as highly simplified as the ones used in this experiment (only two objects appeared in the visual scene, differing in a maximum of two features). Moreover, any effects of over-specification might also emerge on the following noun region, while Engelhardt and colleagues only focused on the adjective. In a similar vein, Davies and Katsos (2013) found evidence that over-specification was dispreferred by listeners as indicated by the lower ratings and longer response times for over-specified compared to minimally-specified utterances. Material in this study, however, comprised expressions containing evaluative and size adjectives, which are known to invoke a contrastive interpretation (cf. Sedivy et al. 1999; Sedivy 2003, 2005).

Other offline and online experiments offer evidence in the opposite direction, namely that over-specification facilitates comprehension. Arts et al. (2011a), for instance, showed that referential redundancy might, in fact, be beneficial to understanding, and ease participants' identification of the target referent. However, they only measured identification times after participants were exposed to the linguistic stimulus, thus no conclusions about the online processing of the referring expression can be drawn. Another ERP study (Tourtouri et al. 2015) provides further support for the notion that over-specification facilitates processing. In this experiment, participants viewed visual stimuli presenting six objects that differed in colour and pattern, and listened to concurrent spoken instructions to locate one of the objects (e.g., 'Find the yellow bowl'), while their EEG was recorded. The presence of a shape competitor on the scene made the instruction minimally-specified, while when no competitor was available the instruction was over-specified. The

authors found no difference between the two conditions in the adjective region, and an attenuated N400 for the over-specified compared to the minimally-specified condition in the noun region. In this study, however, visual displays in the over-specified condition always depicted exactly one object matching the property mentioned in the adjective, but this was not the case in the minimally-specified condition. In other words, when listeners were instructed to 'Find the yellow bowl', there was only one yellow referent in over-specified displays (i.e., the only object that was yellow was the bowl), but two yellow referents in minimally-specified displays (i.e., apart from the bowl, a yellow watering-can was also available). As the authors note, the facilitation observed for the over-specified compared to the minimally-specified instructions may merely be due to the relative predictability of the target referent in the two conditions. That is, in the over-specified but not in the minimally-specified condition, after hearing the adjective participants were able to predict the upcoming word. It is therefore possible that processing of over-specification would have been hindered had the scene included a second yellow object, as was the case in the minimally-specified condition, and even more so if this competitor object fitted Gricean considerations (i.e., was part of a contrast pair, thus making an adjective necessary).

Sedivy et al. (1999) manipulated exactly this factor in a visual-world eye-tracking study using either colour or size prenominal adjectives.[1] They report shorter fixation latencies to the target when it was part of a contrast pair (minimally-specified) compared to when it was not (over-specified). The authors interpreted this finding as evidence that participants readily used pragmatic inferencing to inform their interpretation of the utterance as it unfolded. It is, however, possible that this result was due to the specific experimental task rather than listeners' contrastive interpretation of the adjective. Visual scenes consisted of four objects: a contrast pair differing in one feature (e.g., a yellow and a pink comb), and two singletons, one bearing the same feature as the target (e.g., a yellow bowl) and a distractor object. While the critical instruction mentioned one of the two referents with the shared feature (i.e., one of the yellow objects), it always came second after an instruction that referred

---

[1] Focus stress was also manipulated without, however, yielding significant results.

to one object in the contrast pair. Therefore, an alternative interpretation of the results is that participants were faster to fixate the target when a contrasting object was available (that is, when the instruction was minimally-specified) because their attention was already allocated to the contrast pair. Two additional experiments in which the critical instruction came first yielded similar results, but these studies used scalar adjectives such as 'tall', which inherently invoke a comparison between the members of a contrast pair.

In sum, there is conflicting evidence regarding the comprehension of over-specifications, with some studies suggesting that over-specification hinders comprehension and others indicating a facilitation. This evidence, however, comes from experiments that vary in the size of the referent set, adjectives used, and crucially whether a competitor object fitting Gricean expectations was available in the visual scene. Each of these factors may have contributed to the observed effects.

The current study

The goal of the current study was to explore how the distributional properties of the visual context may (a) influence the comprehension of over-specifications (i.e., whether an adjective's entropy reduction potential influences comprehension above and beyond specificity), and (b) affect the tendency of speakers to include redundant adjectives in their utterances.

In Experiment 1, we investigated the influence of referential specificity and entropy reduction on visually-situated comprehension by orthogonally manipulating these factors. In order to assess processing effort, we measured the index of cognitive activity (ICA)—a direct measure of cognitive load (see end of "Referential entropy reduction")—as well as eye movements as participants followed auditory instructions to locate objects in a visual scene. While the instructions always included a prenominal adjective, we manipulated whether the intended referent was a singleton (over-specified reference) or was part of a contrast set (minimally-specified reference), in order to assess whether listeners compute Gricean pragmatic inferences online and whether their comprehension of the expression is adversely affected when expectations based on those inferences are not met. As in Sedivy et al. (1999), both types of referents (singleton and contrasted) were available in the scene regardless of

whether instructions were minimally-specified or over-specified. In addition, we examined whether the rate of referential entropy reduction in the expression would further influence processing, and whether this influence is additive to any effects of specificity. We turn to this point in the next section.

Concerning production, Experiment 2 evaluated whether the entropy reduction potential of a property (colour or pattern) in the referential space would influence speakers' redundant mention of this property. In other words, speakers may over-specify for a feature of the target referent not only because it stands out, but also based on the extent to which it reduces listener uncertainty about which object is the intended referent. For instance, speakers may be inclined to redundantly use an adjective such as 'blue' to identify a singleton object, not only because the colour blue is a salient property, and therefore easy to refer to, but also because it may help narrow down the referential space: If the set of objects that 'blue' selects is smaller than the set of other objects, the redundant mention of 'blue' before the noun would rapidly restrict the search space and at the same time distribute the effort of target identification over a longer sequence of linguistic elements. If, however, the blue objects outnumbered other objects, 'blue' would not be as effective as before in reducing uncertainty (the number of remaining referential candidates after hearing 'blue' would in this case be greater than before). Although a few recent studies have considered similar notions, such as *discriminability*, and their effects on referential over-specification (Koolen et al. 2015; Fukumura 2018; Vogels et al. 2019), none of these studies directly manipulated such factors.

Thus, Experiment 2 investigates whether and how the distributional properties of the visual scene influence the production of referential over-specification by carefully manipulating the potential of a word to reduce entropy (uncertainty regarding the target referent; cf. Hale 2006; Frank 2013). Identifying which property is more entropy—reducing in order to include it in a description is arguably more demanding for the speakers, than just relying on simple heuristics, such as mentioning the most salient feature. Our hypothesis, therefore, is that over-specifications that include the most informative property—in terms of uncertainty reduction—aim at making visual search more effective for addressees and thus facilitate referential communication. As this hypothesis rests
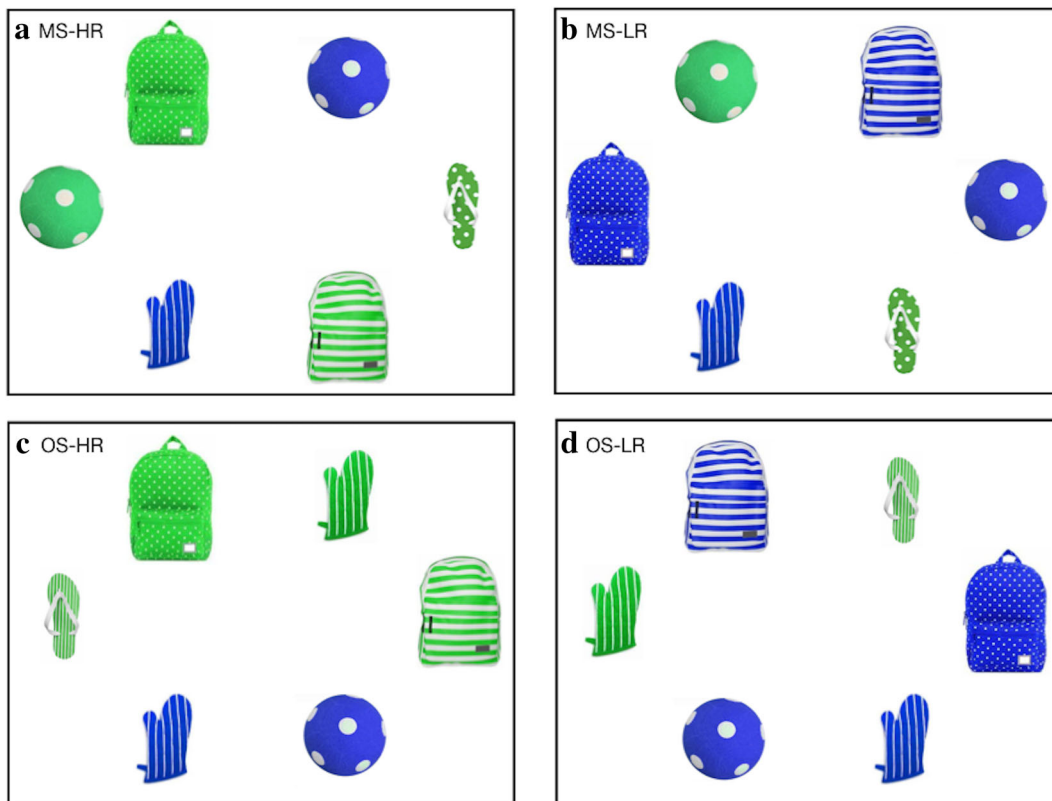
Fig. 1 Experiment 1. Sample visual displays from a colour experimental item, paired with the instruction 'Find the blue ball'. When the display presented a shape competitor (cf. the green ball in **a**, **b**), the instruction was minimally specified (MS). When the target object was of unique shape (cf. one ball in **c**, **d**), the instruction was over-specified (OS). Additionally, when fewer objects matched the mentioned feature (cf. two blue objects in **a**, **c**), the reduction of referential entropy on the adjective was high (HR) compared to when more objects with that feature were present (cf. four blue objects in **b**, **d**), and the adjective resulted in a low reduction of entropy (LR). See the online version of this article for colour figures

upon the extent to which over-specification inhibits or facilitates comprehension processes, we first turn to comprehension, before testing these predictions in production.

Referential entropy reduction

In situated communication, the visual and linguistic context similarly influence listeners' expectations for the upcoming linguistic material in an unfolding utterance (e.g., Altmann and Kamide 1999; Knoeferle et al. 2005; Tanenhaus et al. 1995). For example, when a listener hears 'Find the blue' while immersed in a visual environment such as the one in Fig. 1a, he expects either of two objects to be mentioned next,[2] the ball or the oven mitt. In other words, in this context 'blue' reduces the set of potential referents from 6 to 2 objects and thus drastically reduces listener's uncertainty about the target referent.

We use Shannon's entropy (Shannon 1948), given in (1) below, to quantify this uncertainty regarding the intended referent (*referential entropy*).

$$H(X) = -\sum P(x)log_2 P(x) \qquad (1)$$

In the visual context of Fig. 1a, at 'Find the' (i.e., before any information about the target becomes available), all objects are equally likely to be referred to and referential entropy is *2.58 bits*, as determined by Eq. (1).

---

[2] Throughout the article we use the male pronoun to refer to the listener, and the female pronoun to refer to the speaker, as is convention.

For communication to be successful, the speaker must provide enough information for the listener to reduce this uncertainty to zero. In other words, the listener's mental representation of what the target referent is must move from a state of maximum entropy to a state of minimum entropy, so that by the end of the utterance he will be able to unambiguously identify this object. As the referring expression unfolds over time, incoming words (potentially) contribute to the reduction of referential entropy. This reduction is measured by ΔH, given in (2) below, and is the difference in referential entropy between two consecutive states of the listener's representation (or two consecutive words in the utterance, w-1 and w).

$$\Delta H_w = H_{w-1} - H_w \qquad (2)$$

That is, when 'Find the blue' is uttered in the context of Fig. 1a, referential entropy at 'blue' is 1 bit, and 'blue' reduces entropy by $\Delta H_{blue} = 1.58$ bits. On the other hand, if the expression is 'Find the green', referential entropy at 'green' is 2 bits, and 'green' contributes to the reduction of entropy by $\Delta H_{green-} = 0.58$ bits. That is, while the prenominal adjective in both cases contributes to the reduction of referential entropy, it does so to *differing degrees*, depending on the size of the referential domain each adjective selects. Thus, in situated communication, information conveyed by a word does not only depend on its probability to occur in a particular (visual and linguistic) context (surprisal), but also on the amount of uncertainty about the target referent that this word reduces (cf. Hale 2003, 2006; Frank 2013, for entropy reduction as a measure of processing difficulty outside visually-situated communication).

