COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 45 (2021) e13071

ISSN: 1551-6709 online
DOI: 10.1111/cogs.13071

Rational Redundancy in Referring Expressions: Evidence from Event-related Potentials

Elli N. Tourtouri, Francesca Delogu, Matthew W. Crocker

Department of Language Science and Technology, Saarland University

Received 1 December 2020; received in revised form 29 October 2021; accepted 3 December 2021

Abstract

In referential communication, Grice's Maxim of Quantity is thought to imply that utterances conveying unnecessary information should incur comprehension difficulties. There is, however, considerable evidence that speakers frequently encode redundant information in their referring expressions, raising the question as to whether such overspecifications hinder listeners' processing. Evidence from previous work is inconclusive, and mostly comes from offline studies. In this article, we present two event-related potential (ERP) experiments, investigating the real-time comprehension of referring expressions that contain redundant adjectives in complex visual contexts. Our findings provide support for both Gricean and bounded-rational accounts. We argue that these seemingly incompatible results can be reconciled if common ground is taken into account. We propose a bounded-rational account of overspecification, according to which even redundant words can be beneficial to comprehension to the extent that they facilitate the reduction of listeners' uncertainty regarding the target referent.

Keywords: Bounded-rationality; Common ground; Entropy reduction; ERPs; Gricean maxims of Quantity; Overspecification; Referring expressions

1. Introduction

Establishing reference to an entity in the visual environment is often integral to everyday communication. Imagine, for instance, that you are helping a friend tidy up the nursery. It is

Present address: Elli N. Tourtouri, Psychology of Language Department, Max Planck Institute for Psycholinguistics, The Netherlands.

Correspondence should be sent to Elli N. Tourtouri, Max Planck Institute for Psycholinguistics, Wundtlaan 1, 6525 XD Nijmegen, The Netherlands. E-mail: elli.tourtouri@mpi.nl

likely that your interaction will (for the most part) revolve around locating toys lying around on the floor. That is, in situations where communication concerns the immediate visual scene, the speaker's primary goal is to direct the listener's attention toward some visually co-present referent. Listeners, therefore, need to rapidly map the incoming linguistic input onto the visual context, in order to understand the speakers' utterance and identify the intended object. For instance, if in our example your friend said "*Hand me the blue ball*" in the presence of a blue and a green ball, mapping her utterance to the visual context would be uncomplicated;[1] the noun ("*ball*") refers to the type of the desired object, while the adjective ("*blue*") distinguishes between the two same-type objects by mentioning the property in which they differ.

In situated communication, however, utterances are often not that succinct, as speakers frequently use information that is not strictly required for the identification of a target referent (see Engelhardt et al., 2011, who estimate that redundant information is encoded between 10% and 60% of the time). This issue has concerned psycholinguistics for decades, since philosopher Paul Grice introduced his theory of conversation (Grice, 1975), as researchers have tried to explain why rational speakers are so frequently redundant. In the present article, we provide evidence that redundancy may in fact benefit listeners in the identification of referents, in line with a bounded-rational account of communication, which predicts that any word that reduces listeners' *uncertainty* about the target referent may facilitate referential processing (see also Tourtouri et al., 2019).

1.1. Gricean quantity and the comprehension of overspecifications

It is generally thought that the human cognitive system has evolved so as to optimize behavior to achieve specific goals (Anderson, 1990, 1991; see for review Chater & Oaksford, 1999). This means that in order to solve a given problem, rational agents are likely to choose a course of action approximating the optimal solution. Talking is in Grice's view "a special case or variety of purposive, indeed rational, behavior" (Grice, 1975, p. 47). According to the *Gricean theory* (Grice, 1975), optimal referring expressions should convey only necessary information (*minimal specifications*). Specifically, the two maxims that fall under the category of Quantity predict that for communication to be successful speaker's utterances should carry no less information than required for the purposes of the exchange (first maxim), but also no more information than required (second maxim). To illustrate this, if in the above example your friend intended to specify one out of two blue balls, the same expression as before ("*Hand me the blue ball*") would be *underspecified* (US), violating the first maxim of Quantity, as the color adjective ("*blue*") alone would not be specific enough to uniquely identify the target referent. In this case, you—that is, the listener—might try to reason about this lack of information, and, if there is no evidence that your friend is in fact uncooperative, you might infer that, for example, only one of the two balls is visible from where she stands. In any case, underspecifications lead to referential failure, as the listener is in no position to resolve reference (Davies & Katsos, 2013; Engelhardt et al., 2006).

If a single (blue) ball was present in the room, the expression "*Hand me the blue ball*" would still be inappropriate, in this case because it violates the second maxim of Quantity; the adjective would be redundant. Such *overspecifications* may leave listeners confused regarding the purpose of this redundancy. That is, listeners might be left wondering why the adjective was used even though the noun alone would suffice. Redundant information may, therefore, engage addressees in unintended pragmatic reasoning, which will ultimately need to be canceled. Because in situated contexts the visual input rapidly informs listeners' incremental language processing (Allopenna et al., 1998; Altmann & Kamide, 1999; Eberhard et al., 1995; Knoeferle et al., 2005; Tanenhaus et al., 1995), it is possible that such pragmatic inferences are generated in real time, and may lead to comprehension difficulties (cf. Sedivy et al., 1999). Thus, mapping your friend's overspecified (OS) utterance onto the visual context might be problematic because, as her utterance unfolds over time, it becomes clear that not all information therein was necessary. Despite the redundancy, however, it is possible to establish reference, as contrary to underspecifications the information for identifying the intended referent is in fact encoded in the utterance. Thus, redundancy may not fully disrupt comprehension, but it might be "merely a waste of time" (Grice, 1975, p. 46). *Rational* speakers should, therefore, produce minimal descriptions in order to identify a target referent.

Contra the Gricean view—which deems only minimal specifications to be optimal—speakers do frequently overspecify (Davies & Katsos, 2013; Deutsch & Pechmann, 1982; Engelhardt et al., 2006; Gatt et al., 2017; Mangold & Pobel, 1988; Koolen et al., 2011, 2013, 2016; Pechmann, 1989; Rubio-Fernández, 2016, 2019; Tarenskeen et al., 2015; Tourtouri et al., 2019; van Gompel et al., 2019; Vogels et al., 2019, i.a.). Although the Gricean theory does not make predictions regarding the psychological reality of violating the maxims (for discussion see Geurts & Rubio-Fernández, 2015; Noveck & Reboul, 2008), it does have implications for rational listeners to the extent that they expect speakers to observe the conversational principle and the maxims that follow from it (Grice, 1989). The question is, therefore, raised: How does the use of redundant adjectives influence listeners' referential processing? Empirical studies have to date provided mixed evidence (Arts et al., 2011a; Brodbeck et al., 2015; Davies & Katsos 2013; Engelhardt et al., 2006, 2011; Fukumura & Carminati, 2021; Rehring et al., 2021; Rubio-Fernández, 2020; Sedivy et al., 1999; Tourtouri et al., 2019, i.a.). These studies, however, vary in important aspects, such as the referential set size (mostly using limited contexts of up to four referents), or the kind of adjectives used redundantly (usually color and size adjectives). This variability potentially contributes to the observed discrepancy in the results. Furthermore, earlier work has largely used offline or behavioral dependent measures, making it impossible to assess the real-time comprehension of redundancy.

To our knowledge, only Engelhardt et al. (2011) have used event-related potentials (ERPs) to directly probe the online processing of overspecifications. In their study, participants were presented scenes depicting two shapes, and an accompanying spoken instruction to look at one of them. When the shapes differed (e.g., a star and a circle), a prenominal (color or size) adjective was redundant and elicited more negative amplitudes compared to a condition where the same spoken instruction was paired with a scene of two similar shapes (e.g., two stars), and the prenominal adjective was, therefore, necessary. The topography (centro-parietal) and timing (270–570 ms) of the effect were suggestive of an N400. The N400 is a negative-going deflection of the EEG signal peaking around 400 ms after the onset of a critical word, generally thought to reflect the degree to which the context supports semantic processing;

larger N400 amplitudes are associated with increased processing difficulty (for a review, see Kutas & Federmeier, 2000, 2011). The authors, therefore, concluded that overspecifications hindered listeners' processing. It is, however, possible that this is actually not an effect of redundancy, but rather reflects that listeners are capable of pragmatic reasoning in such highly simplified visual contexts. Nonetheless, the authors do not report results on the subsequent noun, where processing may also be influenced by the redundancy (cf. Rubio-Fernández, 2020; Sedivy et al., 1999; Tourtouri et al., 2019). Another ERP study (Brodbeck et al., 2015) has found evidence that overspecifications in fact benefit comprehension. In this experiment, however, the visual and linguistic stimuli were not presented simultaneously, and the results, therefore, do not speak directly to real-time referential processing.

In sum, in visually situated communication, Gricean (i.e., rational) listeners should expect utterances to convey only necessary information, and redundancy may lead them to generate unintended pragmatic inferences, which will ultimately need to be canceled, potentially impeding processing. Empirical evidence is, however, conflicting, with some studies finding that overspecifications hinder comprehension, and others that they may even have a facilitatory effect, while the variability in the stimuli and methodologies employed leaves room to alternative interpretations of these results.

1.2. Bounded-rational overspecifications

The optimal solution to a given problem may not always be attainable, for example, because crucial information is unavailable, because it requires highly complex calculations or a significant amount of time to be computed, etc. Rational agents may, therefore, resort to an approximate solution that can be achieved given such bounds on cognitive resources (Simon, 1955; see also Chase et al., 1998; Gigerenzer, 1997). Such a *bounded-rational* approach may be better suited to explain the use of OS expressions in situated contexts.

In situated communication, the immediate visual scene is part of the speakers' and listeners' common ground (Clark & Marshall, 1981; Clark, 1996). Given what information is in common ground, speakers formulate assumptions about the listeners' knowledge, based on which they infer whether their utterance will be well understood, and such assumptions can affect the initial stages of language production (Brennan & Clark, 1996; Clark, 1996; Clark & Brennan, 1991; Clark & Marshall, 1981; Galati & Brennan, 2010; Heller et al., 2012). It is, therefore, possible that cooperative speakers produce redundant expressions for the benefit of their bounded-rational conversational partners. In order to assess whether speakers' overspecifications are aimed to facilitate comprehension processes we first need to understand how they influence listeners' comprehension effort.

Recent accounts of communication that are based on information theory (Shannon, 1948) have associated the cognitive *effort* for processing a word in a sentence to the *informativity*—or information load—of that word (Hale, 2001, 2003; Levy, 2008). Information-theoretic metrics, such as surprisal and entropy reduction, can be used as quantitative estimates of the informativity of linguistic events—phonemes, words, utterances, etc.—in terms of their probability to occur in specific contexts (cf. Aylett & Turk, 2004; Crocker et al., 2016; Fenk-Oczlon, 2001; Genzel & Charniak, 2002; Jaeger, 2010; Mahowald et al., 2013). Generally speaking, *surprisal* quantifies the expectancy of a word in its context, while *entropy reduction* measures the degree to which a given word decreases uncertainty about what is being communicated (for comparisons see Frank, 2013; Linzen & Jaeger, 2016; Venhuizen et al., 2019). These metrics have been successfully used as linking hypotheses connecting theories of comprehension to observed behavioral findings. Surprisal theory predicts that the cognitive effort associated with processing a word is proportional to surprisal on that word, that is, its negative log probability given the prior context (Hale, 2001; Levy, 2008; Smith & Levy, 2013). The entropy reduction hypothesis, on the other hand, suggests that the effort associated with processing a word is directly proportional to the reduction in entropy (i.e., uncertainty about the sentence continuation) induced by that word (Hale, 2003, 2006). Linking comprehension to production processes, the Uniform Information Density (UID) hypothesis (Jaeger, 2010; Levy & Jaeger, 2007) proposes that, due to cognitive resource limitations, peaks in the amount of information conveyed by words in an utterance can increase listeners' incremental processing effort. As a consequence, rational speakers should choose encodings that distribute information (i.e., listeners' cognitive effort) as evenly as possible across their utterances. Speakers can thus ensure that the listeners' comprehension system will not be overloaded with information at any point during the utterance. A bounded-rational account of overspecification would, therefore, predict that redundant words potentially facilitate referential processing, to the extent that they allow for the distribution of informationally dense messages across longer sequences of words.

The application of such information-theoretic notions of language processing has, nevertheless, been mostly language-centric, that is, focused on how word-by-word processing is affected by the preceding linguistic context. Tourtouri et al. (2019) extended the notion of entropy reduction into visually situated comprehension, in order to measure the informativity of (redundant) words in different visual contexts and evaluate the predictions of the bounded-rational account. Their study employed eye-tracking to assess the listeners' overt attention, and the index of cognitive activity (ICA; Marshall, 2000, 2002) as a direct measure of cognitive effort (increased ICA values are associated with increased cognitive workload; for the use of ICA relative to linguistic stimuli see Demberg & Sayeed, 2016; Sekicki & Staudte, 2018; Vogels et al., 2018). The two measures yielded contrasting results, which further varied depending on the kind of adjective (color or pattern) used in the expression: Reduced cognitive effort (lower ICA) was observed on OS relative to minimally specified (MS) nouns after both color and pattern adjectives. At the same time, however, listeners were more likely to interpret pattern (but not color) adjectives contrastively (with pattern adjectives, adjective-matching referents that were part of a contrast pair received more anticipatory fixations during the adjective region compared to adjective-matching singleton referents). That is, participants expected that pattern adjectives would be used for distinguishing referents of the same type, and not redundantly. Interestingly, the rate of entropy reduction induced by the prenominal adjective was also found to influence listeners' cognitive workload (i.e., increased ICA for high compared to low entropy reduction rate on both the adjective and noun), and this influence was independent of whether the adjective was necessary or redundant.

In the present study, we probe the real-time comprehension of referentially redundant expressions with ERPs in order to evaluate the predictions of the bounded-rational account.

As detailed above, the only other ERP study (Engelhardt et al., 2011) has found an effect of overspecification (N400) on the adjective, but it is possible that this effect was amplified by the simplistic visual contexts, and/or that the subsequent noun is also influenced by redundancy (cf. Sedivy et al., 1999). Thus, we first aimed to establish the electrophysiological responses that are elicited in complex visual scenes on both the adjective and the noun, by comparing overspecifications not only to minimal descriptions but also to underspecifications (Experiment 1). We then follow up on Tourtouri et al. (2019) using ERPs in order to examine the effects of redundancy and entropy reduction on the real-time comprehension of referring expressions, and disentangle the contradictory findings of that study (Experiment 2).

