Acta Psychologica 226 (2022) 103558

Contents lists available at ScienceDirect

Acta Psychologica

journal homepage: www.elsevier.com/locate/actpsy

Ratings of name agreement and semantic categorization of 247 colored clipart pictures by young German children

Linda Sommerfeld[a,*], Maria Staudte[b], Jutta Kray[a]

[a] Department of Psychology, Saarland University, Germany
[b] Department of Computational Linguistics and Phonetics, Saarland University, Germany

ARTICLE INFO

Keywords:
Name agreement
Semantic categorization
Picture naming
Picture ratings
Children
Age differences

ABSTRACT

Developmental and longitudinal studies with children increasingly use pictorial stimuli in cognitive, psychologic, and psycholinguistic research. To enhance validity and comparability within and across those studies, the use of normed pictures is recommended. Besides, creating picture sets and evaluating them in rating studies is very time consuming, in particular regarding samples of young children in which testing time is rather limited. As there is an increasing number of studies that investigate young German children's semantic language processing with colored clipart stimuli, this work provides a first set of 247 colored cliparts with ratings of German native speaking children aged 4 to 6 years. We assessed two central rating aspects of pictures: Name agreement (Do pictures elicit the intended name of an object?) and semantic categorization (Are objects classified as members of the intended semantic category?). Our ratings indicate that children are proficient in naming and even better in semantic categorization of objects, whereas both seems to improve with increasing age of young childhood. Finally, this paper discusses some features of pictorial objects that might be important for children's name agreement and semantic categorization and could be considered in future picture rating studies.

1. Introduction

Pictures of objects are a basic instrument for investigating developmental changes in memory, perception, and language processing in cognitive, (neuro-)psychological, and psycholinguistic studies (e.g., Ballesteros et al., 2007; Bonin et al., 2003; Landrum, 1997; Mecklenbräuker et al., 2001; Reales & Ballesteros, 1999). In the latter, pictorial objects are an important component of the so-called Visual World Paradigm. Here, adults and children look at computerized visual scenes of pictorial objects and listen to recorded sentences at the same time, while their eye-movements are recorded. Based on the evaluation of their eye movements, it is then possible to draw conclusions about their language processing (Altmann & Kamide, 1999). In recent years, psycholinguistic studies using the Visual World Paradigm often have shown that during language processing, adults and children leverage their semantic knowledge to actively anticipate upcoming words, rather than just passively receiving them (for a review, see Pickering & Gambi, 2018). For instance, when hearing a sentence like "The man *eats* soon the cake.", adults and children anticipatorily look at the picture of a *cake* rather than the picture of a *tractor* upon hearing the verb *eat* and before the end of the sentence. This indicates that they anticipate how the sentence might proceed based on the verb already. Reason for this is that the verb *eat* is semantically constraining, thus limits the number of verb arguments that plausibly complete the sentence to those words belonging to a particular semantic category, *edible* objects in this example (Altmann & Kamide, 1999; Andreu et al., 2013; Ankener et al., 2018; Mani & Huettig, 2012, 2014; Nation et al., 2003).

Thus, pictorial objects are a beneficial tool to investigate anticipatory language processing (for a review, see Pickering & Gambi, 2018). This is an important methodological advantage for psycholinguists, since research on anticipatory language processing recently attracts attention, particularly regarding its development in young childhood (Andreu et al., 2013; Borovsky et al., 2012; Borovsky & Creel, 2014; Mani & Huettig, 2012, 2014; Nation et al., 2003). This may result from the fact that early childhood is a critical period of language development (e.g., Al-Harbi, 2020; Lenneberg, 1967), and that anticipation could, at least in part, explain the speed and accuracy of language processing from early childhood on (Mani & Huettig, 2012).

However, conclusions about anticipatory language processing (or other research phenomena such as memory or perception) are most valid if pictorial stimuli are normed or rated for particular age ranges to ensure that the presented pictorial objects (a) are recognized in the

* Corresponding author.
E-mail address: lindas@coli.uni-saarland.de (L. Sommerfeld).

https://doi.org/10.1016/j.actpsy.2022.103558
Received 7 July 2021; Received in revised form 2 March 2022; Accepted 6 March 2022
Available online 16 April 2022

intended way and (b) can be understood as members of the verb-induced semantic categories (Borovsky et al., 2012; Borovsky & Creel, 2014; Nation et al., 2003). Otherwise, we can expect rather mixed findings across developmental studies, due to differences in the used stimulus materials, which in turn hamper the comparability across studies as well as the drawing of strong empirical conclusions (Berman et al., 1989; Bonin et al., 2003; Cycowicz et al., 1997; Sanfeliu & Fernandez, 1996; Severens et al., 2005; Snodgrass & Vanderwart, 1980). It is important to note that skipping picture rating and using existing norms of corresponding verbal materials instead is no solution. This is because some studies using picture stimuli address groups of children who cannot read yet, so that the corresponding verbal or written ratings of pictures may not be appropriate for them. Furthermore, the processing of words and pictorial objects differs remarkably. For instance, adults and children show shorter naming times for words than for pictures (Greenham & Stelmack, 2001; Theios & Amrhein, 1989; Valente et al., 2016), whereas at least adults categorize pictures faster than words (Durso & Johnson, 1979; Potter & Faulconer, 1975). This difference in task performance between pictorial objects and verbal materials could result from the fact that the lexical-semantic access depends on a variety of features of pictorial objects, such as their quality or complexity as well as the frequency or age of acquisition of their names (e.g., Bonin et al., 2003; Kosslyn & Chabris, 1990; Sanfeliu & Fernandez, 1996; Snodgrass & Vanderwart, 1980).

In sum, pictorial stimuli should be evaluated in a rating study before the actual experiment is conducted to ensure the validity and comparability of the results (e.g., Snodgrass & Vanderwart, 1980). This is common practice not only in the field of language processing research (e.g., Ankener et al., 2018; Borovsky et al., 2012; Borovsky & Creel, 2014; Staudte et al., 2021). However, creating new stimulus materials and evaluating them in rating studies is very time consuming, in particular regarding samples of young children in which testing time is rather limited (Bonin et al., 2003; Sanfeliu & Fernandez, 1996; Snodgrass & Vanderwart, 1980).

The main aim of the present study is therefore to provide an initial set of 247 colored cliparts with their ratings of recognition and semantic categorization for German children aged between 4 and 6 years. The recognition ratings of colored objects are generally useful for experimental researchers examining developmental changes in memory and perception (cf. Baadte & Meinhardt-Injac, 2019; Cycowicz et al., 2001; Pezdek, 1987; Will et al., 2021), while the ratings on semantic categorization are of particular interest for psycholinguistic researchers assessing developmental changes in (anticipatory) language processing (Andreu et al., 2013; Borovsky et al., 2012; Borovsky & Creel, 2014; Mani & Huettig, 2012, 2014; Nation et al., 2003).

Surprisingly, normative data for the recognition of pictorial stimuli are still sparse (Bonin et al., 2003). To date, a few data sets with recognition norms for pictorial stimuli exist for adult samples in different languages (e.g., English norms: Saryazdi et al., 2018; Snodgrass & Vanderwart, 1980, French norms: Bonin et al., 2003, German norms: Schröder et al., 2012, Multi-language norms: Duñabeitia et al., 2018; Kremin et al., 2003). For children, there are English norms for 5- to 10-year-olds (Berman et al., 1989, $n = 40$; Cycowicz et al., 1997, $n = 32$) and Portuguese norms for 5- to 7-year-olds (Pompéia et al., 2001, $n = 36$). Moreover, Wang et al. (2014) collected picture norms for 4- to 6-year-old Chinese Mandarin native speakers ($n = 66$) and Cannard et al. (2005) normed pictures for 3- to 8-year-old French native speaking children ($n = 960$). However, to date there exist no ratings for semantic categorization of pictorial objects for children as well as no recognition ratings for young German children.