Hale's (2006) entropy reduction hypothesis linked the reduction of entropy to processing difficulty, suggesting that the effort associated with processing a word should be directly proportional to this word's contribution to the reduction of uncertainty about the rest of the sentence, quantified in bits of information. According to this hypothesis, addressees should experience some difficulty at each entropy reduction point (i.e., on every word in a sentence), but they should encounter greater difficulty the more bits of information this word reduces. This prediction was tested with reading times, both using corpora (Frank 2010, 2013; Wu et al. 2010) and in a self-paced reading experiment (Linzen and Jaeger 2016). Results showed that the rate of entropy reduction brought about by a word was a significant predictor of processing difficulty on that word, with higher reduction resulting in longer reading times. One recent visual world study (Ankener et al. 2018) tested the effects of entropy reduction on the processing of an object noun, based on the selectional restrictions of a preceding verb. That is, when the verb selected fewer objects in the visual scene (high entropy reduction), processing was facilitated on the subsequent noun, as indexed by ICA and visual attention. However, contra the entropy reduction hypothesis, no differences in processing effort were found after the high reduction of entropy on the verb itself.

In the current research, we examine the influence of referential entropy reduction on processing in visually-situated contexts, and seek to determine how the degree of reduction effected by an adjective may modulate listeners' comprehension processes and explain the use of over-specification by speakers. To estimate processing effort we used the index of cognitive activity (ICA), which in Ankener et al. (2018) resulted in reliable results (but see also Demberg and Sayeed, Experiment 7; Sekicki and Staudte 2018; Vogels et al. 2018 for the use of ICA in visual world studies). The ICA is a direct measure of cognitive load that is based on pupillary response. Fluctuations of pupil size index cognitive effort in a variety of tasks, including language processing (e.g., Engelhardt et al. 2010; Frank and Thompson 2012; Just and Carpenter 1993; Scheepers and Crocker 2004). However, changes in the lighting conditions of the environment are also responsible for pupil dilation. The ICA (Marshall 2000) measures cognitive workload by separating variation in pupil size due to cognitive effort and due to light reflex, while also accounting for random noise. The small and rapid pupil dilations that remain are associated with higher cognitive workload (Marshall 2002). Demberg and Sayeed (2016) showed, for example, that the ICA is sensitive to linguistic manipulations such as ungrammaticality, with conditions related to higher processing demands resulting in higher ICA values. They also demonstrated that the ICA is particularly suitable for the visual world paradigm since it is robust to the change of fixation positions and can thus complement the standard visual attention metrics in order to assess cognitive effort during linguistic processing.

Experiment 1

Experiment 1 aimed to establish whether referential over-specification impedes or facilitates comprehension, and also whether this is further modulated by the rate of entropy reduction in the expression. We recorded participants' ICA and eye movements as they attended to audio instructions to locate a referent in a visual scene (e.g., 'Find the blue ball' in German, combined with displays such as those in Fig. 1). While the instruction was held constant, scenes differed in whether the intended referent belonged to a contrast set (cf. Fig. 1a, b, where a shape competitor is available) or it was a singleton (cf. Fig. 1c, d, where there is no shape competitor). Thus, depending on the visual context, the prenominal adjective was either *necessary* or *redundant*, and the description *minimally-specified* (MS) or *over-specified* (OS), respectively. In addition to specificity, we manipulated entropy reduction, that is the number of objects that matched the adjective (cf. two blue objects in Fig. 1a, c and four blue objects in Fig. 1b, d). Thus, the adjective restricted the set of potential referents to a greater or lesser degree, contributing to a high reduction (HR) of referential entropy (1.58 bits in Fig. 1a, c) or a Low Reduction (LR) of referential entropy (0.58 bits in Fig. 1b, d), respectively. Importantly, this reduction resulted in a smaller (1 bit) or larger (2 bits) amount of residual entropy, respectively, to be eliminated at the noun. In the analyses below, we report ICA values as a measure of comprehension difficulty, fixation probabilities as a measure of visual attention, and response times for comparisons to prior studies.

We considered two regions of interest: the adjective, and the noun. Note, however, that in the adjective region only the entropy reduction manipulation is of interest, because at this point in the utterance participants were not yet able to determine whether the unfolding expression was minimally- or over-specified. Based on the entropy reduction hypothesis (Hale 2006), we expected to find effects of processing effort at each reduction point, with higher reduction resulting in increased processing difficulty. More specifically, ICA values on the adjective should be higher in HR compared to LR conditions. In contrast, ICA values on the noun should be lower in HR compared to LR conditions, since residual entropy on the noun in the HR condition should be low due to the previous high

reduction of entropy on the adjective. It is, however, possible that we only observe an effect on the noun, as in Ankener et al. (2018), where a verb that selected for fewer objects did not itself elicit increased ICA values, but did nevertheless result in lower processing effort on the subsequent noun. Finally, if redundant prenominal adjectives facilitate processing by reducing referential entropy, this should be manifest in an interaction between specificity and entropy reduction, with a larger benefit (lower ICA values) in the OS-HR condition (cf. Fig. 1c).

Anticipatory eye movements triggered by the adjective might, however, reveal how listeners interpret the prenominal adjective (cf. Weber et al. 2006). Within each level of specificity, displays either contained one *contrast object* that fitted a contrastive reading of the adjective (cf. the blue ball in Fig. 1a, b and the blue mitt in Fig. 1c, d), or one *singleton object* that did not match a contrastive reading (cf. the mitt in Fig. 1a, b and the ball in Fig. 1c, d). If listeners are Gricean (i.e., if they assume that an adjective identifies a contrast rather than simply providing redundant information), then the adjective should trigger more anticipatory eye movements towards the contrast object compared to the singleton object.

Regarding specificity, the Gricean account predicts greater processing effort on the noun (higher ICA values) in OS compared to MS conditions. In contrast, the bounded-rational view does not predict such a difference; redundancy may be preferred because it distributes information (i.e., processing effort) across a longer sequence of linguistic elements. As visual attention (proportion of fixations) is primarily informative regarding expectations of upcoming material, we do not expect it to reveal anything on the noun beyond correct identification of the target.

Methods

Participants

Twenty-four native speakers of German (mean age = 25, 17 female), with normal or corrected-to-normal vision and no colour blindness were recruited through the Saarland University Psycholinguistic Group's participant database. Participants were compensated 7 Euros for their participation.

Materials

Pictures of 30 everyday objects (e.g., mugs, bowls, etc.) were used to create the visual displays. The objects differed in colour (red, blue, green) and pattern (dotted, striped, checkered). Both colour and pattern were, therefore, used as distinguishing features to make sure that any effects would not merely be due to colour salience, but also to increase visual complexity and avoid tuning the task to one feature. Pattern was chosen over size, which is more commonly used (cf. Engelhardt et al. 2011; Sedivy et al. 1999), because pattern, like colour, is an intrinsic property of the object and does not invoke a comparison with other objects in the context. We, thus, made sure that preference for a contrastive reading of the adjective would be due to the manipulation and not to the contrastive nature of size adjectives. GIMP (Version 2.8.10) was used to adjust colour hue and brightness and match them across objects. The pictures were then submitted to an offline picture naming task measuring naming agreement for the objects. Twenty-four independent participants were presented with the object pictures in all colours and patterns (distributed across 8 lists) and were asked to provide a description including colour and pattern. Only objects with naming agreement of 80% or higher were then used to create the visual stimuli.

Overall, 660 visual displays were created, of which 480 were used to construct the experimental items, and the rest were used in the fillers. Experimental items were the combination of 4 displays and one spoken instruction (cf. Fig. 1). Displays in one experimental item were essentially four versions of the same display, counterbalancing the target position within the item (cf. the position of the blue ball in Fig. 1), and the colour and pattern per object type throughout the experiment. This gave rise to 120 experimental items, half of which were paired with colour instructions (colour items), and the other half with pattern instructions (pattern items; cf. Fig. 10 in the Appendix). All experimental displays were created in a way that neither the target feature nor the target referent would be identifiable before hearing the critical words. To this end, six objects were used per display in two colours and two patterns. Two of the objects were singletons, and the rest were paired in two contrast sets, such that they could potentially serve as an over-specified or minimally-specified referent,

respectively, with either a colour or a pattern instruction. Furthermore, because determiners in German are marked for gender, only same-gender objects were used in each display, to make sure that the determiner would not reveal the target and that the first point of entropy reduction would always be the adjective. Similarly, no phonological competitors appeared in the same scene, so that adjective onset would always be the first point of disambiguation across items.

Filler displays differed from experimental displays in several respects. First, 105 filler displays depicted only four objects, thereby introducing some variation in the stimuli set while also making the 6-object experimental displays more complex relative to the filler trials. Furthermore, half of the filler items were minimally-specified, and the other half were either over- or under-specified (with a higher proportion of over- relative to under-specifications). In this way, we introduced more variation in the stimuli requiring the listener to be more attentive (as it could be the case that reference could not be resolved), while maintaining a lower proportion of over-specifications as is normally found in language use (cf. Engelhardt et al. 2006). Moreover, all filler displays apart from the under-specified ones contained a set of three same-shape objects (e.g., three balls) differing for both colour and pattern, thus making the use of a second adjective necessary for disambiguation. Under-specified fillers were similar in structure to the experimental displays, but failed to establish reference (e.g., 'the green rucksack' when two objects fit the description; cf. Fig. 1a and c). Twelve fillers were used as practice items in a familiarisation session before the experiment.

Experimental displays were paired with spoken instructions containing a prenominally modified referring expression like 'Find the blue ball' in German ('Finde den blauen Ball'), while filler instructions could mention one, two or no modifiers. The order of mention of colour and pattern adjectives was counterbalanced in the two-modifier fillers. Audio stimuli were recorded with Cubase AI5 in a soundproof booth by a female native speaker of German. Speech was continuous, and no artificial pauses were inserted between words. Sentences were then annotated for adjective and noun onsets using Praat (Version 5.3). Mean word duration was 397.2 ms ($SD = 49.6$) for colour adjectives, 605.1 ($SD = 75.1$) for pattern adjectives, and 557.2 ms ($SD = 75.7$) for the nouns.

Stimuli were divided into 4 lists of 288 trials so that one version of an item was in each list, and no participants saw more than one condition of a given item. Lists were pseudo-randomised for each participant, making sure that at least one filler appeared between consecutive experimental items, and items of the same condition did not appear more than two times in a row. The experiment was implemented and run using E-prime 2.0 (Psychology Software Tools, Inc., Pittsburgh, PA, USA).

Procedure

Participants' eye movements were tracked at a rate of 250 Hz using an SMI RED 250 eye tracker (SensoMotoric Instruments GmbH, Berlin, Germany) attached to the bottom of a 22-inch Dell monitor. After participants gave informed consent, they read the instructions, and they were seated at a distance of approximately 60 cm in front of the monitor. A chinrest was used to minimise head movements. A familiarisation phase was first administered, during which the experimenter gave feedback after each trial, to make sure that the task was clear before the experiment began. Each experimental session was divided into 4 blocks, in between which participants could take short breaks. Calibration was performed at the beginning of each block. On average, participants needed 40 min to complete the experiment.

Visual stimuli were presented at a resolution of 1680 × 1050 pixels. At the beginning of each trial a cross appeared in the middle of the display for a period controlled by the experimenter. After that, the objects appeared while the cross remained on the screen for another 500 ms. The audio instruction was played 1500 ms later. After the end of the instruction, the objects remained on the screen for a wrap-up period of 500 ms. At the end of the trial, a prompt screen appeared asking participants to indicate which side of the screen the target referent was on, or whether it was not possible to tell (under-specified fillers) by pressing the corresponding button on a response pad in front of them.

Data analysis

We analysed the ICA, gaze probabilities as well as response times in two time windows, after adjective and after noun onset. For all analyses, we fitted (generalised) linear mixed models (lme4 package; Bates et al. 2015) in R (version 3.5.1; R Core Team 2018) including entropy reduction and specificity as well as the Feature (colour vs. pattern) of the target referent as fixed factors, and crossed random intercepts and slopes for participants and items. All factors were contrast coded, with positive contrast coding (0.5) for the levels of HR, MS and colour, and negative contrast coding (− 0.5) for LR, OS and pattern. Whenever the maximal models did not converge, we simplified the random effects structure as suggested by Barr et al. (2013). All analyses included only trials with correct responses.

Response times Response times (RTs) were time locked to the onset of the prompt display. Analyses were carried out on log-transformed response times using linear mixed models.

