

Overspecification of small cardinalities in reference production

Natalia Zevakhina and Elena Pasalskaya*

Abstract. This paper presents experimental evidence for overspecification of small cardinalities in reference production. The idea is that when presented with a small set of unique objects (2, 3 or 4), the speaker includes a small cardinality while describing given objects, although it is overinformative for the hearer (e.g., “three stars”). On the contrary, when presented with a large set of unique objects, the speaker does not include cardinality in their description – so she produces a bare plural (e.g. “stars”). The effect of overspecifying small cardinalities resembles the effect of overspecifying color in reference production which has been extensively studied in recent years (cf. Rubio-Fernandez 2016, Tarenskeen et al. 2015). When slides are flashed on the screen one by one, highlighted objects are still overspecified. We argue that one of the main reasons lies in subitizing effect, which is a human capacity to instantaneously grasp small cardinalities.

Keywords. overspecification; redundancy; informativeness; semantics; pragmatics; psycholinguistics; reference production; numerals; color adjectives

1. Introduction. It is well-known that sometimes speakers tend to convey more information than required. In doing so, they use redundant linguistic expressions. Imagine a situation where there is only one cup available in the visual context shared by the speaker and the addressee. In such a situation, the speaker utters (1), where she specifies the color of an object, even though the addressee can identify the object without paying attention to its color. In other words, the speaker could produce (2) where she only names the object itself without specifying its color. Obviously, sentence (1) conveys more information than sentence (2). Why did the speaker produce the overinformative sentence (1) instead of uttering the minimally informative sentence (2)?

- (1) Give me the blue cup, please.
 (2) Give me the cup, please.

There have been proposed two accounts for this puzzle: speaker-oriented and hearer-oriented. According to the speaker-oriented account, producing overinformative utterances requires less cognitive effort from the speaker’s part than producing minimally informative utterances (Pechmann 1989). To illustrate, in a situation of describing an object that possesses several attributes, the speaker need not think of which attributes are relevant to communicate. Rather, she produces those attributes which have come to her mind and she does not waste time and efforts to make certain computations with respect to the relevance. According to the hearer-oriented account,

* We would like to thank Polina Makarova, Timofey Morunov, Veronika Prigorkina, Alina Schipkova, Elina Sigdel for their help in searching for participants of both experiments and presenting them the materials. We acknowledge the discussion of hypotheses with Alex Dainiak and Nina Kazanina. We appreciate the remarks and suggestions from the audiences of *The 8th Experimental Pragmatics Conference (XPrag) 2019* (University of Edinburgh) and the audience of *The 1st Conference “Experiments in Linguistic Meaning” (ELM) 2020* (University of Pennsylvania), online due to Covid-19. Authors: Natalia Zevakhina, HSE University, Russian Federation (nzevakhina@hse.ru) & Elena Pasalskaya, HSE University, Russian Federation (lena@ales.ru). This article is an output of a research project implemented as part of the Basic Research Program at the National Research University Higher School of Economics (HSE University).

overspecification helps hearers identify relevant objects more rapidly, accurately and, therefore, more efficiently (Mangold & Pobel 1988, Rubio-Fernandez 2016).

Some attributes enable overspecification, whereas some others do not. Recent research on attribute redundancy in reference production has demonstrated that color is much more likely to be used redundantly than size, material, shape, pattern, location and orientation (Belke and Meyer 2002, Arts et al. 2011, Brown-Schmidt and Konopka 2011, Mangold and Pobel 1988, Gatt et al. 2007, Gatt et al. 2013).

The contrast between color and other attributes is accounted for in terms of absoluteness and salience (Belke and Meyer 2002, Koolen et al. 2011, Tarenskeen et al. 2015 among others). The absoluteness of color means that color is not contingent upon the speaker, the addressee, a situation, etc. For instance, if the observer sees a red object, she does not have to take into consideration colors of surrounding objects, she can merely report ‘red’. On the contrary, size is relative since the degree of size of a given object is evaluated by the observer among size degrees of surrounding objects. In case a given object does not differ from surrounding objects in terms of size, size is not likely to be reported.

As for the salience, color is salient because of its high visual perceptibility. It is one of the features of preattentive analysis (Trick 1992) and it is computed early in visual processing (Livingstone and Hubel 1988). Other attributes (such as shape, material, pattern or size) are not so salient and, therefore, are less likely to be reported.