2. Experiment 1

The goal of Experiment 1 was (a) to determine whether overspecifications hinder real-time comprehension or not, and (b) to identify the neural underpinnings of referential processing in situated communication. More specifically, we evaluated whether the presence of a redundant prenominal adjective incurs a processing cost as is predicted by the Gricean account, according to which adjectives should be used only when necessary. Under the Gricean account, therefore, rational listeners should expect adjectives only in relation to objects that belong in pairs, and interpret them contrastively (cf. Sedivy et al., 1999). Even though it is uncontroversial that referential ambiguity disrupts comprehension (listeners consistently rate US expressions as worse than minimal descriptions; cf. Davies & Katsos, 2013; Engelhardt et al., 2006), the real-time comprehension of underspecifications has been scarcely examined. Previous research has established a neural marker of referential ambiguity in discourse, the Nref effect: a frontally distributed sustained negativity associated with referents that are ambiguous between potential antecedents in prior discourse (Nieuwland & Van Berkum, 2008a, 2008b; Van Berkum et al., 1999). In visual contexts, (indirect) evidence from an ERP study on perspective taking (Sikos et al., 2019) suggests that referential underspecification also manifests as an Nref effect for the (US) noun. To our knowledge, however, no previous research has directly assessed listeners' neural responses to underspecification in situated comprehension. The secondary goal of this experiment was, therefore, to identify the neurophysiological marker of underspecifications in situated comprehension.

The experiment employed a referential communication task (Krauss & Glucksberg, 1977; Krauss & Weinheimer, 1964), combining spoken sentences with visual scenes. The EEG was recorded as participants listened to instructions like "*Find the yellow bowl*" in German ("*Finde die gelbe Schüssel*") coupled with visual displays such as the ones in Fig. 1. The combination of one instruction with four visual scenes generated four experimental conditions. In the MS condition, which served as a baseline, the adjective (yellow) was necessary for the identification of the target referent (the bowl), because a second object of the same type but different color was also present (Fig. 1a). In the OS condition, "*yellow*" was redundant, because the bowl was singleton, and could be identified by mention of its type alone (Fig. 1b). In the US condition, the adjective was not sufficient for target identification, as it selected two objects of the same type without disambiguating between them (see the bowls in Fig. 1c). Finally,

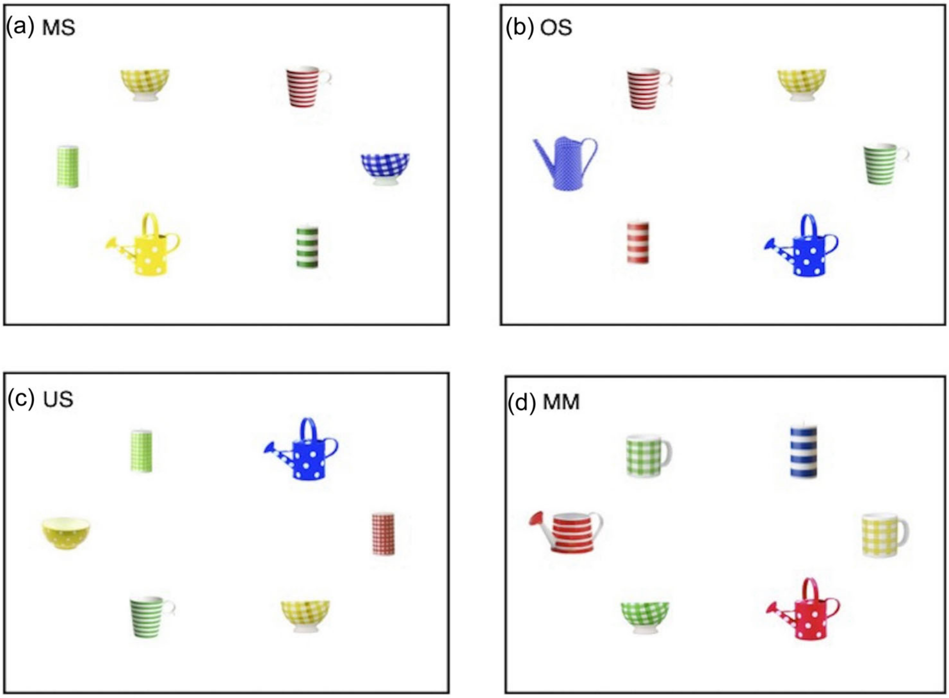

Fig 1. Experiment 1: Sample visual stimuli for color item, paired with the auditory instruction "*Find the yellow bowl*." The instruction was minimally specified (MS) in (a) where another bowl of a different color was available, thus making the use of "*yellow*" necessary. It was overspecified (OS) in (b) where the bowl was singleton, and "*yellow*" was redundant. In (c) and (d), the instruction failed to resolve reference: In (c) it was underspecified (US), because it did not distinguish between the two bows of the same color, while in (d) there is a mismatch (MM) between the information in the linguistic expression – asking for a yellow bowl—and the visual scene—where the only bowl available is not yellow.

we also included a mismatch (MM) condition, where the referent identified by the adjective (see the yellow mug in Fig. 1d) did not match the one mentioned by the noun (see the bowl in Fig. 1d). Thus, the MM condition served as a negative baseline—a case of *explicit* referential failure.

In Engelhardt et al. (2011), visual scenes always presented two referents differing in a maximum of two features (i.e., color/size and shape), and their findings may be limited to such simplified contexts in which redundancy is striking. In the current experiment, we therefore increased the number of referential candidates to six, as well as the number of features across which objects differed in any one visual context to three (color, pattern, and type). Apart from color, pattern was used as a distinguishing feature because—like color, but unlike size—it is an intrinsic property of the object it modifies and does not invoke a comparison with other referential candidates. We, thus, ensured that any preference for a Gricean (i.e., contrastive) interpretation of the adjective would be due to our manipulation, and not due to the contrastive nature of size adjectives.

According to the Gricean account, processing should be more difficult in the OS compared to the MS condition, possibly indexed by an N400 effect (Kutas & Hillyard, 1980, 1984) that could emerge after adjective onset (it can already be detected at this point that the adjective was not required; see Engelhardt et al., 2011). If, on the other hand, Gricean concerns are not at play in referential processing, no such difference should be observed. Moreover, the US condition was expected to result in a processing cost relative to the MS condition, because underspecifications fail to resolve reference. Of greater interest, however, is what kind of effect this comparison may give rise to. One possibility is that this cost may manifest as an Nref effect reflecting referential ambiguity (Nieuwland & Van Berkum, 2008a, 2008b; Sikos et al., 2019; Van Berkum et al., 1999). This effect may emerge on the adjective, as it is already apparent there that the adjective does not help to discern between the two same-type referents in the US condition (see the two bowls in Fig. 1c; a pattern adjective, e.g., *"checkered"*, would minimally distinguish between them). Alternatively, we may observe an N400 effect on the noun for the US versus the MS condition, as listeners may expect a second adjective to appear, disambiguating the potential target in the US condition. (Filler items were included that used two prenominal adjectives to distinguish between these objects). The MM condition served as a negative baseline, allowing us to test whether referential failure due to underspecification is similar to explicit referential failure due to a mismatch between the linguistic and visual input. The MM condition was predicted to yield a larger N400 compared to the MS condition on the noun, which was unexpected after the adjective in the MM condition (see the mug in Fig. 1d), but not in the MS condition. More crucial, however, is whether the US condition would result in a response similar to that in the MM condition, reflecting that they both fail to resolve reference, or to a qualitatively different one, reflecting the different nature of this failure. Alternatively, both comparisons (US vs. MS, and MM vs. MS) could also elicit a P600 effect, possibly indexing some kind of interpretation reanalysis (e.g., Kim & Osterhout, 2005; Osterhout & Holcomb, 1992) or mental updating (e.g., Brouwer, Fitz, & Hoeks, 2012; Hoeks, Stowe, Hendriks, & Brouwer, 2013).

2.1. Method

2.1.1. Participants

Thirty-three Saarland University students (average age = 25, 25 female) participated in the experiment after giving written informed consent, and were monetarily compensated for their participation. All participants were right-handed, native speakers of German with normal or corrected-to-normal vision, and no problems with color perception.

2.1.2. Materials

Pictures of everyday use objects (e.g., mugs, bowls, etc.) were used to create the visual stimuli. The objects differed in color (blue, green, red, yellow) and pattern (checkered, dotted, striped). GIMP (Version 2.8.10) was used to adjust the color hue and brightness of the object pictures. The pictures were then submitted to an offline picture naming study measuring naming agreement for the objects. Twenty-four independent participants were presented with the object pictures in all colors and patterns (distributed across eight lists), and were asked to provide descriptions including color and pattern. Thirty-two objects with inter-rater agreement 80% or higher were then employed to create the visual stimuli (see Table A1 and A2 in Appendix A for a full list of the objects used).

In total, 640 visual stimuli were created, of which 512 were used to create the experimental items, 128 were used for the fillers and 12 for practice trials. Experimental items were the combination of four displays with a single spoken instruction (cf. Fig. 1). This gave rise to 128 experimental items, half of which were paired with color instructions (color items) and the other half with pattern instructions (pattern items). All displays were created in a way that neither the target feature nor the target referent could be identified before hearing the adjective and noun in the accompanying instruction. To this end, six objects were used per display: two pairs of objects for the MS and US conditions, and two singletons for the OS and MM conditions. Fig. 1a shows the display for the MS condition, where the target referent (the yellow bowl) belongs in a contrast pair with an object of the same type and pattern but of different color (the blue bowl), thereby making the use of an adjective necessary for its identification. Four distractor objects filled the remaining positions in the colors and patterns that allowed these objects to function as target referents in the other three conditions. That is, the red mug (see Fig. 1a) could be the target referent in a potential OS condition, because "*mug*" would suffice for target identification, and the use of an adjective would be redundant. The two green candles (see Fig. 1a) could be referents in a potential US condition, since the color adjective ("*green*") would not be necessary nor sufficient for disambiguating between them. Lastly, the yellow watering-can (see Fig. 1a) could be the target in a potential MM condition. The rest of the color item displays were created in the same way, with the display structure allowing all objects potentially to be the target referent in different conditions. Pattern displays were created in a similar way, with pattern being the mentioned feature (see Fig. 1 in Supporting Information). That is, in the MS condition the target referent differed from its competitor in the mentioned feature; in the OS condition the target referent was the single object in the display bearing the mentioned feature; in the US condition there were two objects bearing the mentioned pattern but differing in color; in the MM condition, the referent did not carry the pattern mentioned in the expression, even though there was one object available with this feature. Because determiners in German are marked for gender, only same-gender objects were used in experimental displays, thus ensuring that the determiner would not reveal the target referent. Similarly, no phonological competitors (e.g., "*Schüssel*" vs. "*Schürze*") appeared in the same visual scene, so that the adjective onset would be the earliest point of disambiguation.

The apparent inconsistency between color and pattern items (i.e., that four colors were present in color displays, while only three colors were present in pattern displays) was counterbalanced in the fillers, where three-color displays were paired with color instructions and four-color displays with pattern instructions. Hence, across trials, there were no cues leading up to the target object, the condition, or the distinguishing feature before adjective onset. Fillers also counterbalanced the target position, since in experimental trials the target referent always occupied one of the four innermost positions. Thus, in filler trials target objects occupied each position as many times as necessary to ensure that across trials every object appeared as target in each position equally frequently. Furthermore, instructions in filler trials were the MS counterparts of the OS and US experimental instructions. That is, in half of the fillers we used the MS versions of the OS critical instructions (e.g., "*Find the bowl*" instead of "*Find the yellow bowl*" in Fig. 1b); the other half of the fillers were based on the US critical trials, but always contained adequate information for identifying the target referent (e.g., "*Find the yellow checkered bowl*" instead of "*Find the yellow bowl*" in Fig. 1c). Finally, we counterbalanced the target color and pattern, making sure that across trials every object appeared as target referent in all four colors and three patterns equally frequently.

The auditory stimuli were recorded with neutral intonation by a native female speaker of German in a sound isolated cabin using Cubase AI5. All instructions started with the words "*Finde den/die/das*" ("*Find the*" with the definite article in accusative in the masculine, feminine, and neuter gender). Critical instructions continued with one prenominal adjective mentioning either color (e.g., "*gelbe*", i.e. "yellow") or pattern (e.g., "*karierte*", i.e., "checkered"), followed by the head noun (e.g., "*Schüssel*", i.e., "bowl"), while filler instructions had two or zero adjectives. As speech was continuous, recordings were later annotated for adjective and noun onsets and durations using PRAAT (Version 5.3). Mean adjective duration was 481.3 ms ($SD = 32$ ms) and mean noun duration 557.2 ms ($SD = 75.7$ms).

Stimuli were divided into four lists of 256 trials using a Latin Square design, so that only one version of an item was on each list, and no one participant saw more than one condition of a given item. Lists were pseudo-randomized for each participant, so that no more than two experimental items were consecutive, and that, even when a filler interfered, critical trials of the same condition were not adjacent. The experiment was implemented and run using the E-PRIME software (Psychology Software Tools, Inc., Pittsburgh, PA, USA).

2.1.3. Procedure

Participants were seated in a sound-isolated and electromagnetically shielded cabin at a comfortable distance from a 1,680 × 1,050 resolution monitor. After they read the instructions for the experiment, they were presented with displays of each object in all colors and patterns accompanied by prerecorded audios of each object type. Before the experimental session, participants were presented with 12 practice trials, which aimed at familiarizing them with the task: to indicate whether the target referent appeared on the left side or on the right side of the screen (MS and OS conditions), or whether it was not possible to determine its position (US and MM conditions). More specifically, participants were asked to press one of four buttons to indicate their response: two buttons labeled LEFT and RIGHT indicated the side of the screen where the target referent appeared on, and two buttons with a question mark indicated that it was not possible to specify the target object's location. This means that a correct response in the MS and OS conditions would require participants to press either the LEFT or the RIGHT button, while a correct response in the US and MM conditions would require them to press one of the two question mark buttons (either one of them). When it was made sure that the participant understood the task, the experimental session began. One session was divided into eight blocks of 32 trials, in between which participants could take short breaks, and its average duration was 70 min.

A trial started with a 2.5 s preview of the visual scene. After this time, a cross appeared in the middle of the screen, and 500 ms later the spoken instruction started. At the offset of the audio stimulus the cross disappeared, while the objects remained on the screen for a wrap-up period of 500 ms. Participants were required to fixate the cross while it remained on the screen. In each trial, the display was visible for a total of approximately 5 s. At the end of each trial, a screen appeared prompting participants to perform the task. Responses were given in the form of button presses in a Cedrus response pad (Cedrus Corporation, San Pedro, California, USA).

Participants' EEG was recorded from 26 Ag/AgCl electrodes placed on the scalp according to the standard 10–20 system. The EEG signal was amplified with a BrainAmps DC amplifier (Brain Products, GmbH, Munich, Germany), and digitized at a sampling rate of 500 Hz. Eye movements and blinks were monitored by electrodes placed on the outer canthus of each eye, and above and below the right eye. Impedances were kept below 5 kΩ throughout the experiment.