Index of cognitive activity (ICA) To calculate the ICA we used the BeGaze^{TM} software equipped with the ICA Module (SensoMotoric Instruments GmbH, Berlin, Germany) and Workload RT (EyeTracking, Inc., Solana Beach, CA, USA). Since the ICA values output by the BeGaze^{TM} software are too coarse-grained for the type of effects we expect, we used the ICA Coefficients to compute ICA values per 100 ms (see Demberg and Sayeed 2016, for more details). Data points with a pupil diameter smaller than 2.5 *SD* per participant were eliminated, and a mean ICA value for both eyes was calculated. We compared mean ICA values across conditions within a window of 600 ms starting from the middle of each region (cf. Sekicki and Staudte 2018).

Fixations Eye-tracking data were pre-processed as follows. First, because the objects used in the visual displays could differ in size (cf. rucksack vs. mitt), areas-of-interest were calculated per object as the surface that the object covered on the screen in pixels plus 30 pixels around it. Next, fixations shorter than 80 ms were pooled with the immediately preceding or following fixation, if the distance between them was smaller than 12 pixels; otherwise they were excluded from the analysis. Finally, trials with recording problems (e.g., miscalibrations, track loss, etc.) were excluded from the analysis. For the analysis in the adjective region, to account for the difference in duration of colour and pattern adjectives, we considered a region from 200 ms before adjective offset until
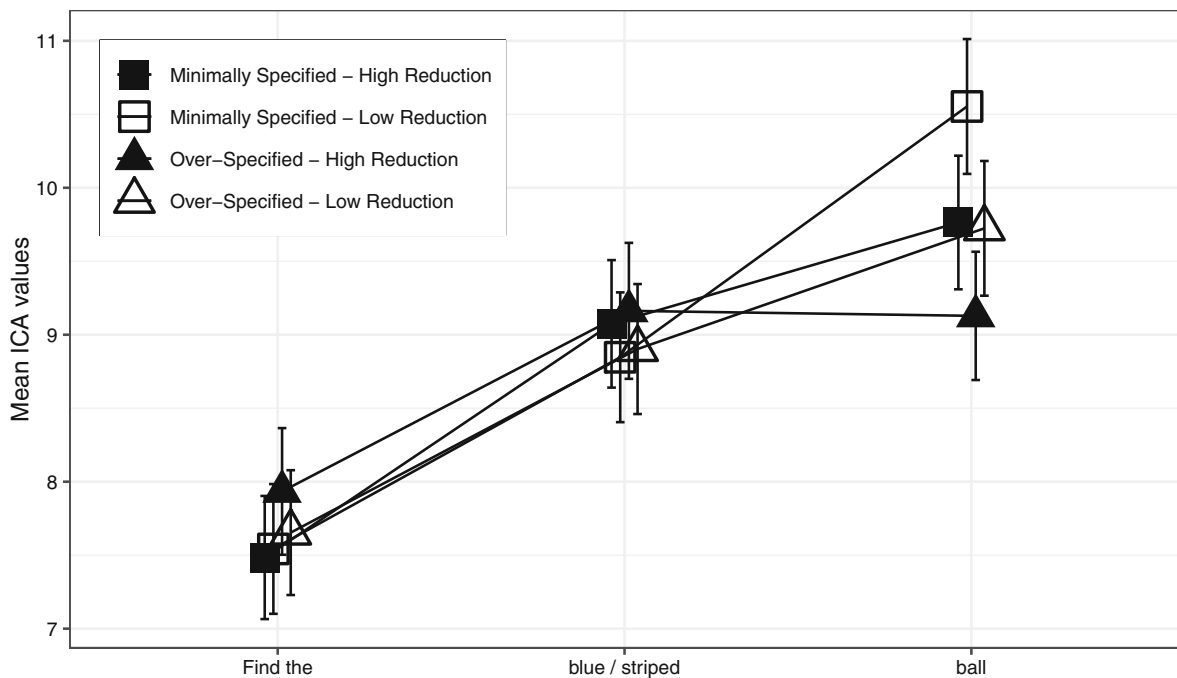
Fig. 2 Experiment 1. Mean ICA values in each condition per region for colour and pattern items combined. A high reduction of entropy (High Reduction; filled shapes) on the adjective resulted in higher ICA values in the adjective region and lower values in the noun region, relative to the Low Reduction conditions (empty shapes). Note that in the noun region specificity and feature also modulated the ICA (see Fig. 3)

200 ms after noun onset,[3] since it is known that it takes around 200 ms to plan and execute a saccade (Matin et al. 1993). As discussed above, the specificity manipulation is not relevant for the adjective, as it is based on information given on the noun. We therefore collapsed across MS and OS conditions, and coded looks to the singleton vs. the contrast objects to estimate whether participants assigned a contrastive reading to the prenominal adjective. For the analysis of eye movements during the noun, we were interested in the influence of specificity and entropy reduction on fixating the target referent, and not in possible early effects (anticipatory eye movements are analysed in the adjective region). We therefore considered fixations that started between 300 and 800 ms after noun onset. In both regions, we considered mean log-gaze probability ratios (cf. Knoeferle and Kreysa 2012) of participants' fixations to (a) the singleton over the contrast object in the adjective region and (b) the target over the competitor object in the noun region. A

positive ratio for (a) would indicate that the singleton object was more likely to be fixated over the contrast object, and a positive ratio for (b) that the target object was more likely to be fixated over the competitor object. Negative values should be interpreted in the opposite way (i.e., as more looks to the contrast object in the adjective region and as more looks to the competitor object in the noun region). A score of zero would indicate no differences in the probability with which each object was fixated. Because the log ratios are based on aggregation, it is not possible to include crossed random effects of participants and items in the same model. We, therefore, fitted separate linear mixed effects models over participants and over items.

Results

Response times

All of the factors included in the model significantly influenced RTs. Participants were faster to give a response in HR (611 ms, $SD = 374$) compared to LR conditions (659 ms, $SD = 397$; $\beta = -0.0796$, $SE =$

---

[3] Recall that speech was continuous, and therefore adjective offset corresponds to noun onset.
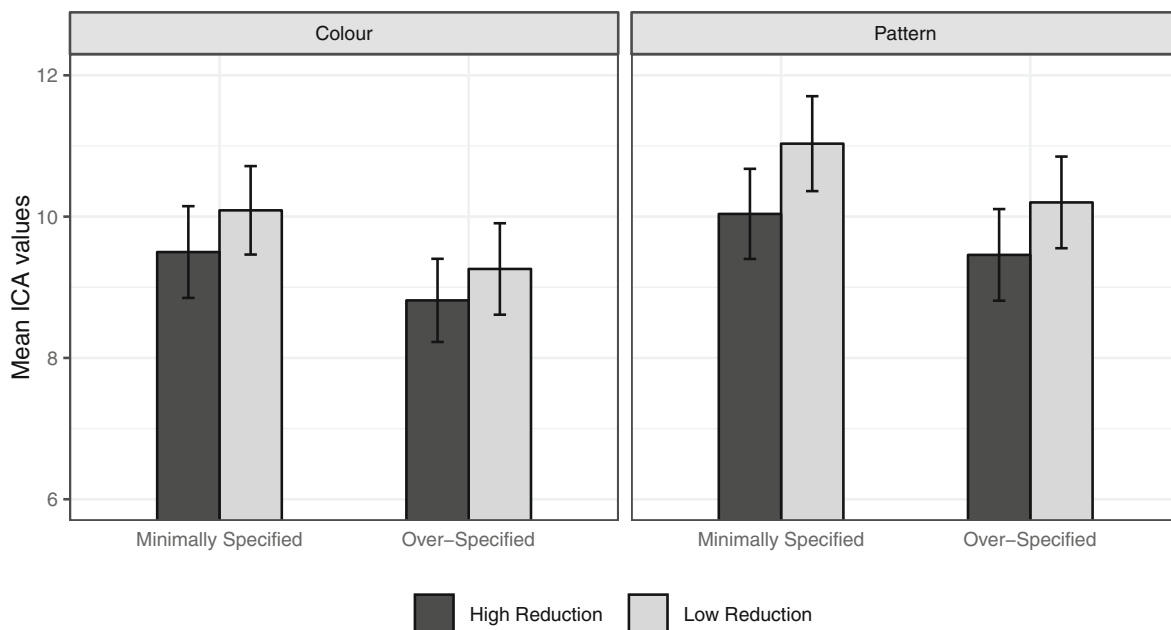
Fig. 3 Experiment 1. Mean ICA values in each condition in the noun region. While ICA values were in general higher in pattern (right panel) than in colour (left panel) items, in both cases a high reduction of entropy (High Reduction; black bars) on the adjective resulted in lower cognitive effort on the noun compared to the Low Reduction conditions (grey bars). Moreover, there was a facilitation for Over-Specified (rightmost bars) compared to Minimally Specified (leftmost bars) descriptions, such that redundant adjectives resulted in lower cognitive effort on the noun

0.0155, t = − 5.14, p < 0.001), and faster in OS (614 ms, $SD = 372$) compared to MS conditions (656 ms, $SD = 398$; $\beta = 0.058$, $SE = 0.016$, t = 3.755, p < 0.001). Faster responses were further observed when the mentioned feature was colour (570 ms, $SD = 323$) compared to pattern (703 ms, $SD = 432$; $\beta = -0.192$, $SE = 0.027$, t = − 7.217, p < 0.001). In addition, the three-way interaction between entropy reduction, specificity and feature significantly influenced RTs ($\beta = 0.135$, $SE = 0.062$, t = 2.181, p < 0.05). We followed up this interaction by fitting separate models for colour and pattern items, and we observed similar results. In colour items, RTs were faster in HR (545 ms, $SD = 306$) compared to LR conditions (594 ms, $SD = 338$; $\beta = -0.086$, $SE = 0.020$, t = − 4.235, p < 0.001), and faster in OS (555 ms, $SD = 323$) compared to MS conditions (584 ms, $SD = 323$; $\beta = 0.053$, $SE = 0.020$, t = 2.651, p < 0.01). Similarly in pattern items, RTs were faster in HR (679 ms, $SD = 423$) vs. LR conditions (726 ms, $SD = 439$; $\beta = -0.073$, $SE = 0.023$, t = − 3.147, p < 0.01), and faster in OS (676 ms, $SD = 409$) vs. MS conditions (729 ms, $SD = 452$; $\beta = 0.064$, SE = 0.023, t = 2.773,

p < 0.01). The entropy reduction × specificity interaction was marginally significant ($\beta = -0.078$, $SE = 0.046$, t = − 1.688, p = 0.092), such that RTs were slower in the MS-LR condition.

ICA

Adjective In the adjective time window (see Fig. 2), the entropy reduction manipulation was found to significantly influence cognitive effort, with higher ICA values in HR vs. LR conditions (β = − 0.026, $SE = 0.013$, $z = -2.068$, p = 0.039). The effect of feature and the interaction between the two factors did not reach significance (p > 0.05).

Noun All of the factors significantly affected participants' cognitive workload in the noun region (Fig. 3). Specifically, we again observed a significant effect of entropy reduction, this time with higher ICA values in LR compared to HR conditions ($\beta = -0.073$, $SE = 0.023$, $z = -3.160$, p < 0.01). Furthermore, specificity and feature were also found to be significant predictors of cognitive load, with higher ICA values for MS compared to OS conditions (β = 0.079, $SE = 0.026$, $z = 3.069$, p < 0.01), and

for colour compared to pattern items ($\beta = -0.076$, $SE = 0.022$, $z = -3.372$, $p < 0.001$). None of the interactions reached significance ($p > 0.05$).

Log-gaze probabilities

Adjective As mentioned above, the specificity manipulation is not relevant in the adjective window (see "Data analysis"). We therefore collapsed across specificity, and included only entropy reduction and feature as fixed factors in the models. We computed log-gaze probability ratios comparing fixations to the *singleton* and *contrast* objects. Table 1 presents the results of this analysis. As indicated by the significant intercept (both by participants and by items), upon hearing the adjective participants were more likely to fixate the contrast object over the singleton object (see negative coefficient). This viewing pattern seemed to be modulated by an interaction between the rate of entropy reduction and the mentioned feature, which we followed up with separate analyses for colour and pattern items. In colour items (Fig. 4), none of the comparisons reached significance; there was only a marginal effect on the intercept in the by-participants analysis. In pattern items (Fig. 5), the contrast object was more likely to be fixated over the singleton, and this effect seemed to be stronger in HR vs. LR conditions.

