Young children’s Spontaneous Comprehension of Various Symbol-Referent Relationships in the Graphic Domain

Gregor Kachel1, Daniel Haun2, & Manuel Bohn1

1 Leuphana University

2 Max-Planck-Institute for Evolutionary Anthropology

Author note

*Ethics, consent and conflict of interest*: This study confirms with recognized standards (e.g. the Declaration of Helsinki) and was approved by an internal ethics committee at the Max-Planck-Institute for Evolutionary Anthropology. Informed consent has been obtained from all participants. The authors declare no conflict of interest. *Scientific Integrity and Openness*: The data and code necessary to reproduce the analyses presented here are publicly accessible, as are the materials necessary to attempt to replicate the findings. Analyses were also pre-registered. Data, code, materials, and the preregistration for this research are available at the following URL XXX Repo XXX. *Acknowledgments*: We are thankful to Susanne Mauritz for her help in the organization of the study and to Valerie Jurgenson and Cynthia Pones for help with data collection. We would like to thank Anne Deiglmayr for hosting this project in her research group and for her continuous support. Finally, we are very thankful to all parents and children participating in the study. Gregor Kachel was supported by the German Research Foundation (Deutsche Forschungsgemeinschaft) under project number 429220405.

The authors made the following contributions. Gregor Kachel: Conceptualization, Funding Acquisition, Project Administration, Investigation, Methodology, Data Curation, Formal Analysis, Visualization, Writing - Original Draft Preparation, Writing - Review & Editing; Daniel Haun: Resources, Writing - Review & Editing; Manuel Bohn: Methodology, Software, Formal Analysis, Validation, Writing - Review & Editing, Supervision.

Correspondence concerning this article should be addressed to Gregor Kachel, Universitätsallee 1, C1.008a, 21335 Lüneburg. E-mail: gregor.kachel@leuphana.de

Abstract

One or two sentences providing a **basic introduction** to the field, comprehensible to a scientist in any discipline. Two to three sentences of **more detailed background**, comprehensible to scientists in related disciplines. One sentence clearly stating the **general problem** being addressed by this particular study. One sentence summarizing the main result (with the words “**here we show**” or their equivalent). Two or three sentences explaining what the **main result** reveals in direct comparison to what was thought to be the case previously, or how the main result adds to previous knowledge. One or two sentences to put the results into a more **general context**. Two or three sentences to provide a **broader perspective**, readily comprehensible to a scientist in any discipline. Abstract must be less then 120words

*Keywords:* Graphic Communication, Iconicity, Analogical Reasoning, Gestalt Principles, Pragmatics, Emerging Literacy, Symbolic Literacy, Symbol-Referent-Relationship

*Word count:* Child Development Max 40 pages // PNAS 1,500–2,000 words

Young children’s Spontaneous Comprehension of Various Symbol-Referent Relationships in the Graphic Domain

# Introduction

not ready yet…

[See googledocs for intro and discussion.](https://docs.google.com/document/d/1nLlkREwo1JXngqy9Dfyhx5kdV4ue9w0fRJ61O8gDNIc/edit?usp=sharing)

# General Methods

All three studies presented below share the same methods and analyses. For the convenience of the reader, common features of the procedure, participant recruiting and stimulus design are reported first before discussing the three studies respectively. All studies were preregistered online prior to data collection (cf. [Study 1](https://aspredicted.org/SJT_H7F), [Study 2](https://aspredicted.org/L2H_XC7), and [Study 3](https://aspredicted.org/DR4_B4B)).

## Data Collection and Setup

In order to continuously trace the development of symbolic competences across the preschool years, data collection aimed to test two children per month of age between the third and the seventh birthday for a total of 96 participants per study while balancing male and female participants. As children participated on the basis of availability and data were collected by several experimenter teams visiting different institutions in parallel, the resulting final samples slightly exceed this preregistered minimum sample size. The final sample approximates an equal distribution of male and female participants and aligns with conventions in the field in providing a minimum of 24 participants per study and year of age (cf. Appendix A figure 5). All participants were recruited in *MASKED FOR REVIEW*, a medium-sized middle-European city, and came from a predominantly white population of middle to high income families. They were contacted via a database of participants for child development studies to which their parents had voluntarily signed up. Children were tested in day- and after schoolcare for the most part, and occasionally in the lab or at home. The studies were reviewed and approved by an internal ethics committee at the *MASKED FOR REVIEW*. Data collection took place from June 2022 to February 2023.

## Setup and Procedure

During test sessions, a child and an experimenter sat down together to play a picture-book-style hiding game presented on a touch-screen laptop. Verbal instructions were played back by the experimental script. Experimenters supervised children during data collection and occasionally reacted with a fixed set of verbal prompts when children were not following the presentation (“Oh look, the game continues!”). Test sessions always took place in a quiet separate room. See figure 1 for an illustration of the setup.



*Figure* 1. *Setup.* Experimenters were sitting behind the children in order to not distract them while supervising data collection.

*Familiarisation.* Experimenters invited the participants to join a hiding game and initially instructed them to follow the narration of the story. First, the presentation introduced a cartoon monkey. This character then placed two cups on the bottom left and right side of the screen. After holding up a banana, one of the barriers was lifted, the banana was placed underneath one of the cups and the barrier was lowered. Children were now prompted to touch the hiding place and in doing so the barrier of their choice was lifted to reveal the banana again if they chose correctly. The experimental script played back prerecorded feedback upon children’s choice (“yes, great job!”; “No, that’s not it. Let’s try again!”) during the familiarization (cf. Appendix B, figure 7 A). In order to succeed during familiarization, children solely had to remember where the item went and touch this part of the screen a few seconds later. To ensure that children were familiar with the goal of the game and the touch interface, they first had to complete a set of four to eight familiarization trials with a success rate of 75%. In case a child did not reply correctly in three out of four trials, another four familiarization trials were provided. If the child was correct in six out of eight trials, she was included in the main sample. Children that did not succeed during familiarization were allowed to participate but their data was not submitted to analysis. These children are reported below as failing the familiarization phase.

*Test.* The main phase of the study commenced with announcing that the cartoon character had an idea for a new game. The narration conveyed that children were not allowed to see where the banana would be hidden, but that the monkey would help them find it. Hence, the cartoon character was established as a knowledgeable and benevolent partner in a cooperative coordination game. The hiding sequence was identical to the familiarization phase, however the placement of the banana was concealed by a barrier covering the lower half of the screen. After the hiding phase, the monkey then held up a piece of paper and a pencil. Pencil movement and a short scribble sound indicated that the monkey was drawing. Children were reminded that the monkey was going to help them. Finally, participants were prompted with the phrase “Where is the banana?” and the monkey’s drawing was placed in the center between the two barriers. The drawing served as a cue to guide children’s choice. In the most basic experimental condition in study one (*Representation*), each hiding place, for example, showed either a solid blue circle or square and the paper displayed a simple outline drawing of one the shapes. Upon making a choice by touching the hiding places, children received no feedback and there was no reveal animation. Rather, children’s choice was acknowledged with neutral feedback (“Ah, thank you”) leading over to the next trial (cf. Appendix B, figure 7 B).