Notably, pictorial objects need to be normed for each language separately, because the way speakers ascribe properties to pictures (e.g., picture names) differs across languages and cultures (Bonin et al., 2003; Dimitropoulou et al., 2009; Pompéia et al., 2001; Sanfeliu & Fernandez, 1996). Also, it is important to consider picture norms of children in the age range of interest, because it differs remarkably across age how individuals process pictorial objects (Borovsky et al., 2012; Borovsky & Creel, 2014; Nation et al., 2003). For example, the way object pictures are named has been shown to vary between age groups, such as kindergarten and school children (e.g., Berman et al., 1989; Cycowicz et al., 1997; Pompéia et al., 2001). This might be due to age-related differences in language and visual experience between younger and older children (Cycowicz et al., 1997).

The present work is therefore the first to provide a public data set of pictorial objects with their ratings on recognition as well as on semantic categorization for young German-speaking children between 4 and 6 years. We believe that these ratings are very helpful for experimental researchers designing studies for examining developmental changes in memory, perception, and language in early childhood. As an indicator of recognition, we collected ratings for the most popular aspect of pictorial objects, that is name agreement (Pompéia et al., 2001). Name agreement (NA) is defined as the extent to which participants agree on a particular name to refer to an object (Bonin et al., 2003). Name agreement ensures that an object is familiar to children and that its pictorial appearance is recognized, thus elicits the intended object representation (Borovsky et al., 2012; Borovsky & Creel, 2014). Name agreement is high when the number of alternative names that participants provide for a given object is low (Snodgrass & Vanderwart, 1980). Most researchers measure name agreement in two ways (e.g., Bonin et al., 2003; Snodgrass & Vanderwart, 1980): First, they compute the percentage of participants producing the most common name of a pictorial object (which is called “common NA” in this study). Second, they determine the information statistic H index (Shannon, 1948), which takes into account the alternative names that participants produce for a concept (Ghasisin et al., 2014; Snodgrass & Vanderwart, 1980). Since naming of pictorial stimuli is very variable in early childhood (Cycowicz et al., 1997), we additionally provide a more liberal measure of name agreement. Therefore, we calculated the percentage of children producing the intended name or a similarly correct synonym of a picture which we term “extended NA”. This information can be useful if general recognition of pictorial objects and their properties is of higher interest than producing the most common picture name, as for instance in studies on language processing in combination with pictorial stimuli (e.g., Ankener et al., 2018; Borovsky et al., 2012; Borovsky & Creel, 2014; Huettig & Mani, 2016; Mani & Huettig, 2012, 2014).

We also collected for the first time ratings of semantic categorization, which we understand as the classification of objects as members of intended semantic categories (cf. Federmeier & Kutas, 1999). Here, semantically constraining verbs that only allow for a limited number of arguments (e.g., eat), served to constrain semantic categories (e.g., edible). That is, all edible objects would be members of the semantic category *edible*. Such a semantic classification of picture stimuli is very useful for psycholinguistic research on anticipatory language processing (e.g., Ankener et al., 2018; Borovsky et al., 2012; Borovsky & Creel, 2014; Huettig & Mani, 2016; Mani & Huettig, 2012, 2014). Here, pictorial objects assessed in terms of their semantic properties can answer the question to what extent children process language anticipatorily based on semantic properties of the given linguistic and visual information (Pickering & Gambi, 2018). This recently attracted attention because anticipation seems to affect the speed and accuracy of language comprehension (Mani & Huettig, 2012), while being modulated by various developmental features such as a children's age or language experience in terms of vocabulary size and reading skills (Borovsky et al., 2012; Borovsky & Creel, 2014; Mani & Huettig, 2012, 2014; Nation et al., 2003).

Norming studies with adults often also collect data on naming latencies (e.g., Bonin et al., 2003; Snodgrass & Vanderwart, 1980), which is prone to be even more variable in young children, and therefore was not measured in the present study. Also, because testing time with young children is limited (Brewer et al., 2013) and because they are not familiar with the concepts of visual complexity, age of acquisition, and familiarity, we collected no judgements about these features of our stimuli.

Before selecting the pictorial stimuli for our study, we looked at the stimulus sets of existing large-scale norming studies. Most of them provided norms (e.g., name agreement, naming latencies, familiarity, visual complexity, age of acquisition) only for black and white line drawings (e.g., Berman et al., 1989; Bonin et al., 2003; Cycowicz et al., 1997; Snodgrass & Vanderwart, 1980). However, the ecological validity of study results increases with increasing iconicity of the used picture stimuli (Reuter et al., 2020), reflecting the extent to which images resemble the real-world objects they represent (Saryazdi et al., 2018). Because colors are a key indicator of a picture's iconicity (Moreno-Martínez & Montoro, 2012), many research on language processing (e.g., Andreu et al., 2013; Ankener et al., 2018; Ankener, 2019; Nation et al., 2003; Staudte et al., 2021) or memory and perception (e.g., Cycowicz et al., 2001; Will et al., 2021) is conducted with colored picture stimuli. To be more precise, these studies often use colored clipart pictures as stimuli. This might be due to the fact that colored cliparts combine the claim of (a) being a highly realistic representation of the intended object, while (b) containing a minimum of deflecting details (Saryazdi & Chambers, 2018). Thus, the use of colored clipart stimuli may reduce possible additional effort in participant's object recognition which might enable them to focus on the main task of a given experimental situation. Therefore, we decided to rate colored clipart pictures in our study.

In sum, the main goal of this work was to provide ratings on recognition and semantic categorization for a large data set of pictorial objects for young German children aged between 4 and 6 years that are not available in the literature so far. These ratings can be useful for experimental researchers designing developmental studies to determine age changes in cognitive functioning, in general, and in anticipatory language processing, in particular. To achieve this goal, we first selected 247 colored objects from open sources and assigned them to visual scenes of four objects whereas the number of objects fitting to a semantic category varied across the scenes. Children aged 4 to 6 were instructed to first name all objects of a scene and then to select all objects fitting to a semantic category. As comparable to other studies in the field (e.g., Borovsky et al., 2012; Borovsky & Creel, 2014; Nation et al., 2003; Staudte et al., 2021), we collected data from 40 children and determined common name agreement, extended name agreement, *H* index, and semantic categorization averaged and for each age group separately (see Supplement 1).

2. Material and methods

2.1. Participants

Forty kindergarten children (mean age = 5.5 years, range 4.0 to 6.8, *SD* = 0.82) participated in the study, whereas 12 children were 4 years old, 18 of them were 5 years old, and 10 children were 6 years old. They all were German native speakers without any reading and writing experience. Data from seven additional children were excluded due to other native languages ($n = 4$) and problems during testing ($n = 3$). Children were recruited via flyer and newspaper advertisement. The study was approved by a local ethics committee of Saarland University. All parents filled in a consent form and received 10 Euro as compensation.