Noun The specificity manipulation becomes relevant during the noun region, as it is at this point that the target referent is mentioned. We, therefore, considered fixations to the *target* vs. the *competitor* object, and specificity was included as a predictor in the models. The results of these analyses are presented in Table 2. Even though the analysis by participants resulted only in a marginally significant three-way specificity × reduction × feature interaction, and no other comparison reached significance, several significant effects were found in the by-items analysis. First, there was an effect of specificity with more looks to the target over the competitor object in OS vs. MS conditions. We also found an effect of reduction such that the target was more likely to be fixated than the competitor object in HR vs. LR conditions, and an effect of feature with more fixations to the target object in colour vs. pattern items. Additionally, there was a significant specificity × feature interaction with more fixations to the target object in the OS condition for colour items. We followed up the interactions by performing separate analyses for colour and pattern items. In the colour items (Fig. 6), the by-participant analysis resulted only in a marginally significant effect of specificity, with more looks to the target object in OS conditions. The by-items analysis revealed a significant effect of specificity in the same direction and a significant effect of reduction with more looks to the target over the competitor object in HR vs. LR

**Table 1** Experiment 1 results—adjective region

	By participants				By items			
	β	SE	t	p	β	SE	t	p
All items								
Intercept	− 0.243	0.081	− 3.000	0.006**	− 0.168	0.077	− 2.176	0.031*
Reduction	− 0.111	0.119	− 0.926	0.364	− 0.093	0.139	− 0.664	0.508
Feature	0.197	0.112	1.756	0.091.	0.228	0.154	1.481	0.141
Reduction:feature	0.583	0.301	1.938	0.064.	0.483	0.279	1.732	0.086.
Colour								
Intercept	− 0.144	0.073	− 1.985	0.059.	− 0.053	0.111	− 0.482	0.632
Reduction	0.183	0.178	1.030	0.313	0.149	0.210	0.710	0.481
Pattern								
Intercept	− 0.342	0.112	− 2.882	0.008**	− 0.283	0.107	− 2.626	0.011*
Reduction	− 0.402	0.205	− 1.960	0.062.	− 0.335	0.182	− 1.838	0.071.

Coefficients, SE, *t*- and *p*-values for log-gaze ratios of fixations to singleton vs. contrast objects. *** $p < .001$, ** $p < .01$, * $p < .05$, . $p < .1$
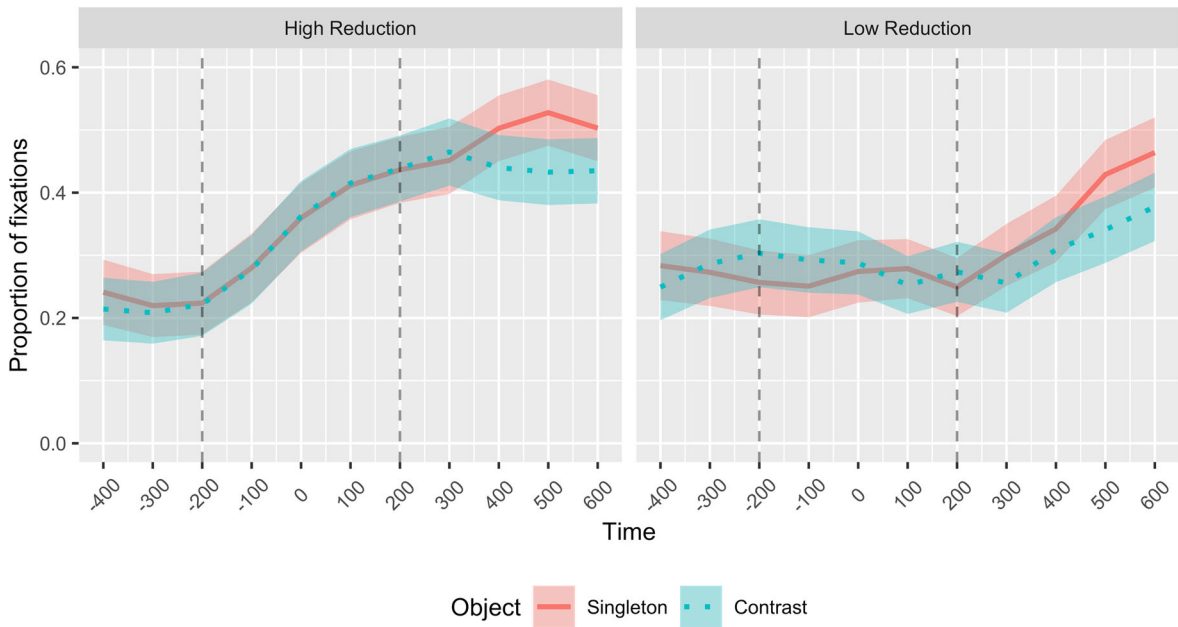
Fig. 4 Experiment 1. Proportion of fixations to the singleton vs. the contrast object in colour items for the High Reduction (left) and Low Reduction (right) conditions. Noun onset is at zero, and the analysis window was from − 200 ms until 200 ms around noun onset (dashed lines). The shaded bands represent 95% CI
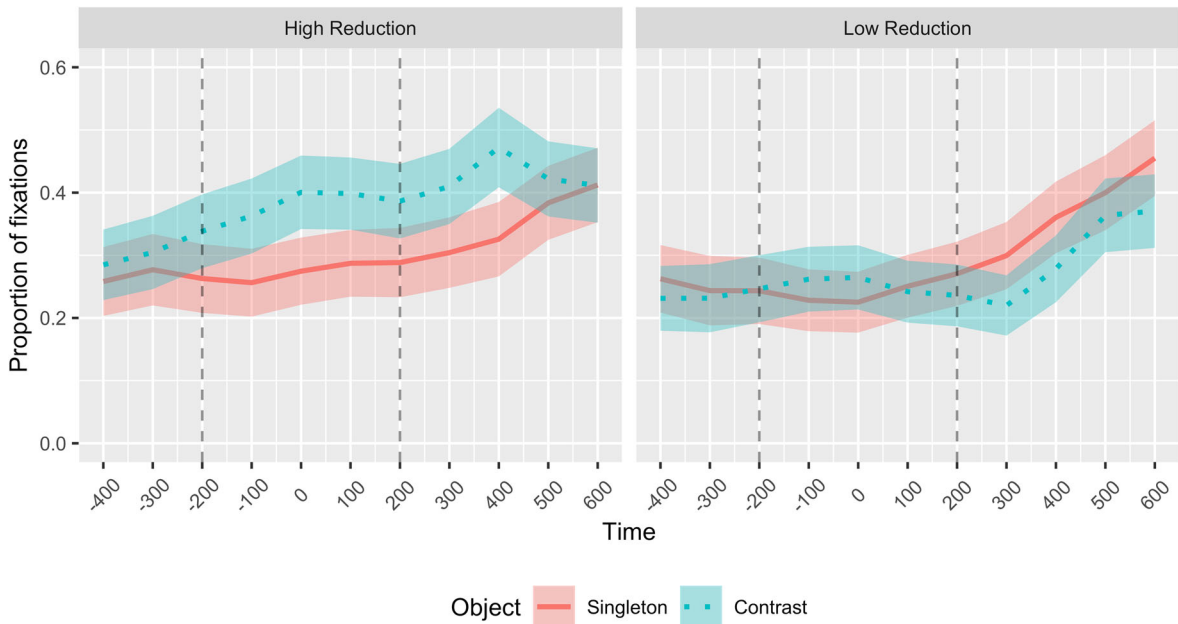
Fig. 5 Experiment 1. Proportion of fixations to the singleton vs. the contrast object in pattern items for the High Reduction (left) and Low Reduction (right) conditions. Noun onset is at zero, and the analysis window was from − 200 ms until 200 ms around noun onset (dashed lines). The shaded bands represent 95% CI

conditions. In the pattern items (Fig. 7), both by-participants and by-items analyses resulted in a specificity × reduction interaction, which was significant and marginally significant, respectively. This interaction seemed to be driven by a smaller log ratio in the MS-LR condition (see Table 3).

Table 2 Experiment 1 results—noun region

|	By participants				By items			
	β	SE	t	p	β	SE	t	p
All items								
Intercept	0.773	0.056	13.825	< 0.001***	0.745	0.0367	20.358	< 0.001***
Specificity	− 0.097	0.066	− 1.463	0.157	− 0.138	0.054	− 2.550	0.012*
Reduction	0.070	0.056	1.345	0.192	0.129	0.065	1.999	0.048*
Feature	0.052	0.062	0.830	0.415	0.159	0.073	2.173	0.032*
Spec:red	0.192	0.120	1.606	0.122	0.153	0.123	1.238	0.218
Spec:feat	− 0.165	0.163	− 1.016	0.320	− 0.229	0.108	− 2.120	0.036*
Red:feat	0.023	0.101	0.226	0.823	0.128	0.130	0.989	0.325
Spec:red:feat	− 0.239	0.129	− 1.849	0.065.	− 0.324	0.247	− 1.315	0.191
Colour items								
Intercept	0.798	0.071	11.283	< 0.001***	0.825	0.056	14.765	< 0.001***
Specificity	− 0.179	0.099	− 1.815	0.083.	− 0.252	0.084	− 2.990	0.004**
Reduction	0.082	0.072	1.138	0.267	0.193	0.087	2.213	0.031*
Spec:red	0.073	0.162	0.447	0.659	− 0.009	0.177	− 0.055	0.957
Pattern items								
Intercept	0.747	0.056	13.243	< 0.001***	0.666	0.047	14.067	< 0.001***
Specificity	− 0.014	0.110	− 0.128	0.899	− 0.023	0.067	− 0.346	0.730
Reduction	0.058	0.073	0.802	0.430	0.065	0.096	0.687	0.495
Spec:red	0.311	0.141	2.213	0.037*	0.314	0.171	1.833	0.072.

Coefficients, SE, t- and p-values for log-gaze ratios of fixations to the target vs. fixations to the competitor object. *** $p < .001$, ** $p < .01$, * $p < .05$, . $p < .1$

Discussion

In this experiment, we aimed to assess whether comprehension of over-specified expressions is hindered or facilitated relative to minimally-specified expressions, and whether the rate at which referential entropy is reduced in the expression further affects processing. In the noun region, we found no evidence that over-specification hinders comprehension. Participants' ICA values were in fact lower in OS vs. MS conditions, indicating that over-specification does not adversely affect comprehension, but if anything over-specification facilitates comprehension. These findings are further supported by the log-gaze probabilities and RTs. Unsurprisingly, participants looked more towards the target than the competitor object after hearing the noun, but this effect was modulated by specificity, such that looks to the target vs. the competitor object were more likely when the noun followed a redundant vs. necessary adjective. Furthermore, participants' RTs in this task were also faster in OS vs. MS conditions. In the adjective region,

anticipatory looks to the singleton vs. contrast objects were expected to reflect participants' interpretation of the adjective. Whereas there is some evidence that contrast objects were fixated more than singletons (supporting the Gricean account), this only occured with pattern adjectives—which are more difficult to discern than colour. It is therefore possible that this effect is related to the length of pattern adjectives, which in this experiment were on average 200 ms longer than colour adjectives. Thus, with pattern adjectives participants may have had more time to consider which object could possibly be the target referent and to employ Gricean reasoning. Nevertheless, participants' gaze behaviour in the adjective region of colour items, as well as the facilitation found for OS conditions with both colour and pattern items in the noun time window, contradict the Gricean account and support the view that over-specification facilitates comprehension.

