Journal of Experimental Child Psychology 194 (2020) 104782

Contents lists available at ScienceDirect

Journal of Experimental Child Psychology

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

Development of scene knowledge: Evidence from explicit and implicit scene knowledge measures

Sabine Öhlschläger [a,b], Melissa Le-Hoa Võ [a,b,*]

[a] Scene Grammar Lab, Department of Psychology, Goethe University Frankfurt, Theodor-W.-Adorno-Platz 6, 60323 Frankfurt am Main, Germany
[b] Center for Research on Individual Development and Adaptive Education of Children at Risk (IDeA), 60323 Frankfurt am Main, Germany

ARTICLE INFO

Article history:
Received 7 February 2018
Revised 1 December 2019

Keywords:
Scene perception
Cognitive development
Semantics
Syntax

ABSTRACT

In our daily lives, we rely on expectations of where to find objects in a scene. Every morning without conscious reflection, we find the milk in the refrigerator. How do these schemata develop during childhood? In the current study, we investigated the behavioral responses of 72 2- to 4-year-olds in two tasks that measured scene knowledge either directly by asking them to furnish a dollhouse or indirectly by observing their eye movements in a violation paradigm using scene photographs. In addition, we collected language acquisition measures for each child to investigate possible relations between the development of scene knowledge and language abilities. Results for both explicit and implicit measures indicated an increase of performance with age in terms of correct object placement relative to corresponding rooms/locations and a difference in first-pass dwell times between consistent and inconsistent objects. The consistency effect in eye movements was associated with shorter processing times for consistent objects, reflecting stronger predictions for objects in their familiar context/location. A reduction of first-pass dwell times to consistent objects was also predicted by the dollhouse performance measure of scene knowledge. Although strong links to language development could not be found, first indications are discussed together with possible improvements of future studies investigating such a link. In sum, our results imply that scene-related predictions effectively can

* Corresponding author at: Scene Grammar Lab, Department of Psychology, Goethe University Frankfurt, Theodor-W.-Adorno-Platz 6, 60323 Frankfurt am Main, Germany.
E-mail address: mlvo@psych.uni-frankfurt.de (M.L.-H. Võ).

https://doi.org/10.1016/j.jecp.2019.104782

influence implicit and explicit behavior by 4 years of age at the latest, allowing optimized attention allocation in scenes.

Introduction

Imagine that you are at your friend's house for the first time. You would expect the milk to be in the refrigerator and would be surprised to find it in the bathroom. Such scene-related expectations not only are intuitive but also have been subject to empirical studies. From a developmental perspective, this promising field of research might not only help to understand how we learn to orientate ourselves in an increasingly complex environment but also help to elucidate the learning mechanisms in other relevant domains. Here, we focused mainly on scene knowledge but also wanted to test possible links to language because, cognitively, experience in both domains optimizes processing (i.e., reduces processing times) when a stimulus is highly expected from the stimulus's context (Rayner, 1998). To this date, only few studies have investigated this relationship during development (Helo, van Ommen, Pannasch, Danteny-Dordoigne, & Rämä, 2017; Saarnio, 1990, 1993a).

The characterization of rules underlying the placement of objects in scenes has been inspired by the semantic–syntax distinction in language (Biederman, 1977; Biederman, Mezzanotte, & Rabinowitz, 1982). Scene knowledge tells us *what* objects to expect in a scene in general (semantics) and *where* they should be more specifically (here considered as syntax). Based on recent studies (e.g., Öhlschläger & Võ, 2017; Võ & Wolfe, 2013a, 2013b), we considered objects that were incongruent with the global meaning of the scene *semantically* inconsistent and objects that were semantically consistent but inconsistent with regard to their probable location with a scene *syntactically* inconsistent (see Fig. 1 for example scenes taken out of the rated and validated SCEGRAM database; Öhlschläger & Võ, 2017). Although the semantic–syntax distinction in perception is first and foremost metaphorical, there has been empirical evidence on the electrophysiological level that event-related potentials can differentiate between whether an object is in the wrong room or at the wrong location (e.g., Võ & Wolfe, 2013a).

Also during development, first behavioral findings suggested that indeed the information about both object probability and location seems to be cognitively represented during the preschool years, whereas the information about size and depth develops more slowly (Saarnio, 1993b). Inspired by Piagetian theory, it was assumed that scene-related expectations might rely on schemata, that is, cognitive structures created by real-world episodic experience that influence how incoming information is selected, interpreted, and organized (Hock, Romanski, Galie, & Williams, 1978; Mandler & Robinson, 1978; Saarnio, 1990, 1993b). A similar schema approach has also been promoted by cognitivist linguists (Nelson, 1974; Sinclair-de Zwart, 1973). This is why schemata could serve as a theoretical scaffolding for linking expectations in the domains of scene and language processing as different expressions of the same cognitive structures. However, in the past contradicting observations were reported for the link between language and scene knowledge (Helo et al., 2017; Saarnio, 1990, 1993a). To truly establish such a link, a large-scale investigation would be needed. Our study, therefore, focused mainly on the developmental trajectories of scene knowledge as measured using implicit and explicit procedures.

The ability to reproduce a fact in response to a *direct* test—that is, one referring to the question of study (e.g., where do objects belong in a house?)—has been used as an index that this information is represented *explicitly* (Dienes & Perner, 1999). By contrast, inferring the availability of knowledge about a fact *indirectly*, by interpreting the response to a presented stimulus (e.g., fixations to objects in a scene) without referring to the study question, was used as evidence for *implicit* representation (Dienes & Perner, 1999). The link of direct measures to explicit knowledge and of indirect measures to implicit knowledge is not straightforward (Dienes & Perner, 1999). In the following, we use the

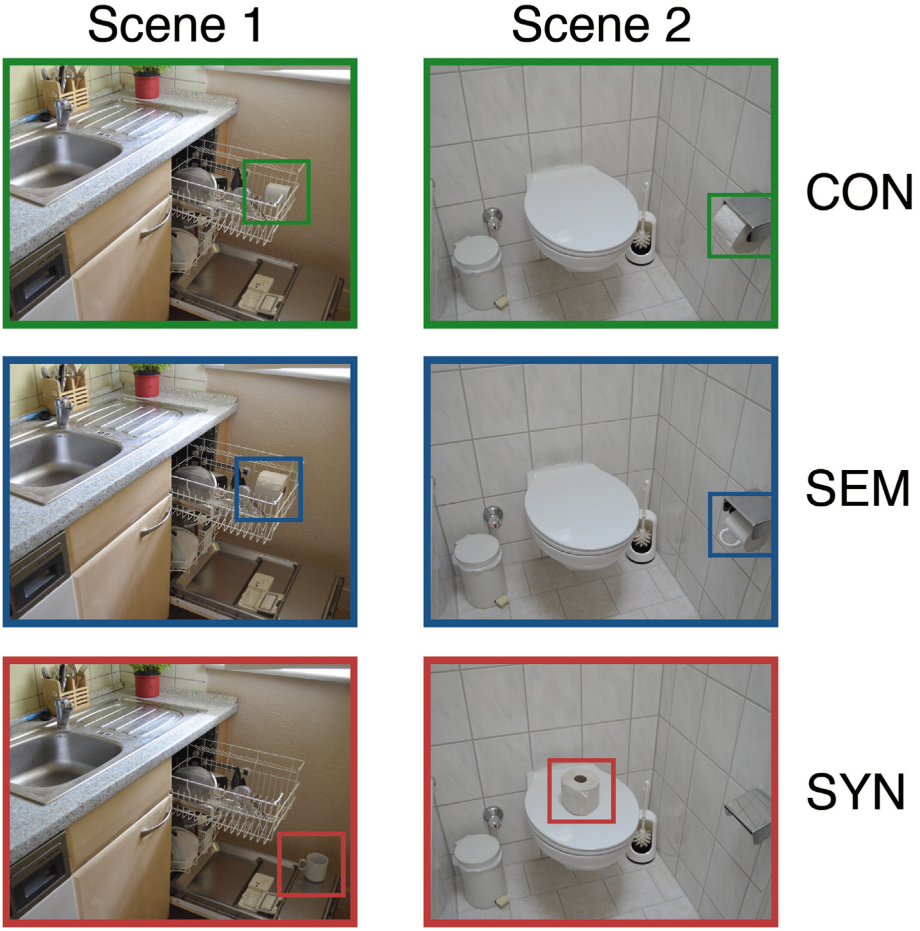

Fig. 1. Two example scenes from the SCEGRAM database (Öhlschläger & Võ, 2017) used in the eye-tracking experiment in the consistent (CON), inconsistent–semantics (SEM), and inconsistent–syntax (SYN) conditions.

terms explicit (direct) and implicit (indirect) to refer to the test measures (i.e., behavioral performance) instead of to the representation of knowledge itself.

As an explicit measure for studying scene schemata during development, the dollhouse has been a helpful tool (Freund, Baker, & Sonnenschein, 1990; Ratner, 1984; Ratner & Myers, 1981). In one study, 3-year-olds initially seemed to have limited access to their scene schemata, which improved when the experimenter reduced the planning demands by labeling and marking the rooms with an object (Freund et al., 1990). Furthermore, due to the independence of verbal expression, the use of the dollhouse allowed revealing that 2-year-olds were able to retrieve the "core defining information" of semantics that exceeded what they were able to express verbally (Ratner, 1984; Ratner & Myers, 1981, p. 365).

As an implicit measure, children's eye movements were recorded during scene exploration, showing that infants, similar to adults, scan a complex visual scene using two different modes (Helo, Pannasch, Sirri, & Rämä, 2014; Helo, Rämä, Pannasch, & Meary, 2016): one early for object localization (ambient mode: short fixations and long saccades) and one late for detailed object feature processing (focal mode: long fixations). For the processing of object–scene inconsistencies, research in adults has extensively demonstrated relatively longer dwell times compared with the consistent control for semantic violations (e.g., De Graef, Christiaens, & d'Ydewalle, 1990; Henderson, Weeks, & Hollingworth, 1999; Öhlschläger & Võ, 2017; Võ & Henderson, 2009) as well as syntactic violations (De Graef et al., 1990; Öhlschläger & Võ, 2017; Võ & Henderson, 2009). Recently, the consistency effect to semantic manipulations was also demonstrated in 2-year-olds, but only when attention was already directed to the critical objects by their high visual saliency (Helo et al., 2017). The authors concluded that children as young as 24 months were able to use their scene schemata.

