Erschienen in: The Journal of Comparative Germanic Linguistics ; 22 (2019), 3. - S. 247-291
https://dx.doi.org/10.1007/s10828-019-09111-y

(The) polar bears are pink. How (the) Germans interpret (the) definite articles in plural subject DPs

Anna Czypionka[1,3] · **Tanja Kupisch**[1,2]

Abstract According to the literature, German optionally allows a definite article with generic nominals, whereas other Germanic languages require a bare nominal (e.g., English *Polar bears are white*). This optionality makes German different from other Germanic languages and more similar to Romance languages, in which definite articles are obligatory with generic nominals in subject positions. Since article use with generic nominals is seen as indicative of an advanced stage of grammaticalization, the question arises whether German has moved towards a more Romance-like stage of definite article use. We present judgment and reaction time data on generic statements. We ran two experiments monitoring the preferred reading of German definites in a nonlinguistic context, i.e., pictures of items showing either prototypical characteristics (e.g., white polar bears) or nonprototypical characteristics (e.g., pink polar bears). Given this nonlinguistic context, participants judged the truth value of auditorily presented sentences with different articles (i.e., *These/Ø/The polar bears are white/pink*). Our results show that demonstratives are interpreted as definite and bare nominals as generic. Contrary to claims in the literature, the definite article is largely interpreted as specific, following the pattern described for other Germanic languages. However, reaction times for definite articles are significantly slower than for demonstratives and bare nominals. We interpret these findings as pointing toward an ongoing change in the semantics of definite articles.

✉ Anna Czypionka
anna.czypionka@uni-konstanz.de

[1] University of Konstanz, Konstanz, Germany

[2] The Arctic University of Norway, Tromsø, Norway

[3] University of Wrocław, Wrocław, Poland

Keywords Genericity · Definiteness · Articles · German · Reaction times · Truth value judgment task

1 Introduction

Our paper is concerned with article use in German and, specifically, the presence or absence of definite articles in generalizing or 'generic' statements, such as (1a) and (1b):

(1) a. Dinosaurier sind ausgestorben. (generic NP, kind-referring NP)
dinosaur.PL *are extinct*
'Dinosaurs are extinct.'

b Katzen schlafen viel. (generic, characterizing sentence)
cat.PL *sleep*.3PL *a.lot*
'Cats sleep a lot.'

Generic statements abstract away from individuals or things, referring to general properties of classes of individuals. As shown in (1), Standard German, like English, typically uses bare plural subject NPs to express genericity. Krifka et al. (1995, 4) differentiate *generic noun phrases* ("kind NPs"), such as (1a), from *characterizing sentences* like those in (2). Our paper is concerned with characterizing sentences, but since the two types of genericity are related, we will also include kind NPs when introducing the background of our study.

There are no linguistic elements whose only function is to mark genericity (Lyons 1999, 181). Many languages use articles, but articles can express other semantic distinctions beyond genericity. Moreover, the way in which articles are used to express genericity varies across languages (see, e.g., Casalegno 1987; Vergnaud and Zubizarreta 1992; Longobardi 1994, 2001; Chierchia 1998 comparing English and Romance languages; as well as Lyons 1999 for a general overview). English and German are generally known to employ bare nouns in order to express a generic meaning (see (1)), while the Romance languages (here exemplified by Italian) standardly use noun phrases with a definite article.[1] In the Romance languages, these definite NPs can be ambiguous between a specific and a generic reading. As shown in (2a) for Italian, the subject NP can refer either to a specific group of cats or to cats in general. Often, the intended generic interpretation can be reconstructed based on extra-linguistic or linguistic context, as in (2b), or by means of adverbs that reinforce the intended reading, as in (2c).

[1] In both Germanic and Romance definite singulars are also possible, with similar restrictions. Again, the details are not relevant for our purposes (see Carlson 1977).

(2) a. I gatti dormono molto. (specific/generic)
ART.PL *cat*.PL *sleep*.3PL *a.lot*
'Cats sleep a lot.'
b. **Ho imparato a scuola** che i gatti dormono molto. (generic)
I learned at school that ART.PL *cat*.PL *sleep*.3PL *a.lot*
'I learned at school that cats sleep a lot.'
c. **In genere/normalmente,** i gatti dormono molto. (generic)
In general/normally ART.PL *cat*.PL *sleep*.3PL *a.lot*
'Generally/normally, cats sleep a lot.'

In English and German, where generic meanings are expressed by bare nouns, another type of ambiguity is found, as bare NPs can also have an existential interpretation besides the generic reading (Diesing 1992). This is illustrated in (3a), where *hungry-looking dogs* corresponds to *some hungry-looking dogs*. In the case of the existential reading, the predicate has an eventive or transitory character; it is clear that no statement about the entire class of (all) hungry-looking dogs is being made. The Italian equivalent is preferable with an indefinite article, as shown in (3b), though speaker opinions vary.

(3) a. Hungry-looking dogs are rummaging in my garbage. (Zamparelli 2002)
b. (Dei) cani dall'aria affamata
INDF.ART.M.PL *dog*.M.PL *from*.DEF.ART.F.SG *air*.F.SG *hungry*.F.SG
stanno rovistando nel mio giardino.
AUX.PROG.3PL *search*.GER *in*.DEF.ART.M.SG POSS.1SG.M.SG *garden*.M.SG
'Hungry-looking dogs are rummaging in my garden.'

In the discussion of the meaning of these two types of NPs, i.e., generic and existential, there have been two influential accounts. According to the first one, bare nouns are proper names of things and the existential reading is derived from the kind-level denotation (Carlson 1977; Chierchia 1998). According to the second account, bare nouns are ambiguous between proper names of kinds and weak indefinites (e.g., Krifka et al. 1995). The details do not matter for our purposes. The relevant point is that generics tend to be expressed by bare nominals in German, that bare nominals can also have an existential meaning, and that which of the two meanings applies depends on the type of predicate.

The use of definite articles with generic plural NPs, as found in Romance, has generally been attributed to an advanced degree of article grammaticalization. Definite articles derive from demonstratives. During the first stage of article grammaticalization, they typically refer to specific individuals or objects that are known or clearly identifiable based on linguistic or extralinguistic context. They cannot refer to abstract concepts, a nonspecific number of objects or classes of objects. In languages using (plural) definite articles with a generic reading, article use has spread beyond specific reference, having lost via semantic bleaching the original, deictic meaning inherent to the demonstratives that they were derived from. There has been a tendency in the literature on the diachronic development of

articles to attribute a more advanced stage of grammaticalization to the Romance languages, as compared to the Germanic languages (Greenberg 1978; Lyons 1999, 337; Longobardi 2001). In some of the literature, the Germanic languages and the Romance languages have been associated with different parametric settings, respectively (e.g., Longobardi 1994; Chierchia 1998). We will outline below that German does not fit this picture, as its definite articles differ from English articles at least with regard to generic subject DPs, where they show some of the properties typically associated with Romance articles. More specifically, they seem to display the same ambiguity that we see in the Romance languages, exemplified by the Italian example above in (2a).[2]

Our major goal in this paper is to establish whether the definite article in German is (still) like its English counterpart, or like articles in the major Romance languages, or in a transitional stage between the two. We further attempt to test methodological tools for examining speakers' preferred article interpretations without asking for explicit grammaticality judgments.

2 Specific and generic references with German plural NPs

2.1 An overview

In general, German patterns like English in terms of article use, i.e., genericity is expressed with bare plural NPs, or NPs with 'zero articles'; see (1b), repeated here as (4a). Specific reference can be expressed by a definite marked plural NP (4b) or by an NP with a demonstrative (4c). Intuitively, the difference between (4b) and (4c) is that the reference is more explicitly specific in (4c).

(4) a. Katzen schlafen viel.
'Cats sleep a lot.'
b. Die Katzen schlafen viel.
'The cats sleep a lot.'
c. Diese Katzen schlafen viel.
'These cats sleep a lot.'

There are many references to German in the theoretical literature on article use, mostly suggesting that the expression of generic reference in German involves optional use of the definite article (e.g., Brugger 1993; Longobardi 1994; Krifka et al. 1995; Chierchia 1998; Dayal 2004; Oosterhof 2008). The examples in (5) through (7) below are taken from the literature. Note that these examples are *not* meant to be representative of German dialects. Krifka et al.'s (1995) examples are indeed explicitly discussed in the context of Standard German.[3]

[2] Note that genericity can also be expressed by means of a singular NP, where German requires a definite article, like English. We will only be concerned with plural definites here.

[3] The authors discuss dialects separately. Here, we are leaving the issue of dialects aside, although they might have an impact on regional varieties of Standard German.

(5) a. dass die Biber intelligent sind (Brugger 1993, 14)
that the beavers intelligent COP.3PL
'that beavers are intelligent'

(6) a. (Das) Gold steigt im Preis.
(the) gold rises in price
'Gold is getting more expensive.' (Krifka et al. 1995: 68)

b. (Die) Pandabären sind vom Aussterben bedroht.
(the) panda.bears COP.3PL by extinction threatened
'Pandas are facing extinction.'
(Krifka et al. 1995: 68; Dayal 2004: 397, 441)

(7) a. (Die) Milch ist weiß. (Oosterhof 2004: 3)
(the) milk COP.3SG white
'Milk is white.'

b. (Die) Stare ernähren sich von Insekten und Früchten. (Oosterhof 2004: 2)
(the) starlings nourish REFL by insects and fruit
'Starlings feed on insects and fruit.'

The examples in (5) through (7) suggest that articles are allowed with both characterizing and kind statements.

In summary, several papers have cited individual examples of definite plural generics in German, sometimes along with suggestions of whether or not the definite article is semantically restricted, albeit without providing quantitative empirical studies that go beyond reporting the intuitions of individual speakers with selected examples.

2.2 Corpus based and experimental data

There have been a number of empirical studies on generic nominals, especially coming from the domain of language acquisition, and such studies have been mostly concerned with English. Gelman and Raman (2003) have shown that English-learning children discriminate different NP types (bare vs. definite) from an early age, though they do make some errors. Pérez-Leroux et al. (2004) also investigated definite marked and bare plural DPs, targeting children and comparing them to adult control groups. They found that younger (4–6 years) and older (6–8 years) children gave a substantial proportion of nontarget generic responses to definite determiners ($\sim$ 70%), while adults performed at ceiling, producing no definite generic errors. By contrast, mean proportions of correct generic responses to bare nouns are high for both adults and children ($\sim$ 90–95%). So, while monolingual English adults have clearly distinct readings for bare and definite plural NPs in English, English-learning children seem to be biased towards a generic reading.

The situation in German appears to be different from English, as German allows generic readings more readily than English. We are aware of only one empirical study testing generic subject NPs with adults: Barton, Kolb and Kupisch (2015)

Table 1 Acceptance of definite marked plural subjects with kind-level (e.g. *are extinct*) and individual-level (= characterizing) statements (Adapted from Barton et al. 2015)

	Hamburg (N)	Berlin (N)	Cologne (W)	Rhein-Main (W)	Freiburg (S)
Kind-level	80.3	90.7	63.9	91.7	77.1
Individual-level	42.5	69.4	54.2	77.1	35.9

asked monolingual German speakers to judge contextualized sentences which contained either a definite or a bare NP (e.g., *Sheep are white* vs. *The sheep are white*). The context had the purpose of biasing towards a generic reading and consisted of short sentences such as *Jedes Kind weiß …* 'Every child knows …' The stimuli were presented both aurally and visually. The authors examined which factors influence the use of articles with generic plural subjects. The participants were 54 monolingual speakers of Standard German from different regions in Germany: North (Hamburg), East (Berlin), West (Cologne and Rhein-Main) and South-West (Freiburg). The test controlled for linguistic factors in the stimulus material (kind-level vs. individual-level predicates) as well as sociolinguistic factors (participants' age and regional background). Some of the stimulus sentences contained a linguistic cue for genericity, such as *in general*. The motivation for comparing speakers from different regions was that comparable studies looking into definite article use with proper names (e.g., *Maria* 'Mary' vs. *die Maria* 'the Mary') had demonstrated that speakers in the Southern and Central West (Stuttgart, Mainz, Cologne) were more inclined to accept articles in these contexts (e.g., Bellmann 1990). The results of Barton et al. (2015) indicated that native German speakers accepted bare subject nominals 99.5% of the time, while also accepting definite plural articles 67.7% of the time. Moreover, Barton et al. confirmed that definite subjects were accepted more often with kind-level predicates than with individual-level predicates: the presence vs. absence of a linguistic cue did not have a significant effect. As shown in Table 1, there were differences between speakers from Northern and Southern Germany. However, these effects disappeared once the oldest speakers, who happened to have the lowest educational level, were excluded. In the end, potential effects of age, regional origin and educational background could not be teased apart, possibly because the number of speakers per region was comparatively low.[4]

Furthermore, it is possible that the nature of the task influenced response patterns: although the speakers were asked to provide spontaneous grammaticality judgments, they might have judged the sentences according to what they thought was correct in Standard German rather than what they would say intuitively, thus accepting more bare nominals than they would have produced themselves. In other words, the method may have created an underestimation of the definites that speakers would have produced in their own speech. In addition, the context

[4] The authors' assessment of dialect proficiency was based on self-assessments and it was ensured that the speakers were originally from the area in which they were tested and had not moved for a longer period of time. The speakers were tested in Standard German, although it was suspected, based on the comments of native German speakers, that the participants' knowledge of the dialect might have an influence on their judgments in Standard German.

explicitly biased speakers towards interpreting definites as generic. What we do not know, therefore, is whether speakers of German would interpret definite plural DPs as generic in the *absence* of a verbal context. Thus, a more neutral way of eliciting judgments, without providing linguistic contexts and based on a more homogeneous speaker population, may shed further light on the speakers' implicit knowledge of article use with generic subject NPs. The same authors have created a truth value judgment task without giving linguistic context, which served as a model for the task we have created below (Kupisch and Barton 2013; Barton 2016). We do not report the detailed results because the tests targeted German-Romance bilinguals and therefore cross-linguistic influence cannot be excluded.