Except for the geometric shapes displayed on the hiding places and the respective drawing, the experimental presentation was identical for all test trials. A single trial lasted roughly 20 to 60 seconds, depending on how swiftly children chose. Each study presented four different experimental conditions with four trials each in a blocked order for a maximum of 16 test trials. Test sessions lasted about 12 minutes in total. The order of conditions was counterbalanced across participants. Children occasionally wished to stop before completing all trials, resulting in minor deviations of the total number of trials per condition that were submitted to analysis. For an overview of the average number of trials participants received in each condition, see tables 2, 4 and 6 in Appendix C. In line with the preregistration, children had to complete a minimum of eight test trials to be included in analysis. Respective exclusions are reported separately for each study.

## Stimuli and Counterbalancing

The set of studies presented here regard communication as a means for solving coordination problems. In the most simple small-world scenario an utterance or symbol, such as a graphic display, provided by a helpful interlocutor should enable an addressee to shift attention to, or help decide for one out of two options that are relevant in a particular practical context and even in the absence of conventions (Wittgenstein, 2009). For the purpose of operationalization, the context in the studies presented below is provided by a game of hide and seek and the options are two hiding places that are distinct by means of the graphic displays they are marked with. The aim was to test at what age children become able to spontaneously use graphic displays employing various dimensions of symbol-referent relationships. For this, the graphic displays presented as referents were designed to saliently differ in one relevant dimension and to be as similar as possible with regard to other surface features. The referent, on the other hand, was a reduced and less straight-lined graphic display akin to a hand drawing that had something in common with one of the referents and is thereby referring to it while it remains distinct with regard to other surface features. In test trials, one of the possible referents serves as a target and the other as a distractor. For counterbalancing, a second referent was designed to refer to the other target, such that across participants the same referents serve equally often as targets and distractors. For an illustration of trial composition, see figure 8 in appendix B. For each of the conditions in the three studies below, four sets of stimuli were designed, consisting of two blue shapes serving as target or distractor, and two drawings that could serve as cues. Each condition covers a particular type of symbol-referent relationship via four stimulus versions with two variations each. For an example, consider figure 9 in appendix B. Panel (A) shows all stimuli for the *Representation* condition. The first column exemplifies a set of targets (a blue square and circle) and referents (outline drawings of a square and circle). During test, participants are presented with four test trials per condition, each composed of the shapes of a single column. During testing, a child sees each trial combination only once and with only one of the two possible cues. Across children, the position (left/right) of the referent, and the identity of the cue are counterbalanced.

## Data Handling and Analyses

In each test trial, participants were prompted to touch one of the two choice options. Choices were logged by the experimental script and directly coded as correct or incorrect. Exclusions of data were solely made on the level of participants with regard to the exclusion criteria reported above. The analyses modeled participants’ binary choices to predict the probability of children interpreting cues correctly and to model how this probability would change as a function of their age. Logistic Bayesian generalized linear mixed models (GLMM) fitted children’s responses (0/1) as a function of their age, the experimental condition and an interaction between trial and condition. Trial and sex were included as fixed effects to be controlled for. Trial number was added as a random slope within subject. To evaluate the relevance of age and condition for children’s performance, a full model was compared with a reduced model lacking the interaction of age and condition using Widely Applicable Information Criterion (WAIC) scores and weights (McElreath, 2018) as well as the difference in Expected Log Predictive Density (ELPD) via the function *loo\_compare*. Furthermore, model estimates were inspected for the different predictors including their 95% Credible Interval (CrI). In each study, the condition hypothesized as the most simple was set as the reference level within conditions to make interpretation of model estimates convenient. All Bayesian models used default priors and were run in Stan (<http://mc-stan.org/>) via the function *brm* of the package *brms* (Bürkner, 2017). To answer the main research question of when children as a group systematically make correct choices in any of the conditions outlined below, we use fitted models to predict the developmental trajectory (with 95% CrI) of group level performance drawn from values of the posterior predicted distribution via the function *fitted*. These trajectories and CrIs were plotted by age. The criterion for settling when children performed above chance was the point at which the lower bound of the 95% CrI for a particular trajectory did no longer overlap with a mid-line demarcating the 50% chance level. All analyses were preregistered prior to data collection. Analyses deviate from the preregistered analyses when comparing models via ELPD differences (Sivula, Magnusson, Matamoros, & Vehtari, 2020). This was simply not as common by the time of preregistration. For the convenience of the reader, we also provide conventional analyses binning participants according to their age in years. To test whether group-level performance was above chance in all experimental groups, two-tailed one-sample t-tests with a chance level set to .5 were computed and are accompanied by Cohen’s *d* as a standardized effect size for significance testing (cf. Appendix C Tables 2, 4 and 6).

# Study 1

Study 1 aimed to establish a baseline for children’s performance and for evaluating task demands by providing the most simple symbol-referent relationship possible, where the cue is a direct representation of the target (*Representation*). From this, three further conditions were derived that were also employing form or shape as a means for establishing reference but that were less iconic by either reducing the amount the amount of information provided (*Pars Pro Toto*), or the amount of similarity between symbol and referent (*Simple Form Analogy*, *Complex Form Analogy).* We hypothesized that as a group children will first succeed with *Representation*, then *Pars Pro Toto*, *Simple Form Analogy* and finally *Complex Form Analogy* (cf. [Preregistration](https://aspredicted.org/SJT_H7F)).

## Stimuli

To make the four conditions in study 1 as comparable as possible, they are all employing the same target stimuli (cf. figure 9, top rows in panels A-B; appendix B). In addition the two referents within a trial can be seen as the round and square equivalents of each other which makes their overall appearance even more similar and aids matching the surface they cover. For the condition *Representation,* the graphical cue is a direct representation, that is an outline drawing, of the referent. The second condition, *Pars Pro Toto,* refers to the targets by means of a part-whole relationship. This is still in principle representational but provides less information and may require children to complete the shapes according to gestalt principles. While this completion is easiest in the first column in panels A to B (ibid.) due to the canonical shapes (square, circle) it is less obvious in columns 2 to 4 (ibid.) with the less iconic shapes albeit they are vertically and horizontally symmetrical on purpose. For comparability, *Pars Pro Toto* uses the same graphical cues as *Representation* but cut in half at a horizontal mid-line. A Stimulus set of an individual trial either uses the top or bottom half, but both variations are counterbalanced across trials. Stimulus variations with a division at the vertical axis were avoided as such cues are likely to have been read as arrows by children of the age that were tested here (Kachel, Bohn, O’Madagain, & Haun, in prep.). Two further conditions aimed to abstract from the original representational symbol-referent relationship by providing graphical analogies in form. In both *Simple Form Analogy* and *Complex Form Analogy* the cue was an abstract line drawing being more round or rectangular, thereby referring to either the round or rectangular equivalent of the target shapes. As this has not been done before in developmental research, our aim was to provide two versions of form analogies both supporting children’s comprehension in distinct ways. In *Simple Form Analogy*, the cues are less dense and therefore more simple to grasp, whereas the more complex versions in *Complex Form Analogy* provide more information. Arguably either variation may support feature extraction. As before, the cues in both conditions are direct equivalents with either round or edgy drawing line progressions. For an overview of all stimuli in Study 1, see figure 9 in appendix B.