Our study set out to look at the relation of such explicit and implicit measures of scene knowledge development while tentatively asking the question of whether language acquisition and scene knowledge development could interact. As an explicit measure of scene knowledge, we assessed how children and adults furnished a dollhouse. As an implicit measure, we recorded children's and adults' eye movements while they were viewing photographs of daily life scenes with object inconsistencies. Our working hypothesis was that knowledge about the correct global context versus the correct location could show different developmental trajectories and that these might coincide with the language abilities of children.

Method

Participants

In total, 96 participants (72 children and 24 adults) were included in the current analysis. The child sample was equally subdivided into three age groups of 2, 3, and 4 years. Half of the children from each age group were exposed to semantic scene–object inconsistencies (semantic: mean age = 42.5 months, *SD* = 10.3, range = 25–59; 20 female; see Table A.1a in Appendix A), whereas the other half were presented with syntactic scene–object inconsistencies (syntactic: mean age = 41.6 months, *SD* = 10.1, range = 25–57; 17 female; see Table A.1a). Children included had normal vision and no neurological disease, as assessed by a parent's questionnaire. All but one child had normal hearing; however, this child showed normal language abilities (see below) and was not excluded. Two 2-year-olds in the syntactic condition were born preterm (i.e., before Week 37 of pregnancy: Weeks 35 and 36); however, birth weight (>2500 g) and/or Apgar score (9/10) of these children were normal, as were their language abilities (see below) according to their unadjusted age, so their data were not discarded from the analysis. Children included in this study were German natives with a monolingual background, and their language abilities were normally developed or above average according to their age, as assessed with standardized language tests. That is, children did not significantly score below 1 standard deviation of the T norm (40–60; see Appendix A.2), at least when taking into account the 95% confidence intervals for the corresponding tests in the Language Development Scales for 3- to 5-year-olds (SETK 3–5; Grimm, Aktas, & Frevert, 2010) and 2-year-olds (SETK-2; Grimm, 2000) (see Appendix A.2). Detailed information about inclusion criteria as well as about those children who were excluded from this study can be found in Table I of Appendix A.1b.

Most of the 3- and 4-year-old children were recruited from local kindergartens and tested using a mobile eye-tracking laboratory set up in a van. Half of the children younger than 3 years were recruited with the support of the developmental psychology department and visited our stationary eye-tracking lab (see "Apparatus" section below) together with their parents. All children took part in two study sessions that took place on different days as close as possible (semantic: mean lag = 6 days, *SD* = 3, range = 1–14 days; syntactic: mean lag = 6 days, *SD* = 4, range = 1–18) and received a gift with a value of about 5 euros as compensation for their participation. Informed written consent was obtained from the parents prior to the participation of their children in the study. The study was conducted in accordance with the Declaration of Helsinki and was approved by the local ethics committee.

The adult control group consisted of 24 volunteers who participated for course credit or financial compensation (a subset of these adult data were used as cross-validation in Öhlschläger & Võ, 2017). Half of the adults were exposed to the semantic inconsistencies (semantic: mean age = 20.5 years, *SD* = 1.9, range = 18–25; 11 female), whereas the other half were presented with the syntactic inconsistencies (syntactic: mean age = 21.8 years, *SD* = 3.5, range = 19–31; 9 female). The adults visited our stationary eye-tracking lab once and underwent the identical procedure as the children (except for the language testing).

Explicit measures of scene knowledge: Dollhouse

Within the scene domain, many questions remain unanswered. Here, we focused on the comparability in the developmental trajectory of the explicit and implicit assessments of scene knowledge and their relatedness.

Procedure

As an explicit measure of scene knowledge, children and adults were asked to equip a wooden dollhouse (Nic Spiel + Art GmbH, Laupheim, Germany) with objects. The dollhouse contained four rooms with a size of 31 × 40 cm each on two floors (see Appendix A.3). The standardized start configuration comprised the following rooms and room-defining objects: bedroom with bed, kitchen with stove, bathroom with shower, and living room with sofa. The dollhouse also contained a children's bedroom created by subdividing a section of the bathroom with an intersection wall,[1] but this room was later ignored for analysis because no statistical information to infer about semantics or syntax for this room category was available (see "Analysis" section below) (Greene, 2013). After being introduced to each room and room-defining object, participants were asked to put the remaining 52 objects where they belonged in the dollhouse. Original instructions and a list of all objects used in the task are available in Appendices A.4 and A.5. We recorded each session on video and photographed the furnished dollhouse (see Appendix A.6).

Analysis

For the analysis, we focused on objects that were characterized as diagnostic and/or informative, more precisely, that were among the top 10 objects of this kind for the four basic categories—bedroom, kitchen, bathroom, and living room—based on the scene statistics from photo databases (Greene, 2013). Diagnosticity describes the probability that a scene is part of a category given that a certain object is present. Mutual information describes the dependency between scene category and the object, which can be given either because object presence is indicative for a scene category or because it is indicative against a scene category (e.g., hydrant not found indoors but rather found outdoors). The following 18 objects included in this study were defined as either diagnostic or informative by Greene (2013): *closet, nightstand, pillow, blanket, kitchen sink,* pot, dinner set, teapot, plate, *sink, toilet, bath rug, toilet paper holder, towel,* toothbrush, *armchair, side table,* and *plant.*[2]

Dollhouse semantics. As a measure of semantic scene knowledge, we calculated the proportion of objects correctly placed in their corresponding room. The chance level or level of guessing was 20%, but objects that could have been placed in the children's bedroom were ignored for analysis because this scene category was not included in Greene (2013) statistics. Objects that were diagnostic/informative in more than one room (e.g., blanket and pillow) were counted as correct in any of the corresponding rooms (e.g., bedroom and living room). We analyzed the proportion of correctly placed objects relative to the number of objects children had placed overall because the 2-year-olds had fewer objects available and not all children placed all the objects. The reader is referred to Table II of Appendix A.1b for further information on exclusion criteria and the number of objects placed by children.

Dollhouse syntax. As a measure of syntactic scene knowledge, we were interested in the inter-object relations. To do this, we first created a sketch of the furnished dollhouse (see Appendix A.7) for each

[1] The nonfixed intersection wall was minimally but visibly misplaced in 15 children in the semantic condition (1 2-year-old, 5 3-year-olds, and 9 4-year-olds) and in only 4 children in the syntactic condition (1 2-year-old, 2 3-year-olds, and 1 4-year-old) and 3 adults (semantics). But this rather randomly affected only one of four rooms.

[2] Note that due to the risk of swallowing, the 2-year olds were handed out only 29 large objects in total, and only the 13 diagnostic/informative objects are displayed in italic.

child using Adobe Photoshop CS (Adobe, San Jose, CA, USA). The sketch was read into MATLAB (The MathWorks, Natick, MA, USA) to define the location for each object by mouse clicking at the position corresponding to the object center. In a consecutive step, we calculated the Euclidian distance between predefined objects that served as anchors (e.g., shower, bathroom sink, toilet, sofa, bed, stove, kitchen sink) for other associated objects (e.g., bath rug, towel, toothbrush, toilet paper, armchair, side table, pillow, blanket, nightstand, pot, tea pot, plate). An object could appear in several pairs (see Appendix A.8). To avoid needing to calculate the distance across room borders, we restricted the distance to objects that were placed in the correct room. Note that a shorter distance indicates better syntax performance in the dollhouse task.

Implicit measure of scene knowledge: Eye-tracking experiment

Apparatus

Stimuli were presented on a 24-inch monitor with a resolution of 1920 × 1080 pixels and a refresh rate of 60 Hz. Scenes subtended a visual angle of about 19° horizontally and 15° vertically. The approximate viewing distance from the screen measured 80 cm. The 3- and 4-year-olds were seated in one of two versions of a child car chair with back support, whereas most of the 2-year-olds were seated in a special infant chair with a foot rest in order to reduce movements. Experimental presentation of stimuli was controlled with MATLAB Version 2013a using the Psychophysics Toolbox Version 3 (Brainard, 1997; Pelli, 1997). Eye movement recordings were monocular using an EyeLink desktop mount eye tracker (SR Research, Kanata, Ontario, Canada) at a sampling rate of 500 Hz in remote mode. Our stationary lab was equipped with a setup comparable to that of our mobile eye-tracking lab with respect to the landmarks mentioned above. Differences concerned the exact eye-tracker version (EyeLink 1000 Plus Version 5.04 instead of EyeLink 1000 Version 5.594) and the experimentation computer (running OSX instead of Windows XP). Only in the mobile eye-tracking lab were the children separated from the screen by a shielded window.

Stimuli

In total, 40 scenes were selected from the SCEGRAM database (Öhlschläger & Võ, 2017) in both the inconsistent and consistent conditions, which were identical between the syntactic and semantic manipulations except for 6 scenes. As can be seen in Fig. 1, SCEGRAM scenes are controlled for object familiarity; each object occurs once in a consistent scene and once in an inconsistent scene (e.g., toilet paper occurs in bathroom and in kitchen), and their consistency was rated by naïve observers. Areas of interest (AOIs) were identical in size and location for the inconsistent–semantic and consistent conditions and were identical in size for the inconsistent–syntax and consistent conditions. To avoid too conservatove eye-tracking thresholds for children, we added a buffer of 75 pixels online to each side of the AOI. Possible differences between conditions were not likely due to low-level salience. The mean saliency rank was calculated within 15 simulated fixations using the Saliency Toolbox (Walther & Koch, 2006) and did not vary between the consistent and inconsistent conditions [semantic: consistent—mean saliency rank = 5.70, SD = 4.96, inconsistent—mean saliency rank = 6.25, SD = 5.49, $t(39) = -0.49$, p = .624, $d = -0.09$; syntactic: consistent—mean saliency = 5.98, SD = 5.23, inconsistent—mean saliency = 7.35, SD = 5.69, $t(39) = -1.56$, p = .126, $d = -0.20$].