In summary, the state of the art is somewhat inconclusive, and there are theoretical, empirical and methodological reasons why more clarification is desirable. From a theoretical point of view, claims made in the literature suggest that definite articles are acceptable with generic subject NPs, including characterizing sentences, which would mean that German has variable means (definite and bare) for expressing genericity. From an empirical and methodological point of view, we need more evidence for the above claims about German. The study by Barton et al. (2015) seems to support the view that German has started to employ the definite strategy, but, as the authors themselves noted, their results might have been compromised by the comparatively heterogeneous participant group with small groups per region. A different approach may shed new light on earlier findings, and establish new paradigms for researching genericity.

2.3 Aims and research questions

The main goal of the current study was therefore to establish whether (or to what extent) sentences with definite plural NPs are interpreted as generic or specific by native speakers of German. If they are interpreted as generic, this would indicate that there is indeed an ongoing process of grammaticalization, as discussed above. To compare the definite plural NP conditions with unambiguously generic and specific sentences, we monitored the interpretation of sentences with bare plural NPs (unambiguously generic), and demonstrative plural NPs (unambiguously specific).

As mentioned above, the interpretation of genericity is often dependent on context, because there are no exclusive linguistic devices to encode genericity (Lyons 1999, 337). For specific reference, a context is indispensable. Put differently, without a context in which particular (specific) objects are present, specific reference would not make any sense. However, in an experimental setting, one cannot construct context sentences that are identical and, at the same time, lead to equally natural follow-up sentences with generic or specific reference (see also Nordemeyer and Frank 2014 for a discussion of the influence of context on the processing of negation). To keep the context as parallel as possible between items and conditions, we therefore opted for the *absence* of a linguistic context in our experimental design, using a stimulus picture instead, so that the only difference between conditions is the type of determiner (or its absence). The stimulus picture makes specific reference plausible, while generic and ambiguous reference are

nevertheless possible. We measured acceptance rates for definite articles in contexts which may be interpreted as generic, comparing them to conditions with bare nominals and demonstratives.

In addition to acceptance rates, we measured reaction times for the different conditions, and analyzed reaction times for judgments associated with the expected interpretation of the sentences. The goal of the reaction time measurements was to assess which condition is associated with higher processing cost in the current task. The results of reaction time measurements can provide a starting point for future studies monitoring the role of task and context for the processing cost associated with different readings. This could prove valuable in the context of language acquisition data which suggest that generic reference is acquired earlier than specific reference (de Villiers and Roeper 1995; Avrutin 1999; Schafer and Villiers 2000; Baauw 2000), and leads to the question of whether one of the two can be considered a 'default'.

In summary, our research questions are the following:

(i) Do native German speakers interpret plural nouns with definite determiners as generic or as specific?
(ii) Do native German speakers interpret bare plural nouns as generic, and nouns with demonstrative determiners as specific, even in the absence of a linguistic context?
(iii) How reliable are the participants' judgments? Do the visual stimuli interfere with the interpretation of the auditive stimuli? (Knowing the answer to these questions will help us assess the relative difficulty of generic and specific interpretations in the current task.)
(iv) Are reaction times longer for specific or for generic statements? This will allow us to assess which of the two conditions is associated with higher processing load (in the current task). The results might be a first step towards explaining whether there is a default interpretation of noun phrases, i.e., the abstract concept or the specific instantiation in the real world.

3 Stimulus material preparation

We used visual stimuli depicting colored animals or fruit, paired with auditory stimuli describing the color of the depicted items in generic or specific statements. For the sake of readability, we present the preparation of the stimulus material in this separate section before presenting the experiments. We wish to stress that when designing the task we wanted to avoid any linguistic context that might bias the interpretation of the DP to either a generic or a specific interpretation. Note that stimulus properties that are treated as *factors* in later analyses are represented in small capital letters (to facilitate distinction from the discussion of these properties in a broader context) and the *levels of the factors* in italics.

3.1 Visual stimuli

The visual stimuli described in this subsection provided the nonlinguistic context for the interpretation of the critical sentences.

Pretest. 33 stimulus pictures were collected from online databases of freely available black-and-white line drawings. All pictures represented easily recognizable fruit or animals with prototypical colors or patterns. To ensure that the pictures were easily recognizable and indeed associated with the colors we had considered prototypical, we performed a pen-and-paper pretest with 18 monolingual native speakers of German. The participants were presented with the line drawings of each item (fruit or animal) printed on paper, and were asked to write down the name of the depicted item and the colors or patterns that they considered to be prototypical for them. Items were used in the final stimulus list only if 75% or more of the participants in the pretest referred to the picture with the same word. This led to the exclusion of three pictures for which various names were given. Then, a prototypical and a non-prototypical color were chosen for each item. A color counted as prototypical if it was among the colors named by the participants in the pretest, and as non-prototypical if never named. For example, the color of lions was usually given as *yellow* (the *canonical* color in the experiment), but also as *beige, yellow–brown,* and *sand*. None of the participants gave the prototypical color of lions as *green* (the *noncanonical* color in the experiment). A full list of the pretested objects, together with proportions of given names and supplied colors, is given in Appendix 1 in Table 10.

The final visual stimulus list consisted of 30 items (i.e., pictures of animals or fruit). Visual stimuli came in two conditions, defined by the factor COLOR-VISUAL, with two levels: *canonical* and *noncanonical* (color). Pictures for each item showed three to five identical drawings of black lines on a white background. For the *canonical* condition, line drawings were filled in with the prototypical color or pattern (e.g., orange carrots, white polar bears, striped tigers). For the *noncanonical* picture condition, line drawings were filled in with a color or pattern that violated the participants' world knowledge about the items' prototypical color (e.g., blue carrots, pink polar bears, checkered tigers). An example of the two picture conditions is given in Fig. 1. A full list of the stimuli with *canonical* and *noncanonical* colors is given in Appendix 1 in Table 11.

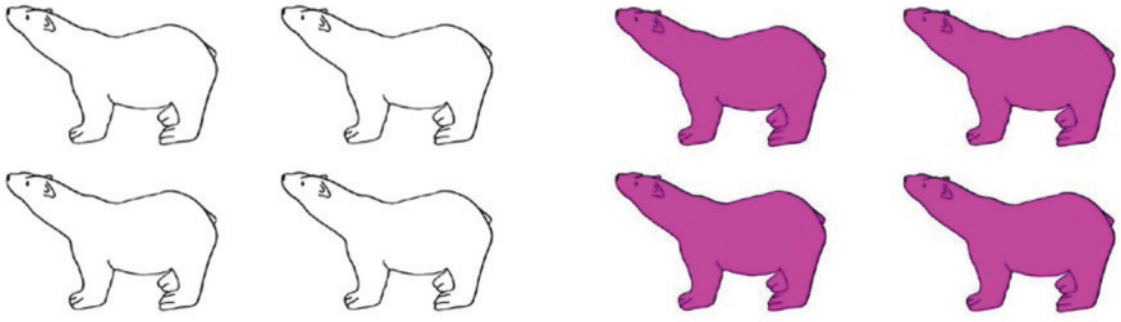


Fig. 1 Example of visual stimulus conditions *canonical* (on the left) and *noncanonical* (on the right)

3.2 Auditive stimuli

Auditive stimuli were auditorily presented statements. All auditive stimulus statements began with a plural NP referring to the objects depicted in the visual stimulus, followed by a copula and then a color or pattern term. Auditive stimuli came in six conditions, defined by two factors. The first factor was DETERMINATION, i.e., the determiner type of the NP, with three levels:

– *zero*, with a bare plural NP, e.g., *Eisbären*
– *definite*, with a plural NP containing a definite article, e.g., *die Eisbären*
– *demonstrative*, with a plural demonstrative, e.g., *diese Eisbären.*

The second factor was COLOR-AUDITIVE, i.e., the color used to describe the NP in the auditive stimulus, with two levels: *canonical* or *noncanonical* (based on world knowledge about an item's prototypical colors, as assessed in the pretest). *Canonical* and *noncanonical* colors in the auditive and in the visual stimuli were identical. The full cross of the two factors led to six auditive stimulus conditions, summarized in Table 2.

All auditive stimulus sentences were recorded as read by a female native speaker of German in a pragmatically neutral intonation. Wave files were recorded with 44,1 kHz sampling and cut off 100 ms before voice onset and 100 ms after voice onset, using Praat (Boersma 2001; Boersma and Weenink 2016).

3.3 Combination of auditive and visual stimuli

In the experiments, auditive and visual stimuli were presented in parallel. The six auditive stimulus conditions outlined above were presented with each of the two visual stimulus conditions (canonical color and noncanonical color), resulting in a total of 12 different possible conditions per item. An overview is given in Table 3, together with the predicted interpretations of the auditive stimuli and the predicted acceptance of the visual-auditive stimulus combinations.

The condition names first give the color of the visual stimulus, then determiner and color term as given in the auditive stimulus, thus following the order in which the information became available to the participants during the experiments. To illustrate: The condition name *noncanonical-zero-canonical* means that a visual

Table 2 Example of all six different auditive stimulus conditions used in the experiment

DETERMINATION	COLOR-AUDITIVE	
	Canonical	*Noncanonical*
Zero	*Eisbären sind weiß.*	*Eisbären sind pink.*
	'Polar bears are white.'	'Polar bears are pink.'
Definite	*Die Eisbären sind weiß.*	*Die Eisbären sind pink.*
	'The polar bears are white.'	'The polar bears are pink.'
Demonstrative	*Diese Eisbären sind weiß.*	*Diese Eisbären sind pink.*
	'These polar bears are white.'	'These polar bears are pink.'

Table 3 Predictions of acceptance. Each row represents one condition, with the three factors defining each condition. Colors for visual and auditive stimuli are *canonical* when they match world knowledge about an item's prototypical color, as assessed in the pretest. They are *noncanonical* when they do not match any color given as prototypical in the pretest

Condition number	COLOR-VISUAL	DETERMINATION	COLOR-AUDITIVE	used in Experiment	Prediction: Interpretation auditive	Prediction: Acceptance of visual + auditive
I	Canonical	Zero	Canonical	Exp1, Exp2	Generic	Yes
II		Zero	Noncanonical	Exp1, Exp2	Generic	No
III		Demonstrative	Canonical	Exp1, Exp2	Specific	Yes
IV		Demonstrative	Noncanonical	Exp1, Exp2	Specific	No
V		Definite	Canonical	Exp1	?	Yes if specific, Yes if generic
VI		Definite	Noncanonical	Exp1	?	No if specific, No if generic
VII	Noncanonical	Zero	Canonical	Exp1, Exp2	Generic	Yes
VIII		Zero	Noncanonical	Exp1, Exp2	Generic	No
IX		Demonstrative	Canonical	Exp1, Exp2	Specific	No
X		Demonstrative	Noncanonical	Exp1, Exp2	Specific	Yes
XI		Definite	Canonical	Exp1	?	No if specific, Yes if generic
XII		Definite	Noncanonical	Exp1	?	Yes if specific, No if generic

stimulus with a noncanonical color (e.g., pink polar bears) was presented, followed by an auditory stimulus sentence beginning with a bare plural ("Ø polar bears…") and ending with the color term that matched the canonical color of the depicted items ("… are white.").

Conditions with canonical visual colors are included to provide a fully balanced design, and serve as control conditions for participants' attention. The judgments of the respective sentences do not show the interpretation of the determiners, since both generic reference (to a prototypical white polar bear) and specific reference (to the visual stimulus depicting a white polar bear) should lead to the same reactions. Therefore, these sentences only provide a general measure of attention.[5] In contrast, conditions with noncanonical visual colors allow us to distinguish between the different interpretations of the determiners: If participants were to judge a particular determiner as specific, they should assume that the sentence is meant to describe the colors in the picture, and thus accept the auditive stimuli with noncanonical colors and reject those with canonical colors. If they were to judge a particular determiner as generic, they should assume that the sentence is meant to describe the prototypical color of the depicted items, and thus reject auditive stimuli with noncanonical colors, and accept those with canonical colors.

In conditions with DETERMINATION either *demonstrative* or *zero* the predictions are clear, as demonstratives should be interpreted as specific and bare nouns as generic. In conditions with *definite* DETERMINATION (conditions V, VI, XI and XII), different response patterns could be predicted: While some earlier studies suggest that these might be interpreted as generic (e.g., Barton et al. 2015; Krifka et al. 1995), this would not match the interpretation of definites in other Germanic languages.