Moreover, the question which factor (absoluteness or salience) has a greater influence on the overspecification, seems to have a negative response: no factor. Take size. It is relative. Size is reported when it becomes contrastive and relevant for the communication (see van Gompel et al. 2019). In other words, it is specified when it is supposed to be salient. For instance, the visual context includes only one big plate, with the rest of the plates being small. However, it is unlikely to be reported when the visual context includes three big and three small objects of one type or three big and three small objects of various types: a plate, a cup, a spoon, etc. Moreover, the degree of the proportion between the objects does not seem to play a role here, as well. To illustrate, Tarenskeen et al. (2015) manipulated size contrast between the objects with a proportion 3:1 but did not receive a significant increase in overspecification of size. Now take pattern. It is absolute and salient. In this respect, it resembles color. However, Gatt et al. (2013), Tarensleen et al. (2015) showed that it is not as salient as color. To summarize, it seems that both factors – absoluteness and salience – determine overspecification in reference communication.

Focusing on color, it is noteworthy to mention several factors that affect color overspecification. Firstly, color is more likely to be overspecified in polychrome contexts than in monochrome contexts (Belke and Meyer 2002, Koolen et al. 2013, Rubio-Fernandez 2016). To illustrate, the speaker is more likely to say *Give me the blue cup, please* in a situation when she is presented with objects of different colors than in a situation when she is presented with objects with the same color, say blue. Secondly, color tends to be overspecified for atypically-colored objects in comparison to variably-colored or stereotypically-colored objects (Westerbeek et al. 2015, Rubio-Fernandez 2016). To illustrate, the probability that the speaker redundantly utters (3) with an atypically-colored wolf is higher than the probability that the speaker redundantly utters (4) or (5) with a variably-colored car or a stereotypically-colored banana respectively.

- (3) Show me a purple wolf, please.
- (4) Show me a red car, please.
- (5) Show me a yellow banana, please.

Thirdly, color is more often overspecified when it is variable than stereotypical for a given category of objects (Sedivy 2003, Rubio-Fernandez 2016). For example, the speaker would produce (4) to a higher degree than (5). Fourthly, color is more likely to be overspecified when referring to objects for which color is more important: e.g., artifacts like clothes or cars are more color-pertinent than geometric objects (Rubio-Fernandez 2016).

Going back to the distinction between color and size, Brown-Schmidt and Konopka (2011) argued for that not only color but also number (or cardinality) is distinct from size. Color adjectives and numerals were reported more frequently and faster than size adjectives in reference communication. They were reported both in contrastive and non-contrastive contexts. On the contrary, size adjectives were reported significantly more often in contrastive contexts. Brown-Schmidt and Konopka (2011) suggested that the reason for this is that, unlike color and number, size is a context-dependent modifier. These findings accord with the following idea circulated in the literature: size is relative (context-dependent), whereas color is absolute (not context-dependent); size is non-salient, whereas color is salient. As for the number, according to the findings by Brown-Schmidt and Konopka (2011), it seems to be absolute and salient, like color. However, this fact was not directly addressed in the literature. Furthermore, it was not clear why number seems to be absolute and salient. Also, it was not clear which numbers were tested in Brown-Schmidt and Konopka (2011). Judging by Figure (2a-b) in Brown-Schmidt and Konopka (2011: 308), the numbers till 5 were involved. On the whole, it seems that what is needed now is a more systematic study of number overspecification in reference production. This is exactly what we do in the present study.

1.1. SUBITIZING. For more than a century (since Bourdon 1908), there has been acknowledged a fast, accurate and confident apprehension of cardinalities of small sets (1-4 or even 1-8) which considerably decreases in cardinalities of large sets. To illustrate, when presented with 3 dots in a display, an observer undoubtedly and very rapidly determines the cardinality of the dots. This capability vanishes when the observer is presented with 15 dots in a display.

The phenomenon of immediately grasping the cardinality of few elements in a given set was coined as *subitizing* in Kaufman et al. (1949), with a reference to Dr. Cornelia C. Coulter (Kaufman et al. 1949: 520) who suggested this term roughly meaning ‘sudden apprehension’, cf. also a similar term *numerousness* in Stevens (1938), Thomas et al. (1999) and Taves (1941). In Kaufman et al. (1949), subitizing was contrasted to *estimation*, that is, an approximate and less accurate apprehension of cardinalities of large sets, cf. also a close term *numerosity* in Stevens (1938), Thomas et al. (1999) and Taves (1941).

The question of what is the threshold for small sets is still debated but there has been a tacit agreement in the literature that the numbers from 1 to 3-4 belong to the subitizing range. Whether the numbers from 5 to 8 belong to the subitizing range is still questionable. The threshold seems to vary from person to person. It is also dependent on a particular experiment setting and some other factors (Akin and Chase 1978, Atkinson, Campbell and Francis 1976, Chi and Klahr 1975, Jensen, Reese and Reese 1950, Mandler and Shebo 1982, Oyama, Kikuchi and Ichihara 1981).