Counterbalancing

To ensure that each object was presented only once to each child, the paired scenes were presented in the same condition; for example, when the kitchen was presented in the consistent condition (e.g., with the cup), then the paired bathroom scene was also presented in the consistent condition (e.g., with the toilet paper) and vice versa. To take into account that object familiarity might change with age, always two children of a similar age (semantic: mean age difference = 17 days, SD = 13, range = 0–45; syntactic: mean age difference = 19 days, SD = 16, range = 0–48) were assigned to experimental lists that were identical despite the inverse assignment of consistent and inconsistent conditions.

Procedure

The identical procedure had been used in the cross-validation of Öhlschläger and Võ (2017). Prior to the start of the experiment, a 5-point calibration and validation was applied using an animated audiovisual fixation target. Drift checks and pauses were performed every 10 trials and could be administered and followed by a recalibration at any time if required during the experiment. The 40 experimental trials started after the participant had performed 2 practice trials. Fig. 2 illustrates the trial sequence, which was identical for both children and adults. A trial was initiated by the onset of an animated fixation spiral presented right or left randomly and equally often. The side of fixation spiral presentation was kept constant for the particular scene across conditions and participants. By looking at the fixation spiral for 0.5 s, the participant initiated the scene presentation. Counting from the first gaze sample detected on the scene, the scene remained visible for 7 s. A 10-s reward video was presented every two scenes on average. The order of the videos was randomly chosen and kept constant across all participants. The participant was instructed to simply view the scenes presented: "You are seeing some pictures, and sometimes Wolle [the sheep] is going to show you a movie!" The instruction did not refer to the inconsistencies.

Analysis

As our measure of eye movements, we focused on first-pass dwell time (FPDT) because it has been shown to reflect semantic influences on object processing (Henderson et al., 1999) and is reliably reported in adults in contrast to first fixation duration (Võ & Henderson, 2009). FPDT is calculated by summing up all fixation durations within the first entry into the critical object AOI until first exit. To investigate the effect of information collected beyond the first visit to the object AOI, the total dwell time (DT) was also calculated (see Appendix B.5). Fixation durations shorter than 100 ms were considered as artifacts (semantic: children 4%, adults 2%; syntactic: children 3%, adults 2%), as proposed by Wass, Smith, and Johnson (2013), and were excluded from the analysis. For trials without fixation on the critical object (first pass), dwell times were entered into the analysis as missing values (semantic: children 14%, adults 2%; syntactic: children 13%, adults 4%).

Standardized language assessments

Each child received detailed language assessments that were video-recorded (see Appendix A.2 for descriptive statistics on all language tests). We focused our analysis on the Concept Classification

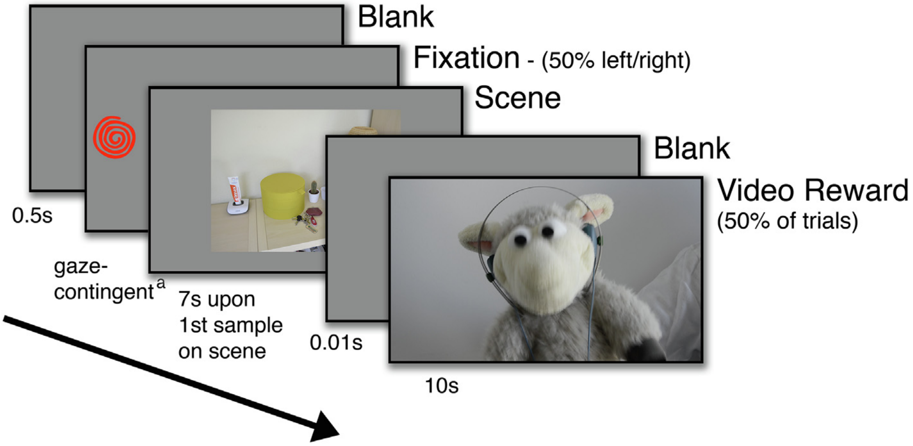

Fig. 2. Trial sequence of the gaze-contingent eye-tracking paradigm. The trial was initiated by the presentation of an animated fixation spiral. When looking at the spiral for 0.5 s, the child initiated the scene presentation. Starting from the first gaze sample detected on the scene, the scene remained visible for 7 s and was followed by a reward video in half of the trials. [a]Looking at spiral for 0.5 s.

subscale of the Patholinguistic Diagnostic Scale for Developmental Language Impairment (PDSS; Kauschke & Siegmüller, 2009). Children were asked to sort cards based on their belongingness to one of five categories (i.e. animals, toys, fruits, clothes, or tools). If there is a semantic–syntax distinction, we would expect concept classification as semantic language ability to specifically predict semantic scene knowledge. The test resulted in two scores by counting the targets correctly sorted IN as well the distractors correctly sorted OUT. These scores could then be transformed to norm values referenced to the corresponding population. This test was the only one identical for children of all age groups. Nevertheless, about half of the 2-year-olds (i.e., 14 of 24; semantic $n = 7$; syntactic $n = 7$) were not motivated or able to successfully complete the test (see Table III of Appendix A.1b for further information on exclusion and nonexclusion). No confidence intervals were reported for this subtest. To consider the full range of inter-individual differences, we did not exclude the five children scoring below a T score of 40, but we tested the effect of exclusion. As a predictor in our analysis, we focused on raw language scores instead of the norm values because we also wanted to include age as a continuous predictor. However, results were not systematically changed using the T norms (see Appendix B.8). As expected, both raw scores—the one for sorting OUT distractors and the one for sorting IN targets—increased with increasing age [OUT: ß = 0.12, $t(55) = 2.41$, $p = .019$; IN: ß = 0.10, $t(55) = 2.54$, $p = .014$] even though they were not replicated when looking at all three age groups [OUT: 3-year-olds–2-year-olds: ß = 1.63, $t(54) = 1.35$, $p = .18$; 4-year-olds–3-year-olds: ß = 1.43, $t(54) = 1.52$, $p = .134$; IN: 3-year-olds–2-year-olds: ß = 2.23, $t(54) = 2.35$, $p = .022$; 4-year-olds–3-year-olds: ß = 0.11, $t(54) = 0.16$, $p = .877$] (see Appendix A.9). We focused on the measure of differentiating the category against other categories (sorting OUT distractors) because the measure of defining a category broad enough to include all members (sorting IN targets) showed a ceiling effect by 3 years of age (see Appendix A.9) so that discriminability was reduced in the older children, who were of special interest to us based on our results on the developmental trajectory of scene knowledge (see below).

Statistical analysis

Single-trial data and analysis scripts can be found at http://sgl.uni-frankfurt.de/suppl/devel/.

We used linear mixed models (LMMs) with crossed random effects for participants and scenes to draw statistical inferences. Models were computed using the lmer program of the *lme4* package (Bates, Mächler, Bolker, & Walker, 2015) with estimates defined to optimize the REML (restricted maximum likelihood) criterion in the R environment for statistical computing and graphics (Version 3.2.4; R Development Core Team, 2012). The corresponding *p* values were determined using the *lmerTest* package (Kuznetsova, Brockhoff, & Christensen, 2017) with Satterthwaite approximations to degrees of freedom. LMMs allow including scenes and participants as random factors within a single analysis and handling unbalanced data (for detailed background, see Baayen, Davidson, & Bates, 2008, and Kliegl, Wei, Dambacher, Yan, & Zhou, 2011).

As fixed effects, we defined between- and within-participant factors *violation type* (semantic vs. syntactic) and *consistency* (consistent vs. inconsistent) and also included between-participant covariates *age* and *language* or *dollhouse scores* as well as their interactions. We included factors as sum contrasts (0.5 vs. −0.5) so that LMM estimates described the difference between the two factor levels each. The intercept represented the grand mean of log-transformed FPDT. Continuous covariates were entered after centering them at their mean. Data from children who did not participate in one measure (e.g., language test) were discarded from analysis prior to centering.

To arrive at a parsimonious random effects structure justified by the data, we used the rePCA program of the *RePsychLing* package (Bates, Kliegl, Vasishth, & Baayen, 2015). The maximum models were constrained in that covariates were entered as additive random effects instead of interactions. Details about the variance components removed from each model can be retrieved from the analysis scripts at http://sgl.uni-frankfurt.de/suppl/devel/.

We post hoc simulated the database power for the effects in LMMs using the *mixedPower* R package (Kumle, Võ, & Draschkow, 2018). The *mixedPower* R package uses the data and original LMM outcome to simulate new data of a sample of identical size as that used to fit the model parameters 1000 times. The power values obtained describe the proportion of how often a specific effect was found in 1000 simulated “runs” of the experiment. A sufficient power is expected at around .80. We also estimated the Bayesian version of the LMMs without specifying priors for fixed or random effects (setting = NULL) using the *blme* R package (Chung, Rabe-Hesketh, Dorie, Gelman, & Liu, 2013). Due to problems with convergence, we estimated the parameters for the intercept only instead of the best models. Results were comparable and can be found in Appendix B.6.

For visualization only, partial effects were removed using the *remef* R package (Hohenstein & Kliegl, 2014). More precisely, the fixed effects of violation type (and age), as well as the random effects due to participants and scenes, were statistically controlled. The FPDT adjusted for these effects was then plotted by consistency (and age) as predicted by language or dollhouse scores or age. All plots were created using the *ggplot2* R package (Wickham, 2009).

FPDT was log-transformed as recommended by the boxcox function of the *MASS* R package (Venables & Ripley, 2002) and by visual inspection of residual distributions. If not mentioned otherwise, the transformation did not significantly change the results.

To investigate the developmental trajectory of our participant-level covariates language and dollhouse scores and their relation, we calculated multiple linear regressions including age at test as a continuous or grouped (difference contrast) predictor.

Results

Developmental trajectory of scene knowledge

Explicit measure

As expected, explicit scene performance improved with increasing age within the children and from children to adults for both dollhouse semantics [children: ß = 0.02, t(86) = 8.22, p < .001; children–adults: ß = −0.36, t(86) = −9.96, p < .001] (see Fig. 3A) and syntax [children: ß = −0.39, t(84) = −7.58, p < .001; children–adults: ß = 4.73, t(84) = 4.91, p < .001] (see Fig. 3B). However, we observed different developmental milestones for dollhouse semantics and syntax; dollhouse semantics were particularly difficult to handle for 2-year-olds and did not reach the adult level. Performance improvements were observed from 2 to 3 years, ß = 0.36, t(85) = 8.38, p < .001, and from 4 years to adulthood, ß = 0.24, t(85) = 6.00, p < .001, but not from 3 to 4 years, ß = 0.04, t(85) = 0.97, p = .334.