2.1.4. Analysis

Offline processing of the EEG data was performed using BrainVision Analyzer 2 (Brain Products, GmbH, Munich, Germany). The EEG signal was filtered (30 Hz high cut-off) and rereferenced offline to the average of the two mastoid electrodes. Single-participant averages were then computed in a 1,000 ms window per condition relative to the onset of the adjective ("*gelbe*") and head noun ("*Schüssel*"), and aligned to a 200 ms prestimulus baseline. Trials were semiautomatically screened offline for eye movements, blinks, electrode drifts, and amplifier blocking. After artifact rejection, eight participants with less than 18 trials per condition were excluded from the analyses. Only artifact-free ERP averages time-locked to the onset of the critical regions entered the analyses.

We report statistics for response accuracy, response times (RTs), and ERPs. All analyses were carried out in R (version 3.5.1; R Core Team, 2018). Response accuracy was analyzed using generalized linear mixed models with a binomial function, and RTs were analyzed with linear mixed models (lme4 package; Bates et al., 2015). In both analyses, Specificity was treatment-coded, with the MS condition as the baseline level, and was included as a fixed factor in the model. Crossed random intercepts and slopes for participants and items were also included in the models. If the maximal model did not converge, the random effects structure was simplified (Barr et al., 2013).

For the ERPs, analyses were carried out in both the adjective region, where Engelhardt et al. (2011) report results, and on the subsequent noun, where processing might also be influenced by the redundancy (cf. Sedivy et al., 1999). Two time-windows were considered a priori: an early time-window (300–500 ms after stimulus onset), where an N400 or Nref effect could manifest (cf. Kutas & Federmeier, 2000, 2011; Nieuwland & Van Berkum, 2008a, 2008b), and a late time-window (600–900 ms after stimulus onset), where a P600 effect could emerge (cf. Hoeks et al., 2013; Kim & Osterhout, 2005). Note, however, that average adjective length was 481 ms, which means that processing of the noun had already started during the late time-window of the adjective region.[2] In the adjective region, we therefore considered only the early time-window. Analyses were carried out as follows: We first performed an omnibus ANOVA (afex package; Singmann et al., 2021) with *Specificity* (levels: MM, MS, OS, US) and *Channel* (15 levels: Fz, Cz, Pz, F3, FC1, FC5, CP1, CP5, P3, F4, FC2, FC6, CP2, CP6, P4)

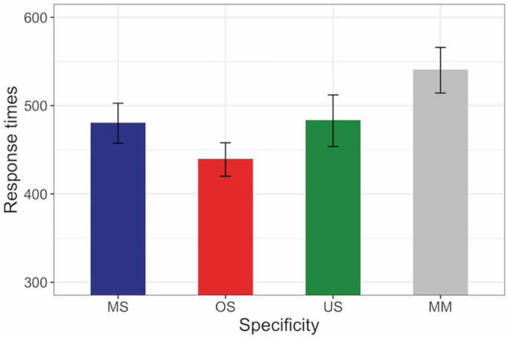

Fig 2. Experiment 1: Response times per condition. Error bars represent 95% CIs.

as predictors. Complex effects and interactions were followed up with pairwise *t*-tests (rstatix package; Kassambara, 2021) with false discovery rate (FDR) correction for multiple comparisons (see Luck, 2014). Where appropriate, we report the Greenhouse–Geisser corrected *p*-values (Greenhouse & Geisser, 1959), with the original degrees of freedom. Generalized eta-squared ($\eta^2 G$) is reported as a measure of effect size. In order to characterize the topographic distribution of any observed effects, we performed separate ANOVAs on midline (Fz, Cz, Pz) and lateral (F3, FC1, FC5, CP1, CP5, P3, F4, FC2, FC6, CP2, CP6, P4) channels. In addition to Specificity, the ANOVAs on the midline and lateral channels included the factor Longitude (levels: anterior, central, posterior). The ANOVAs over lateral channels additionally included the factor Laterality (levels: left, right). The analyses on medial and lateral channels are presented in Supporting Information.

2.2. *Results*

2.2.1. *Response accuracy*

Participants responded correctly at a rate of 90% in the MS condition, 96.1% in the OS condition, 86.8% in the US condition, and 90.4% in the MM condition. Accuracy was higher in the OS compared to the MS condition ($\beta = 1.153$, $SE = 0.252$, $z = 4.563$, $p < .001$), and in the MS compared to the US condition ($\beta = -0.451$, $SE = 0.202$, $z = -2.231$, $p = .026$). No significant differences were found between the MS and MM conditions ($p > .05$). Further analyses included trials with correct responses only.

2.2.2. *Response times*

Response times (RTs) were time-locked to the onset of the prompt display, and analyses were carried out on log-transformed RTs. As is seen in Fig. 2, participants responded faster ($\beta = -0.059$, $SE = 0.026$, $t = -2.23$, $p = .026$) in the OS condition ($M = 439.02$ ms, $SD = 250.4$ ms) and slower ($\beta = 0.085$, $SE = 0.036$, $t = 2.305$, $p = .031$) in the MM condition ($M = 540$ ms, $SD = 348.8$ ms) compared to the MS condition ($M = 480$ ms, $SD = 306.4$ ms). The comparison between the US ($M = 482$ ms, $SD = 356.9$ ms) and the MS conditions did not result in a significant difference ($p > .05$).

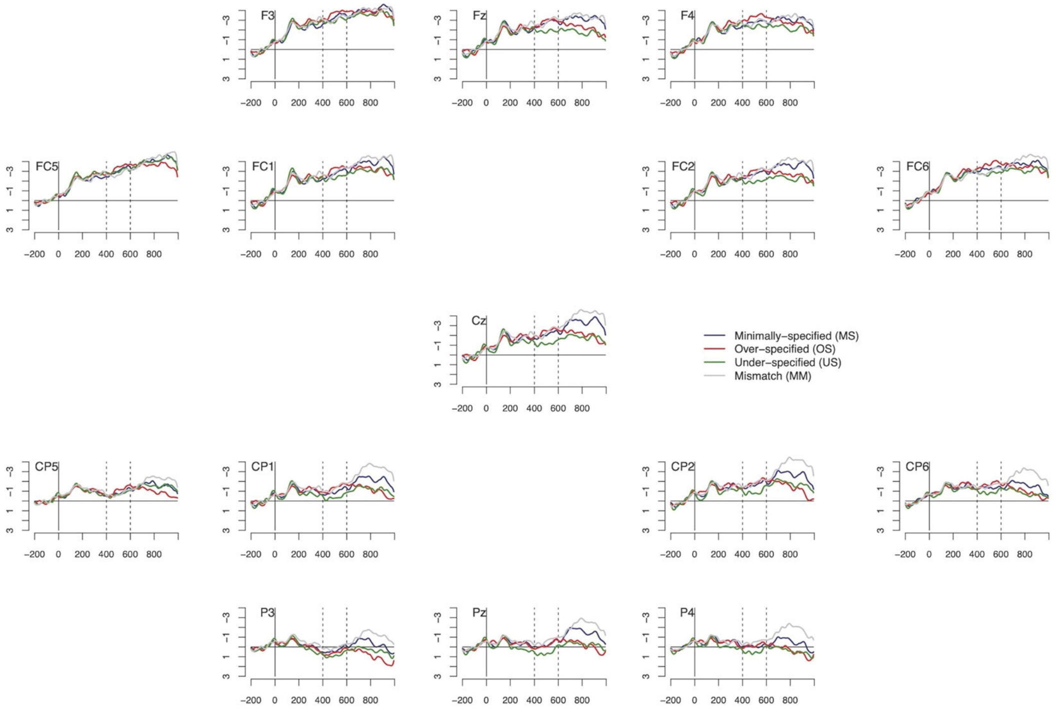

Fig 3. Experiment 1: Averaged ERPs time-locked to the onset of the adjective. Dotted lines represent the 400–600 ms analysis time-window. Negative voltages are plotted upward, in this and all subsequent plots.

2.2.3. ERPs

The omnibus ANOVA in the 300–500 ms time-window did not yield significant effects (see Appendix A). We did, however, observe in the ERP waveforms a positive deflection for the US relative to the MS, OS, and MM conditions starting at approximately 400 ms (see Fig. 3). We therefore carried out further analyses in an overlapping time-window, between 400 and 600 ms. In this time-window, the omnibus ANOVA yielded a significant Specificity × Channel interaction ($F(42,1050) = 1.59$, $p = .011$, $\eta^2 G = .005$). Follow up pairwise comparisons (see Table 1 below) indicated that the differences between the US and each of the other three conditions (MS, OS, and MM) were significant: The US trials elicited a more positive amplitude ($M = -1.28$, $SD = 1.78$) compared to the MS ($M = -1.83$, $SD = 2.05$), the OS ($M = -2.08$, $SD = 2.07$), and the MM ($M = 1.92$, $SD = 2.37$) trials. No other comparisons reached significance ($p > .05$).

In the noun region, the prestimulus baseline correction was performed on an interval (the last 200 ms of the adjective) displaying significant differences between the US and the other three conditions (see above). We therefore left out the US level of the Specificity factor in this region, and performed analyses only on the MS, OS, and MM conditions.[3] Visual inspection of the data indicated a graded negativity peaking at around 400 ms after noun onset, with the MM condition being the most negative and the OS condition the least negative (see Fig. 4).

Table 1
Experiment 1: Pairwise *t*-tests in the adjective and noun region

	Adjective (400–600 ms)			Noun (300–500 ms)			Noun (600–900 ms)		
	t	*df*	*p*	*t*	*df*	*p*	*t*	*df*	*p*
MM vs. MS	−0.451	389	.652	−4.03	389	**< .001**	−0.4	389	7.
MM vs. OS	0.761	389	.536	−10.6	389	**< .001**	−2.73	389	**.02**
MM vs. US	−3.42	389	**.002**	–	–	–	–	–	–
MS vs. OS	1.25	389	.321	−6.92	389	**< .001**	−2.42	389	**.024**
MS vs. US	−3.07	389	**.005**	—	—	—	—	—	—
OS vs. US	−4.32	389	**< .001**	—	—	—	—	—	—

Abbreviations: MM = Mismatch; MS = Minimally specified; OS = Overspecified; US = Underspecified.

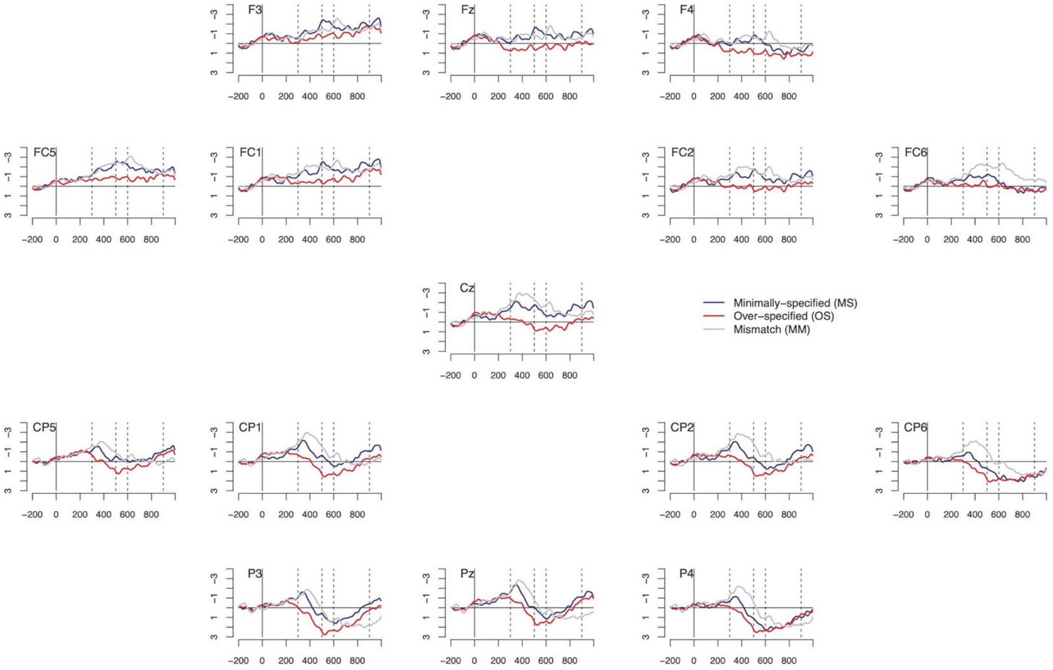

Fig 4. Experiment 1: Averaged ERPs time-locked to the onset of the noun. Dotted lines represent the N400 (300–500 ms) and P600 (600–900 ms) analysis time-windows.

The omnibus ANOVA in the 300–500 ms time-window yielded a significant effect of Specificity ($F(2,50) = 8.74, p < .001, \eta^2G = .09$), and a Specificity × Channel interaction ($F(28,700) = 3.33, p < .001, \eta^2G = .013$). Follow-up pairwise comparisons (see Table 1 above) confirmed that the differences among the MM ($M = -1.72$, $SD = 2.12$), the MS ($M = -0.95$, $SD = 2.08$) and the OS ($M = 0.16$, $SD = 1.58$) conditions were significant. Separate analyses over medial and lateral channels suggested that the effect was broadly distributed and slightly more pronounced over right and central-posterior electrodes (see Supporting Information).

In the 600–900 ms time-window, the omnibus ANOVA yielded a significant Specificity x Channel interaction ($F(28,700) = 3.09$, $p < .001$, $\eta^2G = .006$). Follow-up pairwise comparisons (see Table 1 above) showed significant differences between the OS ($M = 0.41$, $SD = 3.05$) and the MS ($M = -0.22$, $SD = 3.02$) and MM ($M = -0.32$, $SD = 3.14$) conditions. Analyses on midline and lateral electrodes in this time-window indicate a marginal effect for the MM versus OS comparison over right anterior channels (see Supporting Information).

2.3. Discussion

While both over- and underspecifications violate Gricean quantity, as neither provides the precise amount of information necessary, only underspecifications have been shown to be systematically disfavored by comprehenders in offline rating studies (Davies & Katsos, 2013; Engelhardt et al., 2006). To the best of our knowledge, however, no previous work has directly tackled the online processing of underspecifications in visual contexts (see, however, Sikos et al., 2019, for indirect evidence). As for overspecifications, past research has produced mixed results regarding whether referential redundancy is detrimental to comprehension or not (cf. Tourtouri et al., 2019; Engelhardt et al., 2011).