## Participants

A sample of 106 children (M = 59.18 months, SD = 13.58 months, range 36 - 83 months; 51 female) participated in Study 1. In addition, 22 children (11 female) were tested but excluded from analysis for not succeeding during familiarization (N = 13), for not completing at least eight out of 16 test trials (N = 1), or due to being fussy (N = 2). For 4 children, experimenters only learned during testing that they were not fluent enough in German to participate as their families had only recently migrated. Finally, 2 children had to be excluded due to technical issues. For a graphical overview of participants and exclusions across all three studies, see Appendix A figures 5 and 6.

## Analysis

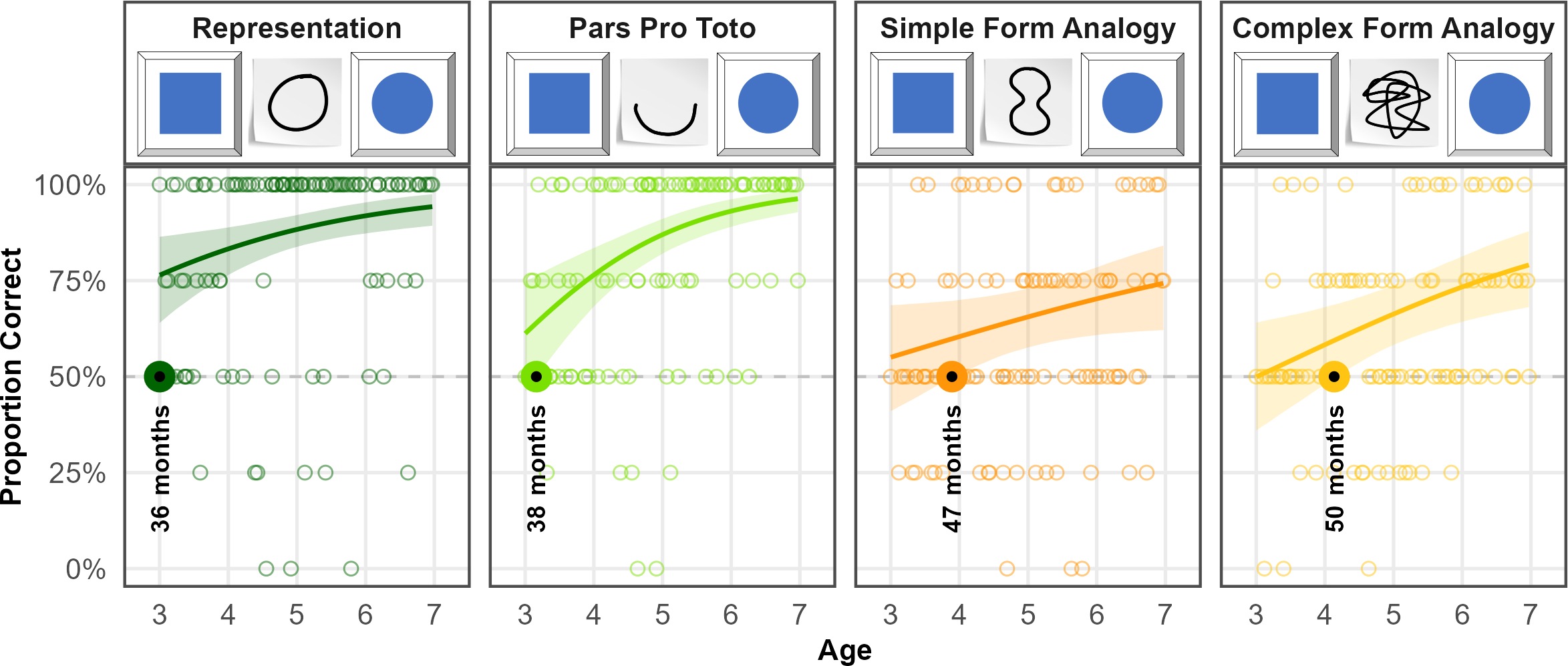
A total of 1688 trials (mean per condition = 422, range: 420 - 424) from 106 participants were submitted for analysis. The full model notation was correct ~ condition\*z.age + z.trial + sex + (z.trial | subid). In addition, a null model lacking the interaction of condition and age was fitted.

## Results

Posterior predictive checks (PPC) for both full and null model indicated excellent fit of observed data and model predictions (see supplement D for more information). Comparing the models using weights based on the Widely Applicable Information Criterion (WAIC) yielded 74.21% of the model weight for the full model, and 25.79% for the null model. Hence, the full model generally had a higher probability of making accurate predictions. Directly comparing the models’ WAIC via expected log predictive density (ELPD) corroborates this (ELPD WAIC; full model = -901.30; null model = -903.69). The standard error of the difference in predictive accuracy (SE = 3.14), however is lower than the difference itself (ELPD diff = -2.38). While the full model slightly exceeds in predictive power, evidence in favor of this model is not decisive. A similar comparison via Leave-One-Out Cross-Validation (LOO) provided essentially the same results. In absence of conclusive evidence for either model, we report the results for the full model in line with the preregistration.

Relative to the *Representation* condition, the *Simple Form Analogy* (beta = -1.39, 95% CrI [-1.75, -1.04) and *Complex Form Analogy* (beta = -1.36, 95% CrI [-1.73, -1.00) have a considerably lower probability of correct responses. The *Pars Pro Toto* condition has no clear difference from the reference condition (beta = -0.13, 95% CrI [-0.54, 0.28). Interaction terms between age and condition were not reliably different from zero. The developmental curves for each condition have essentially similar trajectories. Interaction effects with age were not relevant with the exception of *Pars Pro Toto*. Here, the interaction with age was positive and just above zero (beta = 0.31, 95% CrI [-0.05, 0.66), suggesting that performance increased more steeply across the age range than in the reference condition *Representation*. Generally, participants’ performance improved with age in all conditions (beta = 0.42, 95% CrI [0.13, 0.72). In contrast, trial number has no clear effect on performance (beta = -0.01, 95% CrI [-0.15, 0.12), suggesting no evidence for learning or fatigue throughout the test session.