Importantly, subitizing is not the same as counting small cardinalities. Rather, it involves a separate cognitive mechanism (Revkin et al. 2008). Counting is effortful, error-prone and slow (Trick 1992). It usually takes 250-350 ms per item. In contrast, subitizing is effortless, accurate and rapid. It usually takes 40-100 ms per item. Subitizing has been argued to be a preattentive

mechanism which allows an observer to grasp the cardinality of items without carefully counting them (Trick 1992).

2. Experiment 1.

2.1. HYPOTHESES. Hypothesis 1 was that overspecification of small cardinalities would not differ from overspecification of color. Hypothesis 2 was that overspecification of small cardinalities and overspecification of color would be significantly different from overspecification of large cardinalities.

2.2. PARTICIPANTS. 90 people took part in the experiment (56 females, age range = 17-32, mean age = 21).

2.3. MATERIALS. The experiment had a between-subjects design. In order to verify Hypotheses 1 and 2, we created three conditions: Color, Small Cardinality and Large Cardinality conditions, see Figure 1. In all the three conditions, we used 2x2 pictures/cells presented in one slide, each of which contained various geometric objects (squares, rectangles, crosses, circles, triangles, diamonds, stars, and ovals). The geometric objects were identical within each cell but were different among all cells. In all the conditions, one cell was a target and was highlighted, whereas the other three were distractors. A critical cell took different positions through the experiment: it could be any of the 2 x 2 cells.

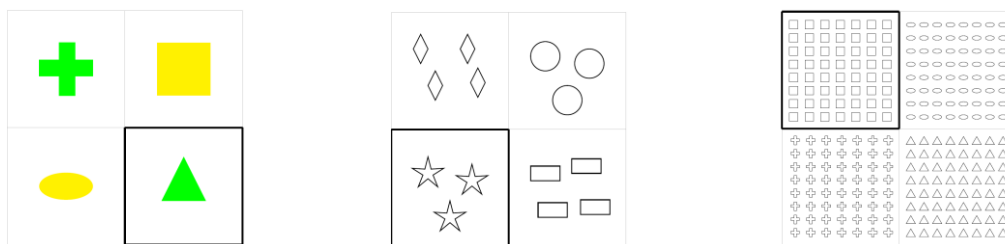


Figure 1: Examples of critical items used in Color, Small Cardinality and Large Cardinality conditions (Experiment 1)

In Color condition, two (out of four) cells comprised objects of one color, whereas two other cells comprised objects of another color. There were three colors: red, green and yellow. In Small Cardinality condition, two (out of four) cells included objects of one small cardinality, while two other cells included objects of another small cardinality. There were three small cardinalities: two, three and four. In Large Cardinality condition, two (out of four) cells had objects of one large cardinality, whilst two other cells had objects of another large cardinality. There were four large cardinalities: 7 x 8 (56), 8 x 7 (56), 8 x 8 (64). Each condition contained 48 critical items, that is, 8 geometric objects x 3 colors x 2 versions in Color condition, 8 geometric objects x 3 small cardinalities x 2 versions in Small Cardinality condition, 8 geometric objects x 3 large cardinalities x 2 versions in Large Cardinality condition.

Depending on a condition, we expected that the participants would use the following ways of referring to objects. In Color condition, there might be either a singular noun (e.g., “a square”) or a color adjective and a noun (e.g., “a red square”). In Small Cardinality condition, the options are either a bare plural noun (e.g., “squares”) or a numeral and a plural noun (e.g., “two squares”). In Large Cardinality condition, there might be either a bare plural noun (e.g., “squares”), a quantifier and a noun (e.g., “many squares”) or a numeral and a noun (e.g., “56 squares”). Importantly, a

singular noun and a bare plural noun are minimal specifications of referring to given highlighted cells. A color adjective and a noun as well as a numeral and a noun are overspecifications. As for combinations of a quantifier and a noun, they do not seem to be minimal specifications. However, they do not seem to be overspecifications either, since adding information of a large amount of some set does not tell anything about a cardinality of such a set, especially if other sets presented in a slide can also be referred to with help of “many” expression. Therefore, it seemed reasonable to treat data with “many” (if they would occur) as minimal specifications.