Experiment 1 investigated the neurophysiological correlates of OS and US referring expressions in visually situated comprehension, when compared to MS (optimal reference) and MM (referential failure) counterparts, all within a single experimental design. Specifically, we measured participants' brain responses to utterances like "*Find the yellow bowl*" (in German) that were OS in the presence of a single bowl, and US when two bowls were available. We compared processing of OS and US utterances to that of their MS counterparts (where a second bowl of a different color was present), as well as to cases of MM between the linguistic and visual input (where the only object in the scene that was yellow was not a bowl).

Our results reveal two effects of interest. First, they indicate that—in the specific visual settings we employed—referential processing is facilitated rather than hindered in the OS relative to the MS condition; as indexed by the decreased N400 amplitude on the noun. This finding is corroborated by behavioral measures, showing that participants' responses were faster and more accurate in the OS compared to the MS condition. Nonetheless, we must interpret this effect with some caution, as we will discuss below. Second, we found that, contrary to US utterances in discourse contexts, which yield an Nref effect (Nieuwland & Van Berkum, 2008a, 2008b), under-specifications in visual contexts elicit a positivity starting at around 400 ms after the onset of the ambiguity (adjective).[4]

More specifically, in the adjective region the ERPs for the MS, OS, and MM conditions overlapped, while the US condition elicited a positivity relative to all three. While this finding was the result of an exploratory analysis (the time-window was selected based on visual inspection), the magnitude of the effect leads us to speculate here on two possible interpretations. Crucially, in both the US and MS conditions the adjective selected two referents, but only in the MS condition these referents were of a *different type* (cf. the bowl and the watering can in Fig. 1a), requiring more information after the adjective to disambiguate the target. By contrast, the referents selected by the adjective in the US condition were of the *same type* (cf. the two bowls in Fig. 1c), and while the adjective allowed listeners to predict the upcoming noun itself (the two bowls were the only yellow entities in Fig. 1c), it did not help to resolve reference—a pattern adjective would have been more appropriate. Other research also reports positivities relative to conditions where information is insufficient. Hoeks, Stowe, Hendriks, and Brouwer (2013), for example, report a broadly distributed positivity, in a time-window roughly 350–900 ms after the onset of the critical word in a sentence that was US in its larger context (partial answer to question). This effect was taken to reflect increased effort in updating the mental representation of what is being communicated. Nieuwland and Van Berkum (2008b), on the other hand, report a positive deflection (late positive component) associated with referentially ambiguous anaphors for a subset of participants, in which the ambiguity did not elicit the Nref effect. The late positive component was interpreted in the context of other positive components, such as the P300, and was linked to task strategies (notably, more positive amplitudes are associated with facilitation in task performance; cf. Fabiani et al., 1987; Magliero et al., 1984). By analogy, the positivity elicited in the US versus MS condition on the adjective may index either of these processes. On the one hand, it might index a process of updating the mental model of what is being communicated. This update could amount to a general expectancy for a disambiguating adjective to appear before the noun, or to the active prediction of what this adjective could be (e.g., "dotted" or "checkered" in Fig. 1c). Alternatively, this positivity may reflect participants' ability to realize that the adjective was not helpful to resolve reference—and the ensuing readiness to give the appropriate response by pushing one of the question-mark buttons when the task prompt appears, and in case no further information was provided (i.e., in this case the positivity would index ease in executing the task). We return to this issue in the discussion of Experiment 2, and in Section 4. It should furthermore be noted that, when compared to the MS condition, the US and MM conditions elicited different effects (see N400 for the MM vs. MS conditions on the noun), suggesting that referential failure due to underspecification is qualitatively different from referential failure due to a mismatch between the visual and linguistic input. Even though the adjective was unnecessary in the MM condition too, it did reduce the set of potential referents to only one object, permitting the prediction of the upcoming noun and the identification of the target referent (which was then falsified by the subsequent noun).

In the noun region, we found a graded negativity for the MM, MS, and OS conditions, peaking at around 400 ms after the onset of the noun, with the MM condition being the most negative, and the OS condition the least negative one. The increased N400 amplitude for the MM relative to the MS condition indexes that processing was hindered in the MM condition, where the noun was not expected after a mismatching adjective (for reviews see Federmeier & Kutas, 1999; Kutas & Federmeier, 2000, 2011). By contrast, the attenuated N400 observed in the OS versus MS conditions reflects that the redundant adjective facilitated processing on the noun. This finding suggests that Gricean considerations (i.e., the computation of contrastive inferences on the adjective) are not reflected on referential processes indexed by the N400—at least not when the visual context is as complex as the contexts used here—and provides tentative support for the bounded-rational account. It is, however, possible that the facilitation observed for the OS condition should not be attributed solely to redundancy; that is, it might not be due to overspecification per se. Rather, the structure of the visual scenes in the OS and MS conditions may have contributed to this effect. That is, in the OS condition the prenominal adjective ("*yellow*") selected exactly one referent (cf. the bowl in Fig. 1b), allowing listeners to formulate precise expectations about the upcoming noun, which were confirmed when the noun was heard. This was, however, not the case in the MS condition, where the adjective selected two referents (cf. the bowl and the watering can in Fig. 1a), which were equally likely to be mentioned by the noun. Thus, the reduced N400 for the OS condition may reflect the *visually determined expectancy* of the noun, which was lower in the OS versus MS conditions (cf. Staudte et al., 2021).

These results, thus, raise the question whether overspecifications would benefit comprehension, even when expectancy is held constant—that is, if a second referent matching the adjective was available in the visual scene, such that the adjective would not be predictive of the noun. If, moreover, the second referent was part of a contrast pair (cf. Sedivy et al., 1999), this would allow us to disentangle between the Gricean and the bounded-rational accounts. Tourtouri et al. (2019) previously examined this question, but the methods used (eye movements and ICA) yielded conflicting results. In Experiment 2, we, therefore, returned to this issue, using ERPs to probe listeners' processing on the adjective and the noun, while we manipulated Specificity keeping word expectancy constant across the MS and OS conditions.

3. Experiment 2

In Experiment 2, we set out to directly investigate whether the traditional Gricean or the bounded-rational account can better explain listeners' comprehension of redundant referring expressions. More specifically, we examined whether overspecifications may benefit comprehension, even in visual contexts that support a Gricean interpretation of the prenominal adjective (i.e., where an adjective-matching referent is part of a contrast set). Word expectancy was, thus, held constant across the MS and OS conditions. In order to directly test the predictions of the bounded-rational account, we furthermore manipulated the rate at which the (redundant) adjective reduced the uncertainty about the target referent (referential entropy) across the utterance. Potential interactions between Specificity and Entropy reduction would indicate whether a benefit for overspecifications is in fact linked to the degree to which the redundant adjective reduces referential entropy.

As in Experiment 1, participants attended to audio instructions to locate a referent, for example, "*Find the blue ball*," combined with visual scenes such as those in Fig. 5. The instruction was held constant within an item, while visual scenes differed in whether the intended referent belonged in a contrast pair (cf. Fig. 5a and b, where a green ball is available) or it was singleton (cf. Fig. 5c and d, where there is no contrast object). Thus, depending on the visual context, the prenominal adjective was either necessary or redundant, and the description was MS or OS, respectively. In addition to Specificity, we manipulated Entropy Reduction, that is, the set size of the referents that matched the instruction at each point in time as the utterance unfolded (cf. two referents matching the adjective in Fig. 5a and c, and four in Fig. 5b and d). More specifically, in all conditions before the adjective was heard (i.e., at "*Find the*"), all objects were potential referential candidates, and referential entropy was $-\log_2(1/6) = 2.58$ bits. The adjective restricted the referential set size to a greater

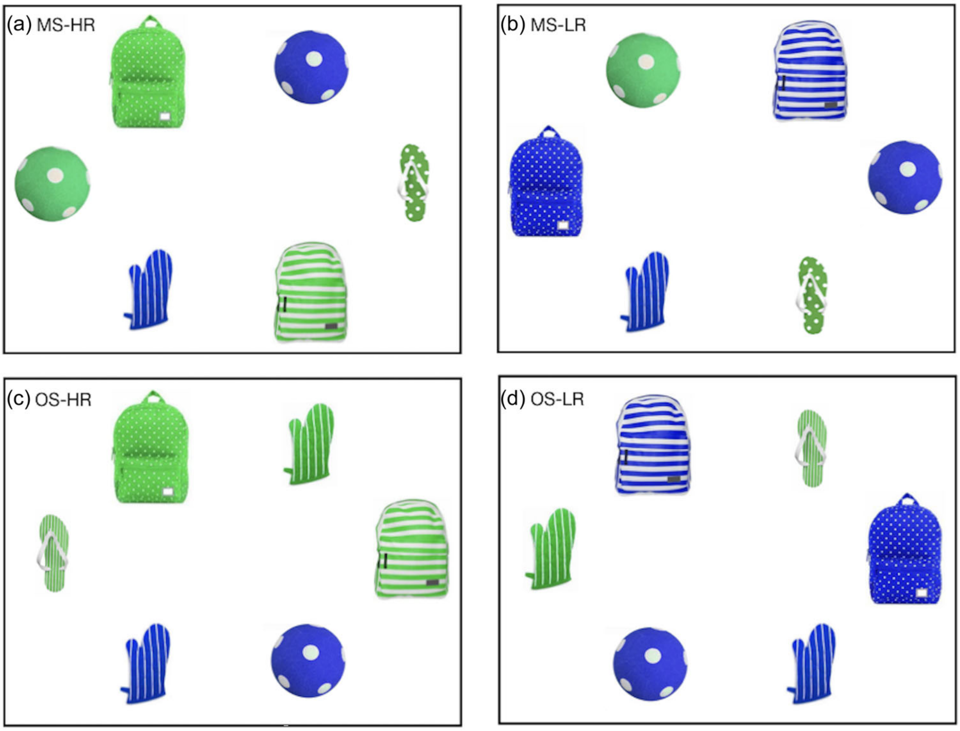

Fig 5. Experiment 2: Sample visual stimuli for a color item, paired with the auditory instruction "*Find the blue ball*." The instruction was minimally specified (MS) in (a) and (b) where a second ball of a different color was present, making the "*blue*" necessary for disambiguation. In (c) and (d) the same instruction was overspecified (OS), because the only ball in the scene was blue, and the adjective was redundant. Conditions, furthermore, differed in the rate at which the (necessary or redundant) adjective reduced referential entropy: In (a) and (c) "*blue*" reduced entropy at a higher rate (1.58 bits), as it identified two out of six potential referents (High reduction; HR). In (b) and (d) "*blue*" reduced entropy at a lower rate (0.58 bits), as it identified four out of six potential referents (Low reduction; LR).

degree when two matching referents were available, contributing to a High Reduction (HR) of referential entropy (1.58 bits in Fig. 5a and c). With four matching referents, on the other hand, the adjective restricted the set size to a lesser degree, contributing to a Low Reduction (LR) of referential entropy (0.58 bits in Fig. 5b and d). Crucially, this reduction resulted in a smaller (1 bit) or larger (2 bits) amount of residual entropy in the HR and LR conditions, respectively, which needed to be eliminated on the noun. We recorded participants' EEG, response times, and response accuracy.[5]

Two regions were considered for analysis: the adjective, and the noun. Note, however, that in the adjective region only the Entropy Reduction manipulation was relevant, because at this point in the utterance it was not feasible to determine whether the unfolding expression was MS or OS. We generally expected to replicate the results of Tourtouri et al. (2019), and observe comparable effects with ERPs. More specifically, we expected to find effects of processing effort at each reduction point (cf. Hale, 2003, 2006): increased difficulty for the HR versus LR conditions on the adjective (potentially manifested as increased N400 amplitude), and lower processing effort for the HR versus LR conditions on the noun (potentially manifested as reduced N400 amplitude; recall that referential entropy is lower on the noun after a high reduction on the adjective). Regarding Specificity, Tourtouri et al. (2019) found more anticipatory fixations to adjective-matching paired versus singleton referents, but only with pattern adjectives (and only in HR conditions). This result was taken to index participants' expectancy that pattern (but not color) adjectives should be used contrastively. This finding was, however, not met with a corresponding effect of increased cognitive workload (i.e., higher ICA) after redundant pattern adjectives; if anything, ICA on the noun was lower after redundant compared to necessary adjectives (i.e., lower in the OS vs. MS conditions independent of the kind of adjective). This discrepancy between the gaze and the ICA data, and between color and pattern overspecifications, is further investigated here, using ERPs to probe the underlying processing on the noun in the two conditions, after both color and pattern adjectives. Any effect of processing effort on the noun (e.g., increased N400 amplitude) in the OS compared to the MS conditions—possibly followed by an effect indicating interpretation reanalysis or updating (e.g., increased P600 amplitude)—would be in-line with the Gricean account. By contrast, such a difference is not predicted by the bounded-rational approach, according to which even redundant words may benefit comprehension, as long as they help to incrementally reduce referential entropy.

3.1. Method

3.1.1. Participants

Forty-four native speakers of German (mean age = 24.2, 19 female), with normal or corrected-to normal vision and no problems with color perception participated in this experiment. Participants gave written informed consent prior to the start of the experiment, and were monetarily compensated for their participation. None of them was previously exposed to the stimuli. The data from one participant were corrupted and not included in the analyses.

3.1.2. Materials

We employed the stimuli from Tourtouri et al. (2019), that is, 660 visual scenes and their corresponding audio files, and created a further 660 visual scenes. In total, 1,320 visual displays were used, 960 of which in the experimental items, and the rest 360 in the fillers. Experimental items were the combination of four visual displays with one spoken instruction. This yielded 240 experimental items, half of which were paired with color instructions (Color items; see Fig. 5), and the other half with pattern instructions (Pattern items; see Fig. 2 in Supporting Information). All experimental displays were created in a way that neither the target feature nor the target referent were identifiable before hearing the critical words. That is, six objects were used per display in two colors and two patterns. Two of the objects were singletons, and the rest were paired in two contrast sets, such that the singleton objects could potentially serve as OS targets, and the contrast objects could serve as MS targets, either with color or with pattern instructions. Only same-gender objects appeared per display, in order to ensure that the determiner would not reveal the target referent, and that the first point of entropy reduction would always be the adjective. Similarly, no phonological competitors were used in the same scene, so that the adjective onset was the disambiguation point in all trials.

Most of the filler displays (210) depicted only four referents. This introduced some variation in the stimuli, and at the same time made the six-referent experimental displays seem more complex by comparison. Half of the filler items were MS, and the other half were either OS or US. We, thus, ensured that listeners would be attentive (as it was possible that reference would not be resolved), while maintaining a lower proportion of overspecifications, as is the case in normal everyday language use. Moreover, all MS and OS filler displays, as well as some of the US displays, contained a set of *three* same-shape referents (e.g., three balls) differing in color and pattern, thus requiring a second adjective for disambiguation. The rest of the US fillers were similar to the experimental displays, but failed to establish reference (e.g., "*Find the green rucksack*" in Fig. 5). Twelve fillers were used as practice trials in a familiarization phase 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 color and pattern adjectives was counterbalanced in the two-modifier fillers. Mean word duration was 397.2 ms ($SD = 49.6$) for color adjectives, 605.1 ms ($SD = 75.1$) for pattern adjectives, and 557.2 ms ($SD = 75.7$) for the nouns.