2.2. Stimulus materials

We created a set of 247 colored clipart pictures which we collected from free clipart libraries (see the picture folder and Supplement 2 in the Supplemental materials for the cliparts and details on the libraries). Some pictures were created via combining and modifying existing cliparts. For better recognizability, some pictures were re-inked and all pictures were cleared of shadows and lettering. Any editing was done with Microsoft "Paint 3D". All pictures were presented computerized at a size of 750 mm × 750 mm.

To measure name agreement and semantic categorization at the same time we assigned four objects to a visual scene (see Fig. 1), with the exception of seven objects that were presented individually on the screen and served as practice items at the beginning of the session to instruct the naming task. The remaining 240 items were sorted in two different types of items which we termed test pictures and filler pictures. For the filler pictures, the objects ($n = 64$) were randomly assigned to the scenes ($n = 16$) and the children's task was only to name the objects. We did this to avoid that the children performed only one task throughout the whole experiment (monotony) and by this get bored and lose interest in the study.

For the majority of 176 so-called test pictures, we collected both name agreement and semantic categorization ratings. The objects and scenes were selected in two steps. In the first step, we searched for 44 semantically constraining verbs, each specifying a particular semantic category (see Supplement 3). Most of these verbs were taken from previous studies on predictive language processing in which these verbs were associated with high close probabilities (cf. Andreu et al., 2013; Ankener, 2019; Mani & Huettig, 2012). The remaining verbs were selected on the basis of face-validity by three German native speakers of our research group. In the second step, we then selected for each verb (e.g., eat) four colored cliparts assumed to be consistent with the verb-induced semantic category (e.g., edible) and familiar to children (e.g., cake, cookie, muffin, and yoghurt). In total, there were 176 test pictures, four of each belonging to the same semantic category. Noteworthy, a small number of verb-object matchings was not very typical (e.g., the verb break was matched with a television picture). This is because we aimed at collecting ratings for a considerable number of semantically constraining verbs paired with four pictorial objects each, that in turn had to be familiar to children. As the number of objects children are familiar with is limited, a few of the verb-object matchings were rather atypical but still plausible (e.g., it is actually possible to break a television). Indeed, this may not be problematic for our data base, as potential users can go through the data and make their own selections.

For the assignment of the test pictures to scenes, we created 44 congruent and 44 incongruent picture scenes. For congruent picture scenes, each four objects belonged to the same semantic category. For instance, we presented four edible objects together (e.g., cake, cookie, muffin, and yoghurt). For incongruent picture scenes, only one object of each scene (e.g., cake) matched to the semantic category (e.g., edible) while the other three objects belonged to another sematic category (e.g., the drivable bus, scooter, tractor). Because testing time is limited for

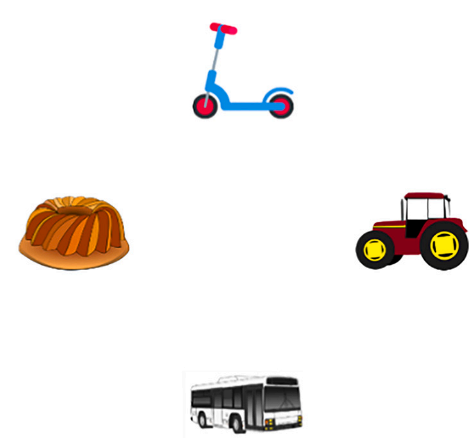

Fig. 1. Example of a picture scene.
Note. After enough time to inspect the scene, the naming task began. Children were asked to name the objects. Here we expected the names "Roller", "Traktor", "Bus", and "Kuchen" (scooter, tractor, bus, cake). For most picture scenes, the categorization task followed. In this example, children were asked to identify those objects that are "edible", and we expected "cake" as answer.

children (Brewer et al., 2013), we divided the scenes into two stimulus lists, each with 22 congruent and 22 incongruent picture scenes. Here, a semantic category that was presented in a congruent scene in list A (e.g., four edible objects) was presented in its incongruent version in list B (e.g., one edible next to three inedible objects). For the filler pictures, 10 randomly chosen filler scenes were presented in both stimulus lists, whereas 6 filler scenes were randomly distributed to only one of the stimulus lists.

In total, the 247 object pictures were divided to two lists of 169 pictures. Each list contained 7 practice pictures, 110 test pictures, and 52 filler pictures. Ninety-one pictures were part of both lists (7 practice pictures, 44 test pictures that were presented in a congruent and in an incongruent scene, 40 filler pictures), and were therefore rated by all 40 children. Seventy-eight pictures were part of only one list each (132 test pictures, 12 filler pictures), and were therefore only rated by 20 children. All 40 participants received only one list. After every second or third scene with test pictures, a scene with filler pictures followed.

2.3. Rating tasks

After naming the seven practice items, participants were presented with 60 scenes, each with four different pictorial objects (see Fig. 1). Once children had inspected a picture scene, the naming task started. Children were instructed to name the objects to the best of their knowledge while the experimenter pointed to them one after the other, always starting with the top object, followed by the next three objects in a clockwise direction. If no name was given, the experimenter asked again "What could that be?" until children produced a name or negated to know it. The naming task took as long as children needed to name all four objects (or to negate to know them) which they did rather quickly after the experimenter pointed at them.

Then, children responded to the categorization task, in which they were asked if none, one, or more objects of a scene were suitable arguments for a certain constraining verb. For example, in the scene presented in Fig. 1, they were asked "Can one eat one or more of those objects and if so, which one(s)?", whereas the German translation of the phrase "which one(s)" is "welche" and does not indicate whether the answer is one or more objects. Here, the children could answer in the order of their choice. If children hesitated, the experimenter engaged them to try an answer only once.

2.4. Procedure

Children were tested individually in one-hour sessions at Saarland University. Parents filled in a consent form, and a form about their children's age and mother language. Then, the experimenter and the child entered the laboratory. After a short adaptation phase, children were asked to take a seat in front of the computer and to look at the picture scenes on the screen. Then, they performed the naming and the categorization task. After each trial, children's answers were noted in a test protocol and the experimenter started the next trial. Participants were given a short break after each third of stimuli (i.e., 20 scenes). At the end, families received the reimbursement.

2.5. Measures

The combination of the naming and the categorization task allowed us to not only provide ratings for certain object pictures but also for verb-object associations, that are often required for psycholinguistic experiments (e.g., Ankener et al., 2018; Borovsky et al., 2012; Borovsky & Creel, 2014; Huettig & Mani, 2016; Mani & Huettig, 2012, 2014).

2.5.1. Name agreement

We measured name agreement in three ways. Extended NA was calculated as the percentage of children who named an object with the intended name or with a similarly correct synonym. Wrong names (being neither the correct name nor a synonym) and "do not know answers" were rated as "wrong answers".

The intended names of the pictures were determined a-priori by three independent German native adult experts of our research group. For instance, regarding the "ball" picture, all three experts distinguished a-priori that the intended name is "ball". Post-hoc, the same experts determined all given alternative answers, i.e., all answers not being identical with the intended picture names, as either a synonym or a wrong name. As synonym we understood an official other correct name of the picture with respect to the intended picture name (verified by https://www.duden.de/synonyme, last access: December 14, 2021). For example, the German word "Karotte", which is a similarly correct name for the German word "Möhre", was rated as synonym for the "Möhre" picture. In addition to that, we rated diminutives (e.g., "Würstchen" instead of "Wurst"), accurate other names of the cliparts (e.g., "Bäume" instead of "Wald"), names including the purpose of the object pictures (e.g., "Turnrucksack" instead of "Beutel"), over-informative names (e.g., "geflochtene Haare" instead of "Haare"), and those names describing the object pictures more accurate or specific as we expected (e.g., "Lavendel" instead of "Blumenstrauß") as correct synonyms. All remaining given answers were rated as wrong names (see Supplement 4 for an overview on given synonyms and wrong answers over all children).