Dollhouse syntax, however, showed a more gradual development and reached the adult level by 4 years of age. Performance improvements were observed from 2 to 3 years, ß = −6.98, t(83) = −5.27, p < .001, and from 3 to 4 years, ß = −3.06, t(83) = −2.63, p = .01, but not beyond 4 years, ß = −1.18, t(83) = −1.02, p = .313.

In sum, our explicit measure revealed that scene knowledge acquisition takes place between 2 and 4 years of age. Semantic performance was particularly bad in 2-year-olds; syntactic performance reached the adult level by 4 years, probably because the dollhouse task emphasized focusing on single objects (see Discussion).

Implicit measure

To investigate the developmental trajectory of implicit scene understanding, we used LMMs with crossed random effects for participants (N = 96: n = 72 children and n = 24 adults) and scenes (n = 46) (see Table 1).

Our main interest was to describe the developmental trajectory within the children: First, we observed that older children dwelled less on the consistent objects compared with the inconsistent objects during the first pass (see Fig. 4). This effect was comparable for semantic and syntactic violations, t = 0.27, p = .788 (see Table 1 and Fig. 5A). The significant Consistency × AgeChild interaction, t = −2.13, p = .033, is illustrated in Fig. 5B, showing the adjusted log-transformed FPDT predicted by age as of consistency with fixed effects of violation type and random effects partialled out. Post hoc power simulations revealed that this effect would be obtained in 55.7% of 1000 repetitions with samples of this size, with a sufficient power being defined as 80%; still, this effect is statistically significant already in this sample size.

The comparison with adults revealed additional information about the development beyond 4 years of age. The consistency effect was stronger in adults (i.e., adults dwelled even less on consistent

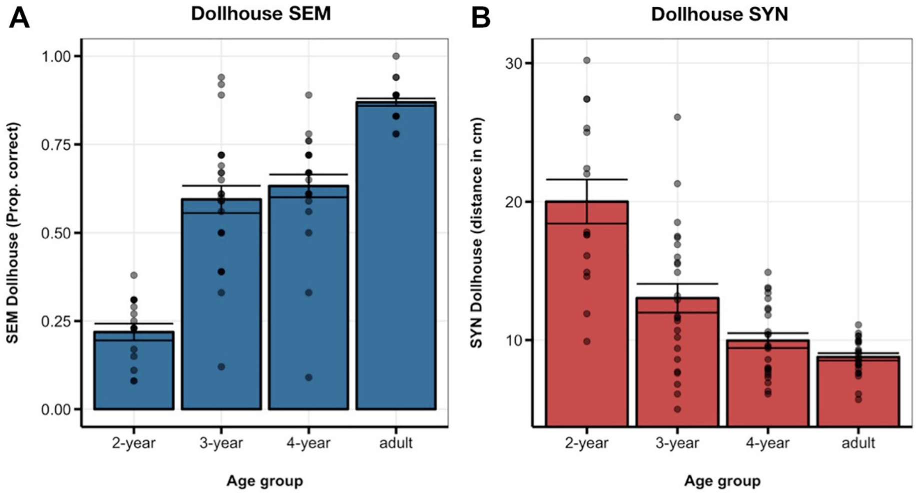

Fig. 3. Dollhouse scores as a function of age groups: 2-, 3-, and 4-year old children and adults. (A) Semantic (SEM) dollhouse: Proportions of objects correctly placed in the according room (larger values indicate better performance). (B) Syntactic (SYN) dollhouse: Distance between objects (smaller values indicate better performance). Error bars depict ± 1 standard error. Points display single-participant means.

Table 1
LMM of FPDT with age of children as continuous predictor and violation type, consistency, and age group as factors.

	LMM				
Fixed effects	Estimate	*SE*	*t* Value	*p* Value	Power
Mean log(FPDT)	6.29	0.0386	162.83	<.001***	
Violation Type	0.0183	0.0542	0.34	.736	.073
Consistency	−0.2214	0.0359	−6.18	<.001***	1
AgeChild (months)	−0.0053	0.002	−2.61	.01*	.738
Age Group	0.1966	0.0493	3.99	<.001***	.971
Violation Type × Consistency	0.0182	0.091	0.20	.842	.055
Violation Type × AgeChild	0.0003	0.004	0.06	.950	.049
Violation Type × Age Group	0.0206	0.0871	0.24	.814	.045
Consistency × AgeChild	−0.0058	0.0027	−2.14	.032*	.557
Consistency × Age Group	0.3047	0.0608	5.01	<.001***	1
Violation Type × Consistency × AgeChild	0.0015	0.0055	0.27	.784	.051
Violation Type × Consistency × Age Group	0.0354	0.1063	0.33	.739	.054

Note. Observations: n = 3401. Groups: participants (n = 96), scene (n = 46). Violation Type: semantic–syntactic; Consistency: consistent–inconsistent; AgeChild: continuous age of children (in months, centered), coded as zeros for adults; Age Group: child–adult. For a model containing the child data only, see Appendix B.7. LMM, linear mixed models; FPDT, first-pass dwell time.

* $p < .05$.

*** $p < .001$.

objects), suggesting the continuation of a developmental trend that yields a more quantitative difference than qualitative difference. Descriptively, the size of the average consistency effect in 4-year-olds was about 160 ms (semantic: 181 ms; syntactic: 137 ms), whereas the effect in adults was twice as big (semantic: 352 ms; syntactic: 231 ms) (see Fig. 4). To test whether this effect was due to a decrease in FPDT, we conducted LMMs separate for the consistent and inconsistent objects. Concordantly, we found that FPDT on consistent objects decreased with increasing age of the children, $ß$ = −0.0082, t = −3.76, p < .001 (see Appendix B.1a) and were even shorter in adults, $ß$ = 0.3468, t = 6.20,

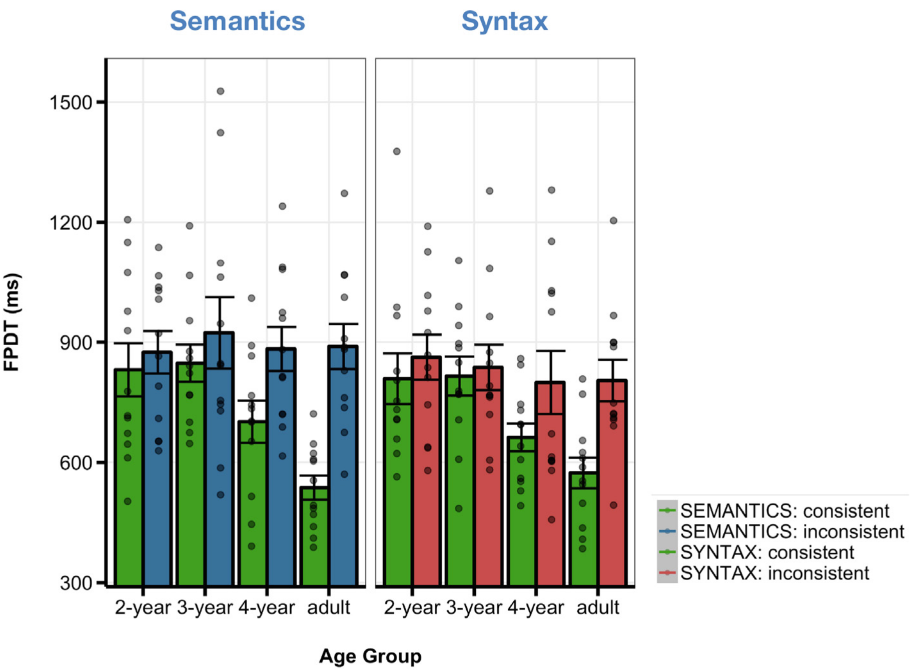

Fig. 4. Mean first-pass dwell time (FPDT) as a function of age group (2-, 3-, and 4-year-old children and adults) by consistency (green = consistent; blue = semantically inconsistent; red = syntactically inconsistent). Error bars depict ± 1 standard error. Points display single-participant means per condition. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

$p < .001$ (see Appendix B.1a), whereas no significant effects were observed for inconsistent objects, all *t*s < |1| (see Appendix B.1b).

In sum, the recorded eye movements revealed a trajectory in scene knowledge within 2- to 4-year-olds that was characterized by a reduction of processing times on the consistent objects rather than an increase of processing times on inconsistent objects. It shows a developmental trend that also seems to be continued beyond childhood, with first-pass dwell times to consistent objects being even more reduced in adults.