Stimuli were divided into four lists of 588 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-randomized 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).

3.1.3. Procedure

The procedure was similar to that of Experiment 1, the only difference being that data was collected in an eye-tracking/ERP laboratory that was not electromagnetically shielded (cf. Footnote 5 above).

3.1.4. Analysis

The EEG data was preprocessed as in Experiment 1; here we additionally used a 0.1Hz low cut-off filter in order to filter out slow drifts. Single-participant averages were computed in a 1,000 ms window per condition relative to the onset of the adjective (*blaue*) and head noun (*Ball*), and aligned to a 200 ms prestimulus baseline. Trials were semiautomatically screened for eye movements, blinks, electrode drifts, and amplifier blocking. After artifact-rejection, the data from 10 participants with fewer than 30 remaining trials per condition were excluded from further analyses, and only artifact-free ERP averages time-locked to the onset of the critical region entered the analyses.

We analyzed participants' ERPs in two regions, after adjective and after noun onset. As in Experiment 1, two time-windows were considered in the noun region: an early (300–500 ms) and a late (600–900 ms) time-window. In the adjective region, only the early time-window was considered, because the average adjective length was 500 ms (397 ms in Color items; 605 ms in Pattern items), and so any later effects should be attributed to processing of the subsequent noun. Response accuracy and reaction times (RTs) in the task were also analyzed. For all analyses, we generally followed the same procedures as in Experiment 1, but further included the *Feature* of the target referent (levels: color, pattern) as a fixed factor in the models. As discussed above, Specificity is not relevant in the adjective region, because it is the subsequent noun which determines whether the expression is OS or MS. For the ERP analyses in the adjective region, we thus included only Entropy Reduction and Feature as fixed factors in the models. In the noun region, Specificity was also included in the analyses. When interactions with Feature were observed, we carried out separate ANOVAs (with FDR correction; afex package, Singmann et al., 2021) for Color and Pattern items in order to better understand and illustrate the effects. As in Experiment 1, separate analyses on midline and lateral channels were performed that allowed the topographic characterization of the observed effects. The same procedure as for the analyses on all channels was generally followed, and complex effects were followed up with pairwise *t*-tests (rstatix package; Kassambara, 2021). The analyses on medial and lateral channels are presented in Supporting Information.

3.2. *Results*

3.2.1. *Response accuracy*

The data from one participant with accuracy less than 75% were excluded from further analyses. All other participants responded accurately at a rate of 93.3% in the MS-HR condition, 92.3% in the MS-LR condition, 96.3% in the OS-HR condition, and 95.8% in the OS-LR condition. Accuracy was higher in the OS compared to the MS conditions ($\beta = -0.703$, $SE = 0.11$, $z = -6.419$, $p < .001$), and in Color compared to Pattern items ($\beta = 1.174$, $SE = 0.172$, $z = 6.816$, $p < .001$). No other effects reached significance ($p > .05$). Only trials with correct responses entered further analyses.

3.2.2. *Response times*

RTs were time-locked to the onset of the prompt display, and analyses were carried out on log-transformed RTs. As is seen in Fig. 6, participants responded faster in the HR ($M = 473$ ms, $SD = 368$) compared to the LR conditions ($M = 514$ ms, $SD = 414$; $\beta = -0.077$, $SE = 0.017$, $t = -4.483$, $p < .001$), and in the OS ($M = 465$ ms, $SD = 347$) compared to the MS conditions ($M = 522$ ms, $SD = 432$; $\beta = 0.086$, $SE = 0.018$, $t = 4.644$, $p < .001$). Faster responses were also found with color ($M = 419$ ms, $SD = 302$) compared to pattern adjectives ($M = 568$ ms, $SD = 454$; $\beta = -0.245$, $SE = 0.036$, $t = -6.847$, $p < .001$).

3.2.3. *ERPs*

In the 300–500 ms time-window of the adjective region, the omnibus ANOVA revealed a marginally significant interaction for Feature × Channel ($F(14,434) = 2.13$, $p = .089$, $\eta^2G = .002$). This effect was followed up with separate analyses for Color and Pattern items

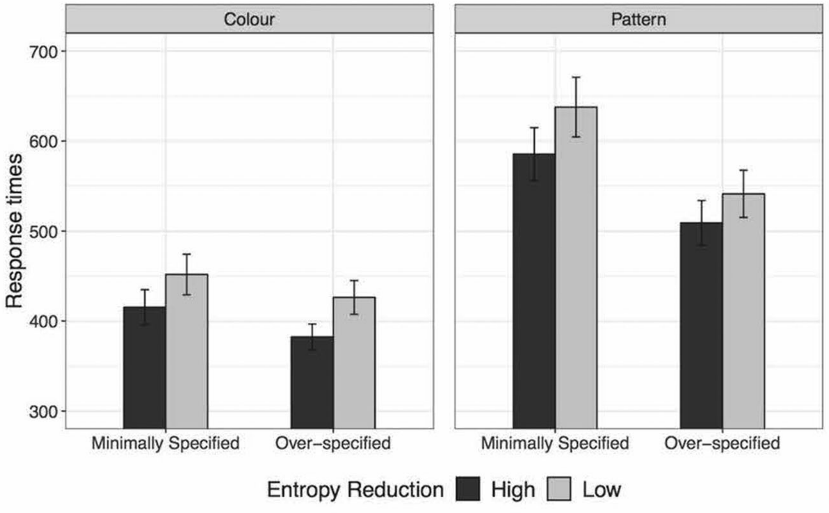

Fig 6. Experiment 2: Response times per condition. Error bars represent 95% CIs.

(see Figs. A1 and A2 in Appendix C), which however did not yield significant results (see Appendix B).

In the noun region, Specificity was also included as a factor in the analyses. In the early time-window (300–500 ms after noun onset), the omnibus ANOVA produced significant effects of Entropy Reduction (F(1,31) = 7.64, p = .01, η^2G = .014), Specificity (F(1,31) = 10.86, p = .002, η^2G = .014), and Feature (F(1,31) = 18.97, p < .001, η^2G = .038), as well as a marginally significant interaction for Entropy Reduction × Feature (F(1,31) = 3.16, p = .085, η^2G = .003). Furthermore, the interaction of Feature × Channel was significant (F(14,434) = 10.18, p < .001, η^2G = .007), and that of Entropy Reduction × Channel was marginally significant (F(14,434) = 2.2, p = .078, η^2G = .001). In the late time-window (600–900 ms), the omnibus ANOVA revealed an effect of Feature (F(1,31) = 42.49, p < .001, η^2G = .05), a three-way interaction of Entropy Reduction × Specificity × Feature (F(1,31) = 4.18, p = .049, η^2G = .003), and a marginally significant interaction of Entropy Reduction × Feature (F(1,31) = 3.05, p = .09, η^2G = .003). The interactions between Channel and Specificity (F(14,434) = 7.18, p < .001, η^2G = .003), and Channel and Feature (F(14,434) = 4.45, p = .003, η^2G = .002) were also significant (see Supporting Information for all results from the omnibus ANOVAs in the early and late time-windows). We followed up these effects with separate ANOVAs for Color and for Pattern items.

For Color items (see Table 2), the effect of Entropy Reduction was significant in the 300–500 ms time-window (see Fig. 7, marked by the first and second dotted lines), with a larger negativity elicited for the LR (M = −3.12, SD = 2.06) compared to the HR (M = −2.07, SD = 2.66) conditions. No other effects or interactions reached significance in this time-window; crucially, there was no significant difference between the MS (M = −2.33, SD = 2.15) and the OS (M = −2.86, SD = 2.59) conditions. In the 600–900 ms time-window (see Fig. 7, marked

Table 2
Experiment 2: ANOVAs on ERPs to the noun in the early (300–500 ms) and late (600–900 ms) time-windows of Color items

	300–500 ms				600–900 ms			
	df	*F*	*p*	η^2G	*df*	*F*	*p*	η^2G
Entropy Reduction	(1,31)	9.12	**.018**	.029	(1,31)	2.8	.201	.006
Specificity	(1,31)	2.45	.245	.008	(1,31)	1.28	.359	.004
Channel	(4,434)	10.49	**<.001**	.047	(4,434)	24.45	**< .001**	.144
Entropy Reduction:Specificity	(1,31)	< 0.001	.997	< .001	(1,31)	2.63	.201	.004
Entropy Reduction:Channel	(4,434)	1.79	.245	.001	(4,434)	1.22	.359	< .001
Specificity:Channel	(4,434)	1.07	.531	< .001	(4,434)	4.14	**.011**	.003
EntRed:Specificity:Channel	(4,434)	0.75	.569	< .001	(4,434)	0.7	.59	< .001

Note: EntRed = Entropy reduction.

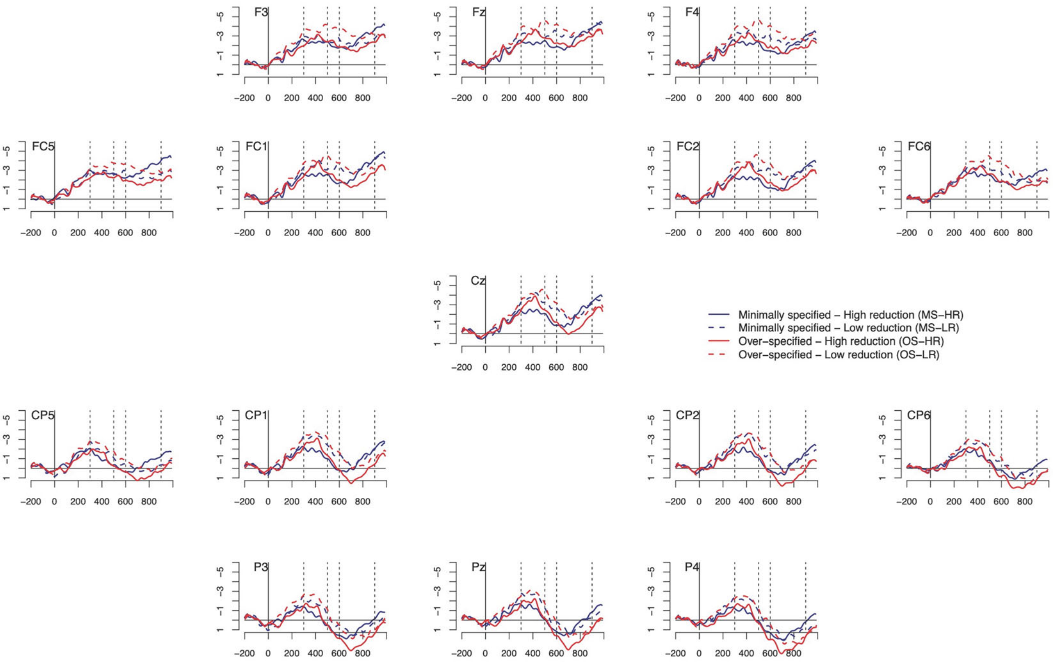

Fig 7. Experiment 2: Averaged ERPs time-locked to the onset of the noun in Color items. The first two dotted lines mark the 300–500 ms time window. The last two dotted lines mark the 600–900 ms time-window.

by the last two dotted lines), only the Specificity × Channel interaction was significant, with OS trials eliciting a more positive response ($M = -0.65$, $SD = 3.27$) compared to MS trials ($M = -1.13$, $SD = 3.01$) in central-posterior channels (see Supporting Information).

For Pattern items (see Table 3), we observed a significant effect of Specificity in the 300–500 ms time-window (see Fig. 8, marked by the first and second dotted lines): A larger negativity was elicited in OS trials ($M = -4.17$, $SD = 2.3$) relative to MS trials ($M = -3.34$,

Table 3
Exp. 2: ANOVAs on ERPs to the noun in the early (300–500 ms) and late (600–900 ms) time-windows of Pattern items

	300–500 ms				600–900 ms			
	df	*F*	*p*	η^2G	*df*	*F*	*p*	η^2G
Entropy Reduction	(1,31)	1.35	.444	.004	(1,31)	0.68	.726	<.001
Specificity	(1,31)	10.53	**.01**	.022	(1,31)	0.11	.746	<.001
Channel	(4,434)	7.28	**.002**	.034	(4,434)	21.59	**<.001**	.133
Entropy Reduction:Specificity	(1,31)	1.03	.444	.002	(1,31)	1.67	.48	.002
Entropy Reduction:Channel	(4,434)	1.34	.444	.001	(4,434)	0.7	.746	<.001
Specificity:Channel	(4,434)	1	.466	<.001	(4,434)	4	**.037**	.003
EntRed:Specificity:Channel	(4,434)	0.59	.619	<.001	(4,434)	0.59	.746	<.001

Note: EntRed = Entropy reduction.

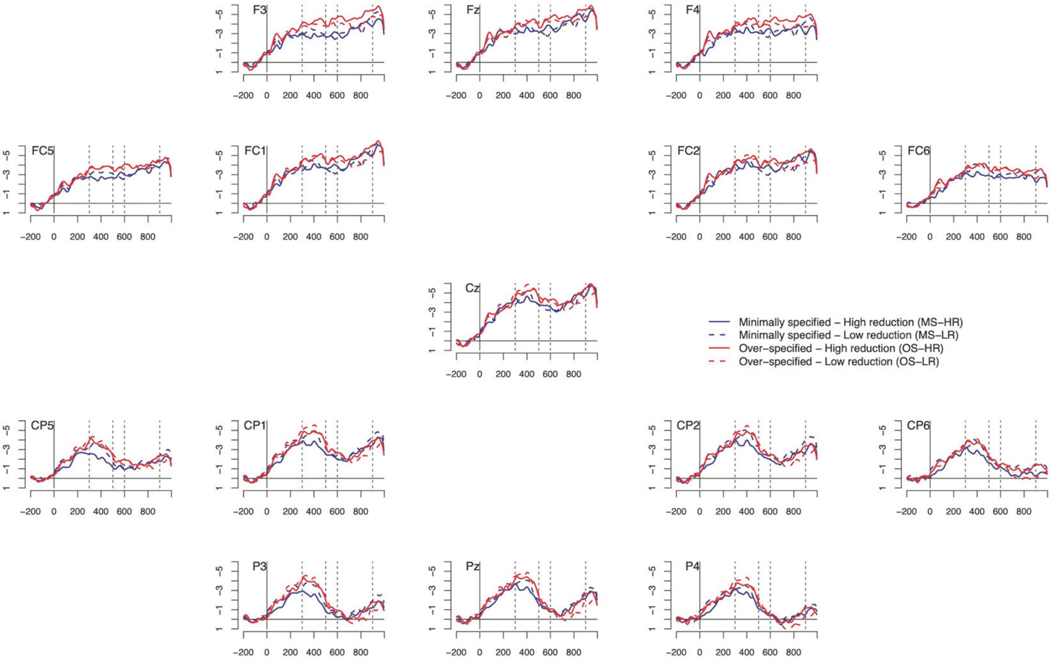

Fig 8. Experiment 2: Averaged ERPs time-locked to the onset of the noun in Pattern items. The first two dotted lines mark the 300–500 ms time window. The last two dotted lines mark the 600–900 ms time-window.