Filler items were images of human faces, tangrams, and artifacts (crockery, furniture, transport and clothes), cf. Figure 2. In parallel to critical items, each filler slide contained four images of two sorts: e.g., two artifacts and two tangrams, two human faces and two artifacts, two tangrams and two human faces. There were 72 filler items that were identical in all the three conditions.



Figure 2: Examples of filler items (Experiment 1)

The idea behind using human faces as fillers was that their description involves more participants’ concentration because they contain many features important to distinguish one face from others (see Koolen et al. 2013): with vs. without beard, with vs. without glasses, hair style, mood of a person, a dress, etc. Tangrams seem to be even more difficult than human faces in reference production. On the contrary, artifacts are easier to be referred to, since their identification is effortless. All the fillers were intentionally left uncolored, that is black-and-white. The idea behind that was that participants were not supposed to concentrate on color of hair, dress, etc.

2.4. PROCEDURE. There were two versions of each condition with a random order of critical and filler items. Importantly, critical and filler items were counterbalanced, so that each pair of critical items was separated with at least one filler. Therefore, each condition formed two experimental lists. Each list had 120 items (48 critical items + 72 fillers). Each list was presented for 30 participants (30 participants for a condition x 3 conditions = 90 participants). The experiment was conducted in the Russian language. Before the experiment, participants were told that they had to describe a highlighted picture to a person who had the same set of pictures but in a different order and who did not know which cell was highlighted. Participants’ task was to describe a highlighted picture to a person so that she understood which cell is referred to. Importantly, after pressing the space key or the right arrow on the keyboard, participants moved from one slide to another one. In the instructions, they were asked to make sure that their interlocutor also changed their slide. Only when the interlocutor confirmed this, a participant could start describing the picture. This was done intentionally to provide participants with some time to carefully examine a given slide. There were a few practice trials (identical to fillers) before the experiment. Because of Covid-19, the experiment was conducted online, via Zoom. Participants gave permission to be audio-recorded. The participants were told that no correct answers are expected. They were instructed not to think too long but not to hurry up.

2.5. RESULTS. 4320 responses (48 critical items x 30 participants x 3 conditions) were received (1440 responses for each condition). However, some of them were excluded due to participants' metaphorical naming objects, mostly in Color condition (e.g., “yellow oval” was described as an antispasmodic pill). Out of 1440 responses in Color condition, 49 responses (3.4%) were excluded. Out of 1440 responses in Small Cardinality condition, 4 responses (0.28%) were excluded. Out of 1440 responses in Large Cardinality condition, 22 responses (1.5%) were excluded. Therefore, 1391 responses for Color condition, 1436 responses for Small Cardinality condition, 1418 responses for Large Cardinality condition were used.

In Large Cardinality condition, 308 responses (out of 1440, 21%) that specified a large amount of objects without a numeral (e.g., “many ovals”) were treated on a par with bare plurals (e.g., “ovals”). In Small Cardinality condition, 2 responses (out of 1440, 0.14%) that specified a small amount of objects without a numeral (e.g., “some ovals”) were treated on a part with bare plurals (e.g., “ovals”).

The results of Experiment 1 are visualized in Figure 3.

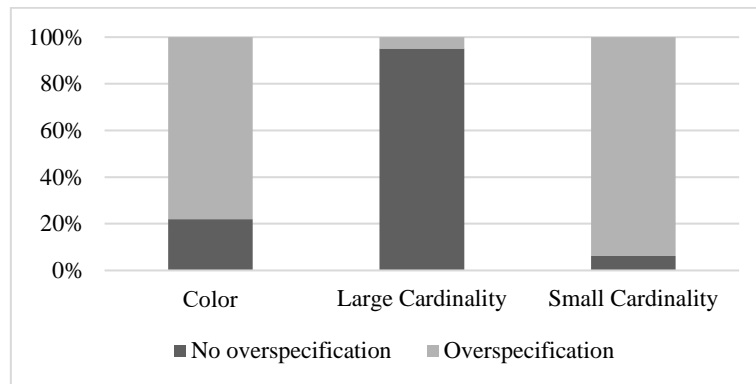


Figure 3: Distribution of (over)specification in Color, Small Cardinality and Large Cardinality conditions (Experiment 1)

A Kruskal-Wallis rank sum test indicated a main effect of Modifier (Color vs. Small Cardinality vs. Large Cardinality): $H(2) = 2642.2, p < 0.0001$. Multiple pairwise comparisons using Wilcoxon rank sum test showed further differences between conditions: Color vs. Large Cardinality ($p < .0001$), Color vs. Small Cardinality ($p < .0001$), Small Cardinality vs. Large Cardinality ($p < .0001$).