Next, common NA was calculated as the percentage of children that named an object with the intended name. All other given names (including synonyms) or "do not know answers" were rated as "wrong answers". In terms of extended NA and common NA, higher values indicate greater name agreement. Lastly, we estimated the information statistic H as shown in Eq. (1) and supposed by Snodgrass and Vanderwart (1980). Here, k refers to the number of different names that children gave per object, and P_i refers to the proportion of children who gave each name. All other given names (including synonyms) were rated as different names. "Do not know answers" were excluded from the calculation. Here, lower H values reflect greater name agreement. If, for example, all children named an object as in the intended way, the H value is 0, whereas higher H values indicate that more alternative names were given. For all three measures of name agreement, plural names and mispronunciations were rated as correct names.

$$H = \sum_{i=1}^{k} P_i log_2\left(\frac{1}{P_i}\right) \quad (1)$$

2.5.2. Semantic categorization

We also calculated the percentage of correct semantic categorizations for 176 pictures to which we assigned a verb-induced semantic category a-priori (see Supplement 3). Here, we asked the children for the object scenes of four objects, whether and which of the depicted objects would match a particular semantically constraining verb (e.g., "Is any of the presented objects edible, and if so which one(s)?"). Correct matching of a picture to a verb was scored as a "correct answer". If a child did not name or point to an object that belonged to the verb or if a child by mistake mentioned that an object did not belong to the verb, this was scored as a "wrong answer". For example, if a picture of a cake was presented alongside three other objects and the crucial semantically constraining verb was "edible", the child answered correctly if it classified the cake as edible. In contrast to this, the child answered incorrectly if it did not name or point to the picture of the cake or mentioned that the cake was not edible.

3. Results and discussion

Supplement 1 presents the 247 pictorial stimuli of this study with their ratings of name agreement and semantic categorization for all children as well as for each age group (4-, 5-, and 6-year-olds) separately. We first present the rating results for name agreement and semantic categorization, followed by post-hoc analyses to control for potential influences of stimulus features. For all analyses we used R

Studio (version 1.2.1335; RStudio Team, 2018). A significance criterion of $p < .05$ was applied and we report confidence intervals of 95% of the mean.

To examine age differences in name agreement and semantic categorization, we ran four linear mixed effects models as implemented in the lme4 library separately for extended NA, common NA, *H* index, and semantic categorization as dependent variables. The independent variable was the age group factor, which was included as a categorical variable in the models because picture ratings of children of different ages (4, 5, and 6 years) are most useful for developmental researchers, who usually plan their empirical studies on the basis of age ranges (cf. Andreu et al., 2013; Borovsky et al., 2012; Borovsky & Creel, 2014; Mani & Huettig, 2012, 2014; Nation et al., 2003; Reuter et al., 2020). The factor age group was effect coded for the two contrasts of most theoretical interest: With the first contrast, we compared 4-year-old versus 5- and 6-year-old children. The second contrast compared 5-year-old versus 6-year-old children. The model also contained random intercepts for items (i.e., the pictures) to capture item-specific influences. All models converged. Their results are reported below. *P*-values were estimated using the Satterthwaite degrees of freedom method as implemented in the lmerTest library, whereas confidence intervals were estimated using the stats library (see Supplement 5 for the contrasts, the models, and their results).

3.1. Name agreement

To first determine relations between the three different measures of name agreement, we computed Pearson's correlation analyses between all children's extended NA, common NA, and *H* index (see Table 1[1]). All measures of name agreement were highly correlated (Cohen, 1988): Extended NA correlated positively with common NA ($r = .82$, $p < .001$) and negatively with *H* index ($r = -.75$, $p < .001$), and both traditional measures of name agreement were correlated negatively ($r = -.90$, $p < .001$).

As can be seen in Table 2[1] and Fig. 2, the percentage of correct recognition of pictures was overall relatively high for extended NA ($M = 84$, $SD = 21$) and somewhat lower for common NA ($M = 77$, $SD = 21$). Results of the linear mixed effects models revealed that 4-year-olds showed lower values for extended NA ($\beta = -6.35$, $SE = 0.94$, $t(492.0) = -6.79$, $p < .001$, CI [−8.19, −4.52]) and common NA ($\beta = -6.15$, $SE = 1.11$, $t(492.0) = -5.56$, $p < .001$, CI [−8.32, −3.98]), as well as higher *H* indices ($\beta = 0.09$, $SE = 0.03$, $t(492.0) = 2.8$, $p = .005$, CI [0.03, 0.16]) compared to the two older age groups (see Supplement 5). Results also revealed that 5 year-old children in comparison to the 6-year-olds showed lower values for extended NA ($\beta = -5.6$, $SE = 1.1$, $t(492.0) = -5.19$, $p < .001$, CI [−7.73, −3.49]) and common NA ($\beta = -6.15$, $SE = 1.28$, $t(492.0) = -4.82$, $p < .001$, CI [−8.66, −3.65]), as well as higher *H* indices ($\beta = 0.17$, $SE = 0.04$, $t(492.0) = 4.33$, $p < .001$, CI [0.09, 0.24]).

Table 1
Results of the Pearson's correlation analyses between all measures.

Measure	1	2	3	4
1. Extended NA	–			
2. Common NA	.82 (< .001)*	–		
3. *H* index	−.75 (< .001)*	−.90 (< .001)*	–	
4. Categorization	.08 (= .948)	.07 (= .948)	−.06 (= .948)	–

Note. NA is the abbreviation for the measure of name agreement.
Cells show correlation coefficients and bracketed Bonferroni-Holm adjusted *p*-values.
* $p < .001$.

Table 2
Mean performance per measure by age group.

Measure	Groups			
	Overall	4-year-olds	5-year-olds	6-year-olds
Extended NA	84 (20.8)	80 (26.3)	83 (22.1)	89 (19.6)
Common NA	77 (25.0)	73 (29.9)	76 (27.0)	82 (24.8)
H index	0.63 (0.74)	0.70 (0.78)	0.69 (0.76)	0.52 (0.69)
Categorization	91 (13.0)	87 (19.2)	91 (13.8)	95 (12.4)

Note. NA is the abbreviation for the measure of name agreement.
Cells show means per measure and bracketed standard deviation.

In sum, the three measures of name agreement were highly correlated and on average the name agreement was quite high for the pictorial data set with better performance for the extended than the common NA measure. Name agreement increased from 4 years to 6 years for all measures, in line with previous findings, underlining that early childhood is a critical period of language development (e.g., Al-Harbi, 2020; Lenneberg, 1967), where children learn new names of objects and their visual representations every day (Carnerero & Pérez-González, 2014).

3.2. Semantic categorization

We first determined the relation between semantic categorization and name agreement (see Table 1). Children's performance in the categorization task did not correlate significantly with extended NA ($r = .08$, $p = .948$), common NA ($r = .07$, $p = .948$), and *H* index ($r = -.06$, $p = .948$). The low correlations between both measures clearly suggest that knowing an object's name is independent of understanding its semantic properties and by this captures different linguistic abilities. This finding is strongly in line with what we know about word learning in general. That is, children typically experience objects and their functional characteristics (several times) before producing their names (Deák et al., 2002; Kemler Nelson, 1999; Kemler Nelson et al., 2000; McGregor et al., 2002). This is also in line with the higher performance in the semantic categorization task than in the name agreement task (see Table 2 and Fig. 2).