SD = 2.16). No other effects or interactions were significant; crucially, there was no difference between the HR condition ($M = -3.6$, SD = 2.52) and the LR condition ($M = -3.9$, SD = 1.98) in this time-window. In the 600–900 ms time-window (see Fig. 8, marked by the last two dotted lines), only the interaction between Specificity and Channel reached significance: A larger positivity was elicited in OS trials ($M = -2.73$, $SD = 3.35$) compared to MS trials

($M = -2.64$, $SD = 2.76$), and this effect was more pronounced over right central-posterior electrodes (see Supporting Information).

3.3. Discussion

Experiment 2 aimed to investigate: (a) whether the rate at which referential entropy is reduced across the utterance influences comprehension, and (b) whether the comprehension of overspecifications is hindered or facilitated relative to that of minimal descriptions, and if any such effects are additive. We recorded participants' brain responses as they listened to instructions to locate a target referent in a visual scene. The instructions always contained a prenominal (color or pattern) adjective, and were MS when the adjective was necessary to identify the target (i.e., the target referent was part of a contrast pair), and OS when the adjective was redundant (i.e., the target referent was singleton). Moreover, the adjective selected either two or four objects, later disambiguated by the noun, thus reducing referential entropy at a higher rate (when two objects were selected; high reduction [HR]) or at a lower rate (when four objects were selected; low reduction, [LR]).

Our results paint an intricate picture. First, the ERP results indicate that the processing of the noun in OS expressions depends on the nature of the redundant prenominal adjective. More specifically, in Pattern items the adjective resulted in an N400 effect on the subsequent noun for the OS compared to the MS conditions. This effect indicates that processing on the noun was more cumbersome when the preceding *pattern* adjective was redundant compared to when it was necessary. In Color items, on the other hand, no such effect was observed; the waveforms for the OS and MS conditions patterned together in the N400 time-window. These findings contrast with the attenuation of the N400 observed for the OS relative to the MS condition in Experiment 1, and suggest that that effect was indeed due to expectancy rather than overspecification per se (i.e., in the OS condition, the noun was highly anticipated, because the redundant adjective selected a unique referent). Furthermore, in both Color and Pattern items, the noun elicited an effect in the P600 time-window for the OS conditions relative to the MS conditions. Under one account, this positivity may reflect reanalysis or error-monitoring processes (Leckey & Federmeier, 2020). That is, listeners initially assigned a contrastive (Gricean) meaning to the adjective, and started to build a representation corresponding to a MS expression, which they had to reinterpret as an overspecification after hearing the noun. On the assumption that such reinterpretation should result in increased cognitive effort, typically manifest in behavioral measures, this account is not corroborated by the behavioral data of this study, which suggest that participants identified the target referent faster (shorter RTs), and more accurately in the OS relative to the MS conditions. These results are in line with the findings of Tourtouri et al. (2019) showing faster and easier identification of the target object in OS compared to MS conditions (shorter RTs, more fixations to the target vs. competitor object, and lower ICA values). Taking into account this data, we offer an alternative interpretation of this positivity as a task-related effect, where the increased positivity is associated with ease in performing the task (cf. Fabiani et al., 1987; Magliero et al., 1984): immediately after the noun, at the end of the trial, participants were required to perform a task, which involved pressing one of two buttons in order to identify the location of the target referent on the horizontal axis (left or right). Task performance, was easier in the OS conditions—where the target referent was singleton, and it was consequently easier to decide which button to press—compared to the MS conditions—where an object of the same type as the target (cf. the green ball in Fig. 5a) was on the other side of the display, possibly disrupting participants' decision-making process.[6] We therefore take it that the positivity for the OS conditions in the late time-window of the noun, in combination with behavioral measures and previous results show that, despite any processing cost evoked by redundant *pattern* adjectives, overspecifications result in easier, faster, and more accurate responses than minimal descriptions across the board.

Second, we found evidence that the rate at which referential entropy is reduced across the utterance influences situated comprehension. While no effects of Entropy Reduction were observed in the adjective region, Entropy Reduction was found to modulate the N400 on the noun, with the LR condition eliciting a larger negativity compared to the HR condition, in line with the Entropy Reduction Hypothesis (Hale, 2003, 2006). As explained above, residual entropy on the noun was larger in the LR conditions following a small reduction of entropy on the adjective, compared to the HR conditions, where entropy reduction on the adjective was higher. Therefore, a larger amount of referential entropy was left to be reduced on the noun in the LR conditions, which resulted in difficulty with processing the noun compared to the HR conditions (N400 effect). Similarly, a high reduction of entropy on the adjective enhanced participants' task performance (faster RTs).

In sum, Experiment 2 examined the processing of OS referring expressions, and the role of referential entropy reduction in situated comprehension. Results showed that overspecifications are processed differently depending on whether the redundant adjective was a color or a pattern adjective. While processing on the noun was more cumbersome after a redundant versus necessary pattern adjective, no such effect was found with color adjectives. The rate of entropy reduction was also found to influence processing, such that a greater reduction of referential entropy due to the (color or pattern) adjective resulted in facilitated processing on the noun.

4. General discussion

In this article, we explored whether, in contrast to strict (Gricean) rationality (Grice, 1975, 1989), an account based on bounded-rationality (cf. Simon, 1955) may better explain the use of redundancy in referential communication, reconciling the seemingly incompatible empirical evidence found in previous research (cf. Arts et al., 2011a; Engelhardt et al., 2006, 2011; Rubio-Fernández, 2020; Sedivy et al., 1999; Tourtouri et al., 2019). We proposed that redundancy in referential descriptions may help to reduce listeners' *uncertainty* about the target referent more efficiently (fast, easily, and reliably). In other words, minimal descriptions may not always be optimal (cf. Garoufi & Koller, 2014); rather, depending on factors such as the complexity of the visual scene—which is part of the speakers' and listeners' common ground (Clark & Marshall, 1981; Clark, 1996)—redundant information may facilitate listeners' cognitive effort in situated referential processing.

According to the Entropy Reduction Hypothesis (Hale, 2003, 2006), the *cognitive effort* required for processing a word in a sentence is associated with the uncertainty about the sentence continuation (quantified as entropy; Shannon, 1948) that is reduced on that word. We extended this account into visually situated comprehension, introducing the notion of *referential entropy* as a measure of the uncertainty regarding the target referent (cf. Tourtouri et al., 2019). For instance, referential entropy at "*blue*" in "*Find the blue* ..." would be lower in a visual scene containing two blue objects compared to a scene with four blue objects, even though in both cases the utterance is exactly the same. The cognitive effort for processing a word relative to the immediate visual scene should be proportional to the degree of referential entropy reduction induced by that word. A *bounded-rational* account of communication would, therefore, predict that any word in an utterance—even if it is redundant—may benefit comprehension, as long as it helps reduce referential entropy, and successfully establish reference. Under this account, overspecifications could be utilized as a means to distribute information across a longer sequence of words, thus reducing the effort associated with processing the referring expression and grounding reference.

In a first experiment, we examined the neurophysiological index of overspecifications in complex visual scenes, on both the redundant adjective and the subsequent noun. Participants were presented with scenes of six objects differing in color, pattern, and type, paired with a spoken instruction to locate one of the objects. The Specificity of the referring expression was manipulated: While the instruction (e.g., "*Find the yellow bowl*") was held constant across conditions, the visual scenes differed, rendering the utterance MS, OS or US. In both the MS and OS conditions, reference was successfully resolved, contrary to the US condition, which resulted in referential failure. In order to establish whether underspecification is qualitatively similar to explicit referential failure due to a mismatch between the visual and linguistic input, a MM condition was also used, where the adjective selected a single referent, which was, however, not mentioned by the noun. Participants' task was to make a button press specifying the location of the target referent on the horizontal axis, or indicating that it was not possible to identify a (single) target location (US and MM conditions). The results presented two important insights. First, they distinguished between two qualitatively different processes associated with referential failure: failure due to the lack of information, and failure due to the mismatch between the linguistic and visual information. The US condition, more specifically, yielded a positivity compared to the MS condition, that started at around 400 ms after the onset of the adjective and was sustained throughout the noun time-window. Based on the results of Experiment 2, this positivity was taken to reflect a task-related effect (an instance of the P300), indexing participants' ability to decide which button they would ultimately press (i.e., the button indicating that the target referent location could not be identified as either left or right). The MM condition, on the other hand, elicited the largest N400 amplitude relative to the MS condition in the noun time-window—likely an effect of low word expectancy (discussed further below). Second—and more central to the concerns of the present article—in complex visual scenes, overspecifications were beneficial to comprehension. Specifically, we found a graded N400 effect for the OS, MS, and MM conditions in the noun region, where the OS condition elicited the most reduced amplitude, and the MM condition the highest amplitude. This effect evidenced that, contra the Gricean account, a redundant prenominal adjective facilitated processing of the head noun. It is, however, possible that this facilitation was a by-product of the display structure in the OS condition, where the adjective (successfully) predicted the subsequent noun, rather than an effect of overspecification per se. This issue was addressed in Experiment 2, where both referents in contrast pairs and singletons matching the adjective were simultaneously available.

Experiment 2 further investigated whether the rate at which referential entropy is reduced across an utterance influences processing, and whether any such effect may be additive to the effects of Specificity. More specifically, in this experiment we examined whether expressions that distribute the reduction of referential entropy more evenly across the utterance may aid processing, and whether effects of Specificity may be observed above and beyond any effects of Entropy Reduction. We used a set of stimuli from a previous eye-tracking study (Tourtouri et al., 2019, Exp. 1), where items combined four visual scenes with one spoken instruction, such as "*Find the blue ball*." In all conditions, (at least) two referents matching the adjective were present: one referent was in a contrast pair, and the other one was singleton. The instruction was MS when it identified one object from within the contrast pair, and OS when it identified the singleton. Referential entropy was reduced at a HR if the prenominal adjective selected two out of six potential referents, or a LR if the adjective selected four out of six referents. Four experimental conditions were generated in this way: MS-HR, MS-LR, OS-HR, and OS-LR. The results were compelling: First, we found that the influence of overspecifications on comprehension differs depending on the kind of redundant adjective used in the expression. Processing of the noun was hindered after a redundant pattern adjective (increased N400 for OS vs. MS conditions), but not after a redundant color adjective (no difference in the N400 between OS and MS conditions). At the same time, however, both kinds of redundant adjectives facilitated listeners' identification of the target referent and their performance in the task (shorter RTs and higher accuracy for OS vs. MS). Lastly, we also observed an increased P600 for the OS relative to the MS condition in both color and pattern items. Under a reanalysis interpretation of the P600, this positivity may index the pragmatically dispreferred status of overspecifications, consistent with the N400 effect for pattern items. Alternatively, a task-related interpretation of the P600 patterns with the behavioral evidence that overspecifications nonetheless facilitate performance of the task. That is, even when the online comprehension was hindered by redundancy, the redundant adjective offered additional cues, allowing listeners to restrict the set of potential referents before the final noun, and easing their task performance. It should be noted here that these results contrast with the findings of Experiment 1, where the N400 amplitude on the noun was attenuated in the OS relative to the MS condition, and suggest that this effect was indeed due to expectancy and not due to overspecification per se (i.e., in the OS condition, the redundant adjective selected a unique referent, and thus the noun was highly anticipated). Regarding Entropy Reduction, the results showed that the rate at which referential entropy is reduced across the utterance further influences comprehension processes. In line with the Entropy Reduction hypothesis (Hale, 2003, 2006), a high reduction of entropy on the (redundant or necessary) adjective facilitated processing on the subsequent noun (reduced N400 for HR vs. LR), and improved participants' target identification times (faster RTs in HR vs. LR).

Testing the same experimental material (see current study, Exp. 2; Tourtouri et al., 2019, Exp. 1) with different measures (ERPs, fixations, ICA, RTs) provides several novel insights, tapping into two distinct aspects of visually situated comprehension: linguistic and situational. The linguistic aspect is related to the *real-time comprehension* of the referring expression, and was indexed by the anticipatory fixations observed on the adjective (Tourtouri et al., 2019, Exp. 1), and the N400 effect on the noun (current study, Exp. 2). The situational aspect is associated with the *grounding* of the referring expression in the visual context, which enables participants to perform the task, and was indexed by the ICA on both the adjective and the noun (Tourtouri et al., 2019, Exp. 1), as well as by the P600 effect on the noun and the overall RTs (current study, Exp. 2).

We have so far argued that redundant adjectives may be used rationally, as a means to manage entropy reduction (and thereby listener cognitive effort) across the utterance. Some previous research has, however, found opposing results (e.g., Davies & Katsos, 2013; Engelhardt et al., 2011), while redundant color and redundant pattern adjectives seem to differ with regard to their contribution to comprehension (see Tourtouri et al., 2019, Exp.1; current study, Exp.2). Does that mean that overspecifications are not always rational? The results of Engelhardt et al. (2011) suggest that processing was hindered on redundant relative to necessary adjectives (N400 for OS vs. MS conditions), while a slowdown in identifying the target object was also observed in the OS condition. Crucially, in their study the visual contexts were highly simplified: only two objects per scene, differing in two features. Additionally, participants were allowed a long preview time (2 sec plus 500 ms with the fixation cross) given how uncomplicated the scenes were. Thus, by the time participants were presented with the spoken instruction, they could actively *predict* how each of the two objects would be referred to, were it to be the target. As a consequence, participants experienced difficulty when they encountered the adjective in the OS condition, as this information was unexpected. For this reason, the N400 effect was elicited on the adjective in Engelhardt et al. (2011), while it was observed on the noun in our experiments; our visual scenes were more complex, and participants likely did not have enough time to make predictions about how the six displayed objects would be referred to.