In addition to the participants' acceptance or rejection, we also measured the reaction times. These reaction times should be interpreted cautiously: We aim at uncovering possible differences in processing load associated with the three NP types in this task, but we refrain from making predictions about the influence of genericity on processing load at this point. Generally, we expect "yes"-answers to be faster than "no"-answers. It is further possible that the presence of the visual stimulus favors specific over generic interpretations. Admittedly, this would represent a task effect, rather than give information about the different processing loads associated with generic and specific interpretations. Still, we will report the outcomes of the reaction time differences together with a careful discussion of potential reasons for our findings. Again, our aim in reporting these results is not to offer a definite answer to the question which reading—generic or specific—is easier to process in general (in the sense of a 'default' reading). Rather, we see our experiment as a first step toward establishing an experimental paradigm for investigating these issues (see Nordemeyer and Frank 2014 for the influence of different types of context on the processing load of negation). For now, we will

[5] To avoid confusing the participants, we did not include filler sentences in the stimulus list. We did so because it would have been very difficult to construct stimuli giving a comparable input (i.e., visual context for an auditorily presented sentence) that did not include specific or generic reference to an NP. Sentences allowing existential readings, on the other hand, could have added interpretation difficulties for the experimental conditions.

interpret our findings as reflecting the relative processing load associated with the different conditions in the currently employed task.

4 Experiment 1

In the first experiment, we used all three auditive stimulus conditions: bare plurals (unambiguously generic), demonstratives (unambiguously specific) and definites (potentially ambiguous between generic and specific readings). The interpretation of bare plurals and demonstratives in the type of sentences we tested is uncontroversial. Demonstratives can only be interpreted as specific. Bare plurals are generally ambiguous between a generic and an existential reading, but previous studies have shown acceptance of a generic reading 99% of the time for the type of predicates we used (Barton et al. 2015). Finally, definite plurals—though having a specific reading in most Germanic languages—are also accepted with generic reference in German (Barton et al. 2015; Krifka et al. 1995). The first experiment will help clarify whether and to what extent definite plural NPs are indeed ambiguous between specific and generic readings, especially if they are presented without contextual cues for one or the other interpretation. It will further assess the interpretation preferences for bare plurals and demonstratives, which we expect to be straightforward, unlike the conditions with definite articles.

4.1 Materials and methods

4.1.1 Stimulus material

The stimulus material consisted of the auditive and visual stimuli outlined above in Sect. 3. For the first experiment, all DETERMINATION conditions, i.e., *zero*, *definite* and *demonstrative*, were used. The combination of visual and auditive stimuli led to the twelve conditions illustrated in Table 4.

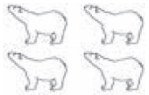
To avoid exposing the participants to too many repetitions of the same pictures, we split the stimulus list. Each list contained half of the conditions of each item with *canonical* COLOR-VISUAL pictures and half of the conditions of each item with *noncanonical* COLOR-VISUAL pictures, balanced between auditory stimulus conditions. Half of the participants saw the first stimulus list, and half saw the second stimulus list. Each participant saw 168 trials, with 14 items per condition.

4.1.2 Procedure

Participants sat in front of a computer screen and held a button box in their hands. They were instructed first orally and then in written form about the procedure. For each trial, an item was presented in one of the twelve conditions included in the stimulus set.

The presentation of visual and auditive stimuli started simultaneously. The visual stimulus (colored line drawings on a white background) was displayed surrounded

Table 4 Stimulus conditions used in Experiment 1, with examples for visual and auditive stimuli

Condition number	Visual stimulus	COLOR-VISUAL	DETERMINATION	COLOR-AUDITIVE	Example auditive	Predicted acceptance
I		Canonical	Zero	Canonical	*Eisbären sind weiß.* 'Polar bears are white.'	*Yes*
II			Zero	Noncanonical	*Eisbären sind pink.* 'Polar bears are pink.'	*No*
III			Demonstrative	Canonical	*Diese Eisbären sind weiß.* 'These polar bears are white.'	*Yes*
IV			Demonstrative	Noncanonical	*Diese Eisbären sind pink.* 'These polar bears are pink.'	*No*
V			Definite	Canonical	*Die Eisbären sind weiß.* 'The polar bears are white.'	*Yes* if specific, *Yes* if generic
VI			Definite	Noncanonical	*Die Eisbären sind pink.* 'The polar bears are pink.'	*No* if specific, *No* if generic
VII		Noncanonical	Zero	Canonical	*Eisbären sind weiß.* 'Polar bears are white.'	*Yes*
VIII			Zero	Noncanonical	*Eisbären sind pink.* 'Polar bears are pink.'	*No*
IX			Demonstrative	Canonical	*Diese Eisbären sind weiß.* 'These polar bears are white.'	*No*
X			Demonstrative	Noncanonical	*Diese Eisbären sind pink.* 'These polar bears are pink.'	*Yes*
XI			Definite	Canonical	*Die Eisbären sind weiß.* 'The polar bears are white.'	*No* if specific, *Yes* if generic
XII			Definite	Noncanonical	*Die Eisbären sind pink.* 'The polar bears are pink.'	*Yes* if specific, *No* if generic

by a black screen, revealing the COLOR-VISUAL condition. For the auditive stimuli, the first word of the sentence (bare NP, article or demonstrative) revealed the DETERMINATION condition. The last word of the sentence (the adjective) revealed the COLOR-AUDITIVE condition. 50 ms after the end of the auditive stimulus file (i.e., 150 ms after the offset of the sentence-final color term), the visual stimulus was replaced by the German sentence *Stimmt dieser Satz?* ('Is this sentence correct?'), presented in green letters on a black screen. Participants responded via presses on a button box. For half of the participants, the right button coded the answer "yes" and the left button the answer "no", for the other half, this order was reversed. The prompt remained on the screen until participants had responded. Individual trials were separated by an 800 ms blank screen. To avoid inducing bias in participants' responses, no feedback to the answers was given.

Two dependent variables were recorded during the experiment: (i) the answer given, and (ii) reaction times for giving this answer, starting from the appearance of the prompt on the screen (150 ms after the end of the auditory presentation of the color term).

Before the experiment, participants were given a training session with four practice trials resembling the critical stimuli. Stimulus presentation and data recordings were performed using the Presentation software by Neurobehavioral Systems Inc. (version 16.1). The experiment lasted for 25–30 min.

4.1.3 Participants

50 participants were tested (25 per stimulus list). All participants were recruited via the Sona Systems participant database of the University of Konstanz. All participants spoke German as their only native language (i.e., they grew up in Germany, their home language was German and they were exposed to no other language before age 6), and had normal or corrected-to-normal vision. Participants reported not being color-blind and/or hearing impaired. Participants were aged between 18 and 31 years, with a mean age of 23 years (SD = 3). 42 participants were female. All participants gave written and informed consent and received four Euros compensation for their participation.

4.2 Results

Results are described as follows: We begin with a report of the acceptance rates per condition for all participants. This allows us to assess if the different determiner types were interpreted as generic or specific. Then, we report the proportion of participants who deviated markedly from the other participants in the unambiguous conditions. Since the results are very clear, the only statistical analyses we report are post hoc comparisons of condition pairs that have small descriptive differences in acceptance rates. We then report reaction times for the participants' answers. Since we are only interested in reaction times associated with the interpretation of unambiguous conditions (i.e., bare nominals and demonstratives), we only look at reaction times for answers that correspond to the expected interpretation (i.e.,

generic and specific, respectively). For definite articles, we look at the reaction times for the interpretation that our participants preferred, as indicated by acceptance rates (i.e., we analyzed reaction times for all sentences where definite articles were interpreted as specific).

The descriptive differences and outcomes of the statistical analyses are summarized in the text, and provided in greater detail in Appendix 3. Data were prepared for statistical analysis in R (R Development Core Team 2017), using core functions and the packages *reshape*, (Wickham 2007), *plyr* (Wickham 2011), and *car* (Fox and Weisberg 2011). Data were analyzed using the packages *lme4* (Bates et al. 2015, *glmer* function for acceptance rates and *lmer* function for reaction times) and *LMERConvenienceFunctions* (Tremblay and Ransijn 2015, *summary* function).

4.2.1 Acceptance rates

An overview of mean acceptance rates and standard deviations over participants per condition is given in Table 5.

In general, the picture from the acceptance rates is clear. With *canonical* visual stimuli (e.g., white polar bears), the participants accepted sentences with *canonical* auditive stimuli, independently of the determiner type. This pattern is compatible with both generic and specific interpretations of the determiners. With noncanonical visual stimuli, where the participants' answers indicate their interpretation of the different determiners, the following picture emerges:

Demonstrative DETERMINATION **conditions** Participants accepted these with demonstratives and noncanonical auditive stimuli, and rejected them with demonstratives and canonical auditive stimuli. In other words, when presented with the picture of pink polar bears, participants accepted *These polar bears are pink*, while rejecting *These polar bears are white.* This pattern of acceptances is compatible

Table 5 Mean acceptance rates per condition over participants, with standard deviations, for Experiment 1. Data from all participants are included

Condition number	COLOR-VISUAL	DETERMINATION	COLOR-AUDITIVE	% Acceptance	S.D. acceptance
I	Canonical	Zero	Canonical	98.43	3.32
II		Zero	Noncanonical	1.14	2.65
III		Demonstrative	Canonical	97.71	4.43
IV		Demonstrative	Noncanonical	1.00	2.50
V		Definite	Canonical	98.14	4.28
VI		Definite	Noncanonical	1.00	2.89
VII	Noncanonical	Zero	Canonical	63.57	36.71
VIII		Zero	Noncanonical	35.43	35.20
IX		Demonstrative	Canonical	4.14	5.61
X		Demonstrative	Noncanonical	96.57	5.63
XI		Definite	Canonical	15.71	19.20
XII		Definite	Noncanonical	85.57	15.44

with a specific interpretation of demonstratives, as expected; acceptance proportions match those with canonical visual stimuli and demonstratives.

***Zero* DETERMINATION conditions** Participants accepted these with bare nouns (zero determiners) and canonical auditive colors in 64% of the cases, and with bare nouns and noncanonical auditive colors in 35% of the cases. In other words, when presented with a picture of pink polar bears, participants accepted *Polar bears are white*, more readily than *Polar bears are pink*. This is surprising, given that in a context like ours, bare plurals are generally thought to signal generic reference. The rates of unexpected answers associated with these two conditions are much higher than for other conditions, and standard deviations are high too, suggesting that the response patterns differ greatly between participants.

***Definite* DETERMINATION conditions** In general, participants accepted these with definite articles and noncanonical auditive stimuli, and rejected them with definite articles and canonical auditive stimuli. In other words, when presented with a picture of pink polar bears, participants accepted *The polar bears are pink* and rejected *The polar bears are white*. This pattern of acceptances is compatible with a specific interpretation of definite articles. While acceptance rates for the *noncanonical-definite-noncanonical* condition are about 10% lower than for the *noncanonical-demonstrative-noncanonical* condition, the rates still indicate a clear preference for interpreting definite articles as signaling specific and not generic reference, at least in the nonlinguistic context employed here.

Zero DETERMINATION conditions with a *noncanonical* COLOR-VISUAL visual stimulus yielded a fairly large number of unexpected responses. While variability in the *definite* articles (ambiguous between generic and specific) was expected, bare plurals should have been interpreted as generic.

A closer look at the data revealed that 14 of the 50 participants behaved very differently from all other participants.[6] One of them provided answers compatible with generic interpretations 93% of the time in the *noncanonical-definite-canonical* condition (condition XI), e.g., accepting *Die Eisbären sind weiß* ('The polar bears are white') together with a picture of pink polar bears. None of the other participants had a similar response pattern. Further support for the idea that this participant interpreted definites as generic comes from the fact that she accepted *noncanonical-definite-noncanonical* (*Die Eisbären sind pink,* 'The polar bears are pink' presented with a picture of pink polar bears) in only in 36% of the cases. In the same condition, thirteen other participants gave unexpected responses in 89% or more of the cases. These participants mostly judged bare plurals as specific when the visual stimulus depicted an item in a noncanonical color; e.g., with a picture of pink polar bears, they accepted *Polar bears are pink* and with a picture of a pink polar bear, they were prone to reject *Polar bears are white*. There are four possible reasons for this behavior: (i) They interpret bare nouns as generic but are more susceptible to interference from the visual stimulus than the others, suggesting that their deviant behavior is the result of a processing error. (ii) These participants interpret bare nouns as specific, which would be a previously unattested interpretation for these

[6] We also checked whether specific items were associated with different behavior than the others; however, none of the items gave any indication of being particularly error-prone.

expressions. (iii) They interpret the visual stimulus along the lines of a 'possible world', one in which polar bears happen to be pink. This would imply that their judgments should not be regarded as errors, but as a reflection of a misguided strategy for completing the task. (iv) They performed the task using a different strategy, namely, to check whether the color of the visual stimulus matched the color term given in the auditive stimulus without paying attention to the determiners. This strategy would have led to the expected answers in all conditions but the ones with *noncanonical* COLOR-VISUAL and *zero* DETERMINATION, which is consistent with what we found.

To assess the influence of these unusual responders on general response patterns, we removed the data from these 14 participants (28% of the total of participants) from the dataset, and reran the analysis. For the sake of readability, the outcome of this additional analysis is presented in Appendix 2. Importantly, the results for the *definite* DETERMINATION conditions were not affected by the removal of these participants.

Post hoc comparisons for condition pairs We did not perform a statistical analysis of all acceptance rates, given that the results are very clear, both for the complete and the reduced participant sets. We did however perform post hoc comparisons between some individual condition pairs in order to check whether seemingly small descriptive differences were statistically significant or not. Since we ran four different comparisons, we set $\alpha = .0125$. A full table of the results for the fixed effect of the models described below is given in Appendix 3 in Table 14.