The results of the linear mixed effects model revealed that 4-year-olds showed poorer semantic categorization of pictures compared to 5- and 6-year-olds ($\beta = -5.74$, $SE = 0.93$, $t(350.0) = -6.17$, $p < .001$, CI [−7.6, −3.91]), and that the 5-year-olds showed poorer semantic categorization than the 6-year-olds ($\beta = -3.5$, $SE = 1.07$, $t(350.0) = -3.26$, $p = .001$, CI [−5.6, −1.40]).

In sum, semantic categorization differs from name agreement as suggested by the low correlations between both measures while semantic categorization seems to develop earlier than object labeling. Performance in semantic categorization also increased in early childhood from 4 years to 6 years which is in line with evidence that young children learn new words and their characteristics every day, including verbs and their semantic constraints (Carnerero & Pérez-González, 2014), hence their growing language experience (Pickering & Gambi, 2018).

3.3. Post-hoc control analyses

Because the lexical and semantic access to pictorial objects depends also on the frequency and age of acquisition of their names (e.g., Bonin et al., 2003; Kosslyn & Chabris, 1990; Sanfeliu & Fernandez, 1996; Snodgrass & Vanderwart, 1980), we ran post-hoc analyses to control whether children's name agreement and semantic categorization ratings were modulated by these factors. In doing so, we consulted the linguistic data base of Schroeder et al. (2015) that provides German word frequency norms for children between 6 and 8 years of age on the basis of a

[1] Table 1 and Table 2 also contain information on the semantic categorization data. See the "Semantic categorization" section below for more details on this data and the corresponding analyses.

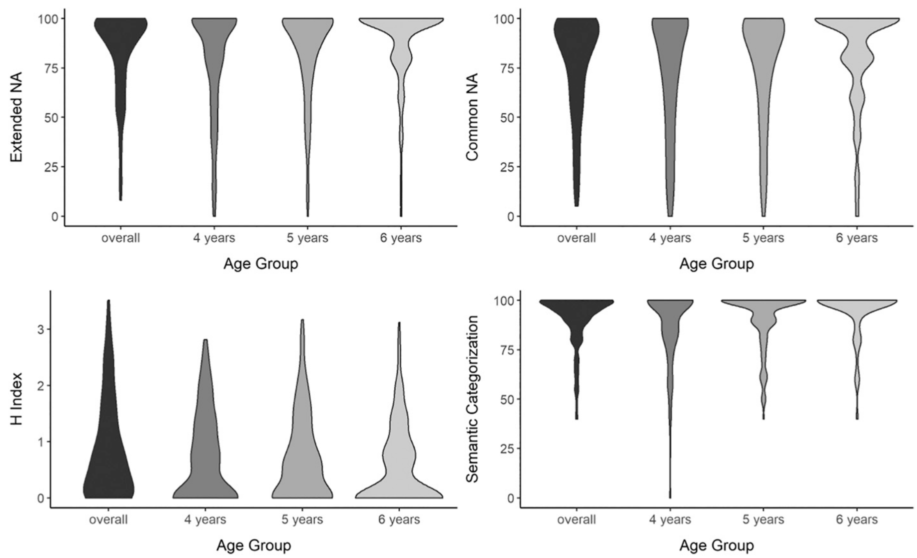

Fig. 2. Violin plots per measure and age group.
Note. Plots show the distribution of the results values per measure for all picture stimuli per age group, i.e., the width of the shaded areas displays the proportion of the data located there.

large corpus of books (n = 218) intended to be read by children of that age range. Here, we extracted the normalized annotated type frequency per million as word frequency variable. Moreover, we consulted the Schröter and Schroeder (2017) data base that provides German age of acquisition norms representing the estimated mean age in years at which a word is acquired. Here, data come from German university students (20–30 years, n = 100), who were asked at which age they believed to have heard or used a word for the first time. We consulted these data bases as they are relatively new and thus may be most representative of the child generation in our sample. Moreover, they cover many pictorial objects of the present work: Schroeder et al. (2015) provide norms for 205 of our 247 pictures, while Schröter and Schroeder (2017) provide norms for 131 of them (see Supplement 1 for the frequency and acquisition norms being available for the pictorial objects of the present study). Thus, we were able to include word frequency norms for 83% and age of acquisition norms for 53% of our pictorial objects into the control analyses.

We ran two post-hoc linear mixed effects models. The first was fitted to the dependent variable extended NA, as there is increasing relevance of picture recognition at a more general level (e.g., Borovsky et al., 2012; Borovsky & Creel, 2014; Huettig & Mani, 2016; Mani & Huettig, 2012, 2014). The second model was run with the dependent measure semantic categorization. Again, the effect coded factor age group as well as random intercepts for the items were added to both models. Finally, we included the z-standardized word frequency and age of acquisition variables as interaction factors to both models (see Supplement 6 for the contrasts, the models, and their results).

The first model revealed that children's name agreement in form of extended NA was still significantly worse for 5-year-olds compared to 6-year-olds ($\beta = -0.28$, $SE = 0.06$, $t(250) = -4.49$, $p < .001$, CI [−0.4, −0.16]), independent of the frequency ($p = .418$) or acquisition age ($p = .332$) of the picture names. Also, 4-year-old's extended NA was still significantly worse than 5- and 6-year-old's extended NA ($\beta = -0.25$, $SE = 0.05$, $t(250) = -4.67$, $p < .001$, CI [−0.36, −0.15]). Again, this was independent of word frequency ($p = .231$) but modulated by age of acquisition ($\beta = -0.15$, $SE = 0.06$, $t(250) = -2.47$, $p = .014$, CI [−0.26, −0.03]). As can be seen in Fig. 3, age differences between 4-year-olds and the two older age groups increased for name agreement with higher age of acquisition. Hence, children's name agreement was not affected by word frequency but modulated by age of acquisition.

The second model revealed that the semantic categorization was still worse in 4-year-olds than in 5- and 6-year-olds ($\beta = -0.41$, $SE = 0.08$, $t(176) = -4.98$, $p < .001$, CI [−0.56, −0.25]), independent of word frequency ($p = .877$) and age of acquisition ($p = .106$). However, when considering these control variables, the difference between the 5- and 6-year-olds in semantic categorization was no longer significant ($p = .105$). Taken together, the fact that the younger children (4-year-olds) were worse in semantic categorization than the older children (5- and 6-year-olds) was robust even when the control variables word frequency and age of acquisition were included in the statistical models. In contrast, this was not the case for the older age groups, here the group difference disappeared. Based on the reported linear mixed effects model it is not possible to say what this finding is due to. Indeed, Pearson correlation analyses between semantic categorization and word frequency respectively age of acquisition reveal information here. Children's semantic categorization was not correlated with word frequency ($r = -.02$, $t(436) = -0.05$, $p = .654$) but negatively correlated with age of acquisition ($r = -.18$, $t(280) = -3.01$, $p = .002$). This suggests that the resolved variance in semantic categorization, originally explained by group differences, goes back at age of acquisition influences.