Moreover, our findings support the entropy reduction hypothesis (Hale 2006) and show that the reduction of uncertainty is a predictor of
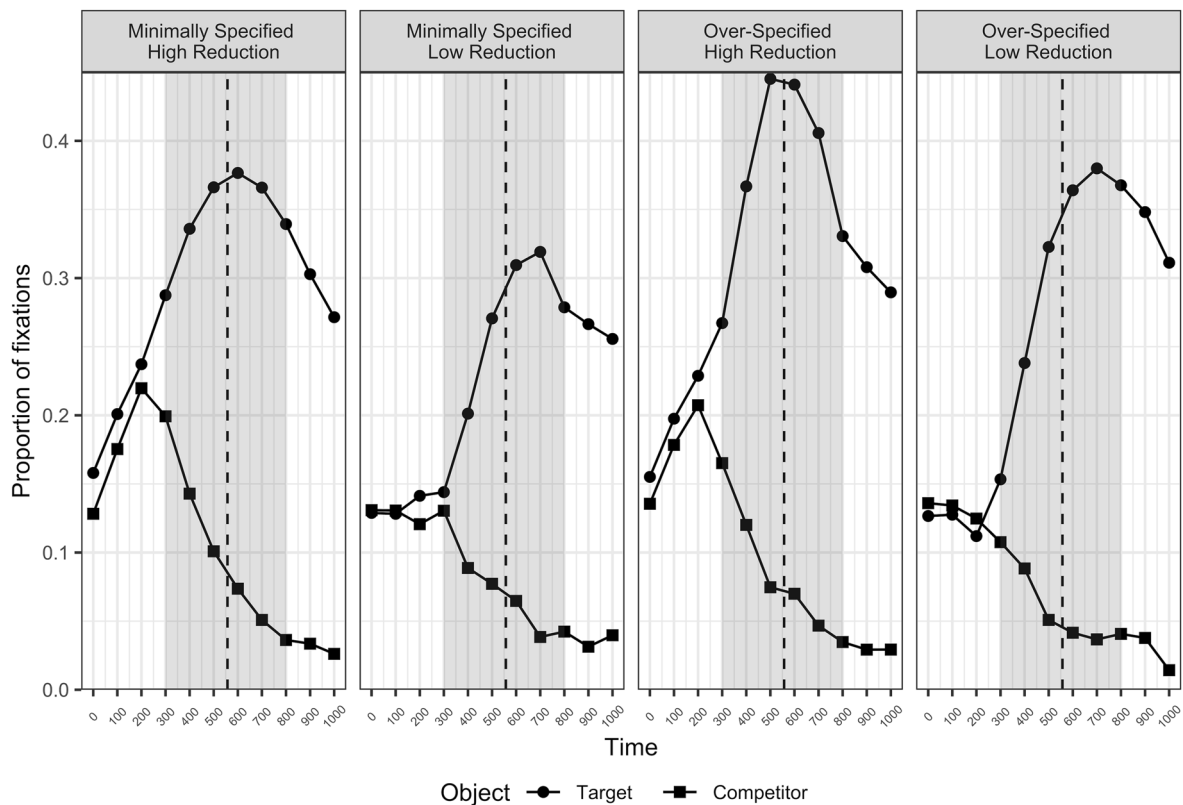
Fig. 6 Experiment 1. Proportion of fixations to the target vs. competitor object in each condition of colour items. Noun onset is at zero. The shaded area represents the analysis window, and the dashed line indicates average noun offset

comprehension difficulty in visually-situated communication. In contrast to Ankener et al. (2018), we found effects of reduction at each reduction point. A high reduction of entropy on the adjective resulted in increased cognitive effort (higher ICA values) in that region, but facilitated processing (lower ICA values) on the following noun; residual entropy on the noun—and the cognitive effort associated with the reduction of this entropy—was correspondingly lower in HR than in LR trials (lower ICA values in HR trials). The facilitation for HR vs. LR conditions was further indexed by the increased likelihood to fixate the target over the competitor object at the noun region in the two conditions, as well as by faster RTs for HR vs. LR conditions.

In sum, Experiment 1 found evidence that referential redundancy benefits processing. Furthermore, a high reduction of entropy on the adjective was found to increase effort in that region (indexed by the higher ICA values for HR relative to LR conditions), but to facilitate processing on the subsequent noun

(ICA values here were lower for HR vs. LR conditions). These effects differed between colour and pattern adjectives, with colour resulting in greater facilitation (main effect of feature with lower ICA values in colour compared to pattern items). We now turn to the question whether speakers are sensitive to these processing concerns and whether they take comprehension effort into account when planning their utterances in situated communication contexts.

Experiment 2

The goal of Experiment 2 was to identify what factors motivate speakers' over-specifications, and whether these factors are primarily associated with egocentric or addressee-oriented (whether Gricean or bounded-rational) concerns. In a referential communication experiment, pairs of participants sat in front of different monitors and collaborated to identify whether the location of a target object on a visual
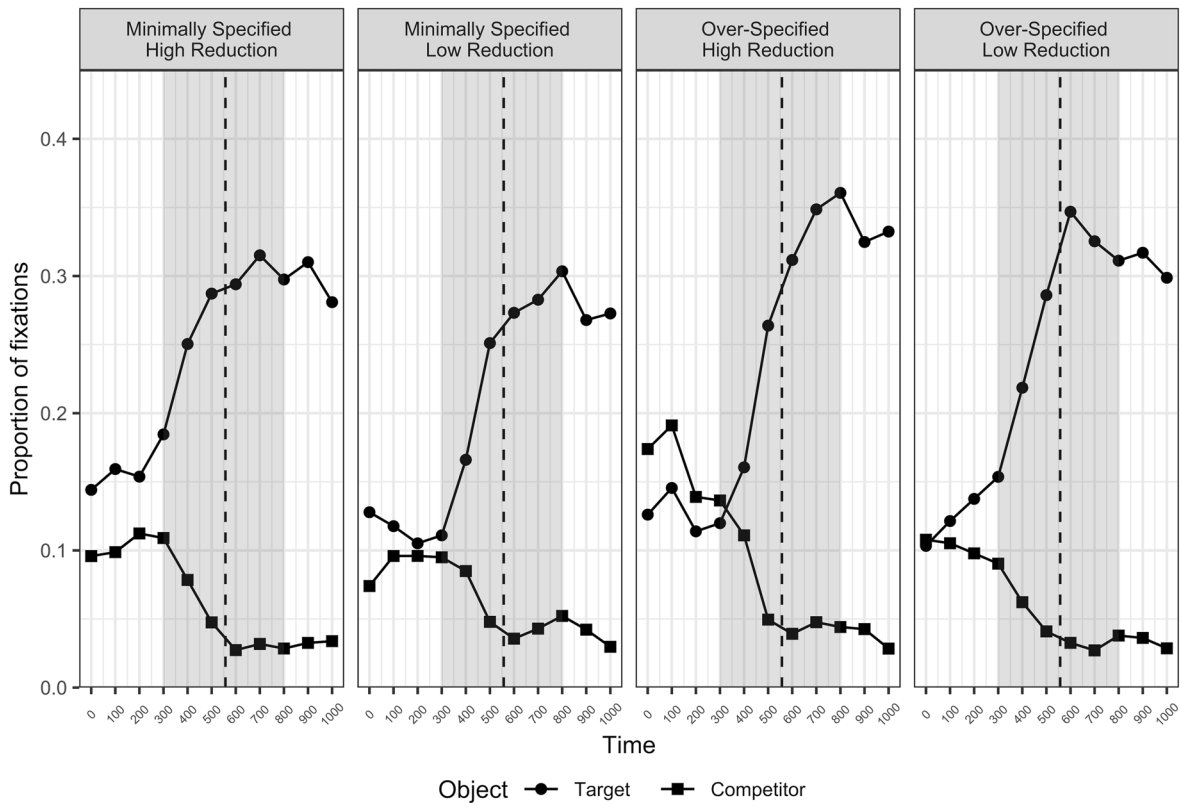
Fig. 7 Experiment 1. Proportion of fixations to the target vs. competitor object in each condition of pattern items. Noun onset is at zero. The shaded area represents the analysis window, and the dashed line indicates average noun offset

Table 3 Experiment 1 results—mean log-gaze probability ratios (SD in parentheses) for fixations to the target over fixations to the competitor object in the noun time window of pattern items

	Mean log ratios by participants	Mean log ratios by items
MS-HR	0.847 (0.729)	0.765 (1.011)
MS-LR	0.632 (0.821)	0.548 (1.038)
OS-HR	0.705 (0.814)	0.636 (1.049)
OS-LR	0.802 (0.800)	0.726 (1.015)

display, such as those in Fig. 8, was the same for both participants. Objects differed in colour and pattern, and in critical trials one feature (adjective) was necessary to identify the intended referent. We manipulated which feature was *necessary* for disambiguation (colour—necessary vs. pattern—necessary), and which feature was more *entropy—reducing* (colour—reducing vs. pattern—reducing vs. equally—reducing). We measured the proportion of over-specifications produced per condition.

The egocentric view holds that production preferences are tuned to minimise speakers' effort,

regardless of the addressees' needs. Therefore, if over-specifications are the result of egocentric production processes, speakers' choices should not be affected by the manipulations described above (i.e., the rate of over-specifications that egocentric speakers produce should be independent of the experimental condition).

Conversely, according to the addressee-oriented view, speakers should prefer structures that ease comprehension for their listeners—both the Gricean and the bounded-rational approaches are in accord with this view. The Gricean account predicts that for
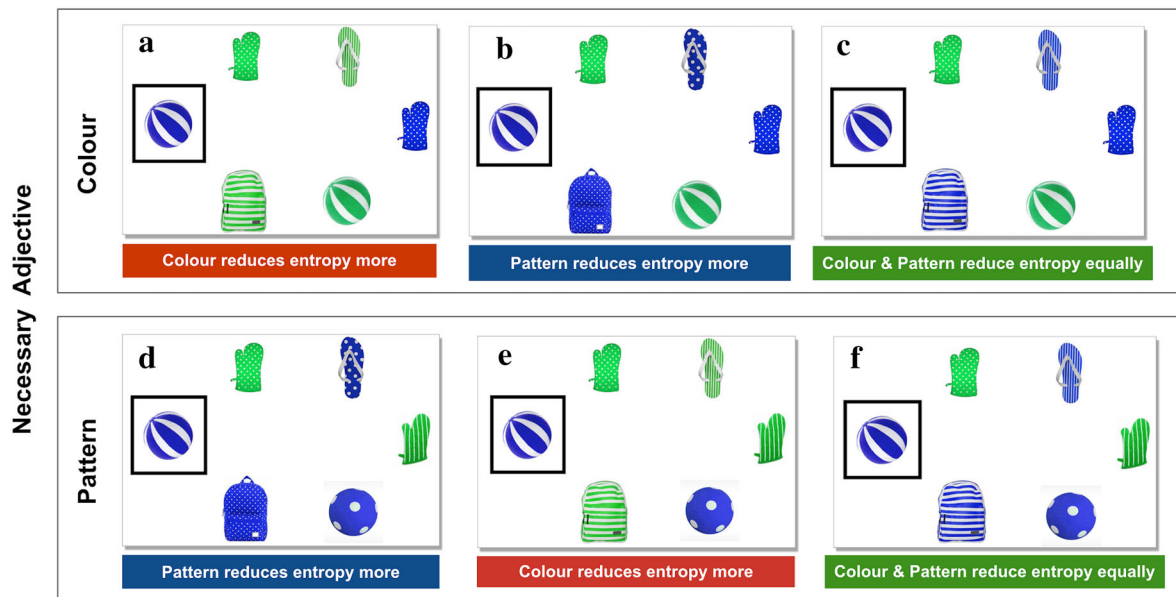
Fig. 8 Experiment 1. Sample visual displays (conditions) from an experimental item. A black frame identified the target object (e.g., the blue striped ball) and was visible only on the Speaker's display. Listeners viewed similar displays, in which the black frame was not present and object positions were mirrored on half of the trials (i.e., the blue striped ball would appear on the right side of their screen). **a**–**c** present conditions where a colour adjective was required for target identification (colour—necessary), and a speaker's question would minimally be 'Is the blue ball on the left?'. **d**–**f** present conditions where a pattern adjective was required (pattern—necessary), and a speaker's question would minimally be 'Is the striped ball on the left?'. In conditions **a** and **d**, the necessary adjective (colour and pattern, respectively) was also the most entropy—reducing adjective; in conditions **b** and **e**, the redundant adjective (pattern and colour, respectively) was the most entropy—reducing adjective, while in **c** and **f** both adjectives were equally entropy—reducing. See the online version of this article for the colour figure

all conditions speakers should prefer to convey the minimal amount of information that is necessary, as this is what would be expected by the listeners. That is, speakers should use the expression 'the blue ball' to refer to the intended referent in the colour—necessary conditions (cf. top panels in Fig. 8) and the expression 'the striped ball' in the pattern—necessary conditions (cf. bottom row in Fig. 8), independent of their entropy reduction potential.

By contrast, the bounded-rational approach predicts that speakers should be more likely to over-specify particularly when the entropy reduction potential of the redundant adjective is higher than that of the necessary adjective. For example, in Fig. 8e 'blue' would be redundant, but it also reduces entropy to a higher degree than the necessary adjective 'striped' ($\Delta H_{blue} = 1.58$ bits vs. $\Delta H_{striped} = 0.58$ bits). Thus, the redundant 'blue' should be used more often in Fig. 8e than in Fig. 8d when the necessary adjective ('striped') is more entropy—reducing. Such production preferences would be in line with the findings from Experiment 1 that listeners favour utterances that

manage entropy more effectively. Finally, colour over-specifications are expected to be more frequent than pattern over-specifications. This prediction is based on the results from Experiment 1 as well as previous research (cf. Sedivy 2003; Rubio-Fernández 2016).

Methods

Participants

Forty-nine pairs of native German speakers, who did not take part in Experiment 1, participated in this experiment. They were randomly assigned to the roles of Speaker (mean age = 23.2, 36 female) and Listener (mean age 24.3, 33 female), and were compensated with 5 Euros for their participation.[4] One pair of

---

[4] In a few cases, two participants could not be scheduled for the same time slot, therefore confederates were used as Listeners. However, the listener's role in this task is minor and should not influence the speakers' results.