Canonical-zero-canonical (I) vs. noncanonical-zero-canonical (VII) The first post hoc test was performed between the conditions *canonical-zero-canonical* (98.43% acceptance) and *noncanonical-zero-canonical* (63.57% acceptance). The difference between these two conditions was analyzed with a binomial generalized linear mixed model (COLOR-VISUAL as fixed effect, participants and items as random intercepts, COLOR-VISUAL as random slope for participants and items). The analysis revealed a statistically significant difference between the conditions ($z = -6.50$, $p < .001$).

Canonical-zero-noncanonical (II) vs. noncanonical-zero-noncanonical (VIII) The second post hoc comparison was performed between the conditions *canonical-zero-noncanonical* (1.14% acceptance) and *noncanonical-zero-nocanonical* (35.43% acceptance). The difference between these two conditions was analyzed using the same model as for the first post hoc test. This analysis revealed a statistically significant difference between the conditions ($z = 6.50$, $p < .001$).

Noncanonical-demonstrative-canonical (IX) vs. noncanonical-definite-canonical (XI) The third post hoc-test was performed to compare the conditions with *noncanonical* visual stimuli and auditive conditions *definite-canonical* (15.71% acceptance) or *demonstrative-canonical* (4.14% acceptance). The difference between these two conditions was analyzed with a binomial generalized linear mixed model (DETERMINATION as fixed effect, participants and items as random intercepts, DETERMINATION as random slope for participants and items). The difference between these two conditions was statistically significant ($z = 3.5$, $p < .001$).

Noncanonical-demonstrative-noncanonical (X) vs. noncanonical-definite-noncanonical (XII) The fourth post hoc test was performed to compare the conditions with *noncanonical* visual stimuli and auditive conditions *definite-noncanonical* (85.57% acceptance) and *demonstrative-noncanonical* (96.57% acceptance), using the same model as for the third post hoc test. This analysis revealed a statistically significant difference between conditions *noncanonical-demonstrative-noncanonical* and *noncanonical-definite-noncanonical* ($z = -3.60$, $p < .001$).

Summary results for acceptance rates (Experiment 1) Bare plurals: Participants accept the sentence *Polar bears are white* in the majority of cases (98.43% acceptance when presented together with a picture of white polar bears, and 63.57% acceptance together with a picture of pink polar bears). The lower acceptance rates for generic interpretation with noncanonical pictures is mainly driven by 14 participants (see above for a discussion of possible strategies, and Appendix 2 for results and analysis without these participants). Importantly, without these participants, acceptance rates rise to 84% with noncanonical pictures (pink polar bears). Acceptance rates for *Polar bears are pink* are close to zero when presented together with a picture of white polar bears, but 35.43% with a picture of pink polar bears (this acceptance rate drops to 17.46% when unusual responders are excluded). Taken together, the acceptance rates for zero determination conditions signal a generic interpretation of bare plurals. However, in our current task, there is an undeniable influence from the visual stimulus, which is more pronounced in some participants than in others. We offered a brief discussion of possible reasons for this influence above, and will return to this issue in the general discussion (Sect. 6).

Demonstratives Participants accept the sentence *These polar bears are white/pink* when the auditive color term matches the color of the visual stimulus (close to 100% for all relevant conditions). They reject the sentences when the auditive color and the visual color do not match (below 5% for all relevant conditions). This is in line with an interpretation of demonstratives as specific. This matches our expectations.

Definites Participants tend to accept the sentence *The polar bears are white/pink* when the auditive color term matches the color of the visual stimulus, and tend to reject it when the auditive and visual colors do not match. The general pattern of acceptance rates for definite articles is similar to the one for demonstratives, indicating that in our current task with a non-linguistic context, definite articles are interpreted as specific. For conditions with *noncanonical* colors in the visual stimuli (e.g., sentences presented together with pictures of pink polar bears), the interpretation of definite articles as specific seems to be less robust than that of demonstratives, with significantly lower acceptance rates for *definite-noncanonical* (*Die Eisbären sind pink*, 85.12% acceptance) than for *demonstrative-noncanonical* (*Diese Eisbären sind pink*, 97.42% acceptance), and significantly higher acceptance rates for definites than demonstratives with *canonical* colors. This contrast in acceptance rates does not change when unusual responders are excluded.

Table 6 Mean reaction times with standard deviations per condition over participants, Experiment 1 (after residual-based outlier removal)

Condition number	COLOR-VISUAL	DETERMINER	COLOR-AUDITIVE	Mean reaction times/ms	S.D. reaction times
I	Canonical	Zero	Canonical	465	292
II		Zero	Noncanonical	475	234
III		Demonstrative	Canonical	432	219
IV		Demonstrative	Noncanonical	443	244
V		Definite	Canonical	465	269
VI		Definite	Noncanonical	521	285
VII	Noncanonical	Zero	Canonical	574	318
VIII		Zero	Noncanonical	584	257
IX		Demonstrative	Canonical	464	210
X		Demonstrative	Noncanonical	451	232
XI		Definite	Canonical	622	265
XII		Definite	Noncanonical	544	303

4.2.2 Reaction times, Experiment 1

Reaction times per condition were analyzed for all responses indicating the expected interpretation, as outlined above.[7] Given that with our stimuli, participants had shown a clear preference for interpreting definite determiners as specific rather than generic, we counted answers to definites as 'expected' answers if they matched the specific interpretation. Before data analysis, reaction times shorter than 100 ms and longer than 6000 ms were removed from the dataset, leading to the removal of 5.8% of the data. Next, unexpected answers were removed from the dataset, leading to the removal of an additional 9.7% of the data. The mean reading times per condition are given in Table 6.

Statistical analysis of reaction times Log-transformed reaction times were analyzed using a linear mixed effects model, using the lmer function of the lme4 package in R. For the first model (Model 1.1), we specified main effects and full interactions of COLOR-VISUAL, DETERMINATION, and COLOR-AUDITIVE as fixed effects, and participant and item as random effects. In addition, DETERMINATION, COLOR-VISUAL and COLOR-AUDITIVE were defined as random slopes for participants. Residual-based outlier removal led to the removal of 0.17% of the data. There were statistically significant main effects of DETERMINATION ($t = -2.88$, $p < .01$ for *definite* compared to *demonstrative* conditions) and for COLOR-VISUAL ($t = 4.14$, $p < .001$), and interactions for DETERMINATION and COLOR-VISUAL (*zero* DETERMINATION vs. *definite* DETERMINATION: $t = -2.325$, $p < .05$); for DETERMINATION and COLOR-AUDITIVE (*demonstrative* vs. *definite* DETERMINATION: $t = -2.25$, $p < .05$;

[7] Removing these unexpected answers allows us to interpret the reaction times with respect to their reading, rather than adding the speculation about the participants' intended interpretation etc. to the discussion.

zero vs. *definite* DETERMINATION: $t = -3.06$, $p < 01$); of COLOR-VISUAL and COLOR-AUDITIVE ($t = -4.79$, $p < .001$). In addition, there was a three-way interaction of DETERMINATION, COLOR-VISUAL and COLOR-AUDITIVE (*demonstrative* vs. *definite* DETERMINATION: $t = 2.84$, $p < .01$; *zero* vs *definite* DETERMINATION: $t = 2.82$, $p < .01$). A full table of the statistical results for the fixed effects after outlier removal is given in Appendix 3 in Table 15. The interaction of COLOR-VISUAL and COLOR-AUDITIVE was pursued separately for each DETERMINATION condition, using a second model (Model 2.1). For Model 2.1, we specified the main effects and interactions of COLOR-VISUAL and COLOR-AUDITIVE as fixed effects, and participant and item as random intercepts, and main effects of COLOR-VISUAL and COLOR-AUDITIVE as random slopes for participant. For *demonstrative* DETERMINATION conditions, there were no statistically significant effects. For *zero* DETERMINATION conditions, there was a statistically significant effect of COLOR-VISUAL (t = statistical results for the fixed effects after outlier removal is 3.81, $p < .001$). For *definite* DETERMINATION conditions, there was a statistically significant main effect of COLOR-AUDITIVE ($t = 3.86$, $p < .001$) and an interaction of COLOR-VISUAL and COLOR-AUDITIVE ($t = 4.58$, $p < .001$). A full table of the statistical results for the fixed effects for Model 2.1 with all three DETERMINATION conditions is given in Appendix 3 in Table 16.

The interaction of COLOR-VISUAL and COLOR-AUDITIVE that was visible for definite articles was pursued using a third model (Model 3.1a and 3.1b). For Model 3.1a, we specified COLOR-VISUAL as fixed effect, participant and item as random intercepts, and COLOR-VISUAL as random slope for participant. For Model 3.1b, we specified COLOR-AUDITIVE as fixed effect, participant and item as random intercepts, and COLOR-AUDITIVE as random slope for participant. *Definite* DETERMINATION conditions were analyzed separately for conditions with COLOR-VISUAL *canonical* and *noncanonical* using Model 3.1a. This analysis revealed a statistically significant difference between *canonical* and *noncanonical* COLOR-VISUAL for conditions with COLOR-AUDITIVE *canonical* ($t = -5.48$, $p < .001$), but not for COLOR-AUDITIVE *noncanonical* ($t = .7$, $p > .5$). *Definite* DETERMINATION conditions were analyzed separately for conditions with COLOR-AUDITIVE *canonical* and *noncanonical* using Model 3.1b. This analysis revealed a statistically significant difference between *canonical* and *noncanonical* COLOR-AUDITIVE for conditions with COLOR-VISUAL *canonical* ($t = -2.31$, $p < .05$), and also for COLOR-VISUAL *noncanonical* ($t = 3.66$, $p > .01$). A full table of the statistical results for the fixed effects for Model 3.1a and 3.1b is given in Appendix 3 in Table 17.

Summary of reaction times (Experiment 1) For *demonstrative* DETERMINATION conditions, no influence of COLOR-VISUAL or COLOR-AUDITIVE was visible on reaction times. This could be due to a floor effect, since reaction times for demonstrative DETERMINATION conditions were faster than for the other DETERMINATION conditions. For *zero* DETERMINATION conditions (bare plural NPs), it took longer to judge a sentence if the color of the presented picture was *noncanonical* than if it was *canonical*. That is, when presented with a picture of a pink polar bear, participants took longer to accept the sentence *Polar bears are white* and to reject the sentence *Polar bears are pink* than they took for the same judgments when presented with a white polar bear. This might be due to the additional workload of having to decide

that the color of the picture should be ignored in this particular condition. The *definite* DETERMINATION conditions that we included in the reaction time measurements were the ones associated with the majority of answers, i.e., with the specific interpretation. However, according to the literature, such DPs could in principle be interpreted as either specific or generic. Sentences from these conditions were associated with longer reaction times than those for the other DETERMINATION conditions. Reaction times were longer with *noncanonical* than with *canonical* COLOR-VISUAL, e.g., judgments were slower when the visual stimulus depicted pink polar bears than when it depicted white polar bears. For both visual stimulus conditions, rejecting a sentence that was inconsistent with the visual stimulus (i.e., *The polar bears are pink* with a picture of white polar bears, and vice versa) took longer than accepting a sentence that was consistent (e.g., *The polar bears are pink* with a picture of pink polar bears, the same for white). We will offer an in-depth dicussion of the results of Experiment 1 together with the results of Experiment 2 in Sect. 6.

In general, the judgments elicited in Experiment 1 matched the expectations formulated above and in the literature. The high proportion of participants who seemed to interpret bare plurals as specific came as a surprise, since bare plurals typically express generic reference in the sentence type we used. We therefore performed a second experiment with identical stimuli, but omitted the conditions with definite articles. In doing so, we reduced the proportion of specific conditions (i.e., those that were mostly interpreted as specific according to Experiment 1) in the whole stimulus set as well as the level of ambiguity. This allows us to assess if participants' judgments and reaction times to bare plurals and demonstratives are influenced by the presence of a potentially ambiguous option.

5 Experiment 2

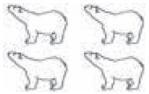
In the second experiment, we used two of the three auditive stimulus conditions: bare plural NPs (typically interpreted as generic) and demonstratives (typically interpreted as specific). We omitted the four conditions with definite articles. The second experiment was designed to control whether the two types are associated with generic and specific interpretations, respectively, also in the absence of the ambiguous option.

5.1 Materials and methods

5.1.1 Stimulus material

The stimulus material consisted of the auditive and visual stimuli outlined above. For the second experiment, only auditive conditions with *zero* and *demonstrative* DETERMINATION were used, omitting the *definite* DETERMINATION conditions (i.e., V, VI, XI and XII in Table 3). The combination of visual and auditive stimuli led to eight different stimulus conditions, illustrated in Table 7. Overall, 28 items were chosen as critical items. Two additional items in two conditions each were chosen as

Table 7 Stimulus conditions used in Experiment 2, with examples for visual and auditive stimuli

Condition number	Visual stimulus	COLOR-VISUAL	DETERMINATION	COLOR-AUDITIVE	Example auditive	Predicted acceptance
I		Canonical	Zero	Canonical	*Eisbären sind weiß.* 'Polar bears are white.'	Yes
II			Zero	Noncanonical	*Eisbären sind pink.* 'Polar bears are pink.'	No
III			Demonstrative	Canonical	*Diese Eisbären sind weiß.* 'These polar bears are white.'	Yes
IV			Demonstrative	Noncanonical	*Diese Eisbären sind pink.* 'These polar bears are pink.'	No
VII		Noncanonical	Zero	Canonical	*Eisbären sind weiß.* 'Polar bears are white.'	Yes
VIII			Zero	Noncanonical	*Eisbären sind pink.* 'Polar bears are pink.'	No
IX			Demonstrative	Canonical	*Diese Eisbären sind weiß.* 'These polar bears are white.'	No
X			Demonstrative	Noncanonical	*Diese Eisbären sind pink.* 'These polar bears are pink.'	Yes

practice trials. The limited number of items relative to the first experiment made it possible to present all stimuli to all participants. Each participant saw 224 items, with 28 items per condition (i.e., all items in all conditions).