Explicit–implicit distinction

To investigate the role of our explicit measure in the prediction of implicit scene understanding, we used LMMs with crossed random effects for participants separate for dollhouse semantics (participants: $n = 65$ children; scenes: $n = 46$) (see Table 2) and dollhouse syntax (participants: $n = 63$ children; scenes: $n = 46$) (see Table 3). Again, the effects of dollhouse semantics, $t = 1.10$, $p = .272$ (see Fig. 6A and Table 2) and dollhouse syntax, $t = -0.65$, $p = .515$ (see Fig. 7A and Table 3) were not specific with respect to semantic or syntactic scene violations. Only dollhouse syntax predicted the consistency effect in eye movements; in line with the age-related eye-tracking results described above, the FPDT to consistent object decreased for children who were able to produce the inter-object relations correctly, $t = 2.01$, $p = .044$ (see Table 3). Fig. 7B illustrates the effects of the adjusted log-transformed FPDT predicted by inter-object distance as of consistency with fixed effects of violation type, age, and random effects partialled out. Post hoc power simulations revealed that this effect would be obtained in 46.7% of 1000 repetitions with samples of this size, with a sufficient power being defined as 80%; still, this effect is statistically significant already in this sample size. However, this

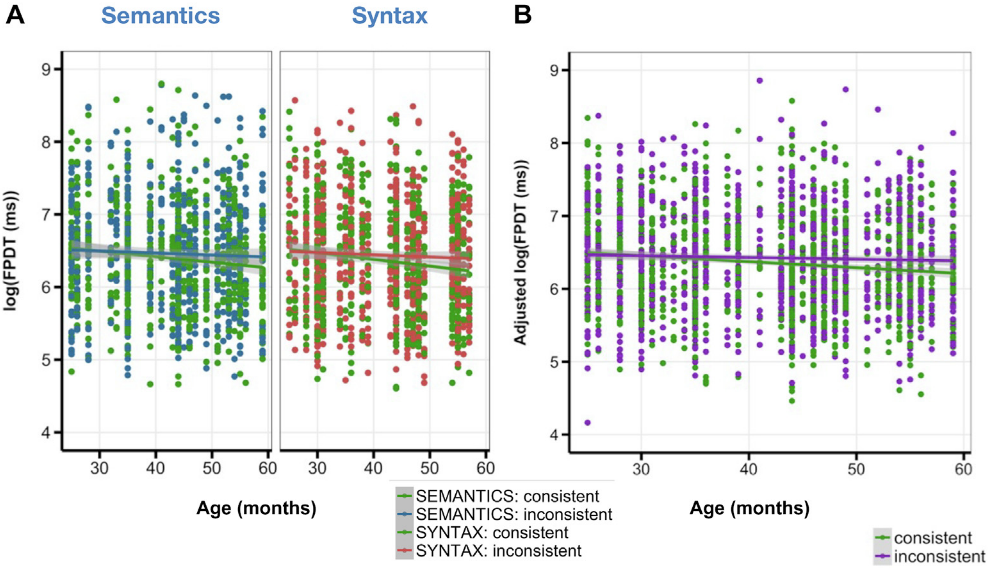

Fig. 5. Regression of log-transformed first-pass dwell time (FPDT) on continuous age within children. (A) Regression for observed FPDT by violation type (semantics vs. syntax) and consistency (consistent vs. inconsistent). (B) Regression with FPDT adjusted for violation type and its interactions as well as random effects based on the linear mixed models. Note that the inconsistent condition in purple does *not* differentiate between violation type (i.e., semantic and syntactic violations) but only represents the inconsistent condition in general. The Consistency × AgeChild interaction was significant, $t = -2.13$, $p = .033$. Shaded areas indicate 95% confidence intervals. Points show FPDT on single trials. (For interpretation of the reference to color in this figure legend, the reader is referred to the Web version of this article.)

effect was not invariant for the log-transformation and was not observable in untransformed data, $ß = 8.83$, $t = 1.12$, $p = .263$.

Dollhouse semantics did not predict the consistency effect in eye movements, $t = 0.44$, $p = .663$ (see Table 2 and Fig. 6B) but rather predicted increased FPDT in general, $t = 2.19$, $p = .033$. At least partly, the missing specificity of the effect in dollhouse semantics might be explained by a randomness introduced by the child bedroom that was later ignored for analysis (see Discussion).

In sum, the relation between explicit and implicit measures for scene knowledge was observed only for the syntactic dollhouse performance. Children who constructed realistic inter-object relations showed stronger consistency effect characterized by shorter FPDT on consistent objects compared with inconsistent objects independent of the type of scene violation. This relation, however, was not invariant for linear transformation.

Concerning the adults, we did not have a specific hypothesis about the relation between FPDT and dollhouse performance, but the interested reader is referred to Appendix B.2b.[3]

Testing possible links to language

Language and explicit measure of scene knowledge

We were interested in the role of language abilities of concept classification in predicting our explicit measures of scene knowledge: proportion correct and inter-object distance (see Fig. 8). Neither the main effect of language [semantic dollhouse: $ß = -0.004$, $t(52) < |1|$; syntactic dollhouse: $ß = -0.06$,

[3] Because the combined analysis for children and adults revealed four- and three-way interactions suggesting that effects of both dollhouse semantics, $ß = 3.21$, $t = 1.72$, $p = .085$, and dollhouse syntax, $ß = 0.1006$, $t = 1.87$, $p = .065$, on FPDT differed as a function of violation type between children and adults (see Appendices B.3a and B.3b), the analysis was conducted separately for both age groups.

Table 2
LMM of FPDT with semantic dollhouse scores (proportions correct) and age as continuous predictors and violation type and consistency as factors.

	LMM				
Fixed effects	Estimate	*SE*	*t* Value	*p* Value	Power
Mean log(FPDT)	6.39	0.0443	144.26	<.001***	
Violation Type	0.0242	0.068	0.36	.723	.169
Consistency	−0.075	0.0441	−1.70	.093°	.811
AgeChild (months)	−0.0098	0.0031	−3.18	.002**	.870
Semantic Dollhouse (proportion correct)	0.2834	0.1292	2.19	.033*	.589
Violation Type × Consistency	0.0178	0.1101	0.16	.872	.217
Violation Type × AgeChild	−0.0034	0.0061	−0.56	.581	.342
Consistency × AgeChild	−0.0067	0.0041	−1.65	.100	.502
Violation Type × Semantic Dollhouse	0.27	0.2587	1.04	.301	.381
Consistency × Semantic Dollhouse	0.0746	0.1712	0.44	.663	.077
AgeChild × Semantic Dollhouse	−0.0115	0.0126	−0.91	.369	.467
Violation Type × Consistency × AgeChild	−0.0024	0.0081	−0.30	.764	.157
Violation Type × Consistency × Semantic Dollhouse	0.3759	0.342	1.10	.272	.136
Violation Type × AgeChild × Semantic Dollhouse	−0.0002	0.0253	−0.01	.993	.230
Consistency × AgeChild × Semantic Dollhouse	0.0036	0.0167	0.21	.831	.130
Violation Type × Consistency × AgeChild × Semantic Dollhouse	0.0152	0.0334	0.46	.649	.227

Note. Observations: n = 2226. Groups: participants (n = 65), scene (n = 46). Violation Type: semantic–syntactic; Consistency: consistent–inconsistent; AgeChild: continuous age of children (in months, centered); Semantic Dollhouse scores: proportions correct (larger values indicate better performance, centered). LMM, linear mixed models; FPDT, first-pass dwell time.

$^{\circ}$ $p < .10$.
* $p < .05$.
** $p < .01$.
*** $p < .001$.

Table 3
LMM of FPDT with syntactic dollhouse scores (inter-object distances) and age as continuous predictors and violation type and consistency as factors.

	LMM				
Fixed effects	Estimate	*SE*	*t* Value	*p* Value	Power
Mean log(FPDT)	6.40	0.044	145.63	<.001***	
Violation Type	0.0336	0.0667	0.50	.616	.094
Consistency	−0.0358	0.0441	−0.81	.420	.107
AgeChild (m)	−0.0053	0.0031	−1.71	.094°	.365
Syntactic Dollhouse (distance in centimeters)	0.0011	0.006	0.18	.854	.051
Violation Type × Consistency	−0.0087	0.1109	−0.08	.938	.066
Violation Type × AgeChild	−0.0009	0.0062	−0.14	.890	.053
Consistency × AgeChild	−0.0009	0.0041	−0.22	.829	.048
Violation Type × Syntactic Dollhouse	−0.0006	0.012	−0.05	.958	.036
Consistency × Syntactic Dollhouse	0.0162	0.008	2.01	.044*	.450
AgeChild × Syntactic Dollhouse	0.0006	0.0005	1.21	.230	.200
Violation Type × Consistency × AgeChild	−0.0016	0.0083	−0.19	.846	.053
Violation Type × Consistency × Syntactic Dollhouse	−0.0105	0.0161	−0.65	.515	.084
Violation Type × AgeChild × Syntactic Dollhouse	0	0.0009	−0.02	.985	.053
Consistency × AgeChild × Syntactic Dollhouse	0.0009	0.0006	1.42	.156	.212
Violation Type × Consistency × AgeChild × Syntactic Dollhouse	−0.0013	0.0012	−1.04	.300	.130

Note. Observations: n = 2160. Groups: participants (n = 63), scene (n = 46). Violation Type: semantic–syntactic; Consistency: consistent–inconsistent; AgeChild: continuous age of children (in months, centered); Syntactic Dollhouse scores: continuous inter-object distances in centimeters (smaller values indicate better performance, centered). LMM, linear mixed models; FPDT, first-pass dwell time.

$^{\circ}$ $p < .10$.
* $p < .05$.
*** $p < .001$.

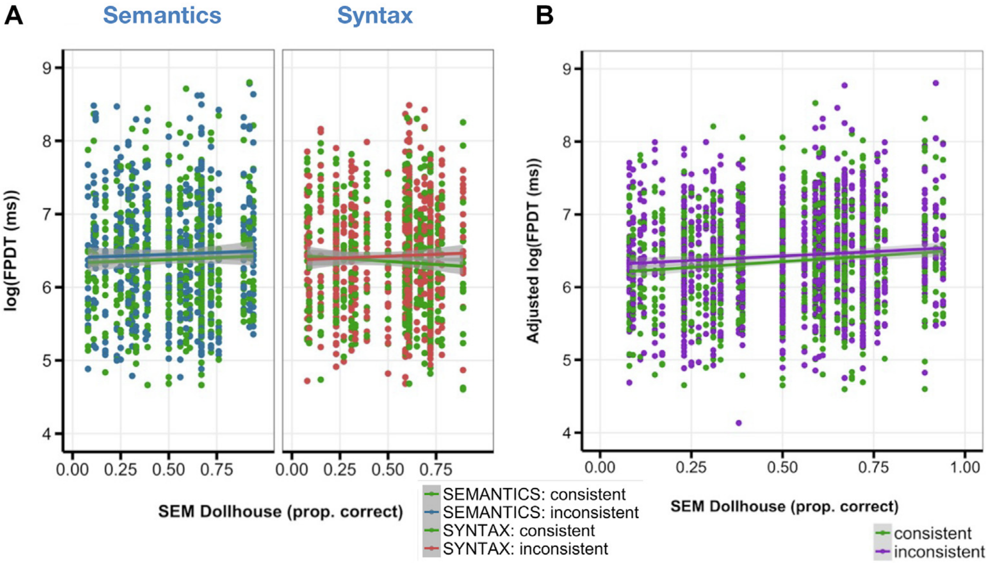

Fig. 6. Regression of log-transformed first-pass dwell time (FPDT) on semantic dollhouse scores (larger values indicate better performance) within children. (A) Regression for observed FPDT by violation type (semantics vs. syntax) and consistency (consistent vs. inconsistent). (B) Regression with FPDT adjusted for violation type and age and their interactions as well as random effects based on the linear mixed models. Note that the inconsistent condition in purple does *not* differentiate between violation type (i.e., semantic and syntactic violations) but only represents the inconsistent condition in general. The Consistency × Semantic Dollhouse interaction was not significant, t = 0.44, p = .663 (see Table 2). Shaded areas indicate 95% confidence intervals. Points show FPDT on single trials. (For interpretation of the reference to color in this figure legend, the reader is referred to the Web version of this article.)