Finally, by tracing when the lower bound of the 95% CrI exceeds the chance level of 50%, it is possible to report when children’s group level performance becomes robustly systematic in favor of the correct choice. In study 1, children perform above chance in the *Representation* condition at least as early as 36 months, which is the lower limit of the age-range. Quickly after, at 38 months, children succeed in the *Pars Pro Toto* condition. In the more abstract conditions *Simple Form Analogy* and *Complex Form Analogy*, preschoolers meet criterion at 47 and 50 months respectively. For a side-by-side comparison of the developmental trajectories in the four conditions of study 1, see figure 2. See appendix C table 1 for a full overview of coefficients for the full model, and table 2 for additional conventional analyses binning participants according to their age in years.



*Figure* 2. *Developmental Trajectories for all Conditions in Study 1.* Panels illustrate an example stimulus combination (distractor, cue, target) and results for the conditions. Coloured lines indicate smoothed mean performance by age. Shaded areas represent 95% CIs. The dashed line demarcates chance level and the dots represent individual means. The coloroured dots and annotation indicate when children’s performance exceeds chance level.

## Discussion

A main finding is that children succeed in *Representation* already and robustly at 36 months of age, which is the lower end of our age-range. Based on the literature, it is reasonable to assume that also in our setup, children would be able to solve the task at hand between the second and third birthday (Callaghan, 1999, 2000; DeLoache & Marzolf, 1992). In the context of the series of studies presented below, this result establishes that the task design and setup are sufficiently clear even for the youngest children in the sample. This is also partially ensured by the 75% criterion of correct choices employed in the familiarization phase. At the same time, the drop out rates due to problems with this familiarization criterion were comparably high at the lower end of the age-range (cf. Appendix A, 6). Testing children younger than three years of age would likely have required additional instructions or slower pacing of the presentation which counters comparability across ages. As it stands, the setup presented here worked perfectly despite providing the exact same task across the considerably large age-range of four years from the end of toddlerhood up into school age. The results for *Representation* further illustrate that already by three years of age children understand a representational symbol-referent relationship as operationalized here - with full-color target shapes and loose outline drawings as cues - in the concrete context of the of the picture-book-style object-choice-task at hand. That is, task demands and framing do not seem to tax children’s performance from three years onwards. Relatedly, it is quite interesting to observe that even within the most simple symbol-referent relationship, a small number of participants between four and six years of age struggled with the task or maybe simply did not care much for making correct choices (cf. dots representing individual performance in 2). This maybe seen as a baseline of failure that puts the numbers of children into perspective that are not succeeding in the more difficult conditions discussed below. Finally, we can observe that the CrI around the estimated mean in the graphic presentation is getting narrower, showing that - as children get older - they are behaving more uniformly. *Representation*, hence, provides also a baseline for the uncertainty to that can be expected with the operationalization we used. For all of these reasons, *Representation* provides a robust and clear conceptual canvass against which the developmental patterns of all other symbol-referent relationships can be discussed.

The first case in question then is *Pars Pro Toto*. Despite using identical targets and - even partially identical cues - as in *Representation* children succeed only at 40 months. The symbol-referent relationship in *Pars pro Toto* is still based in visual similarity but requires a slight inference, namely that a part can stand for a whole. The interaction with age shows that performance improves more quickly with age and that the area covered by the upper and lower CrI bounds shrinks considerably across the age range. *Pars Pro Toto*, hence, stands out as a clear example for a task that children come to master early and within the age-range tested here, that generally is not demanding for preschoolers in general. This contrasts with the less steep developmental trajectories observed in the other two conditions of study 1. *Simple Form Analogy* and *Complex form Analogy* share highly similar developmental trajectories and, when considering the entire age-range, appear equally difficult for participants. They appear as examples for symbol-referent relationships that are demanding for the oldest children in the sample, and may arguably require a second to solve or to describe verbally - even for adults. Establishing reference via an analogy in form is quite unconventional in graphic communication. Both conditions, here, stand out as conditions that are based first and foremost in visual similarity but are conceptually demanding at the same time. With regard to the influence of surface features, group level success occurs slightly earlier with the reduced cues employed in *Simple Form Analogy*. This suggests a slight benefit of a stimuli that are easier to grasp in line the literature on analogical reasoning (Gentner, 1988; Richland, Morrison, & Holyoak, 2006) . Finally, overall results are perfectly in line with the hypothesized developmental succession of group level success coming to pass first with *Representation*, then *Pars Pro Toto*, *Simple Form Analogy* and *Complex Form Analogy*.

# Study 2

The conditions in study 2 continue to abstract from a representational symbol-referent relationship by using different shapes for both cue and target stimuli. The conditions here, draw on gestalt-principles like figure-ground-relationship, continuity and parallelism. The symbol-referent-relationship in *Absolute Position* is established by the respective cue and target sharing the same position with regard to the reference frame they occur in. In *Relative Position* cue and target are each composed of two shapes. Reference is established by the respective shapes being closer or further apart or sharing a position on a horizontal or vertical axis. For *Orientation of Object*, symbol and cue are aligned along the same axis. The final condition, *Orientation of Feature* employs cue and target shapes with a salient feature that is either oriented up- or downward. The preregistered hypothesis, assumed children to perform above chance first in *Orientation of Object*, then *Orientation of Feature*, *Absolute Position* and finally *Relative Position*.

## Stimuli

For comparability, *Absolute Position* and *Relative Position* employed the same distinct shapes as cue and target respectively (circle vs. square, cross). In *Absolute Position* target positions were placed in top and bottom, or central and peripheral conditions in half of trails. For *Relative Position*, two trials present target stimuli that are closer together or further apart, as well as another two trials with target components being aligned on a horizontal or vertical axis. To ensure that cue and target are maximally distinct in all dimensions, cues and target items do not allign on the same axxis. As in study 1, cues and targets are either round and rectangular shapes in these conditions. To convey a sense of direction in *Orientation of Object,* cues and targets feature elongated shapes that are aligned either on a horizontal, vertical, or diagonal axis with rising or falling slope. The shapes are distinct oblong elipses and rectangles or abstract shapes with more or less round or square features. For *Orientation of Feature*, targets and cues are either circles or squares with a salient feature like an opening or a dent. While these shapes and features alternate for the related cue and target, they are aligned up- or downwards. As in study 1, cues in *Orientation of Feature* were oriented on a vertical axis, to counter interpretations as arrow-like cues (Kachel et al., in prep.). For an overview of all stimuli per condition in Study 2, see figure 10 in appendix B.