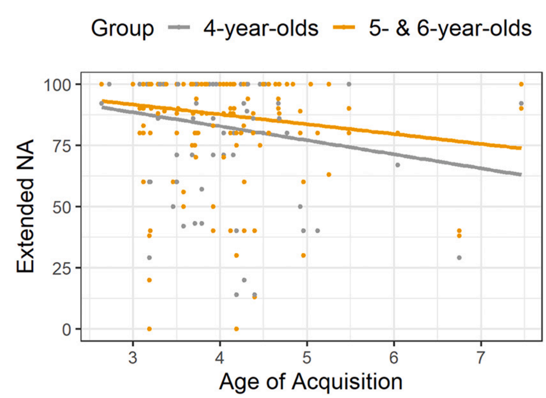

Fig. 3. Linear regression of extended NA on age of acquistion.
Note. Plot of the linear regression of extended NA on the age of acquisition norms for the youngest and the taken together older age groups.

4. General discussion

This study provides 247 colored clipart pictures with their ratings of name agreement and semantic categorization from 4- to 6-year-old German native speaking children. The evaluation of the children's picture ratings revealed that both picture naming and semantic categorization was largely successful in young children, whereas the performance in both measures increased with increasing age throughout early childhood. Also, picture naming and categorization were independent of each other, whereas semantic categorization developed earlier than object naming.

This study is the first to provide children's agreement on picture names at a more general level of recognition because we also assessed an extended measure of name agreement which not only includes the intended object names but also their synonyms as correct answers. As extended NA was highly correlated with the standard measures of name agreement (common NA, *H* index), we conclude that extended NA also represents children's name agreement, but in a more liberal way. However, more liberal ratings of picture naming are highly relevant for an increasing number of studies focusing picture recognition at a more general level as for instance studies on children's language processing in combination with pictorial objects (e.g., Borovsky et al., 2012; Borovsky & Creel, 2014; Huettig & Mani, 2016; Mani & Huettig, 2012, 2014).

As previous picture norming studies (e.g., Snodgrass & Vanderwart, 1980), we also evaluated why children's ratings may not have been successful for some pictures. In doing so, we aimed at gaining insights about which features of pictorial objects may be relevant for a successful recognition and categorization in young children. This can provide important insights for the planning of future studies with pictorial stimuli in this age range. Our post-hoc control analyses revealed that the age of acquisition of a picture, as determined by young adults, influences age differences in name agreement and semantic categorization. Name agreement is lower for pictures with higher age of acquisition norms which is more pronounced in 4-year-olds than in the two older age groups. Moreover, age differences in semantic categorization even disappeared for the older age groups when age of acquisition norms were considered. However, note that the age of acquisition norms for our pictorial stimulus set was with very few exceptions in the intended age range or lower so that our stimulus selection process which required the assigning of four target pictures to constrained verbs by face-validity was successful.

In our post-hoc analysis we also evaluated the impact of word frequency on children's name agreement and semantic categorization and found that word frequencies did not modulate age differences in both measures. This is in line with other studies revealing no effect of word frequency on at least picture naming, when also considering age of acquisition norms (Carroll & White, 1973; Chalard et al., 2003; Morrison et al., 1992). However, it should be noted that the word frequency norms that we included in our control analyses were collected from books of school children aged 6 to 8 years. As our sample was younger in age we cannot fully rule out that the frequency of words have influenced our empirical findings. Moreover, information about the frequencies of pictures of a semantic concept would have been more appropriate here as a control variable which to our knowledge does not exists.

In addition to that, we explored the alternative answers of the children in the naming task (see Supplement 4), as they might provide insights on possible reasons for the poor naming and categorization of some pictures. For example, children had considerable problems to name drinks (e.g., juice, lemonade, and tea). Instead of the drinks themselves, they often named their containers (e.g., cup, glass, and mug), which is plausible as these pictures represent both the drinks and their containers at the same time. Thus, the ambiguity of pictorial objects seems to play a considerable role for name agreement, as ambiguous pictures can activate multiple object representations and thus multiple object names. However, for some studies, this information can also be helpful. This is because some research questions may aim at examining how children name ambiguous pictures, or how this changes with increasing age, as this could provide insight into which aspects of a particular picture stimulus attract children's attention. In any case, the ambiguity of pictures should be considered when using pictures as stimuli in empirical studies. Alternative answers for rather few pictures also revealed that they were not so easy to recognize as we first thought. For instance, the “butter” picture was often labelled as “cheese” and the “frame” picture as “drawing” (see Supplement 4). Hence, the recognizability of those two pictures seems not sufficient and therefore resulted in poor name agreement.

Indeed, when referring to the alternative answers, it should be noted that some of our classifications of synonyms and wrong answers may slightly go back to regional differences in Germany. This is because the experts who made the classification lived in the same federal state (Saarland) during study conduction. This could explain, for instance, why “Saarbahn” was accepted as a synonym for “Zug” or why “Kaffeestücken” was not accepted as a synonym for “Keks” or “Croissant”. However, this is not necessarily problematic for users of our picture ratings, since we provide them along with the list of synonyms and wrong answers, and since other researchers would probably go through the ratings and the list to make their own choices.

Finally, although the semantic categorization performance was very high with a mean percentage of correct categorizations of about 91%, some children were not sure about the meaning of a given verb. For instance, these children did not know the exact meaning of the verbs “clean” and “explore”. Here, we cannot rule out the possibility that participant characteristics, especially regarding socio economic factors, might have affected children's understanding of the verb-induced semantic categories. Aside from these shortcomings the ratings for name agreement and semantic categorization were generally high and therefore our pictorial stimulus set provides a useful data base for developmental researchers when designing their studies.

5. Conclusion

This study is the first to provide ratings of name agreement for 247 and ratings of semantic categorization for 176 colored clipart pictures from 40 German native speaking children aged between 4 and 6 years. The present work can facilitate the preparation of empirical studies in which not only cognition and perception, but in particular semantic language processing of young German children in combination with pictorial stimuli is investigated. By this, our study contributes to the general aim to enhance the validity and comparability within and across studies.

Data availability statement

Preliminary data are available from the corresponding author on reasonable request.

Ethics and consent statement

This study was conducted in accordance with the recommendations of the American Psychological Association, with written informed consent for participation and publishing by legal guardians of all subjects. Approval for this study was obtained from a local ethics committee of Saarland University. The procedures used in this study adhere the tenets of the Declaration of Helsinki.

CRediT authorship contribution statement

All authors initiated and framed the research questions. Linda Sommerfeld conducted and analyzed the study. All authors edited, read, and approved the final manuscript.

Open practices statement

The data and materials for the experiment reported here is available (see Supplemental materials), and the experiment was not preregistered.

Declaration of competing interest

The authors declare that the research was conducted in the absence of any financial or personal relationships that could be construed as a potential conflict of interest.

Acknowledgements

This study was funded by the Deutsche Forschungsgemeinschaft (DFG) — Project-ID 232722074 — SFB 1102.

Supplemental materials

The picture stimuli of the current work and all corresponding supplements can be downloaded at Harvard Dataverse (https://doi.org/10.7910/DVN/EU0IJC).