5.1.2 Procedure

The procedure was identical to that in Experiment 1. The experiment lasted about 15–20 min.

5.1.3 Participants

33 participants were tested. All were recruited via the Sona Systems participant database of the University of Konstanz. Participants were aged 18–31 years (mean age 22.6 years, SD = 3.0 years). All had German as their only native language and normal or corrected-to-normal vision. Participants reported not being color-blind and not having a hearing impairment. Four participants were left-handed, three were male. All gave written and informed consent. They received six Euros compensation for their participation.

5.2 Results

Results are reported following the same pattern as outlined for Experiment 1, i.e., acceptance rates and reaction times. Data preparation and analysis were performed in parallel to Experiment 1.

5.2.1 Acceptance rates

Acceptance rates for all conditions were calculated for all participants. An overview of mean acceptance rates and standard deviations per condition is given in Table 8.

The picture from the acceptance rates is very clear. In conditions with canonical visual stimuli, participants accepted all sentences with canonical auditive stimuli

Table 8 Mean acceptance rates per condition over participants, with standard deviations, for Experiment 2. Data from all participants are included

Condition number	COLOR-VISUAL	DETERMINATION	COLOR-AUDITIVE	% Acceptance	S.D. acceptance
I	Canonical	Zero	Canonical	98.05	3.10
II		Zero	Noncanonical	4.22	15.48
III		Demonstrative	Canonical	98.59	2.35
IV		Demonstrative	Noncanonical	0.76	1.95
VII	Noncanonical	Zero	Canonical	75.97	31.92
VIII		Zero	Noncanonical	27.16	33.06
IX		Demonstrative	Canonical	2.71	4.73
X		Demonstrative	Noncanonical	97.19	3.19

(e.g., *These polar bears/Polar bears are white* on seeing a picture of white polar bears, as in Experiment 1). In conditions with noncanonical visual stimuli, the participants' answers indicate their interpretation of the sentences. Here, the following picture emerges:

Demonstratives As in Experiment 1, participants accepted demonstratives with noncanonical auditive stimuli, and rejected them with canonical auditive stimuli. In other words, when presented with the picture of pink polar bears, participants accepted *These polar bears are pink* and rejected *These polar bears are white*. This means that demonstratives are interpreted as specific; acceptance proportions match those with canonical visual stimuli and demonstratives.

Zero Participants accepted bare nouns with canonical auditive colors in 76% of the cases and with noncanonical auditive colors in about 27% of the cases. In other words, when presented with the picture of pink polar bears, participants accepted *Polar bears are white* 76% of the time and *Polar bears are pink* 27% of the time. The general tendency of answers matches our expectations that *zero* DETERMINATION conditions are interpreted as generic. However, given that *zero* DETERMINATION conditions are supposed to signal generic reference in the current context, it is surprising that participants accepted the sentence *Polar bears are pink* 27% of the time. While this proportion of unexpected answers is lower than in the identical conditions in Experiment 1, it is still unexpectedly high. Again, error rates are high for *zero* DETERMINATION conditions with noncanonical visual stimulus backgrounds, indicating that the responses to these conditions vary a lot between participants. Before we proceed, we take a closer look at the unexpected results for *zero* DETERMINATION conditions, as we did for Experiment 1.

A closer look at the data revealed a distinctly different behavior for four of the 33 participants. These participants responded differently from our expectations in 89% or more of the cases in the *zero* DETERMINATION conditions, interpreting bare plurals as specific when the visual stimulus depicted an item in a noncanonical color; in other words, when presented with a picture of pink polar bears, these participants accepted the sentence *Polar bears are pink* and rejected *Polar bears are white* more than 98% of the time. This response pattern (found in 4 participants) is consistent with a specific interpretation of bare plurals, and the deviation from our expectations is similar to the one found in Experiment 1. However, the proportion of unexpected responders was lower in Experiment 2 than in Experiment 1, possibly because there was no potentially distracting condition (i.e., that with definite articles).

To assess the influence of these unusual responders on general response patterns, we removed the data from these 4 participants (12% of the total of participants) from the dataset, and analyzed the reduced participant set. For the sake of readability, acceptance rates and statistical analyses for the reduced participant set are presented in Appendix 2. Importantly, omitting the participants with unusual response patterns did not change the general pattern of the results for demonstratives.

Post hoc comparisons for condition pairs We did not perform a statistical analysis of all acceptance rates, given that the results are very clear. We did however perform post hoc comparisons between some individual condition pairs in order to see whether small descriptive differences were statistically significant or

not, and whether they should be discussed. A full table of the results for the fixed effects of the models described below is given in Appendix 4 in Table 18.

Canonical-zero-canonical (I) vs noncanonical-zero-canonical (VII) The first post hoc test was performed between the conditions *canonical-zero-canonical* (98.05% acceptance) and *noncanonical-zero-canonical* (75.97% acceptance). The difference between these two conditions was analyzed with a binomial generalized linear mixed model (COLOR-VISUAL as fixed effect, participants and items as random intercepts, COLOR-VISUAL as random slope for participants and items). The analysis revealed a statistically significant difference between the conditions ($z = -4.54$, $p < .001$).

Canonical-zero-noncanonical (II) vs. noncanonical-zero-noncanonical (VIII) The second post hoc comparison for Experiment 2 was performed between the conditions *canonical-zero-noncanonical* (4.22% acceptance) and *noncanonical-zero-nocanonical* (27.16% acceptance). The difference between these two conditions was analyzed using the same model as for the first post hoc test. This analysis revealed a statistically significant difference between the conditions ($z = 5.41$, $p < .001$).

Summary results for acceptance rates (Experiment 2) Acceptance rates show that bare plurals (*zero* DETERMINATION conditions) are interpreted as generic and demonstratives as specific. This matches our expectations based on intuition and the literature and also the findings of Experiment 1. For zero determination conditions, the interpretation is slightly less robust for conditions with noncanonical visual stimuli than for canonical visual stimuli (see statistical analysis directly above). For conditions with *noncanonical* colors in the visual stimuli (i.e., sentences presented together with pictures of pink polar bears), the interpretation of bare plurals as generic is slightly less robust than expected, but more robust than in Experiment 1, with a smaller proportion of participants interpreting bare plurals as specific. This indicates that the interpretation of bare plurals is more susceptible to distracting factors (color of the visual stimulus, presence of a potentially ambiguous condition in the stimulus set) than the interpretation of demonstratives.

5.2.2 Reaction times

Reaction times per condition were analyzed for all responses indicating the expected interpretation, as outlined above (see Table 3). Reaction times shorter than 100 ms and longer than 6000 ms were removed from the dataset, leading to the removal of 7.8% of the data. Next, unexpected answers were removed from the dataset, leading to the removal of an additional 8.3% of the data. The mean reading times per condition are given in Table 9.

Statistical analysis of reaction times Log-transformed reaction times were analyzed using a linear mixed effects model, using the lmer function of the lme4 package in R. For the first model (Model 1.2, identical to Model 1.1. used in Experiment 1) we defined the main effects and full interactions of COLOR-VISUAL, DETERMINATION, and COLOR-AUDITIVE as fixed effects, and participant and item as random effects. In addition, DETERMINATION, COLOR-VISUAL and COLOR-AUDITIVE were

Table 9 Mean reaction times with standard deviations per condition over participants, Experiment 2 (after residual-based outlier removal)

Condition number	COLOR-VISUAL	DETERMINER	COLOR-AUDITIVE	Mean reaction times/ms	S.D. reaction times
I	Canonical	Zero	Canonical	342	146
II		Zero	Noncanonical	352	137
III		Demonstrative	Canonical	331	137
IV		Demonstrative	Noncanonical	341	135
VII	Noncanonical	Zero	Canonical	391	166
VIII		Zero	Noncanonical	429	201
IX		Demonstrative	Canonical	377	205
X		Demonstrative	Noncanonical	350	145

defined as random slopes for participants. Residual-based outlier removal led to the removal of 0.14% of the data. This analysis revealed a statistically significant main effect of DETERMINATION ($t = 4.15$, $p < .001$) and of COLOR-AUDITIVE ($t = 2.35$, $p < .05$), and statistically significant interactions of DETERMINATION and COLOR-VISUAL ($t = -2.22$, $p < .05$), DETERMINATION and COLOR-AUDITIVE ($t = -3.11$, $p < .01$), COLOR-VISUAL and COLOR-AUDITIVE ($t = -2.57$, $p < .05$) and a statistically significant three-way interaction of DETERMINATION, COLOR-VISUAL and COLOR-AUDITIVE ($t = 2.02$, $p < .05$). A full table of the statistical results for the fixed effects after outlier removal is given in Appendix 4 in Table 19.

The interaction of COLOR-VISUAL and COLOR-AUDITIVE was pursued separately for each determiner condition, using a second model (Model 2.2, identical to Model 2.1 used in Experiment 1). For *demonstrative* DETERMINATION conditions, there was a statistically significant main effect of COLOR-AUDITIVE ($t = 2.32$, $p < .05$), and a statistically significant interaction of COLOR-VISUAL and COLOR-AUDITIVE ($t = -2.55$, $p < .05$). For *zero* DETERMINATION conditions, there were statistically significant main effects of COLOR-VISUAL ($t = -3.27$, $p < .01$) and of COLOR-AUDITIVE ($t = -2.14$, $p < .05$). A full table of the statistical results for the fixed effects is given in Appendix 4 in Table 20.

The interaction of COLOR-VISUAL and COLOR-AUDITIVE that was visible for *demonstrative* DETERMINATION conditions was pursued using a third model (Model 3.2a and 3.2b, identical to Models 3.1a and 3.1b in Experiment 1). *Demonstrative* DETERMINATION conditions were analyzed separately for conditions with COLOR-AUDITIVE *canonical* and *noncanonical* using Model 3.2a. This analysis revealed a statistically significant difference between *canonical* and *noncanonical* COLOR-VISUAL for conditions with COLOR-AUDITIVE *noncanonical* ($t = 2.10$, $p < .05$), but not for COLOR-AUDITIVE *canonical* ($t = -1.25$, $p > .2$). *Demonstrative* DETERMINATION conditions were analyzed separately for conditions with COLOR-VISUAL *canonical* and *noncanonical* using Model 3.1b. This analysis revealed a statistically significant difference between *canonical* and *noncanonical* COLOR-AUDITIVE for conditions with COLOR-VISUAL *noncanonical* ($t = 2.10$, $p < .05$), but not for COLOR-AUDITIVE

canonical ($t = -1.25$, $p > .2$). A full table of the statistical results for the fixed effects for Model 3.2a and 3.2b is given in Appendix 4 in Table 21.

Summary of reaction times (Experiment 2) In general, reaction times were longer for conditions with *noncanonical* COLOR-VISUAL than with *canonical* COLOR-VISUAL, and longer for *zero* DETERMINATION conditions than for *demonstrative* DETERMINATION conditions. For conditions with *noncanonical* COLOR-VISUAL with *zero* DETERMINATION, it took longer to reject sentences with *noncanonical* COLOR-AUDITIVE (*Polar bears are pink*) than to accept sentences with *canonical* COLOR-AUDITIVE (*Polar bears are white*). For conditions with *noncanonical* COLOR-VISUAL and *demonstrative* DETERMINATION, it took longer to reject sentences with *canonical* COLOR-AUDITIVE (*These polar bears are white*) than to accept sentences with *noncanonical* COLOR-AUDITIVE (*These polar bears are pink*). Taken together, this shows that, at least in our task, accepting true statements is faster than rejecting wrong statements, and that this judgment is in turn faster for demonstratives than for bare plurals. At this point, we cannot say if this is because generic statements with bare plurals are more difficult to process per se, or if the current task adds the extra workload of ignoring the visual stimulus to these conditions.

6 Discussion and conclusions

6.1 Interpretation of plural nouns: definites, demonstratives and bare plurals

As expected, both experiments showed that demonstrative plural NPs are interpreted as specific, and bare plural NPs as generic. Furthermore, our first experiment showed that—at least in the current paradigm—definite plural NPs are interpreted as specific, not as generic. This fits the interpretation of definite plural NPs in other Germanic languages, and is in contrast to claims in the literature, according to which definite plural articles can have a generic reading (e.g., Brugger 1993; Longobardi 1994; Krifka et al. 1995; Chierchia 1998; Dayal 2004; Oosterhof 2008). However, while the pattern of acceptance rates is very similar for the demonstrative and the definite article conditions in Experiment 1, the specific interpretation of definite articles seems to be less consistent than that of demonstratives (and less consistent than the interpretation of definites by English-speaking adults, see Gelman and Raman 2003). This could be due to the fact that definites have come to be optionally interpreted as generics, if the context allows for this interpretation. Another possibility would be that the specific interpretation of definites is obligatory, but less robust than for demonstratives, and that the difference between acceptance rates for demonstratives and definites is due to errors, rather than 'true' interpretation preferences. We cannot distinguish between the two explanations at this point, although we will point out below that the reaction time data lends support to the first explanation.[8]

[8] One single participant in Experiment 1 seemed to consistently interpret definite plural articles as generic. It is possible that this person had a Romance-type grammar either from the outset (of L1 acquisition) or because she adopted it later in life by means of contact with L2 speakers of German with a Romance language as L1. Since this is an individual speaker this must remain a matter of speculation.