$t(52) < |1|]$ nor the interaction with age [semantic dollhouse: ß = −0.001, $t(52) < |1|$; syntactic dollhouse: ß = −0.01, $t(52) < |1|]$ was significant. For dollhouse semantics, no trends of a systematic relation to the language score were observed. But for dollhouse syntax, in line with the results on the implicit measure (see below), only the 4-year-olds showed a numeric trend in the expected direction; the better the language, the more precisely the children produced the inter-object relations in the dollhouse, ß = −0.37, $t(21) = -1.80$, p = .086. This relation, however, was not observed when excluding the two 4-year-olds who performed below the range of what is considered as normal language performance, ß = −0.40, $t(19) = -1.33$, p = .201. To sum up, we did not observe a robust relation between language and our explicit measure of scene knowledge measure.

Language and implicit measure of scene knowledge

To investigate the role of language skills for concept classification in the prediction of implicit scene understanding, we used an LMM with crossed random effects for participants (n = 57 children) and scenes (n = 46) (see Table 4). In addition, in the relation to categorical semantic language, the consistency effect of eye movements was not modulated by the type of violation (semantics or syntax) with considering age, t = 0.17, p = .865, or without considering age, t = −0.56, p = .576 (see Table 4 and Fig. 9A). Fig. 9B shows the effects of the adjusted FPDT predicted by language skills as of consistency and age with fixed effects of violation type and random effects partialled out and provides the source for the Consistency × AgeChild × CC OUT (continuous language score of sorting out of children) interaction, t = −1.89, p = .059: Only in the 4-year-olds did the consistency effect increase with better language ability. This effect was again driven by a decrease in the FPDT to the consistent object. Post hoc power simulations revealed that this effect would be obtained in 33.1% of 1000 repetitions with samples of this size with a sufficient power being defined as 80%; still, this effect showed a trend already in this sample size. Note that this interaction was significant in the non-log-transformed model,

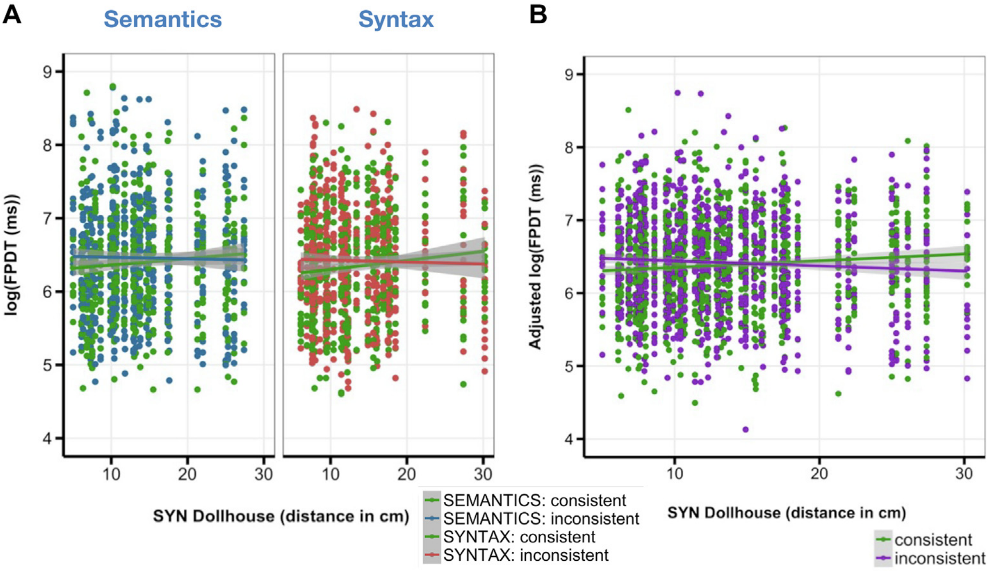

Fig. 7. Regression of log-transformed first-pass dwell time (FPDT) on syntactic (SYN) dollhouse scores (distance in centimeters; smaller values indicate better performance) within children. (A) Regression for observed FPDT by violation type (semantics vs. syntax) and consistency (consistent vs. inconsistent). (B) Regression with FPDT adjusted for violation type and age and their interactions as well as random effects based on the linear mixed models. Note that the inconsistent condition in purple does *not* differentiate between violation type (i.e., semantic and syntactic violations) but only represents the inconsistent condition in general. The Consistency × Syntactic Dollhouse interaction was significant, $t = -2.01$, $p = .044$ (see Table 3). Shaded areas indicate 95% confidence intervals. Points show FPDT on single trials. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

$ß = -2.02$, $t = -2.09$, $p = .037$. The large dropout of the youngest age group might have resulted in effect size values that likely underestimate the true effect size.

When excluding the five children who performed below the range of what is considered as normal language performance, this effect was still significant for both models on log-transformed data, $ß = -0.003$, $t = -2.37$, $p = .018$, and untransformed data, $ß = -3.38$, $t = -2.60$, $p = .009$. Overall, therefore, this implies that our effects seem rather underestimated than overestimated under actual conditions.

Taken together, our data do not seem to show a strong link between scene knowledge and language acquisition. However, in line with the developmental trajectory of implicit scene understanding by 4 years of age, children showed a relation between the efficiency of their language and scene processing independent of whether syntax or semantics were violated. Note that this relation should by no means be interpreted as causal; rather, it can only be taken as a first indication that there might be a relationship between both scene perception and language abilities during development.

Discussion

In this study, we set out to study the developmental trajectories of scene knowledge with respect to the semantic–syntax and explicit–implicit distinctions. Semantics and syntax in eye movements showed very comparable effects, and children with strong scene-based expectations in the implicit measure were good in constructing object locations relative to other objects in the explicit measure. Furthermore, we revealed important milestones in the development of scene knowledge between 3 and 4 years in both explicit and implicit measures.

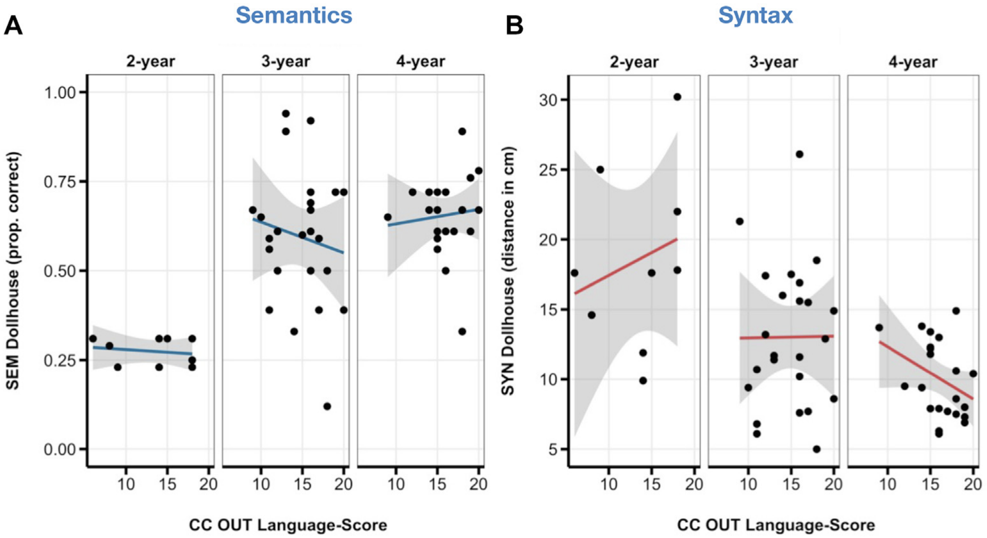

Fig. 8. Regression of dollhouse on language scores (Concept Classification [CC] of sorting OUT; larger values indicate better performance) calculated separately for children aged 2, 3, and 4 years. (A) Regression for observed scores of the proportion of objects correctly placed in the appropriate room (larger values indicate better performance). (B) Regression for observed scores of the inter-object distance (smaller values indicate better performance). Shaded areas indicate 95% confidence intervals. Points indicate single-participant means. SEM, semantic; SYN, syntactic.

Table 4
LMM of FPDT with raw language scores and age as continuous predictors and violation type and consistency as factors.

	LMM				
Fixed effects	Estimate	*SE*	*t* Value	*p* Value	Power
Mean log(FPDT)	6.38	0.0433	147.4	<.001***	
Violation Type	0.0447	0.0654	0.68	.4974	.087
Consistency	−0.0664	0.037	−1.79	.0789°	.334
AgeChild (months)	−0.0044	0.0028	−1.58	.1209	.259
Language CC OUT (raw)	−0.0012	0.0082	−0.14	.8879	.064
Violation Type × Consistency	0.0491	0.1013	0.48	.6297	.077
Violation Type × AgeChild	0.0018	0.0056	0.32	.7516	.057
Consistency × AgeChild	−0.005	0.0037	−1.33	.1821	.186
Violation Type × Language CC OUT	−0.0221	0.0164	−1.35	.1837	.191
Consistency × Language CC OUT	−0.0061	0.0109	−0.56	.5755	.065
AgeChild × Language CC OUT	−0.0005	0.0007	−0.62	.5397	.089
Violation Type × Consistency × AgeChild	0.0065	0.0074	0.88	.3774	.112
Violation Type × Consistency × Language CC OUT	−0.0122	0.0217	−0.56	.5762	.061
Violation Type × AgeChild × Language CC OUT	−0.0021	0.0015	−1.42	.1630	.212
Consistency × AgeChild × Language CC OUT	−0.0019	0.001	−1.89	.0588°	.331
Violation Type × Consistency × AgeChild × Language CC OUT	0.0003	0.002	0.17	.8650	.041

Note. Observations: *n* = 1965. Groups: participants (*n* = 57), scene (*n* = 46). Violation Type: semantic–syntactic; Consistency: consistent–inconsitent; AgeChild: continuous age of children (in months, centered); CC OUT: continuous language score of sorting out of children (higher values indicate better performance, centered). LMM, linear mixed models; FPDT, first-pass dwell time.