## Participants

A total of 99 three- to seven-year-old children (M = 60.04 months, SD = 13.69 months, range 36 - 83 months; 49 female) participated. In addition, a total of 13 children (6 female) were tested but excluded from analysis for failing familiarization (N = 9), being fussy (N = 1), not being fluent enough in German to follow the instructions (N = 1), or due to technical issues (N = 2).

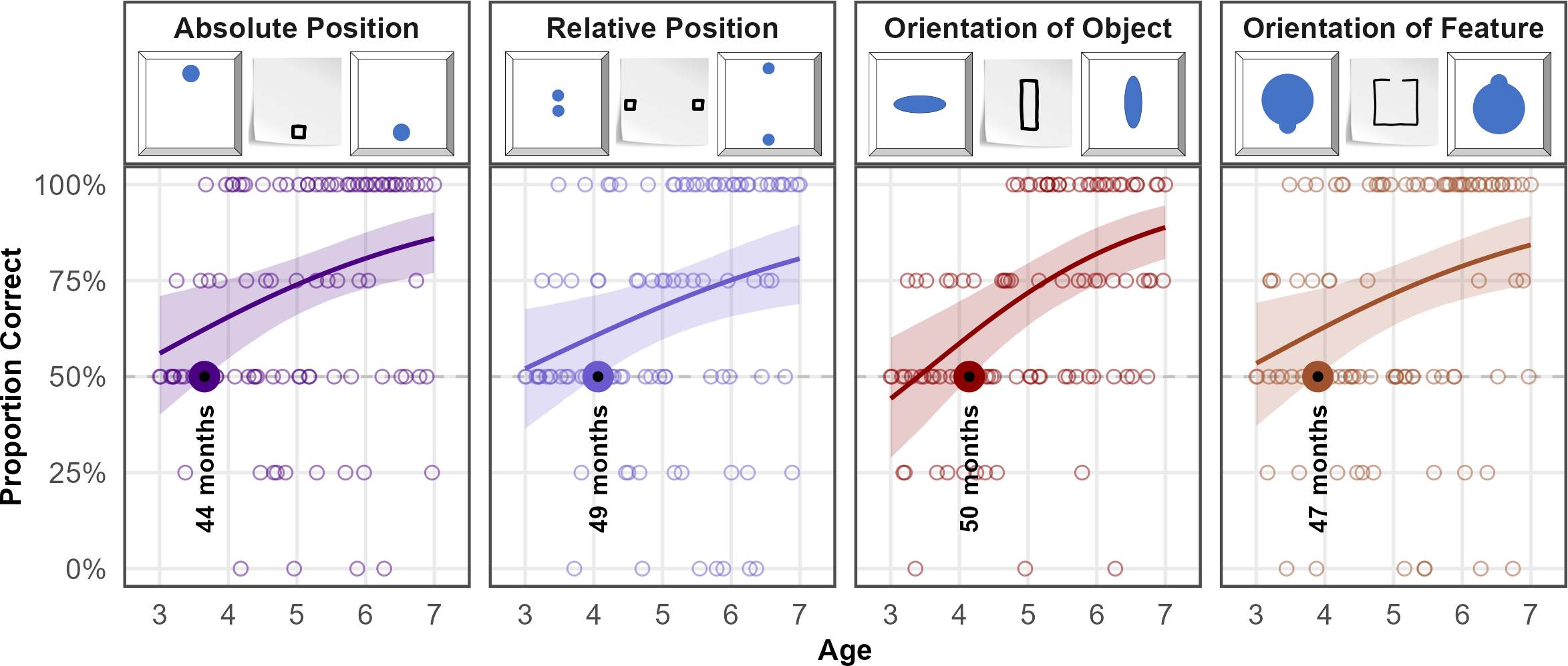
## Analysis

A total of 1561 trials (mean per condition = 390.25, range: 388 - 393) from 99 participants were submitted for analysis. The full model notation was correct ~ condition\*z.age + z.trial + sex + (z.trial | subid). In addition, a null model lacking the interaction of condition and age was fitted.

## Results

For both full and null model, PPCs indicate excellent fit of observed data and model predictions. To compare model performance, we evaluated the WAIC estimates. The null model showed a slightly better predictive performance (ELPD = -887.99) compared to the full model (ELPD = -889.73). WAIC values also indicate better performance of the null model (WAIC = 1,775.99) over the full model (WAIC = 1,779.45). However, the difference between models (ELPD Diff = -1.73) falls within the bounds of uncertainty (SE = 1.74), suggesting no advantage in predictive accuracy. Hence, the preregistered analyses using the full model is reported below.

Across all conditions, performance improved with both age (beta = 0.40, 95% CrI [0.14, 0.68) and slightly with trial number (beta = 0.26, 95% CrI [0.10, 0.43). Relative to *Absolute Position*, children were generally less likely to correctly solve *Relative Position* (beta = -0.28, 95% CrI [-0.64, 0.09). Performance in *Orientation of Object* (beta = -0.10, 95% CrI [-0.47, 0.28) and *Orientation of Feature* (beta = -0.12, 95% CrI [-0.49, 0.24) was not substantially different from the reference category when considering the full age range. Interaction terms between age and condition, including *Relative Position* and age (beta = -0.05), were not credibly different from zero, suggesting similar developmental patterns for all conditions. Tracing the lower bound of the 95% CrI against the 50% chance level (cf. figure 3), the model establishes that children master the condition *Absolute Position* at 44 months, making it the easiest task in study 2. Then in quick succession, children succeed in *Orientation of Feature* at 47 months, *Relative Position* at 49 months and *Orientation of Object* at 50 months. For a side-by-side comparison of the developmental trajectories, see figure 3. For an additional conventional analysis binning participants according to their age in years, see table 4 in appendix C. See appendix C table 3 and table 4 for an overview of model coefficients.



*Figure* 3. *Developmental Trajectories for all Conditions in Study 2.* Panels illustrate example stimulus combinations (distractor, cue, target) and results for all conditions. Coloured lines indicate smoothed mean performance by age. Shaded areas represent 95% CrIs. The dashed line demarcates chance level and the dots represent individual means. The coloured dots and annotation indicate when performance exceeds chance level.

## Discussion

We hypothesized that children as a group would perform above chance earlier with *Absolute Position* than with *Relative Position* due to the latter obviously requiring children to integrate a higher number of items. Results confirm this assumption with a clear 5-months offset in the age of success. We hypothesized further that children would succeed earlier with *Orientation of Object* than *Orientation of Feature*, due to the necessity of evaluating the composition of a target item rather than its general alignment with the cue. Here, we find the opposite to be true. The impression of direction or congruence between cue and target appears to be much more salient for children in graphic stimuli with a salient feature pointing out such as in *Orientation of Feature*, than an overall alignment in the horizontal, vertical or diagonal orientation. However, an overall picture emerges from the results of study 2. With stimuli that are not primarily drawing on representational symbol-referent relationships but that are based in more conceptual Gestalt principles, children generally come to perform above chance closely around the fourth birthday. Arguably, all four of these conditions still have an iconic aspect to them with regard to the overall gestalt of the displays. Study 3 set out to further reduce the amount of iconicity and investigate fully abstract symbol-referent relationships.