German–French bilinguals was not included in the analysis due to French language dominance.

Materials

Eighteen of the object pictures from Experiment 1 were used to create the visual stimuli. One experimental item comprised six versions of one display (conditions; cf. Fig. 8), which differed in whether the mention of colour or pattern was required for disambiguation of the target referent (colour—necessary vs. pattern—necessary), and which of the features was more entropy—reducing compared to the other one (colour—reducing vs. pattern—reducing vs. equally—reducing). As in Experiment 1, critical displays always contained six objects. The target referent was paired with another object of the same type (cf. the balls in Fig. 8), which differed from the target either in colour (cf. Fig. 8a–c) or in pattern (cf. Fig. 8d–f). A competitor object that shared the necessary feature with the target referent was included and was also part of a contrast pair (cf. the mitts in Fig. 8). This was done so that the use of a redundant adjective would not allow listeners to select the target immediately after hearing the adjective. Another two objects were included that differed in colour and pattern depending on the entropy reduction condition. That is, they differed from the target referent in the necessary feature, when this feature was more entropy—reducing than the other one (colour in colour—necessary, cf. Fig. 8a; pattern in pattern—necessary, cf. Fig. 8d); they shared the necessary feature with the target referent when the other feature was more entropy—reducing (pattern in colour—necessary, cf. Fig. 8b; colour in pattern—necessary, cf. Fig. 8e); they shared both features with the target referent when they were equally entropy—reducing (cf. Fig. 8c, f). A total of 216 displays were thus created making up 36 experimental items.

Experimental displays were intermixed with material from another experiment that functioned as fillers for the current experiment. These displays used the same object pictures and display structure, but the target referent was always a singleton and the distractor objects could either be in pairs or were also singletons. In order to increase variability, another set of 144 filler items was also constructed. Filler displays depicted either six or four objects, again differing in colour and pattern. The target referent in filler items was either part of a set of three same-type objects or was a singleton. Thus fillers required either two adjectives for disambiguation or none. Additionally, filler trials varied in whether one of the target properties was more entropy—reducing than the other or both properties reduced entropy to an equal extent.

Overall, 576 visual scenes were created, half of which were then flipped on the vertical axis and were used only on the Listeners' display. The Listeners therefore saw half of the items in the same display configuration as the Speaker and half in a mirrored configuration. Stimuli were distributed into six lists following a Latin square design, so that only one version of an item appeared in a list, and so that participants were exposed to only one condition of each item. Lists were pseudo-randomised so that two trials from the same condition never appeared in a row, and at least one filler and/or one trial from the other experiment intervened between two consecutive experimental trials. E-prime 2.0 (Psychology Software Tools, Inc., Pittsburgh, PA, USA) was used to implement and run the experiment.

Procedure

Speakers and Listeners sat on opposite sides of a glass window separating two adjacent rooms. They each had a 1680 × 1050 resolution monitor in front of them, and used a microphone and headphones to communicate via an audio link. Participants saw displays containing the same objects, but their position on the vertical axis was flipped on half of the trials. They were instructed to imagine taking part in a long-distance call, where they needed to establish whether they share the same visual domain with their partner or not. Their task was to identify whether an object that was designated to the speaker was on the same side of the screen for both of participants. More specifically, after a 3 s preview time a target object was indicated by a black frame (cf. Fig. 8) on the Speaker's screen only, and a sound was played in order to indicate to the Listener that the target had been revealed to the Speaker. The Speaker then had to ask the Listener which side of his screen the target object was on. For example, on the top panels of Fig. 8 a question containing minimal information would be 'Is the blue ball on the left?'. The Listener's task was to respond 'Yes' or 'No' by pressing a button on a response pad.

Listeners were allowed to ask for further information, if necessary. Feedback was given after each trial in the form of a bell (for correct responses) or buzzer (for incorrect responses) and was audible to both participants. Crucially, in order to encourage participants to collaborate rather than perform two disjointed tasks, participants were told that they only had a limited amount of time to complete each trial, and they were requested to produce grammatical utterances.

One experimental session proceeded as follows. When participants came in the lab, they were first asked to give informed consent, and then they were randomly assigned to each role. The roles were described as 'Information-seeker' for the Speaker and 'Information-giver' for the Listener, so that participants' behaviour would not be confined by the speaker/listener distinction. After participants read the instructions corresponding to their role, the Experimenter orally explained their tasks in order to ensure that the instructions were clear to both participants, and that they understood that their tasks converged in a common goal. They were then presented with a preview of the objects that would appear during the experiment, in displays arranged by object type and showing all colour and pattern combinations. During this phase, Speakers were asked to name out loud the object type on each display. They were next shown to their seats and completed a practice block. The experiment began after it was confirmed that both participants understood the task. The experimenter remained in the same room as the Speaker during the experimental session, in order to make sure that the Speaker did not use truncated sentences (e.g., 'blue ball left'). Participants reported that the presence of the experimenter did not affect their performance. Each experimental session lasted approximately 30 min.

Data coding

Speakers' utterances were transcribed and annotated. Audio files from one speaker were corrupted and not further processed. Minimally-specified utterances were coded as '0' and constituted 58.8% of all trials, and over-specified utterances were coded as '1' and constituted 39.1% of all trials. Under-specified utterances (e.g., 'Is the ball on the left' in Fig. 8a–c) were 2.1% overall. Data from two speakers who produced a high rate of under-specifications (more than 15%) were excluded from further analyses. [5]Trials containing self-repairs (e.g., repairs of the adjective or the noun), or revisions of the utterance structure (e.g., providing more/less information after an initial question) were excluded from analyses (5.48%). Moreover, we did not consider trials in which descriptions could not be clearly classified (e.g., 'the green-black umbrella' instead of 'the checkered umbrella') (1.65%). Finally, trials with under-specified utterances were also excluded (1.61%). To account for potential priming effects from the previous trial (e.g., that over-specification on a particular trial could be the result of priming from the immediately preceding filler that required two adjectives), we further excluded trials in which the same number of adjectives were used in the same word order as in the immediately preceding trial (12.36%).[6]

Analysis

Proportions of over-specifications were analysed using generalised linear mixed models with the lme4 package (Bates et al. 2015) in R (version 3.5.1; R Core Team 2018). The models included crossed random intercepts for both participants and items, and random slopes for the necessary and the entropy—reducing feature. Factors were treatment coded, with pattern as reference level for the necessary feature, and colour as reference level for the entropy—reducing feature. When the maximal models did not converge, the random effects structure was simplified (Barr et al. 2013).

Results

According to the egocentric view, Speakers' use of redundancy should be unaffected by our manipulations. Indeed, 16 participants were found to over-specify the majority of the time. Interestingly another 10 participants over-specified regularly, but only for colour. Based on this pattern of results, we categorised participants into three groups depending on their

---

[5] A normal rate of under-specification is under 5% (cf. Davies and Katsos 2013; Engelhardt et al. 2006; Koolen et al. 2011; Koolen et al. 2013; Koolen et al. 2015, among others).

[6] These trials were excluded after Group1 was created (see 'Results' below), as participants in this group almost always over-specified, rarely changing word order.
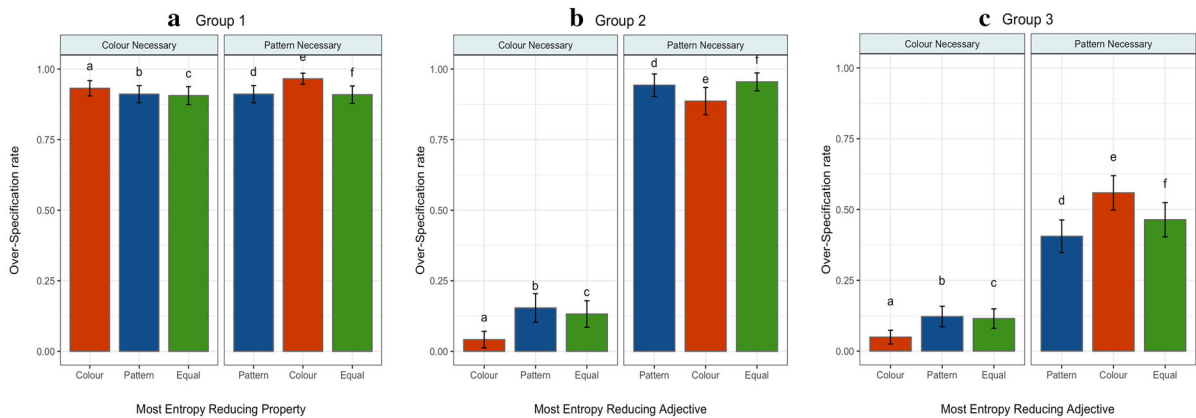
Fig. 9 Experiment 2. Proportions of over-specifications produced in each group of participants. Left panels present the over-specification rates in the colour—necessary conditions (i.e., over-specifications for pattern) and right panels present the over-specification rates in the pattern—necessary conditions (i.e., over-specifications for colour). Bars are coloured based on which feature was the most entropy—reducing one (red for colour, blue for pattern, and green for equal), and are ordered based on condition indices (see also Fig. 8). In Group 1, the rate of over-specification was higher than 80% in all conditions, independent of which feature was necessary or more entropy—reducing. In Group 2, speakers over-specified in more than 80% of pattern—necessary trials (i.e., they regularly over-specified for colour), but rarely over-specified in the colour—necessary conditions (i.e., rarely over-specified for pattern). This behaviour was independent of whether colour was more entropy—reducing than pattern. In Group 3, speakers again over-specified more in the pattern—necessary than the colour—necessary conditions (i.e., more colour over-specification than pattern over-specification), but this was modulated by the entropy—reduction potential of the two features: colour over-specifications were more frequent when colour was more entropy—reducing than pattern. See the online version of this article for the colour figure

general pattern of OS use. Group 1 included speakers (N = 16) who produced both adjectives more than 80% of the time (Fig. 9a), Group 2 consisted of those (N = 10) who produced redundant colour adjectives more than 80% of the time (Fig. 9b), and Group 3 consisted of the remaining participants (N = 16) (Fig. 9c). Analyses were performed per group.[7]

Results from all groups are summarised in Table 4. In Group 1 none of the comparisons reached significance. Over-specifications were equally frequent for both colour and pattern in all conditions (Fig. 9a). In Group 2, only the necessary feature was found to be significant, with more over-specifications in the pattern—necessary than in the colour—necessary conditions (i.e., redundant colour adjectives were used more frequently than redundant pattern adjectives; cf. Fig. 9b). In Group 3, the necessary feature again resulted in a significant effect, with more over-specifications in the pattern—necessary than in the colour—necessary condition (i.e.,

more over-specifications for colour than for pattern), but further comparisons were also found to be significant. In particular, regarding the entropy—reducing factor, the comparison between colour—reducing (the reference level) and pattern—reducing yielded a marginally significant effect indicating a higher proportion of over-specifications when colour reduced entropy more than pattern (cf. red and blue bars, respectively, in Fig. 9c). Moreover two necessary × entropy—reducing interactions were found (see Table 4). The difference in over-specification rate between pattern—necessary and colour—necessary was larger for colour—reducing than pattern—reducing conditions (cf. the difference between red and blue bars in the two panels of Fig. 9c). Similarly, the difference in over-specification rate between pattern—necessary and colour—necessary was larger for colour—reducing compared to equally—reducing conditions (cf. the difference between red and green bars in the two panels of Fig. 9c).

Discussion

In Experiment 2, we aimed to evaluate whether the factors that were found to influence comprehension in Experiment 1 (i.e., target feature and entropy

---

[7] Three of the originally 48 speakers produced minimal descriptions more than 90% of the time. While these participants can be considered as highly Gricean, the current study was concerned with understanding why speakers over-specify. Therefore, these participants were excluded from analyses.

Table 4 Experiment 2 results

	Group 1				Group 2				Group 3			
	β	SE	z	p	β	SE	z	p	β	SE	z	p
Intercept	3.862	0.702	5.498	< 0.001***	2.588	0.677	3.822	< 0.001***	0.255	0.274	0.929	0.353
Necessary												
Colour	− 0.747	0.763	− 0.978	0.328	− 6.159	1.136	− 5.422	< 0.001***	− 3.266	0.574	− 5.687	< 0.001***
Entropy—reducing												
Pattern	− 1.115	0.734	− 1.517	0.129	0.896	0.922	0.973	0.33	− 0.65	0.349	− 1.86	0.063.
Equal	− 1.139	0.735	− 1.549	0.121	0.994	0.91	1.092	0.275	− 0.374	0.352	− 1.064	0.287
Necessary:entropy-reducing												
Colour:pattern	0.856	0.962	0.889	0.374	0.577	1.262	0.457	0.647	1.618	0.708	2.284	0.022*
Colour:equal	0.685	0.953	0.719	0.472	0.34	1.266	0.269	0.788	1.274	0.71	1.796	0.072.