There is some evidence from other experiments which used linguistic contexts where speakers were more inclined towards a generic reading of definites (e.g., Pérez-Leroux et al. 2004). In turn, this suggests that the preference of generic over definites found in these earlier experiments is also dependent on the type of context and experimental task, and does not necessarily reflect a default status for generic expressions. The issue of whether unambiguously specific or generic expressions can be assigned a default status thus remains unclear; for now, it seems that measurable preferences depend heavily on the type of task and context employed.

Bare plurals were usually interpreted as generic in both experiments. However, their generic interpretation seemed to be less consistent than the specific interpretation of demonstratives. In addition, a subgroup of the participants in both experiments showed interpretation patterns that indicate a specific, rather than generic interpretation of bare plurals. While this was only the case for a subgroup of participants, and these participants did not change the interpretation of the definite determiners, this finding for bare plurals was unexpected. Different explanations are possible: First, it could be the case that the interpretation of bare plurals as generics is less strict (more ambiguous) than that of demonstrative determiners as specific. This would indicate that the unexpected answers are not mistakes, but that, for some reason, some participants favored the usually dispreferred interpretation. This is unlikely, because there is no indication in the literature on Germanic that bare plurals can be interpreted as specific. The second possibility is that bare plurals unambiguously signal generic reference in our stimuli, but that the judgment task is more difficult than for demonstrative determiners because the salient visual stimulus providing the background has to be ignored. It is possible that the participants who exhibited this unusual interpretation pattern were simply more susceptible to interference from the visual stimulus than the other participants. A third possibility is that a certain proportion of the responders developed response strategies that were far removed from their ordinary interpretation of different determiner types, and opted to check for a match between the colors in the visual and the auditive stimulus, without paying too close attention to the determiner type. Our current setup does not allow us to conclusively decide between the latter two options. However, the fact that the rate of unusual responders dropped from 28% in Experiment 1 to 12% in Experiment 2 seems to suggest that this pattern of responses has to do with the stimulus material. It could be that in Experiment 1, either the presence of an ambiguous option (definite articles) or the higher proportion of potentially specific utterances (2/3 of the material in Experiment 1, but only 1/2 of the material in Experiment 2) contributed to the high rate of unexpected responders.

6.2 Reaction times

Reaction times for demonstratives were faster than for bare plurals in both experiments. At this point, we cannot say whether this difference is due to an effect of our experimental paradigm (e.g., reflecting the additional workload of ignoring the salient visual stimulus in the generic bare plural conditions, or else reflecting the support for demonstrative readings since the visual stimuli provide a relevant context) or whether it suggests that specific reference is in general processed more

easily than generic reference. Data from experiments like ours should be compared with data from other experiments on the processing of genericity, using a different kind of context, before a clearer picture emerges.[9] For child language acquisition, there is evidence that children grasp part/whole relations before unique reference (Schafer and Villiers 2000; Avrutin 1999) and that they overuse expletive determiners (de Villiers and Roeper 1995; Baauw 2000), i.e., determiners that do not function to establish specific reference. Based on that, Matthewson et al. (2001) proposed an acquisition sketch according to which there would be a progression from N to NP to DP, with each of the forms containing further distinctions. Parallel to many other arguments in acquisition, they advanced the hypothesis that the child looks progressively for distinctions that expand the syntactic/semantic tree toward more 'specificity', with familiarity/uniqueness marking the 'most specific' option. The child moves down this list from N to NP to DP (which is like "moving up the tree"), with N referring to kinds and generic entities, NP to specific entities and DP to familiar/unique entities. Their proposal could be taken to imply, or at least that is our interpretation, that generic reference is more basic than specific reference.

In the first experiment, reaction times for definite articles are longer than for demonstratives, and slightly longer than for bare plurals. This is interesting, given that the reaction times for definite articles that we analyzed exclusively belonged to instances that were interpreted as specific. This suggests that it is not specificity per se that is associated with faster processing in our task (as could be concluded from reaction time data of the second experiment in isolation). As noted above, we can only offer a tentative interpretation of our reaction times here. With this caveat in mind, we interpret the longer reaction times for definites as indicating that their interpretation as specific is preferred but optional, and that comprehenders must make a choice that is not necessary in the other conditions. Future studies using different linguistic and nonlinguistic contexts will allow us to learn more about the role of context in the interpretation of potentially ambiguous constructions, and the influence that context has on the processing load associated with different interpretations.

What do the data tell us with regard to potentially ongoing grammaticalization processes? In other words, is the German definite in the process of losing its original deictic meaning component and becoming more like Romance definites which can be used without specific reference? Our data on preferred interpretations would suggest that German definites are still far from the Romance stage, although the longer reaction times with definite articles as compared to demonstratives might be interpreted in terms of an incipient process towards such a stage (suggesting that definites are indeed ambiguous, and that a choice is necessary for their interpretation that is unnecessary for demonstratives and bare plurals).

[9] Studies on the processing of negation suggest that truth value judgments for negative sentences are associated with higher processing cost than for positive sentences, but that this difference disappears with a context leading participants to expect negative sentences (see Nordemeyer and Frank 2014 and references therein).

6.3 Methods

In contrast to earlier studies (Barton 2016; Pérez-Leroux et al. 2004), we avoided a linguistic context with explicit reference to generic or specific interpretations because we suspected that such a context would bias the interpretation in the direction of generic reference.[10] Instead, we opted for a nonlinguistic context in the form of easily recognizable pictures. This context made sentences with specific reference plausible (demonstrative conditions would not have made any sense without a context), while not making generic reference implausible. However, it could be argued that generic interpretation made it necessary to ignore the visual stimulus, while no such work was necessary for the specific interpretations. The analysis of the error rates shows that conditions with bare plural NPs are more error-prone than other conditions when presented together with a noncanonical visual stimulus, but not in combination with canonical visual stimuli. We therefore assume that this is not due to enhanced difficulty of generic interpretation per se, but rather represents a task effect in the current paradigm. As explained above, some participants seemed to be more susceptible to this interference of visual stimuli with generic interpretation than others. Since our main goal was to monitor the interpretation of definites, rather than quantify interference of visual and linguistic stimuli, we opted to remove data from these susceptible participants for the separate analyses. Importantly, the removed participants did not exhibit unusual patterns of definite interpretations, so we are confident that we did not alter the interpretation of the critical condition too much by this exclusion. For the design of future experiments, it will be important to keep in mind this influence of visual stimuli on linguistic interpretation, and also that differences between participants are to be expected.

In contrast to Barton et al. (2015), regional variation in determiner use was not a focus of the current study, and we did not explicitly assess the dialect and sociolinguistic background of our participants. However, our participant group was generally homogenous: All were aged between 18 and 35 years and students at the University of Konstanz. Therefore, we can safely assume they speak standard German in addition to any dialect they may speak, and have a comparable level of formal education. This participant homogeneity is crucial for our ambition to measure reaction times, and for the development of a paradigm that allows us to assess processing difficulties associated with different interpretations.

6.4 Outlook

Future studies could address the following points. First, we should aim to establish a paradigm in which generic interpretations are favored and compare the outcomes to our current paradigm. Second, it would be desirable to assess whether the unexpected specific interpretation of bare plurals was caused by the task, or whether

[10] Gelman and Raman (2003) also used pictures instead of a linguistic context in an English acquisition study. However, the focus of their study was on the different factors used in acquisition of genericity, and their linguistic expressions did not include potentially ambiguous options, since English definites are usually considered to unambiguously encode specific reference.

the interpretation of bare plurals is more flexible than expected. Third, it would be interesting to use the current paradigm to monitor specific and generic reference in languages other than German. Our data support the widely held idea that German definite plurals are preferentially interpreted as specific in the absence of a linguistic context that would support a generic reading. In Romance languages, however, definite plurals are more readily interpreted as generic. Would this interpretation also be susceptible to interference from nonmatching visual stimuli, just like the interpretation of German (ambiguous) definites and (unambigous) bare plurals? Finally, the comparison between Experiment 1 and Experiment 2 shows that the presence or absence of an ambiguous option (German definite plurals) does not alter the general pattern of acceptance rates, but that a subset of comprehenders is more susceptible than others to an influence of the ambiguous option. The presence of an ambiguous option did not alter the influence of determiners or visually and auditorily presented colors on reaction times (although there was a general slowdown in the presence of an ambiguous option, the tendencies for the different conditions were the same). These findings suggest that a comparison between a language like German (with a truly ambiguous/underspecified option) and a language with explicit specific and generic reference can be meaningful, provided that the experimental setup is right.

Acknowledgements The authors would like to thank Luana d'Agosto and Miriam Geiss for stimulus preparation and recording, Jana Neitsch for providing her voice for the auditory stimuli, Anja Arnhold for Praat support, Oleksiy Bobrov for Presentation programming, and Luana d'Agosto, Miriam Geiss, Sarah Zander, Christina Gozebina, and Marc Meisezahl for data acquisition, and Leo Vrana for proofreading.

Appendix 1: Stimulus material

Results of the pretest for visual stimulus preparation

See Table 10.

Table 10 Results of the pretest for stimulus preparation. The table depicts English translations of the names, the German names used in the experiment, what percent of the participants suggested the intended name, prototypical colors suggested by the participants, and whether the item was included in the stimulus material

Name			Colors given by participants	Included in stimuli
English	German	%	English	
seagull	Möwe	33	–	No
plum	Pflaume	56	Purple, blue, yellow	No
raspberry	Himbeere	72	Purple, red	No
dalmatian	Dalmatiner	78	Black and white, spotted, mottled	Yes
tangerine	Mandarine	78	Orange	Yes
raven	Rabe	78	Black	Yes
eggplant	Aubergine	89	Purple, violet, 1x black	Yes
polar bear	Eisbär	89	White	Yes
flamingo	Flamingo	89	Pink, fuchsia	Yes
swan	Schwan	89	White	Yes
carrot	Karotte	94	Orange	Yes
cherry	Kirsche	94	Red	Yes
ladybird	Marienkäfer	94	Red and black, dotted, red with black/white dots	Yes
banana	Banane	100	Yellow	Yes
elephant	Elefant	100	Grey	Yes
strawberry	Erdbeere	100	Red	Yes
frog	Frosch	100	Green	Yes
fox	Fuchs	100	Brownish red (∼ auburn), red, orange, red and white, orange-red, reddish	Yes
cucumber	Gurke	100	Green	Yes
koala	Koala	100	Grey, grey-brown	Yes
crocodile	Krokodil	100	Green, green–brown	Yes

Table 10 continued

Name			Colors given by participants	Included in stimuli
English	German	%	English	
chick	Küken	100	Yellow	Yes
lion	Löwe	100	Yellow, orange, brown, light brown, sand, beige	Yes
rhinoceros	Nashorn	100	Grey	Yes
hippopotamus	Nilpferd	100	Grey	Yes
panda	Panda	100	Black and white	Yes
penguin	Pinguin	100	Black and white, black-white-yellow	Yes
pig	Schwein	100	Pink, fuchsia, skin-colored	Yes
sunflower	Sonnenblume	100	Yellow, yellow and green	Yes
tiger	Tiger	100	Yellow, black, striped	Yes
tomato	Tomate	100	Red	Yes
zebra	Zebra	100	Black and white stripes	Yes
lemon	Zitrone	100	Yellow	Yes

List of items in the visual stimulus set with canonical and noncanonical colors

See Table 11.

Table 11 Full list of referents used in the visual and auditory stimuli. The German names of referents and the colors in the *canonical* and *noncanonical* visual and auditive stimulus conditions are given, together with their English translations

Name		Canonical color		Noncanonical color	
German	English	German	English	German	English
Aubergine	Eggplant	lila	purple	rot	red
Banane	Banana	gelb	yellow	blau	blue
Dalmatiner	Dalmatian	gepunktet	spotted	orange	orange
Eisbär	Polar bear	weiß	white	pink	fuchsia
Elefant	Elephant	grau	grey	orange	orange
Erdbeere	Strawberry	rot	red	grau	grey
Flamingo	Flamingo	rosa	pink	gelb	yellow
Frosch	Frog	grün	green	pink	fuchsia
Fuchs	Fox	rotbraun	red-brown	gelbgrün	yellow-green
Gurke	Cucumber	grün	green	blau	blue
Karotte	Carrot	orange	orange	blau	blue
Kirsche	Cherry	rot	red	lila	purple
Koala	Koala	grau	grey	orange	orange
Krokodil	Crocodile	grün	green	weiß	white
Küken	Chick	gelb	yellow	grün	green
Löwe	Lion	gelb	yellow	rot	red
Mandarine	Tangerine	orange	orange	schwarz	black
Marienkäfer	Ladybird	gepunktet	spotted	gestreift	striped
Nashorn	Rhinoceros	grau	grey	pink	fuchsia
Nilpferd	Hippopotamus	grau	grey	gelb	yellow
Panda	Panda bear	schwarzweiß	black and white	rotbraun	red and brown
Pinguin	Penguin	schwarzweiß	black and white	grüngelb	green and yellow
Rabe	Raven	schwarz	black	grün	green
Schwan	Swan	weiß	white	lila	purple
Schwein	Pig	rosa	rose/pink	blau	blue
Sonnenblume	Sunflower	gelb	yellow	blau	blue
Tiger	Tiger	gestreift	striped	kariert	checkered
Tomate	Tomato	rot	red	braun	brown
Zebra	Zebra	gestreift	striped	kariert	checkered
Zitrone	Lemon	gelb	yellow	lila	purple

Appendix 2: Acceptance rates for the reduced participant sets

Acceptance rates for the reduced participant set, Experiment 1

To assess the influence of the unusual responders on general response patterns, we removed the data from these 14 participants (28% of the total of participants) from the dataset.