# Study 3

The final study features symbol-referent relationships that are based on analogies in size or number without cues and targets sharing similarities in other visual aspects, and while avoiding to draw on conventions in graphic communication. In order to correctly interpret the symbol-referent relationships employed here, children need to assign or extract conceptual order in graphical displays. Such processes can be fostered or obstructed by surface level features of the stimuli at hand which makes it harder to assess whether the age at which children master a task - our primary aim of investigation - depends more on the ability to generally form symbol-referent relationships from analogies or from children’s developing ability to process visual complexity. The two base conditions of study three are *Size of Object* and *Size of Number.* Here the cues refer to a target shape as a whole via an analogy in size or number. They are complemented by the conditions *Size of Feature* and *Number of Feature* in which the cues refer to a salient aspect of the target stimuli. In these cases, extracting conceptual information is arguably more demanding. The overall performance across the ages as well as the relative offset in the age at which children solve a task based on symbol-referent relationships targeting objects or their features, can then serve to evaluate such surface-level effects across two different domains. Prior to data collection, we hypothesized that children will succeed earlier with symbol-referent relationships based in size than in number, and that children will succeed earlier when cues refer to a target object per se rather than a salient feature thereof.

## Stimuli

In *Size of Object* the targets in a each trial are identical shapes that are either small or large with regard to the reference frame they are presented in. To make cue and target shapes as distinct as possible they are again employing either squares or circles respectively or fully abstract shapes such as a random scribble or straight lines (cf. figure 11, A). For comparability, *Size of Feature* employs the exact same cues as *Size of Object*, but features target shapes with either a relatively large or small void, opening, or protrusion. In *Number of Object*, cues and targets are composed of different shapes with some being simple line drawings. The displayed quantities were one against three, and two against four. As such they are easy to grasp and distinguish even for young children. Special attention was paid to the arrangement of the objects to ensure that the cue and target objects do not share visually similarity by forming a similar pattern, which is difficult as such small number arrays lend themselves to being grouped into canonical shapes by Gestalt principles such as proximity and closure. This issue was addressed by presenting the targets depicting the magnitudes three and four as if outlining irregular shapes. By contrast, the corresponding cues were aligned along a vertical of horizontal axis. For comparability, *Number of Feature* employs the same cues as *Number of Object*. To evoke a sense of quantity in the closed forms serving as referents, the target shapes in *Number of Feature* either have salient protrusions, or partial areas resulting from incisions. For an overview of all stimuli presented in Study 3, see figure 11 in appendix B.

## Participants

A total of 99 three- to seven-year-old children (M = 59.88 months, SD = 13.44 months, range 36 - 83 months; 55 female) participated. In addition, 23 children (7 female) were tested but excluded for low performance during familiarization (N = 12), for not completing at least eight out of 16 test trials (N = 1), or being fussy (N = 3). Further exclusions were necessary due to language problems (N = 4) and technical issues (N = 3).

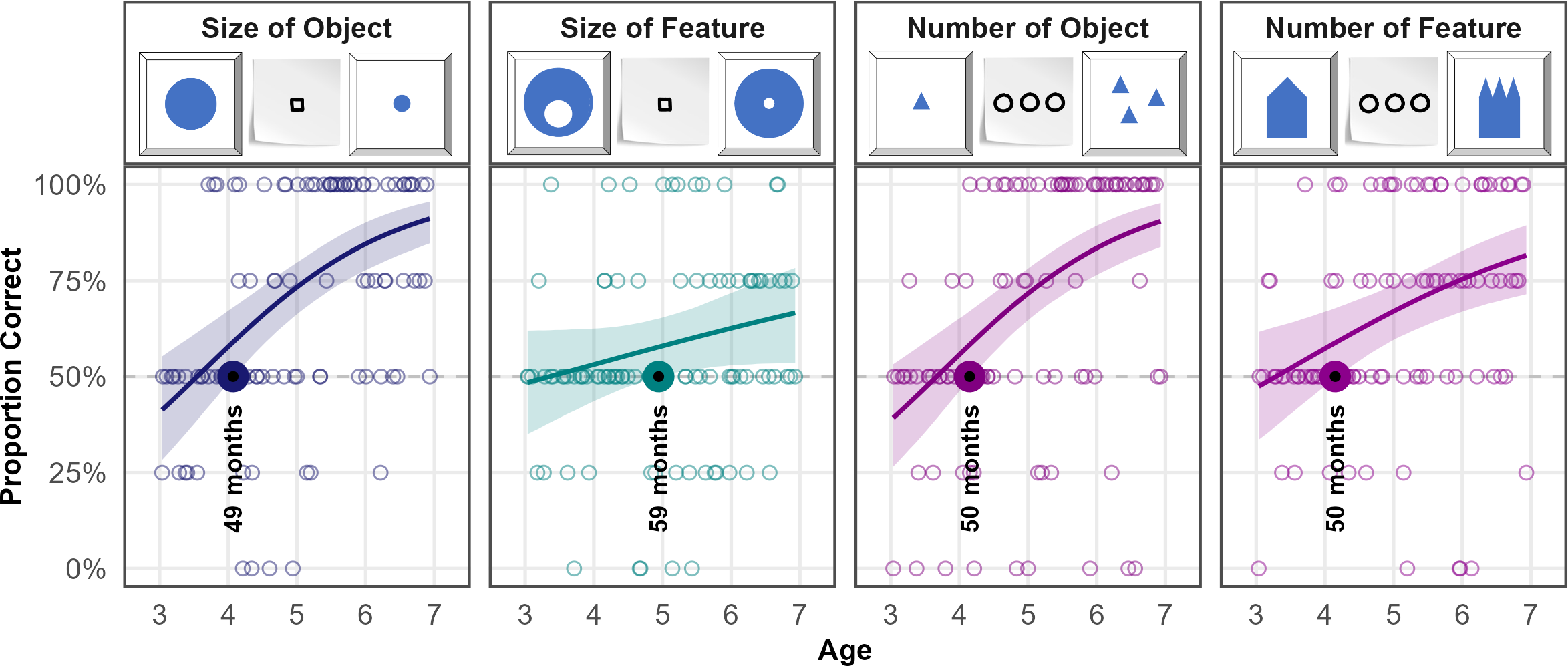
## Analysis

For study three, 1559 trials (mean per condition = 389.75, range: 388 - 392) from 99 participants were submitted for analysis. Data were analyzed both with a full model (correct ~ condition\*z.age + z.trial + sex + (z.trial | subid)) and a null model lacking the interaction of condition and age.