Coefficients, *SE*, *z*- and *p*-values for the comparisons of over-specification rates between the reference (pattern—necessary and colour—reducing) and other levels of the fixed factors and their interactions. *** $p < .001$, ** $p < .01$, * $p < .05$, . $p < .1$

reduction on the adjective) would modulate speakers' use of over-specification. In a referential communication task, we manipulated whether colour or pattern was necessary to identify the target referent, and which of these features was more entropy—reducing (colour vs. pattern vs. equal). We measured participants' over-specification rate in each condition. The higher overall rate of minimally-specified referring expressions (59%) compared to over-specified referring expressions (39%) clearly demonstrates that speakers are able to produce and often do produce minimal descriptions. What we are interested in understanding, however, is under what circumstances they over-specify.

The first finding of interest was that speakers adopted different production strategies. We therefore split participants in groups according to their general pattern of over-specification. We first identified a group (Group 1) that over-specified more than 80% of the time with both pattern adjectives (i.e., in the colour—necessary conditions) and with colour adjectives (i.e., in the pattern—necessary conditions). That is, these participants very rarely produced expressions that did not encode both adjectives. This behaviour is in line with the predictions from the egocentric view, that speakers' use of over-specification should not be affected by the experimental manipulations. This result indicates that at least for some speakers, over-specification is a strategy for reducing cognitive load of target identification (e.g., by simply using a template that contains both modifiers, regardless of the visual environment).

Group 2 included participants whose over-specification rate was greater than 80%, but only in the pattern—necessary conditions. That is, they regularly used a redundant colour adjective, independent of its entropy reduction potential. The results from this group are in accord with both the egocentric view and with the audience-design view in that speakers prefer to use redundant colour adjectives more frequently than redundant pattern adjectives. On the one hand, this preference may be due to colour salience, which eases property selection for the speakers. On the other hand, it also facilitates target identification for the listeners, who favour colour over-specifications over pattern over-specifications, as was shown in Experiment 1.

Finally, the remainder of participants were grouped together (Group 3). Results showed that speakers in

this group also over-specified more for colour than for pattern, but their use of redundant adjectives varied with the distributional properties of the visual scene: They over-specified more frequently when the redundant adjective was more entropy—reducing than the necessary adjective. This behaviour matches the predictions of the bounded-rational approach, which argues that speakers should over-specify more when the redundant adjective reduces referential entropy to a higher degree than the necessary adjective. These results also fit the findings from Experiment 1, which showed that listeners favour over-specified expressions, as well as expressions that reduce entropy at a high rate early on, but this preference is greater for colour than for pattern adjectives.

Overall, while individual differences seem to govern production choices, we found evidence for the use of a bounded-rational strategy (Group 3). This strategy appears to take into account the distributional properties of the visual scene in order to ease the listener's task, by producing a redundant adjective more frequently when it helps reduce entropy at a higher rate than the necessary adjective. We further found that egocentric concerns may also be at play in referential communication, and that at least in some cases (Group 1) over-specifications may be for the speaker. A third strategy was also observed (Group 2), in which over-specifications were used independent of condition, but only with colour adjectives. This strategy could be interpreted either under an audience-design or an egocentric view, as colour is a visually salient property and arguably preferred by both speakers and listeners in such tasks. Evidence in support of the Gricean account in this experiment was minimal; only three of our participants systematically used minimal information in all conditions.

General discussion

In this study, we evaluated the hypothesis that in complex visual scenes speakers frequently over-specify—contra the Gricean account—in an effort to distribute referential entropy across a longer sequence of words, as this facilitates comprehension for the listeners. In a comprehension experiment, we examined how the rate of entropy reduction influences comprehension processes of minimally- and over-specified referring expressions. A production

experiment tested whether speakers' use of redundant adjectives is modulated by the extent to which these adjectives reduce entropy.

Previous work is inconclusive regarding whether over-specification impedes or facilitates comprehension. Therefore, Experiment 1 investigated whether the use of a redundant adjective influences the comprehension of a referring expression, and whether this is further modulated by the rate of entropy reduction in the expression. We contrasted the predictions of Gricean accounts (Grice 1975, 1989), which suggest that addressees should encounter comprehension difficulties when speakers use more information than is minimally required, with the predictions of bounded-rational approaches to communication (Hale 2003, 2006; Frank and Jaeger 2008; Jaeger 2010; Levy and Jaeger 2007), which suggest that a redundant adjective may ease processing by distributing entropy reduction across a longer sequence of linguistic units. Our findings are in line with the bounded-rational account, and indicate that both over-specification and a high reduction of entropy before the head noun independently facilitate comprehension, as evidenced by lower ICA values and faster RTs. Listeners were also found to benefit from descriptions mentioning colour over pattern, resulting in lower ICA values and faster RTs.

Experiment 2 then examined whether speakers take these listener preferences into account when planning their utterances. Results showed that at least for some speakers (Group 3) production choices were influenced by the intent to effectively modulate entropy reduction across the utterance, as was our hypothesis. That is, over-specifications were more likely to occur when a property that was not necessary to be mentioned for target identification was more entropy—reducing than the necessary property. In line with the findings from Experiment 1, this strategy was used more with redundant colour adjectives than with redundant pattern adjectives. Thus this provides evidence for the audience-design view that speakers choose expressions that will make comprehension easier for their addressees. How this behaviour would be modulated by increased scene complexity is an open question. Another group of speakers also over-specified more for colour than for pattern but this tendency did not depend on the visual scene (Group 2). This strategy is consistent with the predictions of both the egocentric and audience-design views, because

colour is a salient property and its redundant use is not costly to the speaker while it is beneficial to the addressee. In addition, our findings also provide support for the egocentric view. Some speakers (Group 1) did not vary their use of over-specification and consistently specified both colour and pattern across all trials.

Previous comprehension studies that have found evidence in support of the Gricean account (i.e., that redundancy adversely affects comprehension) have mostly used offline tasks evaluating the addresses' acceptance of over-specifications (e.g., Davies and Katsos 2013) but not their online processing. One ERP study (Engelhardt et al. 2011) reported an N400-like effect on the adjective for over-specified compared to minimally-specified descriptions, which was interpreted as increased cost relative to the processing of the redundant adjective. It is, however, possible that the redundant adjective was not helpful in this particular case because the visual domains were highly simplified, consisting of merely two objects (e.g., a star and a circle). That is, entropy was already low initially, and it need not be smoothed out with the insertion of a redundant word. In the current study, visual displays employed more objects, which differed across more features, and therefore had significantly higher referential entropy.

While some previous research has also shown a processing advantage for colour over-specifications, this has generally been attributed to the special status of colour (e.g., Sedivy 2003). That is, due to its salience, colour is not interpreted contrastively and therefore its redundant use is licensed. In Experiment 1, we found evidence that pattern over-specifications also result in a processing advantage; both ICA values and RTs were lower in OS compared to MS conditions. Unsurprisingly, this advantage was not as strong as for colour adjectives. Speakers' preferences in Experiment 2 were in line with this finding, with participants in both Group 2 and Group 3 over-specifying more for colour than for pattern. This confirms that there is a general preference for the redundant use of colour (e.g., Belke 2006; Belke and Meyer 2002; Koolen et al. 2013, 2015; Rubio-Fernández 2016; Tarenskeen et al. 2015), but because this strategy is also favoured by the listeners, it is not clear whether this should be considered evidence for the egocentric account or the audience-design view. By exploiting features of the visual environment that

stand out, the use of redundant colour adjectives seems to be a good trade-off between speakers' production pressures on the one hand and listeners' processing effort on the other. This may explain why speakers of languages that place adjectival modifiers post-nominally, such as Spanish, also include redundant colour adjectives in their referring expressions (Rubio-Fernández 2016), even though these adjectives come at a point where referential entropy is already reduced to zero (i.e., by the preceding noun). That is, a redundant adjective after the noun may still be useful to the listener, as in complex displays it gives them additional cues to guide their visual search for the target referent.

Some recent production studies (Fukumura 2018; Koolen et al. 2015; Vogels et al. 2019) have shown that the discriminatory power of a property affects the tendency of speakers to redundantly mention this property in their utterances. None of these studies, however, directly manipulated discriminability. In the current work, we considered a related notion, namely the reduction of referential entropy that is brought about by a given property. This notion was inspired by previous research on entropy reduction (e.g., Frank 2013; Hale 2006; Linzen and Jaeger 2016), and by work on how information is distributed across the signal (e.g., Jaeger 2010; Levy and Jaeger 2007), and was adapted to visually-situated contexts. In more recent related work, Ankener et al. (2018) found evidence that entropy reduction influences processing, but in their work a high reduction of entropy on the verb was only manifest as a facilitation on the subsequent noun. By contrast, we found effects of entropy reduction at each reduction point. In other words, a high reduction of entropy on the adjective (whether the adjective was necessary or not for target identification) was associated with increased processing effort in that region, as predicted by the entropy reduction hypothesis (Hale 2006). However, it also resulted in a processing advantage on the noun, because residual entropy was reduced following the adjective. This pattern of effects did not differ for necessary and redundant adjectives, indicating that they are similarly exploited for the reduction of uncertainty. Thus, we have shown that entropy reduction is a predictor of cognitive effort in situated contexts as well, with an even distribution of reduction across the signal resulting in a processing advantage. A question that requires further research, however, is

whether the advantage observed at the noun is due to the fact that the noun was the final word in both Ankener et al. (2018) and the current study, or whether nouns have a special functional status because they directly point to objects in the world.

In sum, the present findings indicate that there is no penalty associated with over-specification for the listeners, contra the predictions of Gricean accounts. Rather, our findings suggest that redundancy and early reduction of entropy across the expression, both facilitate processing. Our production experiment finds evidence for individual differences, with speakers being more or less sensitive to variability in the visual context. One group of participants generally over-specified regardless of the visual context—a strategy that requires minimal effort for the speaker and supports the view that speakers over-specify to ease production processes. A second group over-specified for colour only—regardless of how much colour adjectives reduced entropy—consistent with both speaker and listener preferences for colour over-specifications due to visual salience. Finally, some speakers rationally over-specified—over-specified more when the redundant adjective reduced entropy to a greater extent—and more often with colour adjectives. These findings suggest (a) that redundancy facilitates comprehension, and that speakers behave in a manner generally consistent with this, but varying in their sensitivity to the specifics of the immediate visual context, and (b) that the rate of entropy reduction predicts processing effort in referential communication and possibly explains the preference for over-specified expressions by listeners and some speakers: The inclusion of a redundant adjective distributes entropy reduction (i.e., processing effort) over the utterance.

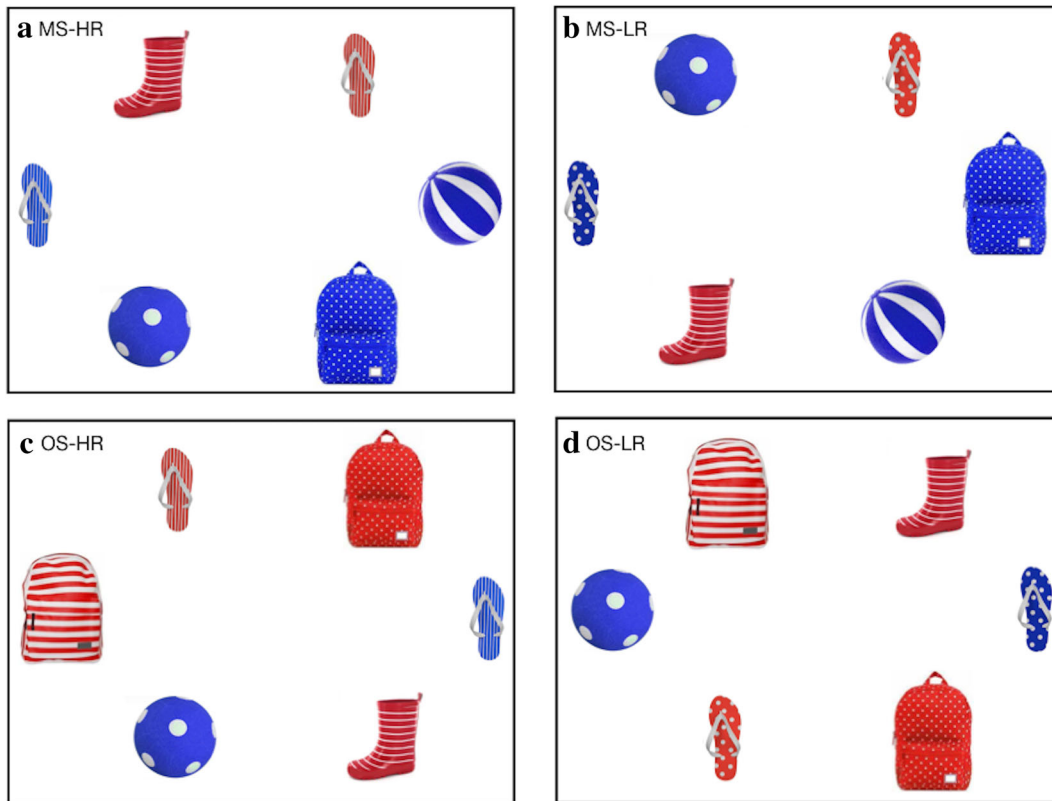**Data availability statement** Data are available from the corresponding author on reasonable request.