An overview of mean acceptance rates and standard deviations per condition for the reduced participant set is given in Table 12.

The general picture from the acceptance rates is clear, and the overall pattern remains similar to the results for the complete participant set. With canonical visual stimuli, participants accepted all sentences that had canonical auditive stimuli. This pattern would be compatible with both generic and specific interpretations of the determiners. For conditions with noncanonical visual stimuli, the participants' answers indicate their interpretation of the determiners:

Demonstrative **DETERMINATION conditions** Participants in the reduced set accepted these with demonstrative determiners and noncanonical auditive stimuli, and rejected them with demonstrative determiners and canonical auditive stimuli. This pattern of acceptances is compatible with a specific interpretation of demonstrative determiners, as expected; acceptance rates are very close to those for the complete participant dataset.

Zero **DETERMINATION conditions** Participants in the reduced set accepted *zero* determiners with canonical auditive colors 84% of the time, and with noncanonical auditive colors 17% of the time. This indicates that they preferred a generic interpretation of bare plurals. Still, the acceptance rate of *noncanonical-zero-canonical* sentences is closer to that of *noncanonical-definite-noncanonical* sentences than that for *noncanonical-demonstrative-noncanonical* (the latter being the unambiguously specific comparison condition). This could indicate that the

Table 12 Mean acceptance rates per condition over participants, with standard deviations, for Experiment 1 (reduced participant set)

Condition number	COLOR-VISUAL	DETERMINATION	COLOR-AUDITIVE	% Acceptance	S.D. acceptance
I	Canonical	Zero	Canonical	98.81	3.19
II		Zero	Noncanonical	1.59	3.01
III		Demonstrative	Canonical	98.41	3.01
IV		Demonstrative	Noncanonical	0.99	2.51
V		Definite	Canonical	98.41	3.46
VI		Definite	Noncanonical	0.60	2.00
VII	Noncanonical	Zero	Canonical	83.73	14.55
VIII		Zero	Noncanonical	17.46	17.37
IX		Demonstrative	Canonical	3.77	4.67
X		Demonstrative	Noncanonical	97.42	4.88
XI		Definite	Canonical	16.87	17.59
XII		Definite	Noncanonical	85.12	14.92

interpretation of zero determiners is less robust than that of demonstrative determiners. Participants tended to interpret zero determiners as generic, but this preference was less reliable than the preference to interpret demonstrative determiners as specific. Another explanation is that the task is more difficult for zero determiners than for the other determiner types, given that the nonlinguistic visual context (the picture of pink polar bears) has to be ignored while judging the generic statement. Thus, higher error rates in the judgment task may not reflect the processing of genericity per se, but rather a task effect. When considering what distinguished the removed participants from the participants remaining in the reduced set, it is possible that the removed participants may simply have been more susceptible to interference from the visual stimulus than the remaining participants.

Definite **DETERMINATION conditions** Participants generally accepted these with definite determiners and noncanonical auditive stimuli, and rejected them with definite determiners and canonical auditive stimuli. This pattern of acceptances is compatible with a specific interpretation of *demonstrative* determiners. Importantly, acceptance rates for *definite* **DETERMINATION** conditions in the reduced participant set are very close to those in the full participant set. (This indicates that the interpretation of *definite* DETERMINATION conditions was robust across participants, and was not influenced by the participants' interpretation of the *zero* DETERMINATION conditions.)

In the following, we present the statistical analysis for the reduced participant set. The results of the analysis for the full participant set are also repeated here for easier comparison. A detailed description of the results tables is given in Appendix 3.

Canonical-zero-canonical (I) vs. noncanonical-zero-canonical (VII) The first post hoc test was performed between the conditions *canonical-zero-canonical* (full participant set: 98.43% acceptance; reduced participant set: 98.81% acceptance) and *noncanonical-zero-canonical* (full participant set: 63.57% acceptance; reduced participant set: 83.73% acceptance). The difference between these two conditions was analyzed with a binomial generalized linear mixed model (COLOR-VISUAL as fixed effect, participants and items as random intercepts, COLOR-VISUAL as random slope for participants and items). The analysis revealed a statistically significant difference between the conditions (full participant set: $z = -6.50$, $p < .001$; reduced participant set: $z = -4.20$, $p < .001$).

Canonical-zero-noncanonical (II) vs. noncanonical-zero-noncanonical (VIII) The second post hoc comparison for Experiment 1 was performed between the conditions *canonical-zero-noncanonical* (full participant set: 1.14% acceptance; reduced participant set: 1.59% acceptance) and *noncanonical-zero-nocanonical* (full participant set: 35.43% acceptance; reduced participant set: 17.46% acceptance). The difference between these two conditions was analyzed using the same model as for the first post hoc test. This analysis revealed a statistically significant difference between the conditions (full participant set: $z = 6.50$, $p < .001$; reduced participant set: $z = 5.72$, $p < .001$).

Noncanonical-demonstrative-canonical (IX) vs. noncanonical-definite-canonical (XI) The third post hoc-test was performed to compare the conditions with *noncanonical* visual stimuli and auditive conditions *definite-canonical* (full participant set: 15.71% acceptance, reduced participant set: 16.87% acceptance)

or *demonstrative-canonical* (full participant set: 4.14% acceptance, reduced participant set: 3.77% acceptance). The difference between these two conditions was analyzed with a binomial generalized linear mixed model (DETERMINATION as fixed effect, participants and items as random intercepts, DETERMINATION as random slope for participants and items). The difference between these two conditions was statistically significant (full participant set: $z = 3.5$, $p < .001$; reduced participant set: $z = 3.88$, $p < .001$).

Noncanonical-demonstrative-noncanonical (X) vs. noncanonical-definite-noncanonical (XII). The fourth post hoc test was performed to compare the conditions with *noncanonical* visual stimuli and auditive conditions *definite-noncanonical* (full participant set: 85.57% acceptance, reduced participant set: 85.12% acceptance) and *demonstrative-noncanonical* (full participant set: 96.57% acceptance; reduced participant set: 97.42% acceptance), using the same model as for the third post hoc test. This analysis revealed a statistically significant difference between conditions *noncanonical-demonstrative-noncanonical* and *noncanonical-definite-noncanonical* (full participant set: $z = -3.60$, $p < .001$; reduced participant set: $z = -3.97$, $p < .001$).

Acceptance rates for the reduced participant set, Experiment 2

An overview of mean acceptance rates and standard deviations per condition for the reduced participant set are summarized in Table 13.

In summary, acceptance rates for the reduced participant set are similar to those for the complete participant set. With canonical visual stimuli, participants accepted all sentences with canonical auditive stimuli, as in Experiment 1.

For conditions with noncanonical visual stimuli, the participants' answers indicate their interpretation of the determiners. Here, the following picture emerges:

Demonstratives Participants in the reduced participant set accepted these with demonstrative determiners and noncanonical auditive stimuli, and rejected them with demonstrative determiners and canonical auditive stimuli. This pattern of acceptances is compatible with a specific interpretation of demonstratives, as

Table 13 Mean acceptance rates per condition over participants, with standard deviations, for Experiment 2 (data from the reduced participant set)

Condition number	COLOR-VISUAL	DETERMINATION	COLOR-AUDITIVE	% Acceptance	S.D. acceptance
I	Canonical	Zero	Canonical	98.03	3.11
II		Zero	Noncanonical	1.60	2.63
III		Demonstrative	Canonical	98.65	2.42
IV		Demonstrative	Noncanonical	0.62	1.67
VII	Noncanonical	Zero	Canonical	82.88	22.25
VIII		Zero	Noncanonical	17.49	21.11
IX		Demonstrative	Canonical	2.59	4.47
X		Demonstrative	Noncanonical	97.04	3.32

expected; acceptance rates are very close to those for the complete participant dataset.

Zero The participants accepted bare nouns with canonical auditive colors in 83% of these cases, and with noncanonical auditive colors in 18% of the cases. Thus, the 29 participants in the reduced participant set preferred a generic interpretation of bare nouns. While the remaining participants generally interpreted bare nouns as generic, they sometimes opted for the other interpretation. We discuss possible reasons for this variability in the general discussion (Sect. 6).

Post hoc comparisons for condition pairs We did not perform a statistical analysis of all acceptance rates, given that the results are very clear, for both the complete and the reduced participant sets. We did however perform post hoc comparisons between some individual condition pairs in order to see whether small descriptive differences were statistically significant or not, and whether they should be discussed. A full table of the results for the fixed effects of the models described below is given in Appendix 4 in Table 18.

Canonical-zero-canonical (I) vs noncanonical-zero-canonical (VII) The first post hoc test was performed between the conditions *canonical-zero-canonical* (full participant set: 98.05% acceptance; reduced participant set: 98.03% acceptance) and *noncanonical-zero-canonical* (full participant set: 75.97% acceptance; reduced participant set: 82.88% acceptance). The difference between these two conditions was analyzed with a binomial generalized linear mixed model (COLOR-VISUAL as fixed effect, participants and items as random intercepts, COLOR-VISUAL as random slope for participants and items). The analysis revealed a statistically significant difference between the conditions (full participant set: $z = -4.54$, $p < .001$; reduced participant set: $z = -4.61$, $p < .001$).

Canonical-zero-noncanonical (II) vs. noncanonical-zero-noncanonical (VIII) The second post hoc comparison for Experiment 2 was performed between the conditions *canonical-zero-noncanonical* (full participant set: 4.22% acceptance; reduced participant set: 1.6% acceptance) and *noncanonical-zero-noncanonical* (full participant set: 27.16% acceptance; reduced participant set: 17.49% acceptance). The difference between these two conditions was analyzed using the same model as for the first post hoc test. This analysis revealed a statistically significant differences between the conditions (full participant set: $z = 5.41$, $p < .001$; reduced participant set: $z = 6.08$, $p < .001$).

Appendix 3: Results of the statistical evaluations for Experiment 1

Analysis of acceptance rates

See Table 14.

Table 14 Summary of the fixed effects in the mixed effects models for acceptance rates, Experiment 1, post hoc comparisons

	Estimate	Std. error	*z*-value	Pr(> |*z*|)
Post hoc 1, *Canonical-zero-canonical* (I) vs. *noncanonical-zero-canonical* (VII). full participant set				
(Intercept)	− 0.93	0.38	− 2.49	0.0128*
Color-visual	− 3.57	0.55	− 6.50	7.92e−11***
Post hoc 1, reduced participant set				
(Intercept)	− 1.74	0.19	− 9.33	< 2e−16***
Color-visual	− 4.08	0.97	− 4.20	2.7e−05***
Post hoc 2, *Canonical-zero-noncanonical* (II) vs. *noncanonical-zero-noncanonical* (VIII). full participant set				
(Intercept)	0.93	0.38	2.49	0.013*
Color-visual	3.57	0.55	6.50	7.92e−11***
Post hoc 2, reduced participant set				
(Intercept)	1.78	0.23	7.72	1.18e−14***
Color-visual	2.37	0.41	5.72	1.06e−08***
Post hoc 3, *noncanonical-definite-canonical* (II) vs. *noncanonical-demonstrative-canonical* (VIII). full participant set				
(Intercept)	2.20	0.33	6.71	2e−11***
Determination	1.69	0.48	3.50	0.00047***
Post hoc 3, *noncanonical-definite-canonical* (II) vs. *noncanonical-demonstrative-canonical* (VIII). reduced participant set				
(Intercept)	2.03	0.33	6.25	4.1e−10***
Determination	2.07	0.53	3.88	.000105***
Post hoc 4, *noncanonical-definite-noncanonical* (II) vs. *noncanonical-demonstrative-noncanonical* (VIII). full participant set				
(Intercept)	− 2.12	0.27	− 7.92	2.36e−15***
Determination	− 1.27	0.35	− 3.60	.000316***
Post hoc 4, *noncanonical-definite-noncanonical* (II) vs. *noncanonical-demonstrative-noncanonical* (VIII). reduced participant set				
(Intercept)	− 2.02	0.27	− 7.44	9.88e−14***
Determination	− 1.70	0.43	− 3.97	7.08e−05***

Analysis of reaction times, Experiment 1

See Tables 15, 16 and 17.