2.5. DISCUSSION. The results of Experiment 1 confirmed Hypothesis 2 but disconfirmed Hypothesis 1. Both small cardinalities (2, 3, 4) and color adjectives were overspecified in reference production to a greater extent than large cardinalities. In this respect, small cardinalities and color resemble each other. A plausible reason for this resemblance might be their absoluteness and salience. However, they were overspecified in a different way, with the proportions for small cardinalities higher than the proportions for color adjectives. This is an unexpected finding in the research field of overspecification in reference production that has mostly concentrated on color and tentatively concluded that it is the most overspecified attribute.

If small cardinalities are absolute and salient, why are they so? A possible reason is that they undergo the subitizing effect. This is what we tested in Experiment 2, where the slides were presented in a flashed mode. Additionally, this might be relevant for metaphors that occurred in the participants' responses. Metaphors have been argued to be time-consuming (cf. Noveck et al.

2001 among others) in comprehension studies. Metaphorical interpretations require longer reaction times than literal interpretations. What about speeded reference production? Would they occur if participants had timings to refer to a flashed picture?

There is one more question. It has been argued that participants demonstrate consistency in reference strategies (cf. Tarenskeen et al. 2015). For example, if they start using a color adjective, they do so almost through the whole experiment. Would they do so when they were presented with a flashed picture and when they had timings to refer to it?

All these questions are addressed in Experiment 2.

3. Experiment 2.

3.1. HYPOTHESES. Hypothesis 3 was that overspecification of small cardinalities presented in a flashed mode would be similar to overspecification of small cardinalities in Experiment 1. Hypothesis 4 was that there would be no metaphorical expressions produced while referring to objects. Hypothesis 5 was that proportions of (over)specification would be consistent through the whole experiment.

3.2. PARTICIPANTS. 31 people took part in the experiment (23 females, age range = 20-30, mean age = 24).

3.3. MATERIALS AND PROCEDURE. Both critical and filler items were identical to the items used in the Small Cardinality condition of Experiment 1. However, the procedure was different. The slides for Small Cardinality condition of Experiment 1 were presented in a flashed mode on the screen, cf. Figure 4. Firstly, participants saw a blank slide with a fixation dot for 500 ms. It was followed by a slide with four cells. Each cell contained a unique set of geometric objects. The cardinalities of two cells were identical and the cardinalities of two other cells were also identical. No cell was highlighted. Such a slide appeared on the screen for 5 000 ms. During this time interval, participants could carefully examine a given slide. After that, participants were presented with the same slide, however, importantly, one of the cells was highlighted. The presentation of such a slide was for a quite short time, only for 200 ms. The reason for this short time interval was that according to Trick (1992) subitizing usually takes 40-100 ms per item, that is on average 70 ms per item. Therefore, the interval would be enough to subitize the cardinality in the highlighted cell. A slide presented for 200 ms was followed by a blank slide appeared on the screen for 5 000 ms. During this slide, participants had to describe the highlighted cell of the previous slide.

Participants were instructed to get ready while being presented with the first slide, to carefully examine four pictures in the second slide, to catch sight of which cell was highlighted in the third slide, and to describe a highlighted cell while being presented with the fourth slide. Participants had to describe highlighted cells to a person who had the same set of pictures but in a different order and who did not know which cell was highlighted. That is, instead of a video presentation, a person had a mere presentation (as in Experiment 1). There were a few practice trials (identical to fillers) before the experiment. Because of Covid-19, the experiment was held online, via Zoom. Participants gave permission to be audio-recorded. The participants were told that there were no correct answers. They were instructed not to think too long but not to hurry up. The video presentation lasted 22 minutes.