° *p* < .10.

*** *p* < .001.

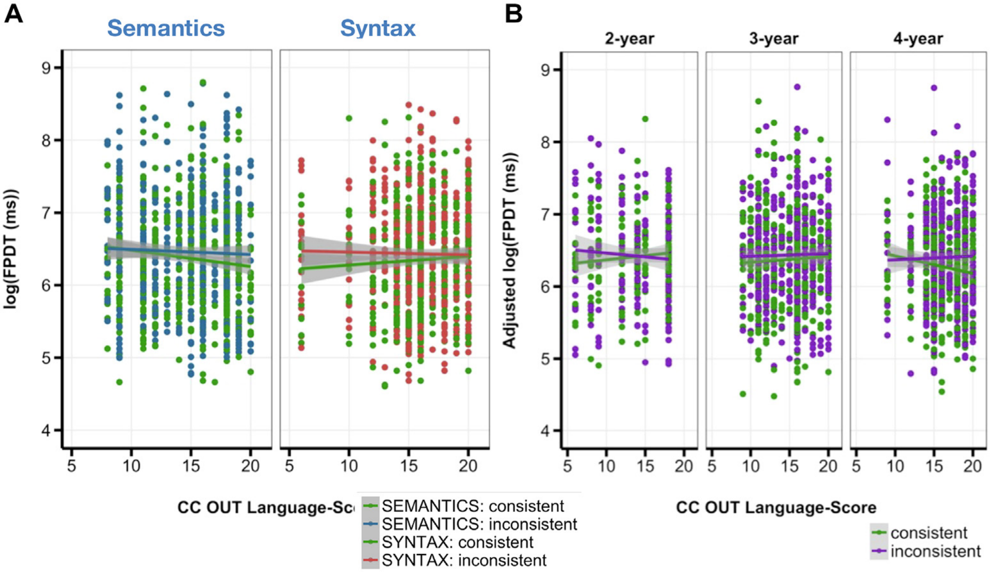

Fig. 9. Regression of log-transformed first-pass dwell time (FDPT) on language Concept Classification (CC) of sorting OUT (larger values indicate better performance) within children. (A) Regression for observed FPDT by violation type (semantics vs. syntax) and consistency (consistent vs. inconsistent). (B) Regression with FPDT adjusted for violation type as well as random effects based on the linear mixed models. Note that the inconsistent condition in purple does *not* differentiate between violation type (i.e., semantic and syntactic violations) but only represents the inconsistent condition in general. The Consistency × AgeChild × CC OUT interaction was not significant, $t = -1.89$, $p = .059$) (Table 4). Shaded areas indicate 95% confidence intervals. Points show FPDT on single trials. (For interpretation of the reference to color in this figure legend, the reader is referred to the Web version of this article.)

Developmental trajectory of scene knowledge

Explicit measure

Our results converge with the findings of other explicit scene knowledge tasks showing an important developmental step between 3 and 4 years for scene syntax (Saarnio, 1990, 1993b). In contrast to this previous work, we did not observe a similar milestone for semantic explicit scene knowledge with both 3- and 4-year-olds performing below adult level (~60%). However, the authors presented the scene categories (two or five) only in isolation, and children were asked for each object whether it fit in one single room or not, resulting in a guessing probability of 50% (Saarnio, 1990, 1993b). By contrast, in our dollhouse, the five room categories were present simultaneously and the children needed to decide for each object in which of the five rooms to put it, resulting in a guessing probability of only 20%.

The finding that children under 5 years of age were not performing at ceiling for putting objects in the correct rooms is also in line with a study that used a dollhouse with similar instructions (Freund et al., 1990). Interestingly, when asked to sort the object by furniture categories, 3-year-olds were even worse than when asked to sort by room categories (Freund et al., 1990), indicating that considering single objects is not enough but that also the spatial relations indeed played a role in our findings.

Implicit measure

We observed that in children by 4 years of age, FPDT to consistent objects was significantly reduced and that this developmental trend was even more pronounced in adults. In line with previous research (Mandler & Robinson, 1978), this qualitatively comparable trajectory relative to adults might indicate that already by 4 years children have formed schemata that can be used to process scene-consistent information similar to adults; without experience with the visual environment, all objects are novel and interesting independent of whether they match their context. During development, we have been repeatedly exposed to objects in their appropriate context so that we build up expectation for these consistent scene–object combinations. Here, we were able to provide evidence that the well-replicated consistency effects in adults for semantic violations (e.g., De Graef et al., 1990; Henderson et al., 1999; Öhlschläger & Võ, 2017; Võ & Henderson, 2009) and syntactic violations (De Graef et al., 1990; Öhlschläger & Võ, 2017; Võ & Henderson, 2009), in accordance with the prediction-based explanations, might actually be driven by a decrease of processing times to consistent objects rather than an increase of processing times to inconsistent objects.

The consistency effects cannot be accounted for by differences in bottom-up saliency, which was controlled for between inconsistent and consistent critical objects. Nevertheless, bottom-up features play a major role for ocular–motor guidance early in development until they become outweighed by more top-down processes (e.g., expectations, goals) with increasing age (Açık, Sarwary, Schultze-Kraft, Onat, & König, 2010; Helo et al., 2017). Furthermore, reflexive saccadic response latencies reach intra-individual stability at 2 and 4 years of age for high- and low-salient targets, respectively (Kooiker, van der Steen, & Pel, 2016). In a study manipulating object–scene inconsistencies in children as young as 24 months, semantic violations became relevant only for highly salient objects in a generally highly salient stimulus set (Helo et al., 2017). In contrast, the saliency for single critical objects of the SCEGRAM database (Öhlschläger & Võ, 2017) used in this study is more moderate (Underwood, Templeman, Lamming, & Foulsham, 2008), so that it might not aid in attracting the children's attention. To conclude, scene schemata can be activated and used to guide eye movements by 4 years of age at the latest. We collected evidence that scene schemata themselves might already be established earlier; that is, the consistency effect did not reveal a clear trajectory when considering all opportunities instead of only the first opportunity to visit the critical object, $ß = -0.0035$, $t = -1.39$, $p = .165$ (see Appendix B.4). Altogether, these observations indicate that before 4 years of age, to apply scene schemata and facilitate information processing, additional clues in terms of bottom-up guidance or additional time for encoding need to be provided.

Explicit–implicit distinction

As a special feature of this study, we tracked the acquisition of scene knowledge simultaneously in an explicit measure and in an implicit measure and were able to show an important developmental milestone from 3 to 4 years. By linking explicit and implicit knowledge, we found a positive relation between the syntactic dollhouse measure and the consistency effect in FPDT that was again characterized by building up expectations for objects in familiar contexts. Consequently, it seems that age-related scene knowledge can be successfully and convergently captured by both tasks. However, we did not observe a similar milestone for semantic explicit scene knowledge, indicating that explicit and implicit measures might not be entirely comparable.

This divergence might be due to the fact that the implicit measure did not need an instruction, whereas for the explicit measure this was of essential importance, especially for semantic performance. To be able to put an object into the correct room, the children need to have understood which room is which. It is possible that the marking of five different rooms was overwhelming, especially for the youngest children. Furthermore, also in previous studies the definition of when children possess "core" semantic knowledge seemed to be extremely dependent on the specific task instruction (Freund et al., 1990; Ratner, 1984; Ratner & Myers, 1981).

Alternatively, the difference between the explicit and implicit measures could be explained by the fact that the dollhouse is lacking gist information, which is of particular importance for semantic processing. A scene's gist refers to the general semantic information, including the scene's basic-level category as well as the corresponding statistical features such as color, texture, and the global spatial layout that can be extracted during a very short first glimpse of the scene (Oliva & Torralba, 2006; Wolfe, Võ, Evans, & Greene, 2011). In our dollhouse, the category information is marked on the object level and, therefore, objects are represented individually, that is, by the selective pathway only. By contrast, in the scene photographs used in our implicit eye-tracking paradigm, the gist information also allows for global processing via the nonselective processing pathway that might be integrated with the selective object information (Wolfe et al., 2011). Given these differences between our explicit and implicit measures, it seems even more astonishing that they were showing similar milestones and relations.

Semantic–syntax distinction

Based on previous literature on explicit scene knowledge, one could have expected an earlier development of semantics than of syntax (Saarnio, 1990, 1993b). At least in our implicit measure, for which we could directly compare gaze durations, we clearly did not observe a qualitative or quantitative differentiation either in the developmental trajectory or in the relation to the semantic language measure. These results have also been confirmed in Bayesian analysis backing up the null effects against a mere lack of power (see Appendix B.6). The lack of behavioral differences between scene semantics and syntax might have at least three reasons. First, our eye movement measure was not sensitive enough. Second, schemata of children from 2 to 4 years have no differential conceptualization of semantics and syntax in scenes; that is, weird is weird no matter whether the global context is wrong for an object or its location. Finally, it might well be that there is no categorical distinction between semantic and syntactic processing in scenes altogether (not in children and maybe not even in adults) and that instead the processing of object meanings and their spatial relations lies on a continuum. Evidence for the first argument is constituted by the fact that also in adults the consistency effect did not differ in magnitude between semantics and syntax, $ß = -0.0134$, $t = -0.11$, $p = .913$ (see Appendix B.5) and that previous results have been mixed (De Graef et al., 1990; Võ & Henderson, 2009). Evidence for the second argument—the sensitivity argument—comes from another study reporting that in children from 3 to 6 years of age semantic and syntactic scene knowledge in various explicit measures (free vs. constrained) were strongly cross-correlated and shared predictive variance (Saarnio, 1993a).

Testing the third reason, at least in children, could involve studying the development of the semantic–syntax differentiation on the electrophysiological level—as it has been shown before in adults (Võ & Wolfe, 2013a)—which might be able to tease apart possible differences in semantic versus syntactic processing.