References

Al-Harbi, S. S. (2020). Language development and acquisition in early childhood. *Journal of Education and Learning, 14*(1), 69–73. https://doi.org/10.11591/edulearn.v14i1.14209

Altmann, G. T., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition, 73*(3), 247–264. https://doi.org/10.1016/S0010-0277(99)00059-1

Andreu, L., Sanz-Torrent, M., & Trueswell, J. C. (2013). Anticipatory sentence processing in children with specific language impairment: Evidence from eye movements during listening. *Applied Psycholinguistics, 34*(1), 5–44. https://doi.org/10.1017/S0142716411000592

Ankener, C. (2019). The influence of visual information on word predictability and processing effort [Doctoral dissertation, Saarland University–Germany]. SciDok — Der Wissenschaftsserver der Universität des Saarlandes. https://doi.org/10.22028/D291-28451

Ankener, C. S., Sekicki, M., & Staudte, M. (2018). The influence of visual uncertainty on word surprisal and processing effort. *Frontiers in Psychology, 9*, 2387. https://doi.org/10.3389/fpsyg.2018.02387

Baadte, C., & Meinhardt-Injac, B. (2019). The picture superiority effect in associative memory: A developmental study. *British Journal of Developmental Psychology, 37*(3), 382–395. https://doi.org/10.1111/bjdp.12280

Ballesteros, S., Reales, J. M., & Mayas, J. (2007). Picture priming in normal aging and alzheimer s disease. *Psicothema, 19*(2), 239–244.

Berman, S., Friedman, D., Hamberger, M., & Snodgrass, J. G. (1989). Developmental picture norms: Relationships between name agreement, familiarity, and visual complexity for child and adult ratings of two sets of line drawings. *Behavior Research Methods, Instruments, & Computers, 21*(3), 371–382. https://doi.org/10.3758/BF03202800

Bonin, P., Peereman, R., Malardier, N., Méot, A., & Chalard, M. (2003). A new set of 299 pictures for psycholinguistic studies: French norms for name agreement, image agreement, conceptual familiarity, visual complexity, image variability, age of acquisition, and naming latencies. *Behavior Research Methods, Instruments, & Computers, 35*(1), 158–167. https://doi.org/10.3758/BF03195507

Borovsky, A., & Creel, S. C. (2014). Children and adults integrate talker and verb information in online processing. *Developmental Psychology, 50*(5), 1600–1613. https://doi.org/10.1037/a0035591

Borovsky, A., Elman, J. L., & Fernald, A. (2012). Knowing a lot for one's age: Vocabulary skill and not age is associated with anticipatory incremental sentence interpretation in children and adults. *Journal of Experimental Child Psychology, 112*(4), 417–436. https://doi.org/10.1016/j.jecp.2012.01.005

Brewer, R., Anthony, L., Brown, Q., Irwin, G., Nias, J., & Tate, B. (2013). Using gamification to motivate children to complete empirical studies in lab environments. In N. Sawhney, E. Reardon, & J. P. Hourcade (Eds.), *IDC '13: Proceedings of the 12th International Conference on Interaction Design and Children* (pp. 388–391). Association for Computing Machinery. https://doi.org/10.1145/2485760.2485816

Cannard, C., Blaye, A., Scheuner, N., & Bonthoux, F. (2005). Picture naming in 3- to 8-year-old French children: Methodological considerations for name agreement. *Behavior Research Methods, 37*(3), 417–425. https://doi.org/10.3758/BF03192710

Carnerero, J. J., & Pérez-González, L. A. (2014). Induction of naming after observing visual stimuli and their names in children with autism. *Research in Developmental Disabilities, 35*(10), 2514–2526. https://doi.org/10.1016/j.ridd.2014.06.004

Carroll, J. B., & White, M. N. (1973). Age-of-acquisition norms for 220 picturable nouns. *Journal of Verbal Learning and Verbal Behavior, 12*(5), 563–576. https://doi.org/10.1016/S0022-5371(73)80036-2

Chalard, M., Bonin, P., Méot, A., Boyer, B., & Fayol, M. (2003). Objective age-of-acquisition (AoA) norms for a set of 230 object names in French: Relationships with psycholinguistic variables, the English data from Morrison et al. (1997), and naming latencies. *European Journal of Cognitive Psychology, 15*(2), 209–245. https://doi.org/10.1080/09541440244000076

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates Publishers.

Cycowicz, Y. M., Friedman, D., Rothstein, M., & Snodgrass, J. G. (1997). Picture naming by young children: Norms for name agreement, familiarity, and visual complexity. *Journal of Experimental Child Psychology, 65*(2), 171–237. https://doi.org/10.1006/jecp.1996.2356

Cycowicz, Y. M., Friedman, D., Snodgrass, J. G., & Duff, M. (2001). Recognition and source memory for pictures in children and adults. *Neuropsychologia, 39*(3), 255–267. https://doi.org/10.1016/S0028-3932(00)00108-1

Deák, G. O., Ray, S. D., & Pick, A. D. (2002). Matching and naming objects by shape or function: Age and context effects in preschool children. *Developmental Psychology, 38*(4), 503–518. https://doi.org/10.1037/0012-1649.38.4.503

Dimitropoulou, M., Duñabeitia, J. A., Blitsas, P., & Carreiras, M. (2009). A standardized set of 260 pictures for modern Greek: Norms for name agreement, age of acquisition, and visual complexity. *Behavior Research Methods, 41*(2), 584–589. https://doi.org/10.3758/BRM.41.2.584

Duñabeitia, J. A., Crepaldi, D., Meyer, A. S., New, B., Pliatsikas, C., Smolka, E., & Brysbaert, M. (2018). MultiPic: A standardized set of 750 drawings with norms for six European languages. *Quarterly Journal of Experimental Psychology, 71*(4), 808–816. https://doi.org/10.1080/2F17470218.2017.1310261

Durso, F. T., & Johnson, M. K. (1979). Facilitation in naming and categorizing repeated pictures and words. *Journal of Experimental Psychology: Human Learning and Memory, 5*(5), 449–459. https://doi.org/10.1037/0278-7393.5.5.449

Federmeier, K. D., & Kutas, M. (1999). A rose by any other name: Long-term memory structure and sentence processing. *Journal of Memory and Language, 41*(4), 469–495. https://doi.org/10.1006/jmla.1999.2660

Ghasisin, L., Yadegari, F., Rahgozar, M., Nazari, A., & Rastegarianzade, N. (2014). A new set of 272 pictures for psycholinguistic studies: Persian norms for name agreement, image agreement, conceptual familiarity, visual complexity, and age of acquisition. *Behavior Research Methods, 47*(4), 1148–1158. https://doi.org/10.3758/s13428-014-0537-0

Greenham, S. L., & Stelmack, R. M. (2001). Event-related potentials and picture-word naming: Effects of attention and semantic relation for children and adults. *Developmental Neuropsychology, 20*(3), 619–638. https://doi.org/10.1207/875656401753549834

Huettig, F., & Mani, N. (2016). Is prediction necessary to understand language? Probably not. *Language, Cognition and Neuroscience, 31*(1), 19–31. https://doi.org/10.1080/23273798.2015.1072223

Kemler Nelson, D. G. (1999). Attention functional properties in toddlers' naming and problem-solving. *Cognitive Development, 14*(1), 77–100. https://doi.org/10.1016/S0885-2014(99)80019-7

Kemler Nelson, D. G., Russell, R., Duke, N., & Jones, K. (2000). Two-year-olds will name artifacts by their functions. *Child Development, 71*(5), 1271–1288. https://doi.org/10.1111/1467-8624.00228

Kosslyn, S. M., & Chabris, C. F. (1990). Naming pictures. *Journal of Visual Languages & Computing, 1*(1), 77–95. https://doi.org/10.1016/S1045-926X(05)80035-7