Table 15 Summary of the fixed effects for Model 1.1, depicting reaction times for Experiment 1, all conditions

	Estimate	Std. error	df	*t* value	Pr(> |*t*|)
(Intercept)	5.96	0.06	63	98.76	< 2e−16***
DETERMINATION *dem.*	− 0.11	0.04	550	− 2.88	0.00417**
DETERMINATION *zero*	0.07	0.04	536	1.59	0.11230
COLOR-VISUAL	− 0.02	0.04	1189	− 0.52	0.60320
COLOR-AUDITIVE	0.15	0.04	4596	4.14	3.59e−05***
DETERMINATION *dem.*: COLOR-VISUAL	− 0.01	0.05	6990	− 0.23	0.81636
DETERMINATION *zero*:COLOR-VISUAL	− 0.12	0.05	6753	− 2.33	0.02011*
DETERMINATION *dem.*:COLOR-AUDITIVE	− 0.11	0.05	6993	− 2.25	0.02438*
DETERMINATION *zero*:COLOR-AUDITIVE	− 0.17	0.06	7001	− 3.06	0.00221**
COLOR-VISUAL:COLOR-AUDITIVE	− 0.24	0.05	6989	− 4.79	1.68e−06***
DETERMINATION *dem.*:COLOR-VISUAL:COLOR-AUDITIVE	0.20	0.07	6978	2.84	0.00449**
DETERMINATION *zero*:COLOR-VISUAL:COLOR-AUDITIVE	0.21	0.07	6993	2.82	0.00475**

Table 16 Summary of the fixed effects for Model 2.1 in Experiment 1. Reaction times are analyzed separately for each DETERMINATION condition

	Estimate	Std. error	df	*t* value	Pr(> |*t*|)
demonstrative DETERMINATION					
(Intercept)	5.86	0.05	61.3	129.81	< 2e−16***
COLOR-VISUAL	− 0.03	0.03	1864.7	− 0.91	0.365
COLOR-AUDITIVE	0.04	0.03	976.1	1.19	0.235
COLOR-VISUAL:COLOR-AUDITIVE	− 0.04	0.05	2488.2	− 0.93	0.350
zero DETERMINATION					
(Intercept)	6.04	0.06	56.5	100.15	< 2e−16***
COLOR-VISUAL	− 0.15	0.04	1579.4	− 3.81	0.000142***
COLOR-AUDITIVE	− 0.01	0.04	1546.2	− 0.27	0.79
COLOR-VISUAL:COLOR-AUDITIVE	− 0.04	0.06	2102.5	− 0.67	0.51
definite DETERMINATION					
(Intercept)	6.00	0.07	55.2	89.80	< 2e−16***
COLOR-VISUAL	− 0.03	0.04	121.6	− 0.59	0.55
COLOR-AUDITIVE	0.15	0.04	526.3	3.86	0.00013***
COLOR-VISUAL:COLOR-AUDITIVE	− 0.24	0.05	2308.9	− 4.58	4.92e−06***

Table 17 Summary of the fixed effects for Models 3.1a and 3.1b in Experiment 1. Reaction times are analyzed for *definite* DETERMINATION conditions, separated by COLOR-AUDITIVE or by COLOR-VISUAL

	Estimate	Std. error	df	*t* value	Pr(> |*t*|)
Color-auditive canonical					
(Intercept)	6.12	0.07	47.6	88.33	< 2e−16***
COLOR-VISUAL	− 0.26	0.05	46.6	− 5.48	1.67e−06***
Color-auditive noncanonical					
(Intercept)	5.97	0.06	49.9	98.60	< 2e−16***
COLOR-VISUAL	− 0.03	0.04	46.7	− 0.67	0.507
Color-visual canonical					
(Intercept)	5.95	0.06	48.6	98.63	< 2e−16***
COLOR-AUDITIVE	− 0.90	0.04	49.9	− 2.31	0.0249*
Color-visual noncanonical					
(Intercept)	5.97	0.06	51.4	97.79	< 2e−16***
COLOR-AUDITIVE	0.15	0.04	324.5	3.66	0.000299***

Appendix 4: Results of the statistical evaluations for Experiment 2

Analysis of acceptance rates, Experiment 2

See Table 18.

Table 18 Summary of the fixed effects for post hoc comparisons for acceptance rates, Experiment 2

	Estimate	Std. error	*z*-value	Pr(> |*z*|)
Post hoc 1, *Canonical-zero-canonical* (I) vs. *noncanonical-zero-canonical* (VII). full participant set				
(Intercept)	− 1.65	0.46	− 3.56	0.00037***
Color-visual	− 2.69	0.59	− 4.54	5.66e−06***
Post hoc 1, reduced participant set				
(Intercept)	− 2.00	0.3200	− 6.24	4.36e−10***
Color-visual	− 2.37	0.52	− 4.61	4.11e−06***
Post hoc 2, *Canonical-zero-noncanonical* (II) vs. *noncanonical-zero-noncanonical* (VIII). full participant set				
(Intercept)	1.35	0.43	3.10	0.00191**
Color-visual	3.13	0.58	5.41	6.22e−08***
Post hoc 2, reduced participant set				
(Intercept)	1.91	0.29	6.66	2.83e−11***
Color-visual	2.33	0.38	6.08	1.19e−09***

Analysis of reaction times, Experiment 2

See Tables 19, 20 and 21.

Table 19 Summary of the fixed effects for Model 1.2, reaction times for Experiment 2, all conditions

	Estimate	Std. error	df	*t* value	Pr(> |*t*|)
(Intercept)	5.62	0.07	35	85.19	< 2e−16***
DETERMINATION	0.13	0.03	239	4.15	4.65e−05***
COLOR-VISUAL	− 0.02	0.03	363	− 0.59	0.55595
COLOR-AUDITIVE	0.06	0.03	318	2.35	0.01942*
DETERMINATION:COLOR-VISUAL	− 0.09	0.04	5640	− 2.22	0.02626*
DETERMINATION:COLOR-AUDITIVE	− 0.12	0.04	5947	− 3.11	0.00185**
COLOR-VISUAL:COLOR-AUDITIVE	− 0.09	0.04	6130	− 2.57	0.01031*
DETERMINATION:COLOR-VISUAL:COLOR-AUDITIVE	0.11	0.05	6129	2.02	0.04348*

Table 20 Summary of the fixed effects for Model 2.2 in Experiment 2. Reaction times are analyzed separately for each DETERMINATION condition

	Estimate	Std. error	df	*t* value	Pr(> |*t*|)
demonstrative DETERMINATION					
(Intercept)	5.62	0.07	34	83.68	< 2e−16***
COLOR-VISUAL	− 0.02	0.03	589	− 0.62	0.5379
COLOR-AUDITIVE	0.06	0.03	117	2.32	0.0221*
COLOR-VISUAL:COLOR-AUDITIVE	− 0.09	0.04	3222	− 2.55	0.0108*
zero DETERMINATION					
(Intercept)	5.75	0.07	34.1	81.81	< 2e−16***
COLOR-VISUAL	− 0.11	0.03	58.2	− 3.27	0.00182**
COLOR-AUDITIVE	− 0.06	0.03	124.1	− 2.14	0.03431*
COLOR-VISUAL:COLOR-AUDITIVE	0.02	0.04	2678.7	0.39	0.69540

Table 21 Summary of the fixed effects for Models 3.2a and 3.2b in Experiment 2. Reaction times are analyzed for *demonstrative* DETERMINATION conditions, separated by COLOR-AUDITIVE or by COLOR-VISUAL

	Estimate	Std. error	df	*t* value	Pr(> |*t*|)
Color-auditive canonical					
(Intercept)	5.68	0.07	33.5	87.60	< 2e−16***
COLOR-VISUAL	− 0.11	0.03	1335.3	− 4.22	2.61e−05***
Color-auditive noncanonical					
(Intercept)	5.62	0.07	32.7	82.48	< 2e−16***
COLOR-VISUAL	− 0.02	0.03	741.4	− 0.60	0.551
Color-visual canonical					
(Intercept)	5.60	0.07	32.3	85.79	< 2e−16***
COLOR-AUDITIVE	− 0.03	0.03	28.6	− 1.25	0.222
Color-visual noncanonical					
(Intercept)	5.62	0.07	32.3	82.77	< 2e−16***
COLOR-AUDITIVE	0.06	0.03	31.0	2.10	0.0445*

References

Avrutin, Sergey. 1999. *Development of the syntax-discourse interface*. Dordrecht: Kluwer.

Baauw, Sergio. 2000. *Grammatical features and the acquisition of reference. A comparative study of Dutch and Spanish*. Utrecht: LOT.

Barton, Dagmar. 2016. *Generische Nominalphrasen bei deutsch-französischer Zweisprachigkeit*. Unpublished PhD dissertation, University of Hamburg.

Barton, Dagmar, Nadine Kolb, and Tanja Kupisch. 2015. Definite article use with generic reference in German: an empirical study. *Zeitschrift für Sprachwissenschaft* 34 (2): 147–173.

Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67 (1): 1–48.

Bellmann, Günter. 1990. *Pronomen und Korrektur*. Tübingen: Niemeyer.

Boersma, Paul. 2001. Praat, a system for doing phonetics by computer. *Glot International* 5 (9/10): 341–345.

Boersma, Paul, and David Weenink.2016. Praat: doing phonetics by computer [Computer program]. Version 6.0.19, retrieved 13 June 2016 from http://www.praat.org/.

Brugger, G. 1993. Generic interpretations and expletive determiners. *University of Venice Working Papers in Linguistics* 3: 1–30.

Carlson, Gregory. 1977. *Reference to kinds in English*. PhD dissertation, University of Massachusetts, Amherst.

Casalegno, Paolo. 1987. Sulla logica dei plurali. *Teoria* 2: 125–143.

Chierchia, Gennaro. 1998. Reference to kinds across languages. *Natural Language Semantics* 6 (4): 339–405.

Dayal, Veneeta. 2004. Number marking and (in)definiteness in kind terms. *Linguistics and Philosophy* 27 (4): 393–450.

de Villiers, J., and T. Roeper. 1995. Barriers, binding and the acquisition of the NP-DP distinction. *Language Acquisition* 4: 73–104.

Diesing, Molly. 1992. *Indefinites (Linguistic Inquiry Monographs 20)*. Cambridge, MA: MIT Press.

Fox, John, and Sanford Weisberg. 2011. *An R companion to applied regression*. Thousand Oaks: Sage.

Gelman, Susan A., and Lakshmi Raman. 2003. Preschool children use linguistic form class and pragmatic cues to interpret generics. *Child Development* 74 (1): 308–325.

Greenberg, Joseph H. 1978. How does a language acquire gender markers? In *Universals of human language, vol 3: Word structure*, ed. Joseph H. Greenberg, Charles A. Ferguson, and Edith A. Moravcsik, 47–82. Stanford: Stanford University Press.

Krifka, Manfred, Francis J. Pelletier, Gregory N. Carlson, Alice ter Meulen, Gennaro Chierchia, and Godehard Link. 1995. Genericity: An introduction. In *The generic book*, ed. Gregory N. Carlson and Francis J. Pelletier, 1–124. Chicago, IL: University of Chicago Press.

Kupisch, Tanja, and Dagmar Barton. 2013. How bilinguals deal with variation: Generic reference in adult German bilinguals. *Studia Linguistica* 67: 1–27. https://doi.org/10.1111/stul.12006.

Longobardi, Giuseppe. 1994. Reference and proper names: A Theory of N-movement in syntax and logical form. *Linguistic Inquiry* 25 (4): 609–665.

Longobardi, Giuseppe. 2001. The structure of DP: Some principles, parameters, and problems. In *The handbook of contemporary syntactic theory*, ed. Marc Baltin and Chris Collins, 562–603. Oxford: Blackwell.

Lyons, Christopher. 1999. *Definiteness*. Cambridge: Cambridge University Press.

Matthewson, Lisa, Tim Bryant, and Tom Roeper. 2001. A Salish stage in the acquisition of English determiners: Unfamiliar 'definites'. *The Proceedings of SULA, University of Massachusetts Occasional Papers in Linguistics* 25: 63–71.

Nordemeyer, Ann, and Mike Frank. 2014. A pragmatic account of the processing of negative sentences. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 36 (36). UC Merced. https://escholarship.org/uc/item/95j3b18w.

Oosterhof, Albert. 2004. Generic noun phrases in Dutch. In *Proceedings of the 20th Scandinavian Conference of Linguistics*, ed. Fred Karlsson, 1–22. Helsinki: University of Helsinki. http://www.ling.helsinki.fi/kielitiede/20scl/Oosterhof.pdf.

Oosterhof, Albert. 2008. *The semantics of generics in Dutch and related languages*. Amsterdam: Benjamins.

Pérez-Leroux, Ana T., Alan Munn, Cristina Schmitt, and Michelle DeIrish. 2004. Learning definite determiners: Genericity and definiteness in English and Spanish. In *28th Boston University Conference on Language Development Proceedings—BUCLD 28 Online Proceedings Supplement*, ed. Alejna Brugos, Linnea Micciulla and Christine E. Smith. http://www.bu.edu/bucld/files/2011/05/28-perez-leroux.pdf. Accessed 29 Apr 2015.

R Development Core Team. 2017. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Schafer, R. and J. de Villiers. 2000. Imagining articles: What *a* and *the* can tell us about the emergence of DP. In *Proceedings of the 24th annual Boston University Conference on Language Development*, eds. S. Catherine Howell, Sarah A. Fish, and Thea Keith-Lucas, 609–620. Somerville, MA: Cascadilla Press.

Tremblay, Antoine, and Johannes Ransijn. 2015. *LmerConvenienceFunctions: Model selection and* post hoc *analysis for (g)lmer models.* https://CRAN.Rproject.org/package=LMERConvenienceFunctions. R (package version 2.10).

Vergnaud, Jean-Roger, and Maria Luisa Zubizarreta. 1992. The definite determiner and the inalienable constructions in French and in English. *Linguistic Inquiry* 23: 97–141.

Wickham, Hadley. 2007. Reshaping data with the reshape package. *Journal of Statistical Software* 21 (12): 1–20.

Wickham, Hadley. 2011. The split-apply-combine strategy for data analysis. *Journal of Statistical Software* 40 (1): 1–29.

Zamparelli, Roberto. 2002. Definite and bare kind-denoting noun phrases. In *Romance languages and linguistic theory 2000: Selected papers from 'Going Romance' 2000*, ed. Claire Beyssade, Reineke Bok-Bennema, Frank Drijkoningen, and Paola Monachesi, 305–343. Amsterdam: John Benjamins.