## Results

PPCs indicated excellent fit of observed data and model predictions in both models. When comparing performance, the full model showed a better fit (ELPD = -941.74) relative to the null model (ELPD = -946.09). The WAIC values also favored the full model (WAIC = 1,883.49) over the null model (WAIC = 1,892.19). Despite the slightly lower WAIC and higher ELPD of the full model, the difference in predictive accuracy (ELPD Diff = -4.35) remains almost within the range of sampling uncertainty (SE = 3.96). In the absence of substantial differences, the full model is reported below in line with the preregistration.

Overall, children’s performance increased with age (beta = 0.70, 95% CrI [0.44, 0.97) and with trial number (beta = 0.21, 95% CrI [0.07, 0.36), indicating general improvement across development and time-on-task. Relative to *Size of Object*, participants were substantially less accurate in *Size of Feature* (beta = -0.71, 95% CrI [-1.07, -0.37). A smaller, marginal effect was observed in *Number of Feature* (beta = -0.31, 95% CrI [-0.67, 0.03). *Number of Object* (beta = -0.08, 95% CrI [-0.44, 0.27) did not differ reliably from *Size of Object*. Age moderated performance less strongly in *Size of Feature* (beta = -0.50) and *Number of Feature* (beta = -0.28), suggesting lower developmental gains compared to *Size of Object*. Generally, the conditions relying on feature-based reference are associated with lower overall performance and weaker developmental gains. The best overview of the relative performance across conditions is provided by plotting the model estimates (cf. figure 4). Children succeed in most conditions just after the fourth birthday. Model estimates indicate group level success in *Size of Object* at 49 months, *Number of Object* at 50 months, and *Number of Feature* at 50 months. The exception to this pattern is *Size of Feature* where children master the task no sooner than 59 months of age. For an overview of model coefficients see table table 5, appendix C. For alternative analyses binning children by year of age, please see table 6 (ibid).



*Figure* 4. *Developmental Trajectories for all Conditions in Study 3.* Panels illustrate example stimulus combinations (distractor, cue, target) and results. Coloured lines indicate smoothed mean performance by age. Shaded areas represent 95% CrIs. The dashed line demarcates chance level and the dots represent individual means. The coloured dots and annotation indicate when performance exceeds chance level.

## Discussion

We hypothesized that children will succeed earlier with analogies in size than in number, and that children will succeed earlier when cues refer to the target objects per se rather than salient features thereof. We find limited support for the hypothesis that the feature conditions are generally more demanding. However, that children succeed a month earlier in *Number of Object* than in *Size of Object* as hypothesized is negligible especially in the context of the four year age-range that is considered here. Taken together the results may better be interpreted as further indicating that children are solving analogy-based symbol-referent relationships just after the fourth birthday, as if - provided that a critical level of reasoning development is attained - children can generally dismiss surface level features and focus on conceptional dimensions (Gentner, 1988; Richland et al., 2006) and flexibly employ their analogical reasoning skills in communicative contexts regardless of the specific conceptual dimensions such as number or size.

The late success at 59 months in *Size of Feature* is slightly ambiguous. *Size of Feature* stands out as the most difficult condition in the series of studies presented here and is definitely quite demanding as it is a highly conceptual symbol-referent relationship and the only condition that requires children to consider proportions within the composition of the target stimulus. However, given that the developmental trajectory is quite flat and the CrI comparably wide, the operationalization or graphical implementation of the concept *Size of Feature* may simply not have been ideal and salient enough. *Size of Object* demonstrates that children generally can grasp form analogies earlier with the exact same cues and *Absolute Position* and *Relative Position* from study 2 indicate that children can also master figure-ground relationships or relational patterns in targets even prior to the fourth birthday. It is reasonable to assume that children could succeed slightly earlier in *Size of Feature* if the opening or bumps representing the size relationship in the target displays would simply be more pronounced.

# Additional Analyses

??? Ich würde erstmal alles weitere fertig machen und die Additional Analyses optional lassen ???

Preregistration: \*An additional exploratory analysis will include a random effect for item level effects (Model: correct ~ task*z.age +z.trial +z.sex +(z.trial|id) +(z.age|item)). Results will help to evaluate the equivalence of items within a task and be reported in the supplements. Due to the low number of individual items within a task we expect this model to be less diagnostic with regard to our main research question and, therefore, will not include the term in the main analysis.*

# Discussion

not ready yet…

[See googledocs for intro and discussion.](https://docs.google.com/document/d/1nLlkREwo1JXngqy9Dfyhx5kdV4ue9w0fRJ61O8gDNIc/edit?usp=sharing)

# References

Bürkner, P.-C. (2017). Brms: An r package for bayesian multilevel models using stan. *Journal of Statistical Software*, *80*, 1–28.

Callaghan, T. C. (1999). Early understanding and production of graphic symbols. *Child Development*, *70*(6), 1314–1324.

Callaghan, T. C. (2000). Factors affecting children’s graphic symbol use in the third year: Language, similarity, and iconicity. *Cognitive Development*, *15*(2), 185–214.

DeLoache, J. S., & Marzolf, D. P. (1992). When a picture is not worth a thousand words: Young children’s understanding of pictures and models. *Cognitive Development*, *7*(3), 317–329.

Gentner, D. (1988). Metaphor as structure mapping: The relational shift. *Child Development*, 47–59.

Kachel, G., Bohn, M., O’Madagain, C., & Haun, D. (in prep.). *Young children’s comprehension of deixis in the graphic domain*.