Testing possible links to language

It is clear that scene processing and language processing are not the same thing. Nevertheless, rules that govern one cognitive process might also be involved in governing another. Therefore, we also investigated possible links between the language and perception domains. Already at this young age, words are not isolated but rather become organized in the mental lexicon based on associative relatedness (e.g., dog–bone) but also on conceptual category relatedness (e.g., dog–cow) (Arias-Trejo & Plunkett, 2013). In this study, we focused on concept classification as an indicator of language abilities, which is also assumed to rely on the experience with real-world objects and, therefore, could represent a possible first common denominator between linguistic knowledge and knowledge in other domains (e.g., world knowledge) (Kauschke & Siegmüller, 2009). We found a positive link that, however, was not significant for either explicit or implicit scene knowledge. These results are corroborated by another study where in a hierarchical regression model nearly all predictive power of category knowledge was accounted for by the shared variance with scene knowledge (Saarnio, 1993a). This is why we interpret the results observed in this study as first indications of shared variance between scene knowledge and conceptual semantic language skills. Because being able to make strong claims about such a link would require large-scale investigations of many children with varying scene and language abilities across many age groups, our results can be taken as only a first step toward thoroughly testing a possible domain generality in future studies.

Limitations and outlook

Based on the lack of previous results that either implicitly or explicitly addressed scene semantics and syntax as well as the specific link between scene knowledge and language, we were not able to conduct an a priori power analysis. It is likely that the 72 children we were able to test (i.e., 36 per semantic vs. syntax condition) were still not enough to encounter sufficient variability in performance in both the scene and language domains. In addition, dropouts are quite common in developmental studies, particularly when using sensitive methods such as eye tracking. However, by making use of state-of-the art LMM analyses, we were able to take into account this imbalance in the design and also to conduct a post hoc power analysis.

Furthermore, it is important to note that we cannot assume a causal relation between language and scene grammar given that the effects we observed are only correlational. Because we did not find any specific effects, it is hard to attribute the effects to language in particular. We believe that the ability to understand both language and scenes might be two different outcomes of one underlying cognitive construct. However, further studies will need to dig deeper by testing more children, including more targeted language tests, and/or by using more fine-grained measures such as event-related potentials.

Acknowledgements

This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) — project number 222641018 — SFB/TRR 135, subproject C7 to MLV.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jecp.2019.104782.

References

Açık, A., Sarwary, A., Schultze-Kraft, R., Onat, S., & König, P. (2010). Developmental changes in natural viewing behavior: Bottom-up and top-down differences between children, young adults and older adults. *Frontiers in Psychology, 1*. https://doi.org/10.3389/fpsyg.2010.00207.

Arias-Trejo, N., & Plunkett, K. (2013). What's in a link: Associative and taxonomic priming effects in the infant lexicon. *Cognition, 128*, 214–227.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*, 390–412.

Bates, D. M., Kliegl, R., Vasishth, S., & Baayen, R. H. (2015). Parsimonious mixed models. Retrieved October 10, 2017, from https://arxiv.org/pdf/1506.04967.pdf.

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48.

Biederman, I. (1977). On processing information from a glance at a scene: Some implications for a syntax and semantics of visual processing. In UODIGS '76: Proceedings of the ACM/SIGGRAPH Workshop on User-Oriented Design of Interactive Graphics Systems (pp. 75–88). New York: ACM Press.

Biederman, I., Mezzanotte, R. J., & Rabinowitz, J. C. (1982). Scene perception: Detecting and judging objects undergoing relational violations. *Cognitive Psychology, 14*, 143–177.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*, 433–436.

Chung, Y., Rabe-Hesketh, S., Dorie, V., Gelman, A., & Liu, J. (2013). A nondegenerate penalized likelihood estimator for variance parameters in multilevel models. *Psychometrika, 78*, 685–709.

De Graef, P., Christiaens, D., & d'Ydewalle, G. R. (1990). Perceptual effects of scene context on object identification. *Psychological Research Psychologische Forschung, 52*, 317–329.

Dienes, Z., & Perner, J. (1999). A theory of implicit and explicit knowledge. *Behavioral and Brain Sciences, 22*, 735–808.

Freund, L. S., Baker, L., & Sonnenschein, S. (1990). Developmental changes in strategic approaches to classification. *Journal of Experimental Child Psychology, 49*, 343–362.

Greene, M. R. (2013). Statistics of high-level scene context. *Frontiers in Psychology, 4*. https://doi.org/10.3389/fpsyg.2013.00777.

Grimm, H. (2000). *Sprachentwicklungstest für zweijährige Kinder (SETK-2)*. Göttingen, Germany: Hogrefe.

Grimm, H., Aktas, M., & Frevert, S. (2010). *Sprachentwicklungstest für drei- bis fünfjährige Kinder (SETK 3–5)* (2nd ed.). Göttingen, Germany: Hogrefe.

Helo, A., Pannasch, S., Sirri, L., & Rämä, P. (2014). The maturation of eye movement behavior: Scene viewing characteristics in children and adults. *Vision Research, 103*, 83–91.

Helo, A., Rämä, P., Pannasch, S., & Meary, D. (2016). Eye movement patterns and visual attention during scene viewing in 3- to 12-month-olds. *Visual Neuroscience, 33*, E014.

Helo, A., van Ommen, S., Pannasch, S., Danteny-Dordoigne, L., & Rämä, P. (2017). Influence of semantic consistency and perceptual features on visual attention during scene viewing in toddlers. *Infant Behavior and Development, 49*, 248–266.

Henderson, J. M., Weeks, P. A., Jr., & Hollingworth, A. (1999). The effects of semantic consistency on eye movements during complex scene viewing. *Journal of Experimental Psychology: Human Perception and Performance, 25*, 210–228.

Hock, H. S., Romanski, L., Galie, A., & Williams, C. S. (1978). Real-world schemata and scene recognition in adults and children. *Memory & Cognition, 6*, 423–431.

Hohenstein, S., & Kliegl, R. (2014). Semantic preview benefit during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*, 166–190.

Kauschke, C., & Siegmüller, J. (2009). *Patholinguistische Diagnostik bei Sprachentwicklungsstörungen (PDSS)*. München, Germany: Elsevier.

Kliegl, R., Wei, P., Dambacher, M., Yan, M., & Zhou, X. (2011). Experimental effects and individual differences in linear mixed models: Estimating the relationship between spatial, object, and attraction effects in visual attention. *Frontiers in Psychology, 1*. https://doi.org/10.3389/fpsyg.2010.00238.

Kooiker, M. J. G., van der Steen, J., & Pel, J. J. M. (2016). Development of salience-driven and visually-guided eye movement responses. *Journal of Vision, 16*. https://doi.org/10.1167/16.5.18.

Kumle, L., Võ, M. L-H., & Draschkow, D. (2018). Mixedpower: A library for estimating simulation-based power for mixed models in R. Retrieved from https://zenodo.org/record/1341048#.XhE_Pmftz5o.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*(13). https://doi.org/10.18637/jss.v082.i13.

Mandler, J. M., & Robinson, C. A. (1978). Developmental changes in picture recognition. *Journal of Experimental Child Psychology, 26*, 122–136.

Nelson, K. (1974). Concept, word, and sentence: Interrelations in acquisition and development. *Psychological Review, 81*, 267–285.

Öhlschläger, S., & Võ, M. L.-H. (2017). SCEGRAM: An image database for semantic and syntactic inconsistencies in scenes. *Behavior Research Methods, 49*, 1780–1791.

Oliva, A., & Torralba, A. (2006). Building the gist of a scene: The role of global image features in recognition. In S. Martinez-Conde, S. Macknik, M. Martinez, J.-. M. Alonso, & P. Tse (Eds.). *Progress in brain research* (Vol. 155, pp. 23–36). Amsterdam: Elsevier.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10*, 437–442.

R Development Core Team (2012). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Ratner, H. H. (1984). Memory demands and the development of young children's memory. *Child Development, 55*, 2173–2191.

Ratner, H. H., & Myers, N. A. (1981). Long-term memory and retrieval at ages 2, 3, 4. *Journal of Experimental Child Psychology, 31*, 365–386.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*, 372–422.

Saarnio, D. A. (1990). Schematic knowledge and memory in young children. *International Journal of Behavioral Development, 13*, 431–446.

Saarnio, D. A. (1993a). Scene memory in young children. *Merrill-Palmer Quarterly, 39*, 196–212.

Saarnio, D. A. (1993b). Understanding aspects of pictures: The development of scene schemata in young children. *Journal of Genetic Psychology, 154*, 41–51.

Sinclair-de Zwart, H. (1973). Language acquisition and cognitive development. In T. E. Moore (Ed.), *Cognitive development and acquisition of language* (pp. 9–25). New York: Academic Press.

Underwood, G., Templeman, E., Lamming, L., & Foulsham, T. (2008). Is attention necessary for object identification? Evidence from eye movements during the inspection of real-world scenes. *Consciousness and Cognition, 17*, 159–170.

Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S* (4th ed.). New York: Springer.

Võ, M. L.-H., & Henderson, J. M. (2009). Does gravity matter? Effects of semantic and syntactic inconsistencies on the allocation of attention during scene perception. *Journal of Vision, 9*. https://doi.org/10.1167/9.3.24.

Võ, M. L.-H., & Wolfe, J. M. (2013a). Differential electrophysiological signatures of semantic and syntactic scene processing. *Psychological Science, 24*, 1816–1823.

Võ, M. L.-H., & Wolfe, J. M. (2013b). The interplay of episodic and semantic memory in guiding repeated search in scenes. *Cognition, 126*, 198–212.

Walther, D., & Koch, C. (2006). Modeling attention to salient proto-objects. *Neural Networks, 19*, 1395–1407.

Wass, S. V., Smith, T. J., & Johnson, M. H. (2013). Parsing eye-tracking data of variable quality to provide accurate fixation duration estimates in infants and adults. *Behavior Research Methods, 45*, 229–250.

Wickham, H. (2009). *ggplot2: Elegant graphics for data analysis*. New York: Springer.

Wolfe, J. M., Võ, M. L.-H., Evans, K. K., & Greene, M. R. (2011). Visual search in scenes involves selective and nonselective pathways. *Trends in Cognitive Sciences, 15*, 77–84.