Furthermore, Davies and Katsos (2013) tested a host of quality adjectives, such as *unbroken* or *modern*, along with more commonly used size adjectives, such as *tall* and *small*. Overspecified expressions were judged to be less natural for describing the target object compared to MS expressions. Such a finding, however, does not suggest that redundancy is generally detrimental to comprehension, but rather indicates that the redundant use of not salient features is dispreferred. Besides, one finding is common in all previous studies: Color adjectives have a special status. Speakers are more likely to redundantly encode color than other kinds of adjectives (see Rubio-Fernández, 2016, 2019; Tarenskeen et al., 2015, i.a.), and listeners are more likely to prefer overspecifications for color than for other features (see Fukumura & Carminati, 2021; Rehring et al., 2021; Sedivy et al., 1999). In Experiment 2, we also observed a preference for color versus pattern redundant adjectives. We do, however, wonder whether redundant pattern adjectives would still be dispreferred, in case patterns were similarly salient across objects; for example, if the dots were equally prominent on all dotted objects, and in the absence of such a salient color manipulation.

These seemingly incompatible results can be reconciled under a bounded-rational account of communication, with common ground at its core. In particular, as MS descriptions are not necessarily optimal, speakers need to consider the common ground, in order to minimize *joint effort* in establishing reference (Clark & Wilkes-Gibbs, 1986). In this sense, speakers are likely to produce *redundant* words in situations that increase the demands for successful interaction, in order to facilitate referential processing for the listeners. Even though it might be more effortful for speakers to produce OS compared to MS utterances (at least in terms of the number of words that need to be articulated), this effort will eventually pay off: Redundancy can facilitate the identification of the target referent, decreasing the likelihood of a misunderstanding that would require speakers to repeat or revise their utterance. Due to common ground, listeners are able to recognize the speakers' intention, and not ascribe other, pragmatic, meaning to the use of redundancy. Visual complexity, which is part of the common ground, affects speakers' choice of referring expressions as well as listeners' incremental interpretation of these expressions. In other words, the distributional properties of the visual scene shape the use and the on-line comprehension of referring expressions. When the demands for successful communication are relaxed (e.g., small referent set), pragmatic inferences may be generated, leading to a contrastive meaning (cf. Davies & Katsos, 2013; Engelhardt et al., 2011; Sedivy et al., 1999); when the demands are increased, the interpretation of the adjective may be tuned to a noncontrastive meaning (cf. Arts et al., 2011a; Brodbeck et al., 2015; Rubio-Fernández, 2020; Tourtouri et al., 2019, Exp.1). Thus, common ground can tune the interpretation of the adjective, affecting comprehension on the noun.

Another bounded-rational account of overspecification is offered by Degen, Hawkins, Graff, Kreiss, and Goodman (2020) within the Rational Speech Acts (RSA) framework (Frank & Goodman, 2012; Goodman & Frank, 2016). In its basic form, RSA models a pragmatic speaker who reasons about a literal listener in producing utterances: The probability with which the speaker will produce an utterance is proportional to the *utility* of that utterance, which in turn depends on utterance *informativeness* (the probability that the listener will identify the target referent based on the speaker's utterance) and *cost* (the effort that the speaker needs to expend in order to produce the utterance). Basic RSA uses Boolean semantics to compute *listener's informativeness* (i.e., assigns 1 to utterances that are true of the target referent, and 0 to utterances that are not true of the target referent). Degen and colleagues (2020) model the production of overspecifications by adopting a continuous semantics, which allows referents to take on values between 0 and 1; for example, instead of insisting that objects as absolutely blue or not blue, it allows them to very with respect to their "blue-ness." By contrast, our bounded-rational account explains referential redundancy as a means of optimizing joint effort given the information that is in common ground. In RSA terms, we propose that modifying the *utterance cost* rather than the semantics of the lexicon can better explain the use of redundancy in referential communication. That is, the cost for producing an utterance is not predetermined and should not be conceived as directly proportional to the number of words in the utterance. Rather, the extra effort that interlocutors would have to expend in case the utterance leads to communicative failure should also be taken into account in determining this cost. Furthermore, our account offers an "ecological" view of overspecifications, where both the speakers' and listeners' perspective as well as their immediate environment are considered in explaining the use of redundancy in referential communication. This view can explain for example, why overspecifications are dispreferred for certain types of adjectives (e.g., pattern) or in simplified visual contexts. While it may be possible for the Degen et al. (2020) model to qualitatively distinguish color and pattern adjectives using their fuzzy semantics approach, their current model is focused on explaining the production of overspecifications. In the absence of explicit linking hypotheses with comprehension measures, it is an open question as to whether their model can account for the full complement of the current results, for example, the neurophysiological indices regarding the comprehension of color versus pattern overspecifications, especially in relation to the degree at which they restrict the set of potential referents in the immediate visual context (cf. high vs. low entropy reduction).

To summarize, we propose that the influence of referential redundancy on communication is not predetermined, but rather depends on the conditions under which communication takes place: Aspects of the common ground, such as the complexity of the visual scene, can both affect the speakers' use of redundant adjectives and adjust the listeners' interpretation of these adjectives. Visual complexity is only one aspect of the common ground that may influence the production and comprehension of overspecifications. Other aspects, such as the importance of the task (cf. Arts et al., 2011b; Maes et al., 2004), or previous experience with/knowledge of the speaker (cf. Brown-Schmidt et al., 2015; Grodner & Sedivy, 2011; Pogue et al., 2016; Ryskin et al., 2019; Vogels et al., 2019), may also be relevant.

In conclusion, even though redundancy may impede the listener's on-line comprehension of a referring expression in some contexts, it also results in a general benefit, as it provides additional visual cues facilitating the grounding of the referring expression in the visual context. Over and above that, redundant words may contribute to the efficient reduction of referential entropy—that is, uncertainty about the target referent—facilitating grounding. In all, these results challenge strict (Gricean) rationality, and support a bounded-rational approach to referential communication, suggesting that optimal expressions are determined based on the conditions in the speakers' and listeners' common ground.

Notes

[1] For clarity, we reserve the feminine pronoun for the speaker and the masculine pronoun for the listener throughout.

[2] Recall that speech was continuous without inserting silence between the adjective and the noun.

[3] Performing baseline correction on an interval that already exhibiting a difference between conditions may have artificially pulled the waveforms together, thereby masking any potential effect between the US and the other conditions. We have therefore not considered the US condition for analyses in this region. Instead, we refer the reader to Fig. 3, where it is visible that the positive shift between the US and the other conditions that started after adjective onset (the first dotted line), is sustained after noun onset (the second dotted line); that is, the US condition elicited a long-lasting positivity that started in the adjective region.

[4] This positivity seems to be long-lasting, extending through the end of the noun region (cf. Fig. 3). See also footnote 3.

[5] For comparability with Tourtouri et al. (2019), we also recorded participants' pupillary activity, based on which the Index of Cognitive Activity (ICA) is computed. No effects were, however, found in this measure, likely because participants had to suppress their eye-movements during the presentation of the auditory stimulus. We, therefore, do not report or discuss ICA results here.

[6] Recall that a task-related interpretation was proposed for the positivity observed for the US versus MS conditions in Experiment 1.

Acknowledgments

This research was supported by the Deutsche Forschungsgemeinschaft (DFG)—SFB/CRC 1102 'Information Density and Linguistic Encoding' (Project C3) awarded to M. W. Crocker (grant number: 232722074).

The first experiment that appears in this article was presented at the annual meeting of the Cognitive Science Society in 2015 (Los Angeles, USA).

Open access funding enabled and organized by Projekt DEAL.

Conflict of Interest

The authors declare no conflicts of interest.

References

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*, *38*(4), 419–439.

Altmann, G. T., & Kamide, Y. (1999). Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition*, *73*(3), 247–264.

Anderson, J. R. (1990). *The rational analysis of thought*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Anderson, J. R. (1991). Is human cognition adaptive? *Behavioral and Brain Sciences*, *14*(3), 471–485.

Arts, A., Maes, A., Noordman, L., & Jansen, C. (2011a). Overspecification facilitates object identification. *Journal of Pragmatics*, *43*(1), 361–374.

Arts, A., Maes, A., Noordman, L., & Jansen, C. (2011b). Overspecification in written instruction. *Linguistics*, *49*(3), 555–574.

Aylett, M., & Turk, A. (2004). The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech*, *47*(1), 31–56.

Barr, D., Levy, R., Scheepers, C., & Tily, H. (2013). Random-effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278.

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

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 1482–1493.

Brodbeck, C., Gwilliams, L., & Pylkkänen, L. (2015). EEG can track the time course of successful reference resolution in small visual worlds. *Frontiers in Psychology*, *6*, 1787.

Brouwer, H., Fitz, H., & Hoeks, J. (2012). Getting real about Semantic Illusions: Rethinking the functional role of the P600 in language comprehension. *Brain Research*, *1446*, 127–143.

Brown-Schmidt, S., Yoon, S. O., & Ryskin, R. A. (2015). People as contexts in conversation. In B. H. Ross (Ed.), *Psychology of learning and motivation* (Vol. 62, pp. 59–99). Waltham, MA: Academic Press.

Chase, V. M., Hertwig, R., & Gigerenzer, G. (1998). Visions of rationality. *Trends in Cognitive Sciences*, *2*(6), 206–214.

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

Clark, H. H. (1996). *Using language*. Cambridge, UK: Cambridge University Press.

Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In L. B. Resnick, J. M. Levine, & S. D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 127–149). Washington, DC: American Psychological Association.

Clark, H. H., & Marshall, C. R. (1981). Definite knowledge and mutual knowledge. In A. K. Joshi, B. L. Webber, & I. A. Sag (Eds.) *Elements of discourse understanding* (pp. 10–63). Cambridge, UK: Cambridge University Press.

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, *22*(1), 1–39.

Crocker, M. W., Demberg, V., & Teich, E. (2016). Information density and linguistic encoding (ideal). *KI - Künstliche Intelligenz*, *30*(1), 77–81.

Davies, C., & Katsos, N. (2013). Are speakers and listeners 'only moderately Gricean'? an empirical response to Engelhardt et al. (2006). *Journal of Pragmatics*, *49*(1), 78–106.

Degen, J., Hawkins, R. D., Graf, C., Kreiss, E., Goodman, N. D (2020). When redundancy is useful: A Bayesian approach to "overinformative" referring expressions. *Psychological Review*, *2020*(4), 591–621.

Demberg, V., & Sayeed, A. (2016). The frequency of rapid pupil dilations as a measure of linguistic processing difficulty. *PLoS One*, *11*(1), 1–29.

Deutsch, W., & Pechmann, T. (1982). Social interaction and the development of definite descriptions. *Cognition*, *11*(2), 159–184.

Eberhard, K. M., Spivey-Knowlton, M. J., Sedivy, J. C., & Tanenhaus, M. K. (1995). Eye movements as a window into real-time spoken language comprehension in natural contexts. *Journal of Psycholinguistic Research*, *24*(6), 409–436.

Engelhardt, P. E., Bailey, K. G., & Ferreira, F. (2006). Do speakers and listeners observe the Gricean maxim of quantity? *Journal of Memory and Language*, *54*(4), 554–573.

Engelhardt, P., Demiral, S., & Ferreira, F. (2011). Over-specified referring expressions impair comprehension: An ERP study. *Brain and Cognition*, *77*, 304–314.

Fabiani, M., Gratton, G., Karis, D., & Donchin, E. (1987). Definition, identification, and reliability of measurement of the P300 component of the event-related brain potential. *Advances in Psychology*, *2*, 1–78.

Federmeier, K. D., & Kutas, M. (1999). A rose by any other name: Long-term memory structure and sentence processing. *Journal of Memory and Language*, *41*(4), 469–495.

Fenk-Oczlon, G. (2001). Familiarity, information flow, and linguistic form. In J. Bybee & P. Hopper (Eds.), *Frequency and the emergence of linguistic structure* (pp. 431–448). Philadelphia, PA: John Benjamins.

Frank, S. L. (2013). Uncertainty reduction as a measure of cognitive load in sentence comprehension. *Topics in Cognitive Science*, *5*(3), 475–494.

Frank, M. C., & Goodman, N. D. (2012). Predicting pragmatic reasoning in language games. *Science*, *336*(6084), 998–998.

Fukumura, K., & Carminati, M. N. (2021). Overspecification and incremental referential processing: An eye-tracking study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. https://doi.org/10.1037/xlm0001015

Galati, A., & Brennan, S. E. (2010). Attenuating information in spoken communication: For the speaker, or for the addressee? *Journal of Memory and Language*, *62*(1), 35–51.

Garoufi, K., & Koller, A. (2014). Generation of effective expressions in situated context. *Language, Cognition and Neuroscience*, *29*(8), 986–1001.

Gatt, A., Krahmer, E., van Deemter, K. & van Gompel, R.P. (2017). Reference production as search: The impact of domain size on the production of distinguishing descriptions. *Cognitive Science*, *41*, 1457–1492.

Genzel, D., & Charniak, E. (2002). Entropy rate constancy in text. In P. Isabelle, E. Charniak, & D. Lin (Eds.), *Proceedings of the 40th annual meeting on association for computational linguistics, ACL '02* (pp. 199–206). Stroudsburg, PA: Association for Computational Linguistics.

Geurts, B., & Rubio-Fernández, P. (2015). Pragmatics and processing. *Ratio*, *28*(4), 446–469.

Gigerenzer, G. (1997). Bounded rationality: Models of fast and frugal inference. *Swiss Journal of Economics and Statistics (SJES)*, *133*(II), 201–218.

Goodman, N. D., & Frank, M. C. (2016). Pragmatic language interpretation as probabilistic inference. *Trends in Cognitive Sciences*, *20*(11), 818–829.

Greenhouse, S., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika*, *24*(2), 95–112.

Grice, P. H. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics (Vol 3), speech acts* (pp. 41–58). Waltham, MA: Academic Press.

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

Grodner, D. J., & Sedivy, J. C. (2011). The effect of speaker-specific information on pragmatic inferences. In *Processing and acquisition of reference* (pp. 239–271). Cambridge, MA: MIT Press.

Hale, J. (2001). A probabilistic earley parser as a psycholinguistic model. In *Proceedings of NAACL '01* (pp. 1–8). Stroudsburg, PA: Association for Computational Linguistics.

Hale, J. (2003). The information conveyed by words in sentences. *Journal of Psycholinguistic Research*, *2*(32), 101–123.

Hale, J. (2006). Uncertainty about the rest of the sentence. *Cognitive Science*, *30*(4), 643–672.

Heller, D., Gorman, K. S., & Tanenhaus, M. K. (2012). To name or to describe: Shared knowledge affects referential form. *Topics in Cognitive Science*, *4*(2), 290–305.

Hoeks, J. C. J., Stowe, L. A., Hendriks, P., & Brouwer, H. (2013). Questions left unanswered: How the brain responds to missing information. *PLoS One*, *8*(10), 1–9.

Jaeger, T. F. (2010). Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology*, *61*(1), 23–62.