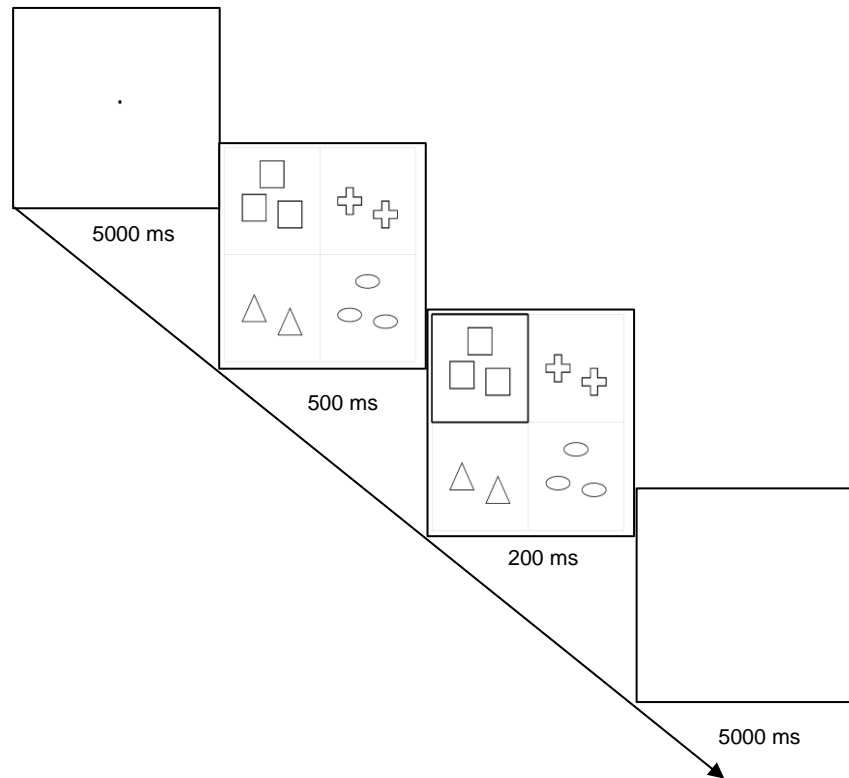


Figure 4: Example of the presentation sequence of a critical item (Experiment 2)

3.4. RESULTS. 1488 responses (48 critical items x 31 participants) were received. However, 90 responses (6.05%) of them were excluded because of problems similar to those occurred in Experiment 1: participants' self-corrections in counting objects (e.g., "three... four stars"), self-corrections in specifying cardinalities (e.g., "squares... three squares"), errors in counting objects (e.g., "five ovals" when a cell with four ovals was highlighted). Interestingly, there were no metaphorical naming of objects. Moreover, there were some technical problems (unstable Internet connection) while presenting the materials to the participants via Zoom. This fact disallowed to record audio responses to some of the critical items. The retained 1398 responses were used.

The results of Experiments 1 and 2 are visualized in Figure 5.

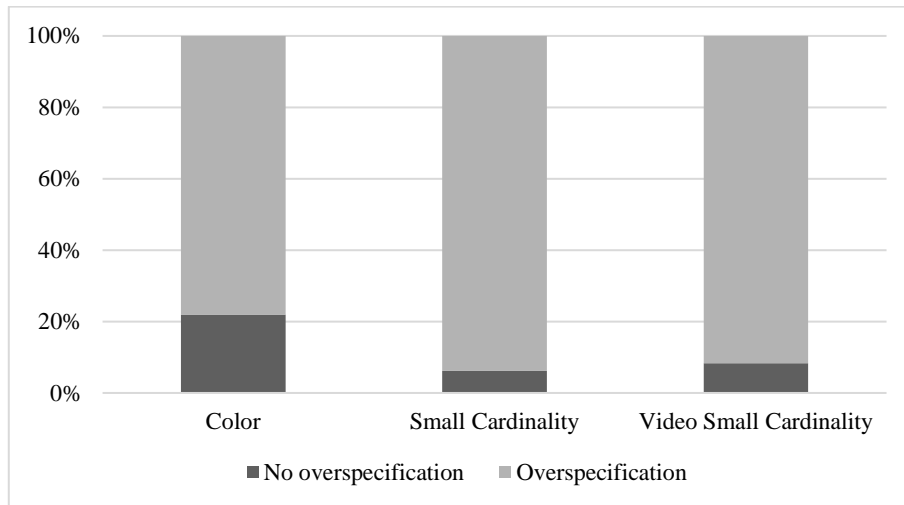


Figure 5: Distribution of (over)specification in Color, Small Cardinality and Video Small Cardinality conditions (Experiments 1 and 2)

A Kruskal-Wallis rank sum test showed a main effect of Condition (Color vs. Small Cardinality vs. Video Small Cardinality) in Experiments 1 and 2: $H(2) = 193.17, p < 0.0001$. Multiple pairwise comparisons using Wilcoxon rank sum test revealed further differences between conditions: Color vs. Video Small Cardinality ($p < .0001$), Small Cardinality vs. Video Small Cardinality ($p < .026$), Color vs. Small Cardinality ($p < .0001$).

Additionally, we tested whether participants were tired during the experiment and whether it affected the results. We calculated proportions of (over)specification in the first part (first 24 critical items) and the second part (next 24 critical items) of the experiment. The results are visualized in Figure 6.