Kremin, H., Akhutina, T., Basso, A., Davidoff, J., De Wilde, M., Kitzing, P., Lorenz, A., Perrier, D., van der Sandt-Koederman, M., Vendrell, J., & Weniger, D. (2003). A cross-linguistic data bank for oral picture naming in Dutch, English, German, French, Italian, Russian, Spanish, and Swedish (PEDOI). *Brain and Cognition, 53*(2), 243–246. https://doi.org/10.1016/S0278-2626(03)00119-2

Landrum, R. E. (1997). Implicit memory effects when using pictures with children and adults: Hypermnesia too? *The Journal of General Psychology, 124*(1), 5–17. https://doi.org/10.1080/00221309709595504

Lenneberg, E. H. (1967). The biological foundations of language. *Hospital Practice, 2*(12), 59–67. https://doi.org/10.1080/21548331.1967.11707799

Mani, N., & Huettig, F. (2012). Prediction during language processing is a piece of cake — But only for skilled producers. *Journal of Experimental Psychology: Human Perception and Performance, 38*(4), 843–847. https://doi.org/10.1037/a0029284

Mani, N., & Huettig, F. (2014). Word reading skill predicts anticipation of upcoming spoken language input: A study of children developing proficiency in reading. *Journal of Experimental Child Psychology, 126*, 264–279. https://doi.org/10.1016/j.jecp.2014.05.004

McGregor, K. K., Friedman, R. M., Reilly, R. M., & Newman, R. M. (2002). Semantic representation and naming in young children. *Journal of Speech, Language, and Hearing Research, 45*(2), 332–346. https://doi.org/10.1044/1092-4388(2002/026)

Mecklenbräuker, S., Hupbach, A., & Wippich, W. (2001). What colour is the car? Implicit memory for colour information in children. *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology, 54A*(4), 1069–1086. https://doi.org/10.1080/02724980143000055

Moreno-Martínez, F. J., & Montoro, P. R. (2012). An ecological alternative to Snodgrass & Vanderwart: 360 high quality colour images with norms for seven psycholinguistic variables. *Plos One, 7*(5), Article e37527. https://doi.org/10.1371/journal.pone.0037527

Morrison, C. M., Ellis, A. W., & Quinlan, P. T. (1992). Age of acquisition, not word frequency, affects object naming, not object recognition. *Memory & Cognition, 20*(6), 705–714. https://doi.org/10.3758/BF03202720

Nation, K., Marshall, C. M., & Altmann, G. T. (2003). Investigating individual differences in children's real-time sentence comprehension using language-mediated eye movements. *Journal of Experimental Child Psychology, 86*(4), 314–329. https://doi.org/10.1016/j.jecp.2003.09.001

Pezdek, K. (1987). Memory for pictures: A life-span study of the role of visual detail. *Child Development, 58*(3), 807–815. https://doi.org/10.2307/1130218

Pickering, M. J., & Gambi, C. (2018). Predicting while comprehending language: A theory and review. *Psychological Bulletin, 144*(10), 1002–1044. https://doi.org/10.1037/bul0000158

Pompéia, S., Miranda, M. C., & Bueno, O. F. A. (2001). A set of 400 pictures standardised for Portuguese: Norms for name agreement, familiarity and visual complexity for children and adults. *Arquivos de Neuro-Psiquiatria, 59*(2B), 330–337. https://doi.org/10.1590/S0004-282X2001000300004

Potter, M. C., & Faulconer, B. A. (1975). Time to understand pictures and words. *Nature, 253*(5491), 437–438. https://doi.org/10.1038/253437a0

Reales, J. M., & Ballesteros, S. (1999). Implicit and explicit memory for visual and haptic objects: Cross-modal priming depends on structural descriptions. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*(3), 644–663. https://doi.org/10.1037/0278-7393.25.3.644

Reuter, T., Dalawella, K., & Lew-Williams, C. (2020). Adults and children predict in complex and variable referential contexts. *Language, Cognition and Neuroscience, 36*(4), 1–17. https://doi.org/10.1080/23273798.2020.1839665

RStudio Team. (2018). RStudio: Integrated Development for R (Version 1.2.1335) [Computer software]. RStudio Inc. http://www.rstudio.com/

Sanfeliu, M. C., & Fernandez, A. (1996). A set of 254 Snodgrass-Vanderwart pictures standardized for Spanish: Norms for name agreement, image agreement, familiarity, and visual complexity. *Behavior Research Methods, Instruments, & Computers, 28*(4), 537–555. https://doi.org/10.3758/BF03200541

Saryazdi, R., Bannon, J., Rodrigues, A., Klammer, C., & Chambers, C. G. (2018). Picture perfect: A stimulus set of 225 pairs of matched clipart and photographic images normed by mechanical turk and laboratory participants. *Behavior Research Methods, 50*(6), 2498–2510. https://doi.org/10.3758/s13428-018-1028-5

Saryazdi, R., & Chambers, C. G. (2018). Mapping language visual referents: Does the degree of image realism matter? *Acta Psychologica, 182*, 91–99. https://doi.org/10.1016/j.actpsy.2017.11.003

Schröder, A., Gemballa, T., Ruppin, S., & Wartenburger, I. (2012). German norms for semantic typicality, age of acquisition, and concept familiarity. *Behavior Research Methods, 44*(2), 380–394. https://doi.org/10.3758/s13428-011-0164-y

Schroeder, S., Wuerzner, K. M., Heister, J., Geyken, A., & Kliegl, R. (2015). childLex — Eine lexikalische Datenbank zur Schriftsprache für Kinder im Deutschen. *Psychologische Rundschau, 66*(3), 155–165. https://doi.org/10.1026/0033-3042/a000275

Schröter, P., & Schroeder, S. (2017). The developmental lexicon project: A behavioral database to investigate visual word recognition across the lifespan. *Behavior Research Methods, 49*(6), 2183–2203. https://doi.org/10.3758/s13428-016-0851-9

Severens, E., Van Lommel, S., Ratinckx, E., & Hartsuiker, R. J. (2005). Timed picture naming norms for 590 pictures in Dutch. *Acta Psychologica, 119*(2), 159–187. https://doi.org/10.1016/j.actpsy.2005.01.002

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal, 27*, 623–656.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory, 6*(2), 174–215. https://doi.org/10.1037/0278-7393.6.2.174

Staudte, M., Ankener, C., Drenhaus, H., & Crocker, M. W. (2021). Graded expectations in visually situated comprehension: Costs and benefits as indexed by the N400. *Psychonomic Bulletin & Review, 28*(2), 624–631. https://doi.org/10.3758/s13423-020-01827-3

Theios, J., & Amrhein, P. C. (1989). Theoretical analysis of the cognitive processing of lexical and pictorial stimuli: Reading, naming, and visual and conceptual comparisons. *Psychological Review, 96*(1), 5–24. https://doi.org/10.1037/0033-295X.96.1.5

Valente, A., Pinet, S., Alario, F.-X., & Laganaro, M. (2016). "When" does picture naming take longer than word reading? *Frontiers in Psychology, 7*, Article 31. https://doi.org/10.3389/fpsyg.2016.00031

Wang, L., Chen, C. W., & Zhu, L. (2014). Picture norms for Chinese preschool children: Name agreement, familiarity, and visual complexity. *Plos One, 9*(3), Article e90450. https://doi.org/10.1371/journal.pone.0090450

Will, P., Merritt, E., Jenkins, R., & Kingstone, A. (2021). The medusa effect reveals levels of mind perception in pictures. *Proceedings of the National Academy of Sciences, 118*(32), Article e2106640118. https://doi.org/10.1073/pnas.2106640118