Kassambara, A. (2021). rstatix: Pipe-friendly framework for basic statistical tests. R package version 0.7.0. https://CRAN.R-project.org/package=rstatix

Kim, A., & Osterhout, L. (2005). The independence of combinatory semantic processing: Evidence from event-related potentials. *Journal of Memory and Language*, *52*(2), 205–225.

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*, *95*(1), 95–127.

Koolen, R., Gatt, A., Goudbeek, M., & Krahmer, E. (2011). Factors causing overspecification in definite descriptions. *Journal of Pragmatics*, *43*(13), 3231–3250.

Koolen, R., Goudbeek, M., & Krahmer, E. (2013). The effect of scene variation on the redundant use of color in definite reference. *Cognitive Science*, *37*, 395–411.

Koolen, R., Krahmer, E., & Swerts, M. (2016). How distractor objects trigger referential overspecification: Testing the effects of visual clutter and distractor distance. *Cognitive Science*, *40*, 1617–1647.

Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Science*, *4*(12), 463–470.

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*(1), 621–647.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, *207*(4427), 203–205.

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, *307*, 161–163.

Krauss, R. M., & Glucksberg, S. (1977). Social and nonsocial speech. *Scientific American*, *236*(2), 100–105.

Krauss, R. M., & Weinheimer, S. (1964). Changes in reference phrases as a function of frequency of usage in social interaction: A preliminary study. *Psychonomic Science*, *1*(1), 113–114.

Leckey, M, Federmeier, KD. (2020). The P3b and P600(s): Positive contributions to language comprehension. *Psychophysiology*, *57*, e13351. https://doi.org/10.1111/psyp.13351

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

Levy, R. P., & Jaeger, T. F. (2007). Speakers optimize information density through syntactic reduction. In B. Schoelkopf, J. C. Platt, & T. Hoffman (Eds.), *Advances in neural information processing systems* (vol. *19*, pp. 849–856). Cambridge, MA: MIT Press.

Linzen, T., & Jaeger, T. F. (2016). Uncertainty and expectation in sentence processing: Evidence from subcategorization distributions. *Cognitive Science*, *40*(6), 1382–1411.

Luck, S. J. (2014). The mass univariate approach and permutation statistics. In *An introduction to the event-related potential technique* (p. online). Retrieved from http://openwetware.org/wiki/Mass_Univariate_ERP_Toolbox

Maes, A., Arts, A., & Noordman, L. (2004). Reference management in instructive discourse. *Discourse Processes*, *37*(2), 117–144.

Magliero, A., Bashore, T. R., Coles, M. G., & Donchin, E. (1984). On the dependence of P300 latency on stimulus evaluation processes. *Psychophysiology*, *21*(2), 171–186.

Mahowald, K., Fedorenko, E., Piantadosi, S. T., & Gibson, E. (2013). Info/information theory: Speakers choose shorter words in predictive contexts. *Cognition*, *126*(2), 313–318.

Mangold, R., & Pobel, R. (1988). Informativeness and instrumentality in referential communication. *Journal of Language and Social Psychology*, *7*(3–4). https://doi.org/10.1177/0261927X8800700403

Marshall, S. P. (2000). Method and apparatus for eye tracking and monitoring pupil dilation to evaluate cognitive activity. *US Patent 6,090,05*. Washington, DC: U.S. Patent and Trademark Office.

Marshall, S. P. (2002). The index of cognitive activity: measuring cognitive workload. *Proceedings of the IEEE 7th Conference on Human Factors and Power Plants*, 7–7. https://doi.org/10.1109/HFPP.2002.1042860

Nieuwland, M., & Van Berkum, J. J. A. (2008a). The interplay between semantic and referential aspects of anaphoric noun phrase resolution: Evidence from ERPs. *Brain and Language*, *106*(2), 119–131.

Nieuwland, M., & Van Berkum, J. J. A. (2008b). The neurocognition of referential ambiguity in language comprehension. *Language and Linguistic Compass*, *2*(4), 603–630.

Noveck, I. A., & Reboul, A. (2008). Experimental pragmatics: a Gricean turn in the study of language. *Trends in Cognitive Sciences*, *12*(411), 425–431.

Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language*, *31*(6), 785–806.

Pechmann, T. (1989). Incremental speech production and referential overspecification. *Linguistics*, *27*(1), 89–110.

Pogue, A., Kurumada, C., & Tanenhaus, M. K. (2016). Talker-specific generalization of pragmatic inferences based on under- and over-informative prenominal adjective use. *Frontiers in Psychology*, *6*, 2035.

R Core Team. (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rehrig, G., Cullimore, R. A., Henderson, J. M., & Ferreira, F. (2021). When more is more: redundant modifiers can facilitate visual search. *Cognitive Research: Principles and Implications*, *6*(1). https://doi.org/10.1186/s41235-021-00275-4

Rubio-Fernández, P. (2016). How redundant are redundant color adjectives? An efficiency-based analysis of color overspecification. *Frontiers in Psychology*, *7*, 1–15.

Rubio-Fernández, P. (2019). Overinformative speakers are cooperative: Revisiting the Gricean maxim of quantity. *Cognitive Science*, *43*(11), e12797.

Rubio-Fernández, P. (2020). Redundant color words are more efficient than shorter descriptions. PsyArXiv. https://doi.org/10.31234/osf.io/gbpt3

Ryskin, R., Kurumada, C., & Brown-Schmidt, S. (2019). Information integration in modulation of pragmatic inferences during online language comprehension. *Cognitive Science*, *43*(8), e12769.

Sedivy, J. C., Tanenhaus, M. K., Chambers, C., & Carlson, G. (1999). Achieving incremental semantic interpretation through contextual representation. *Cognition*, *71*(2), 109–147.

Sekicki, M. & Staudte, M. (2018). Eye'll help you out! How the gaze cue reduces the cognitive load required for reference processing. *Cognitive Science*, *42*(8), 2418–2458.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*(3), 379–423.

Sikos, L., Tomlinson, S., Heins, C., & Grodner, D. (2019). What do you know? ERP evidence for immediate use of common ground during online reference resolution. *Cognition*, *182*, 275–285.

Simon, H. A., (1955). A behavioral model of rational choice. *Quarterly Journal of Economics*, *69*(1), 99–118. https://doi.org/10.2307/1884852

Singmann, H., Bolker, B., Westfall, J., Aust, F., & Ben-Shachar, M. S. (2021). afex: Analysis of factorial experiments. R package version 0.28-1. https://CRAN.R-project.org/package=afex

Smith, N., & Levy, R. (2013). The effect of word predictability on reading time is logarithmic. *Cognition*, *128*(3), 302–319.

Staudte M., Ankener, C. S., Drenhaus, H., & Crocker, M. W. (2021). Graded expectations in visually situated comprehension: Costs and benefits as indexed by the N400. *Psychonomic Bulletin & Review*, *28*, 624–631

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

Tarerskeen, S., Broersma, M., & Geurts, B. (2015). Overspecification of color, pattern, and size: salience, absoluteness, and consistency. *Frontiers in Psychology*, *6*, 1703.

Tourtouri, E. N., Delogu, F., Sikos, L., & Crocker, M. W. (2019). Rational over-specification in visually-situated comprehension and production. *Journal of Cultural Cognitive Science*, *3*(2), 175–202.

Van Berkum, J. J. A., Brown, C. M., Hagoort, P. (1999). Early referential context effects in sentence processing: Evidence from event-related brain potentials. *Journal of Memory and Language*, *41*(2), 147–182.

van Gompel, R. P. G., van Deemter, K., Gatt, A., Snoeren, R., & Krahmer, E. J. (2019). Conceptualization in reference production: Probabilistic modeling and experimental testing. *Psychological Review*, *126*(3), 345–373.

Venhuizen, N. J., Crocker, M. W., & Brouwer, H. (2019). Semantic entropy in language comprehension. *Entropy*, *21*(12), 1159.

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*, *9*, 2276.

Vogels, J., Howcroft, D. M., Tourtouri, E., & Demberg, V. (2019). How speakers adapt object descriptions to listeners under load. *Language, Cognition and Neuroscience*, *35*(1), 78–92.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Supplementary Material
Supplementary Figure S1
Supplementary Figure S2

Appendix A

Table A1
Experiment 1: ANOVAs on the ERPs to the adjective in two overlapping time-windows (300–500 ms, and 400–600 ms after adjective onset)

	300–500 ms				400–600 ms			
	Df	*F*	*p*	η^2G	*df*	*F*	*p*	η^2G
Specificity	(3,75)	0.57	.636	.006	(3,75)	1.54	.212	.013
Channel	(14,350)	15.74	<.001	.157	(14.350)	15.5	<.001	.degr17
Specificity:Channel	(42,1050)	1.25	.135	.003	(42,1050)	1.59	**.011**	.005

Appendix B

Table B1
Experiment2: ANOVAs on the ERPs to the adjective in the N400 (300–500 ms) time-window

	Effect	*df*	*F*	*p*	η^2G
All items	Entropy reduction	(1,31)	0	.951	<.001
	Feature	(1,31)	0.47	.498	.002
	Channel	(14,434)	23.85	<.001	.115
	Entropy reduction:Feature	(1,31)	1.3	.262	.002
	Entropy reduction:Channel	(14,434)	1.05	.388	<.001
	Feature:Channel	(14,434)	2.13	**.089**	.002
	Entropy reduction:Feature:Channel	(14,434)	0.49	.669	<.001
Color items	Entropy reduction	(1,31)	0.37	.545	.002
	Channel	(1,31)	25.16	<.001	.139
	Entropy reduction:Channel	(14,434)	0.79	.545	.001
Pattern items	Entropy reduction	(1,31)	0.57	.682	.002
	Channel	(1,31)	16.14	<.001	.094
	Entropy reduction:Channel	(14,434)	0.57	.684	<.001

Note. The first rows present results for all items, and are followed by results for Color and Pattern items separately.

Appendix C

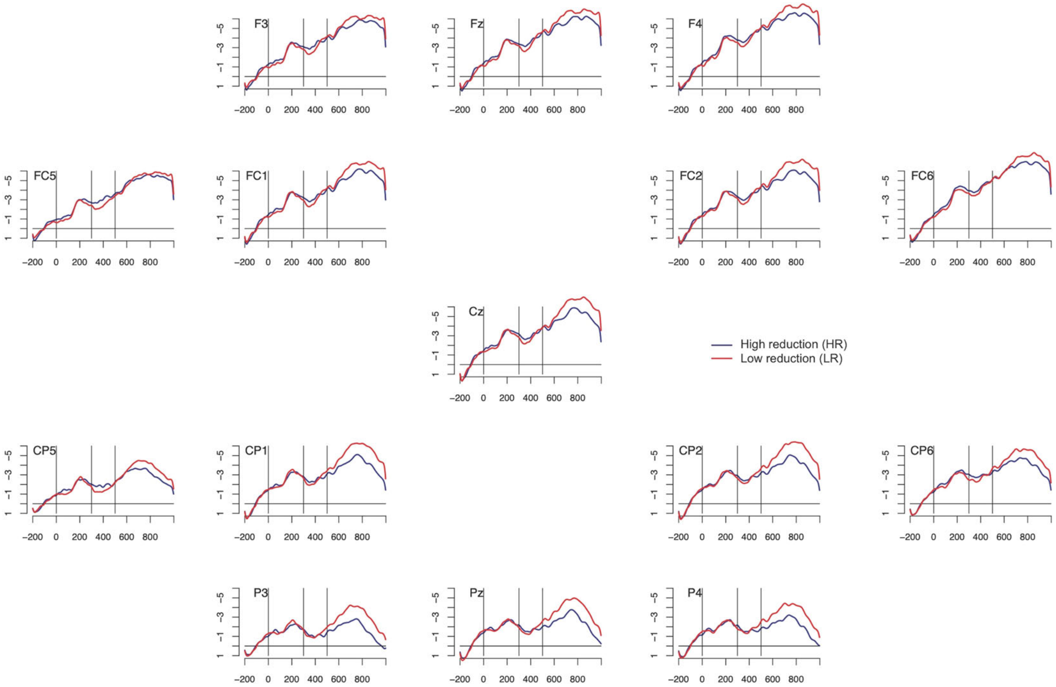

Fig C1. Experiment 2: Averaged ERPs time-locked to the onset of the adjective for High Reduction (blue line) and Low Reduction (red line) conditions in Color items. The subsequent noun started at about 400 ms.

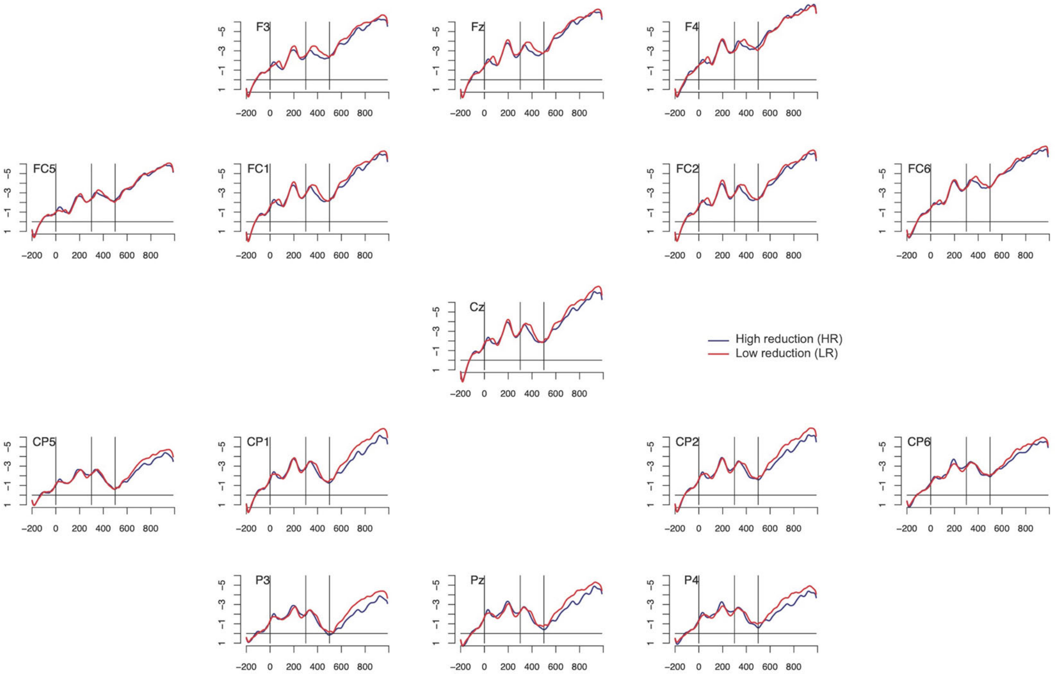

Fig C2. Experiment 2: Averaged ERPs time-locked to the onset of the adjective for High Reduction (blue line) and Low Reduction (red line) conditions in Pattern items. The subsequent noun started at about 600 ms.