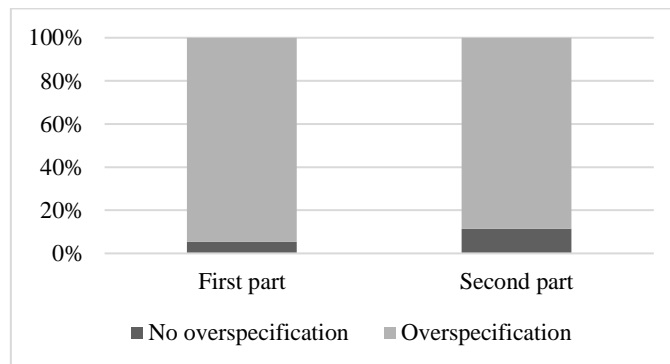


Figure 6: Distribution of (over)specification in Video Small Cardinality condition between the first and second part of the experiment (Experiment 2)

A Wilcoxon rank sum test with continuity correction showed a higher proportion of overspecification in the first part of Experiment 2: $W = 258880, p < 0.0001$.

3.5. DISCUSSION. Due to the significant difference between Small Cardinality vs. Video Small Cardinality conditions, Hypothesis 3 was not confirmed. However, the proportions of overspecification in both cases are visually quite similar and distinct from Color condition (cf. Figure 5). A plausible reason for this might lie in the mode of presentation. Firstly, the experiment con-

tained too many slides (120) and, secondly, the timing of 200 ms for the presentation of a slide with the highlighted was relatively short, even though the previous slide was shown for 5 000 ms.

Be that as it may, the proportion of overspecification in Experiment 2 is still relatively low. This suggests that subitizing plays a role in overspecifying small cardinalities. Subitizing makes small cardinalities salient, and, in this respect, they resemble color that has been argued to be also salient (Brown-Schmidt and Konopka 2011, Tarenskeen et al. 2015 among others).

Moreover, in both experiments, numerals were produced in exact meanings ‘exactly n ’ (cf. the discussion of which meanings are primary for numerals: at-least meanings ‘at least n and possibly $n+1$ ’ vs. exact meanings ‘exactly n ’ – in Papafragou and Musolino 2003, Musolino 2004, Breheny 2008 as well as more recent studies). A possible explanation for this is again subitizing. It seems natural to assume that subitizing yields the exact meanings of the numerals. This finding has the following consequence related to color. Like color, small cardinalities is absolute. Hypothesis 4 was confirmed. There was no metaphorical naming of objects in Experiment 2. It accords with the idea that metaphors are time-consuming and are not produced in reference production under time pressure.

Hypothesis 5 was not confirmed. It suggests that consistency is not appropriate under time pressure.

4. General discussion. The two experiments reported in this paper demonstrated that small cardinalities (till 4) are overspecified in reference production because of the two factors: absoluteness and salience. Due to these factors, small cardinalities resemble color, which has been argued to be absolute and salient (Tarenskeen et al. 2015 among many others). A possible explanation for the salience and absoluteness of small cardinalities is subitizing which has been observed for small cardinalities till 4 (Kaufman et al. 1949 among many other studies). Subitizing makes small cardinalities salient and forces the corresponding numerals to be used in exact meanings (‘exactly n ’) rather than in at-least meanings (‘at least n ’). The present study supports the idea that both absoluteness and salience play a crucial role in overspecification of cognitive domains in reference production. Additionally, the paper has implications for producing metaphor and inconsistency in reference under time pressure.

References

- Akin, Omer & William Chase. 1978. Classification of three-dimensional structures. *Journal of Experimental Psychology: Human perception and performance* 4(3). 397–410.
- Arts, Anja, Maes, Alfons, Noordman, Leo & Carel Jansen. 2011. Overspecification in written instruction. *Linguistics* 49. 555–574.
- Atkinson, Janette, Campbell, Fergus & Marcus Francis. 1976. The magic number 4 ± 0 : A new look at visual numerosity judgements. *Perception* 5. 327–334.
- Belke, Eva & Antje Meyer. 2002. Tracking the time course of multidimensional stimulus discrimination: Analyses of viewing patterns and processing times during “same”-“different” decisions. *European Journal of Cognitive Psychology* 14. 237–266.
- Bourdon, Benjamin. 1908. Sur le temps nécessaire pour nommer les nombres. *Revue philosophique de la France et de l'étranger* 65. 426–431.
- Breheny, Richard. 2008. A new look at the semantics and pragmatics of numerically quantified noun phrases. *Journal of Semantics* 25(2). 93–139.