Compliance with ethical standards

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

Appendix

See Fig. 10.

Fig. 10 Experiment 1. Sample visual displays from a pattern experimental item, paired with the instruction 'Find the dotted ball'. When the display presented a shape competitor (cf. the striped ball in **a**, **b**), the instruction was minimally specified (MS). When the target object was of unique shape (cf. one ball in **c**, **d**), the instruction was over-specified (OS). Additionally, when fewer objects matched the mentioned feature (cf. two dotted objects in **a**, **c**), the reduction of referential entropy on the adjective was high (HR) compared to when more objects with that feature were present (cf. four dotted objects in **b**, **d**), and the adjective resulted in a low reduction of entropy (LR). See the online version of this article for the colour figure

References

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition*. https://doi.org/10.1016/S0010-0277(99)00059-1.

Ankener, C. S., Sekicki, M., & Staudte, M. (2018). The influence of visual uncertainty on word surprisal and processing effort. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2018.02387.

Arnold, J. E. (2008). Reference production: Production-internal and addressee-oriented processes. *Language and Cognitive Processes*. https://doi.org/10.1080/01690960801920099.

Arts, A., Maes, A., Noordman, L., & Jansen, C. (2011a). Overspecification facilitates object identification. *Journal of Pragmatics*. https://doi.org/10.1016/j.pragma.2010.07.013.

Arts, A., Maes, A., Noordman, L., & Jansen, C. (2011b). Overspecification in written instruction. *Linguistics*. https://doi.org/10.1515/LING.2011.017.

Bach, K. (2006). The top 10 misconceptions about implicature. In B. J. Birner & G. Ward (Eds.), *Drawing the boundaries of meaning: Neo-Gricean studies in pragmatics and semantic in honor of Laurence R. Horn* (pp. 21–30). Amsterdam: John Benjamins.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random-effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*. https://doi.org/10.1016/j.jml.2012.11.001.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). *lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-8*. http://CRAN.R-project.org/package=lme4.

Belke, E. (2006). Visual determinants of preferred adjective order. *Visual Cognition*. https://doi.org/10.1080/13506280500260484.

Belke, E., & Meyer, A. S. (2002). Tracking the time course of multidimensional stimulus discrimination: Analyses of viewing patterns and processing times during "same"– "different" decisions. *European Journal of Cognitive Psychology*. https://doi.org/10.1080/09541440143000050.

Brodbeck, C., Gwilliams, L., & Pylkkänen, L. (2015). EEG can track the time course of successful reference resolution in small visual worlds. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2015.01787.

Davies, C., & Katsos, N. (2013). Are speakers and listeners 'only moderately Gricean'? An empirical response to Engelhardt et al. (2006). *Journal of Pragmatics*. https://doi.org/10.1016/j.pragma.2013.01.004.

Demberg, V., & Sayeed, A. (2016). The frequency of rapid pupil dilations as a measure of linguistic processing difficulty. *PLoS One*. https://doi.org/10.1371/journal.pone.0146194.

Deutsch, W., & Pechmann, T. (1982). Social interaction and the development of definite descriptions. *Cognition*. https://doi.org/10.1016/0010-0277(82)90024-5.

Engelhardt, P. E., Bailey, K., & Ferreira, F. (2006). Do speakers and listeners observe the Gricean Maxim of Quantity. *Journal of Memory and Language*. https://doi.org/10.1016/j.jml.2005.12.009.

Engelhardt, P. E., Demiral, Ş. B., & Ferreira, F. (2011). Over-specified referring expression impair comprehension: An ERP study. *Brain and Cognition*. https://doi.org/10.1016/j.bandc.2011.07.004.

Engelhardt, P. E., Ferreira, F., & Patsenko, E. G. (2010). Pupillometry reveals processing load during spoken language comprehension. *The Quarterly Journal of Experimental Psychology*. https://doi.org/10.1080/17470210903469864.

Frank, S. L. (2010). Uncertainty reduction as a measure of cognitive processing effort. In J. T. Hale (Ed.), *Proceedings of the 2010 workshop on cognitive modeling and computational linguistics* (pp. 81–89). Uppsala: Association for Computational Linguistics.

Frank, S. L. (2013). Uncertainty reduction as a measure of cognitive processing load in sentence comprehension. *Topics in Cognitive Science, 5,* 10. https://doi.org/10.1111/tops.12025.

Frank, A., & Jaeger, T. F. (2008). Speaking rationally: Uniform information density as an optimal strategy for language production. In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), *Proceedings of the 30th annual conference of the cognitive science society* (pp. 939–944). Austin, TX: Cognitive Science Society.

Frank, S. L., & Thompson, R. L. (2012). Early effects of word surprisal on pupil size during reading. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 34th annual conference of the cognitive science society* (pp. 1554–1559). Austin, TX: Cognitive Science Society.

Fukumura, K. (2018). Ordering adjectives in referential communication. *Journal of Memory and Language*. https://doi.org/10.1016/j.jml.2018.03.003.

Gatt, A., Krahmer, E., van Deemter, K., & van Gompel, R. P. G. (2017). Reference production as search: The impact of domain size on the production of distinguishing descriptions. *Cognitive Science*. https://doi.org/10.1111/cogs.12375.

Geurts, B., & Rubio-Fernández, P. (2015). Pragmatics and processing. *Ratio, 28,* 446–469.

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

Grice, H. P. (1989). *Studies in the way of words*. Cambridge, MA: Harvard University Press.

Hale, J. T. (2003). The information conveyed by words in sentences. *Journal of Psycholinguistic Research*. https://doi.org/10.1023/a:1022492123056.

Hale, J. T. (2006). Uncertainty about the rest of the sentence. *Cognitive Science*. https://doi.org/10.1207/s15516709cog0000_64.

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

Just, M. A., & Carpenter, P. A. (1993). The intensity dimension of thought: Pupillometric indices of sentence processing. *Canadian Journal of Experimental Psychology, 47,* 310–339.

Knoeferle, P., Crocker, M. W., Scheepers, C., & Pickering, M. J. (2005). The influence of the immediate visual context on incremental thematic role-assignment: Evidence from eye-movements in depicted events. *Cognition*. https://doi.org/10.1016/j.cognition.2004.03.002.

Knoeferle, P., & Kreysa, H. (2012). Can speaker gaze modulate syntactic structuring and thematic role assignment during spoken sentence comprehension? *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2012.00538.

Koolen, R., Gatt, A., Goudbeek, M., & Krahmer, E. (2011). Factors causing over-specification in definite descriptions. *Journal of Pragmatics*. https://doi.org/10.1016/j.pragma.2011.06.008.

Koolen, R., Goudbeek, M., & Krahmer, E. (2013). The effect of scene variation on the redundant use of color in definite reference. *Cognitive Science*. https://doi.org/10.1111/cogs.12019.

Koolen, R., Krahmer, E., & Swerts, M. (2015). How distractor objects trigger referential overspecification: Testing the effects of visual clutter and distractor distance. *Cognitive Science*. https://doi.org/10.1111/cogs.12297.

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., & Jaeger, T. F. (2007). Speakers optimize information density through syntactic reduction. *Advances in Neural Information Processing Systems, 19,* 841–848.

Linzen, T., & Jaeger, T. F. (2016). Uncertainty and expectation in sentence processing: Evidence from subcategorization distributions. *Cognitive Science*. https://doi.org/10.1111/cogs.12274.

Maes, A., Arts, A., & Noordman, L. (2004). Reference management in instructive discourse. *Discourse Processes, 37,* 117–144.

Marshall, S. P. (2000). *Method and apparatus for eye tracking and monitoring pupil dilation to evaluate cognitive activity*. US Patent 6,090,051.

Marshall, S. P. (2002). The index of cognitive activity: Measuring cognitive workload. In *Proceedings of the 2002 IEEE 7th conference on human factors and power plants* (pp. 7.5–7.9). New York: IEEE.

Matin, E., Shao, K. C., & Boff, K. R. (1993). Saccadic overhead: Information-processing time with and without saccades. *Perception and Psychophysics, 53,* 372–380.

Paraboni, I., van Deemter, K., & Masthoff, J. (2007). Generating referring expressions: Making referents easy to identify. *Computational Linguistics*. https://doi.org/10.1162/coli.2007.33.2.229.

Pechmann, T. (1989). Incremental speech production and referential overspecification. *Linguistics*. https://doi.org/10.1515/ling.1989.27.1.89.

R Core Team. (2018). R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. Available at: https://www.R-project.org/.

Rubio-Fernández, P. (2016). How redundant are redundant color adjectives? An efficiency-based analysis of color overspecification. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2016.00153.

Scheepers, C., & Crocker, M. W. (2004). Constituent order priming from reading to listening: A visual-world study. In M. Carreiras & C. Clifton (Eds.), *The on-line study of sentence comprehension: Eyetracking, ERP and beyond* (pp. 167–185). New York: Psychology Press.

Schriefers, H., & Pechmann, T. (1988). Incremental production of noun-phrases by human speakers. In M. Zock & G. Sabah (Eds.), *Advances in natural language generation* (pp. 172–179). London: Pinter.

Sedivy, J. C. (2003). Pragmatic versus form-based accounts of referential contrast: Evidence for effects of informativity expectations. *Journal of Psycholinguistic Research, 32*(1), 3–23.

Sedivy, J. C. (2005). Evaluating explanations for referential context effects: Evidence for Gricean mechanisms in online language interpretation. In J. C. Trueswell & M. K. Tanenhaus (Eds.), *Approaches to studying world-situated language use: Bridging the language-as-product and language-as-action traditions* (pp. 345–364). Cambridge MA: MIT Press.

Sedivy, J. C., Tanenhaus, M. K., Chambers, C. G., & Carlson, G. N. (1999). Achieving incremental semantic interpretation through contextual interpretation. *Cognition*. https://doi.org/10.1016/S0010-0277(99)00025-6.

Sekicki, M., & Staudte, M. (2018). Eye'll help you out! How the gaze cue reduces the cognitive load required for reference processing. *Cognitive Science*. https://doi.org/10.1111/cogs.12682.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*. https://doi.org/10.1126/science.7777863.

Tarenskeen, S., Broersma, M., & Geurts, B. (2015). Overspecification of color, pattern, and size: Salience, absoluteness, and consistency. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2015.01703.

Tourtouri, E. N., Delogu, F., & Crocker, M. W. (2015). ERP indices of situated reference in visual contexts. In *Proceedings of the 37th annual conference of the cognitive science society* (pp. 2422–2427). Austin, TX: Cognitive Science Society.

Vogels, J., Demberg, V., & Kray, J. (2018). The index of cognitive activity as a measure of cognitive processing load in dual task settings. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2018.02276.

Vogels, J., Howcroft, D. M., Tourtouri, E., Demberg, V. (2019). How speakers adapt object descriptions to listeners under load. *Language, Cognition and Neuroscience*, 1–15.

Vonk, W., Hustinx, L. G. M. M., & Simons, W. H. G. (1992). The use of referential expressions in structuring discourse. *Language and Cognitive Processes*. https://doi.org/10.1080/01690969208409389.

Weber, A., Grice, M., & Crocker, M. W. (2006). The role of prosody in the interpretation of structural ambiguities: A study of anticipatory eye movements. *Cognition*. https://doi.org/10.1016/j.cognition.2005.07.001.

Wu, S., Bachrach, A., Cardenas, C., & Schuler, W. (2010). Complexity metrics in an incremental right-corner parser. In J. T. Hale (Ed.), *Proceedings of the 48th annual meeting of the association for computational linguistics* (pp. 1189–1198). Uppsala: Association for Computational Linguistics.