- Brown-Schmidt, Sarah & Agnieszka Konopka. 2011. Experimental approaches to referential domains and the on-line processing of referring expressions in unscripted conversations. *Information* 2. 302–326.
- Chi, Michelene & David Klahr. 1975. Span and rate of apprehension in children and adults. *Journal of Experimental Child Psychology* 19. 434–439.
- Gatt, Albert, Van der Sluis, Ielka, & Emiel Kraemer. 2007. Evaluating algorithms for the generation of referring expressions using a balanced corpus. *Proceedings of the 11th European Workshop on Natural Language Generation*. 49–56.
- Gatt, Albert, Kraemer, Emiel, Van Gompel, Roger & Kees Van Deemter. 2013. Production of referring expressions: Preference trumps discrimination. *Proceedings of the 35th Annual Conference of the Cognitive Science Society*. 483–488.
- Kaufman, E., Lord, M. & T. Reese. 1949. The discrimination of visual number. *American Journal of Psychology* 62. 498–525.
- Koolen, Ruud, Gatt, Albert, Goudbeek, Martijn, & Emiel Kraemer. 2011. Factors causing over-specification in definite descriptions. *Journal of Pragmatics* 43. 3231–3250.
- Jensen, E., Reese, E. & T. Reese. 1950. The subitizing and counting of visually presented fields of dots. *Journal of Psychology* 30. 363–392.
- Livingstone, Margaret & David Hubel. 1988. Segregation of form, color, movement, and depth: anatomy, physiology, and perception. *Science* 240. 740–749.
- Mandler, George & Billie Shebo. 1982. Subitizing: An analysis of its component processes. *Journal of Experimental Psychology: General* 111(1). 1–22.
- Mangold, Roland & Rupert Pobel. 1988. Informativeness and instrumentality in referential communication. *Journal of Language and Social Psychology* 7. 181–191.
- Musolino, Julien. 2004. The semantics and acquisition of number words: integrating linguistic and developmental perspectives. *Cognition* 93(1). 1–41.
- Noveck, Ira, Maryse Bianco, & Alain Castry. 2001. The costs and benefits of metaphor. *Metaphor and Symbol* 16. 109–121.
- Oyama, Tadasu, Kikuchi, Tadashi & Shigeru Ichihara. 1981. Span of attention, backward masking and reaction time. *Perception and Psychophysics* 29(2). 106–112.
- Papafragou, Anna, & Julien Musolino. 2003. Scalar implicatures: experiments at the syntax-semantics interface. *Cognition* 86. 253–282.
- Pechmann, Thomas. 1989. Incremental speech production and referential overspecification. *Linguistics* 27. 89–110.
- Revkin, Susannah, Piazza, Manuela, Izard, Véronique, Cohen, Laurent & Stanislas Dehaene. 2008. Does subitizing reflect numerical estimation? *Psychological Science* 19(6). 607–614.
- Rubio-Fernández, Paula. 2016. How redundant are redundant color adjectives? An efficiency-based analysis of color overspecification. *Frontiers in Psychology* 7. 153.
- Sedivy, Julie. 2003. Pragmatic versus form-based accounts of referential contrast: Evidence for effects of informativity expectations. *Journal of Psycholinguistic Research* 32. 3–23.
- Stevens, Stanley. 1938. On the problem scales for the measurement of psychological magnitudes. *Journal of Unified Science* 9. 94–99.
- Tarenskeen, Sammie, Broersma, Mirjam, & Bart Geurts. 2015. Overspecification of color, pattern, and size: salience, absoluteness, and consistency. *Frontiers in Psychology* 6. 1703.
- Taves, Ernest. 1941. Two mechanisms for the perception of visual numerosness. *Psychology* 37. 1–47.

- Thomas, Roger, Phillips, Julia & Cheryl Young. 1999. Comparative cognition: Human numerosness judgments. *The American Journal of Psychology* 112 (2). 215–233.
- Trick, Lana. 1992. A Theory of enumeration that grows out of a general theory of vision: Subitizing, counting, and FINSTs. In J. Campbell (ed.), *The Nature and Origins of Mathematical Skills*, 257–299. Elsevier Science Publishers B.V.
- Van Gompel, Roger, Van Deemter, Kees, Gatt, Albert, Snoeren, Rick, & Emiel Krahmer. 2019. Conceptualization in reference production: Probabilistic modeling and experimental testing. *Psychological Review* 126(3). 345–373.
- Westerbeek, Hans, Koolen, Ruud, & Alfons Maes. 2015. Stored object knowledge and the production of referring expressions: The case of color typicality. *Frontiers in Psychology* 6. 935.