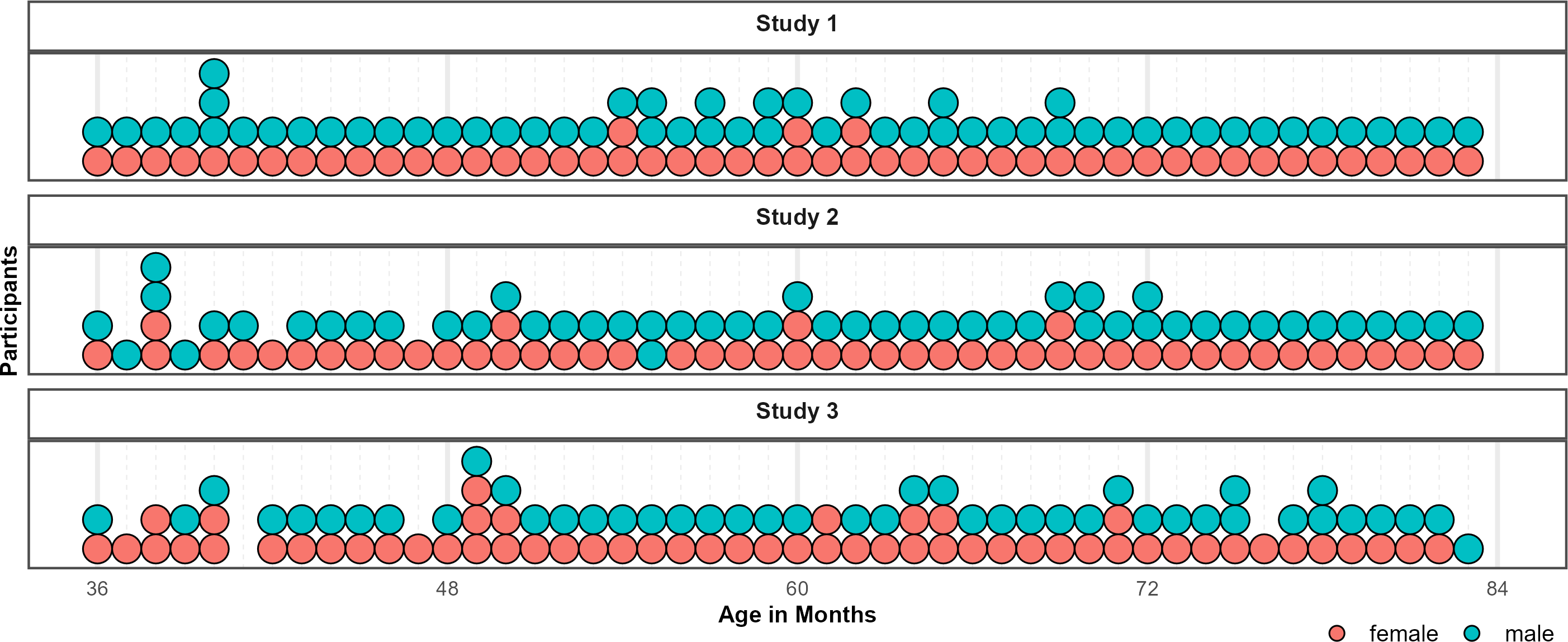
McElreath, R. (2018). *Statistical rethinking: A bayesian course with examples in r and stan*. Chapman; Hall/CRC.

Richland, L. E., Morrison, R. G., & Holyoak, K. J. (2006). Children’s development of analogical reasoning: Insights from scene analogy problems. *Journal of Experimental Child Psychology*, *94*(3), 249–273.

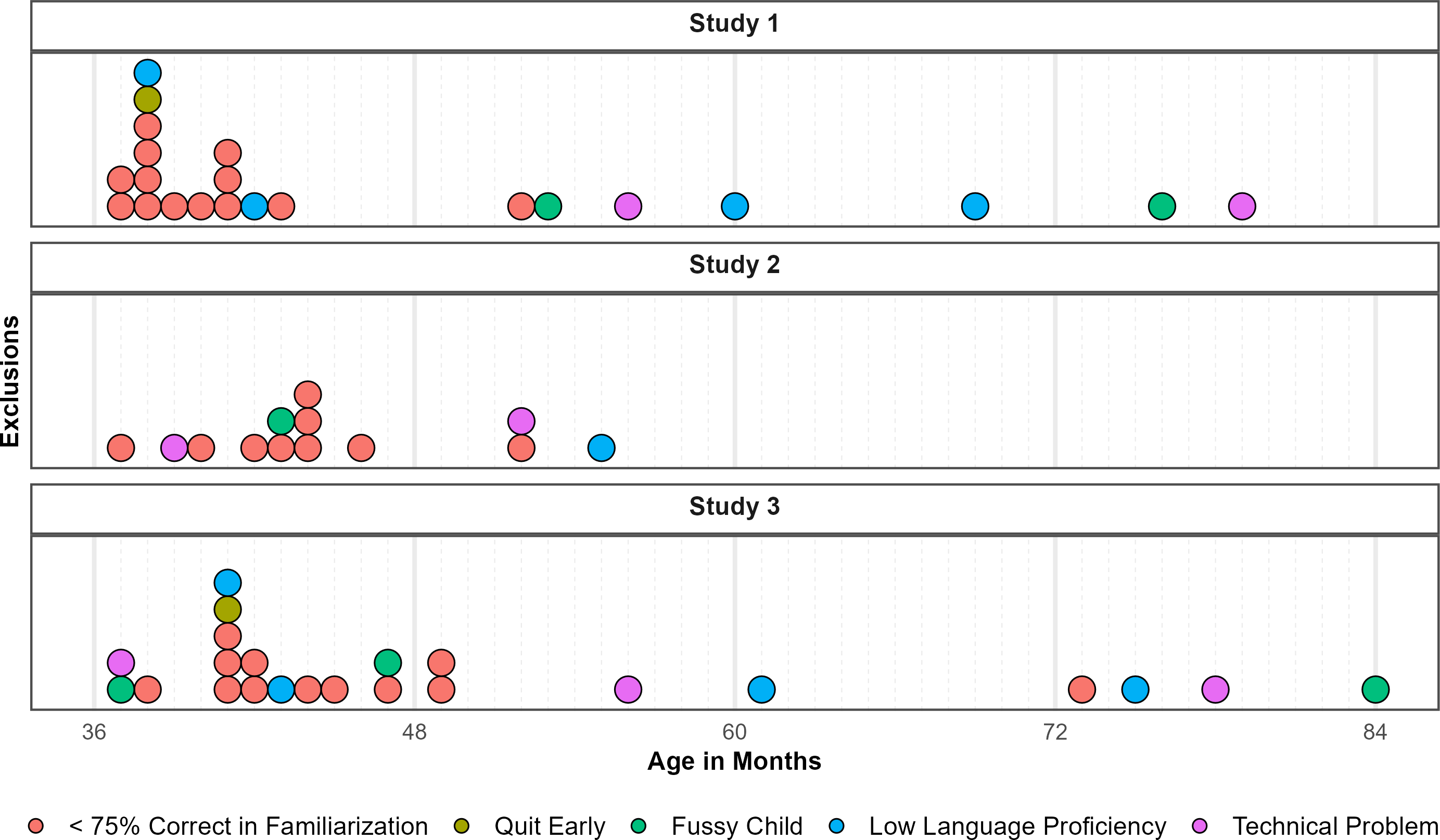
Sivula, T., Magnusson, M., Matamoros, A. A., & Vehtari, A. (2020). Uncertainty in bayesian leave-one-out cross-validation based model comparison. *arXiv Preprint arXiv:2008.10296*.

Wittgenstein, L. (2009). *Philosophical investigations*. John Wiley & Sons.

# Appendix A - Participants and Exclusions

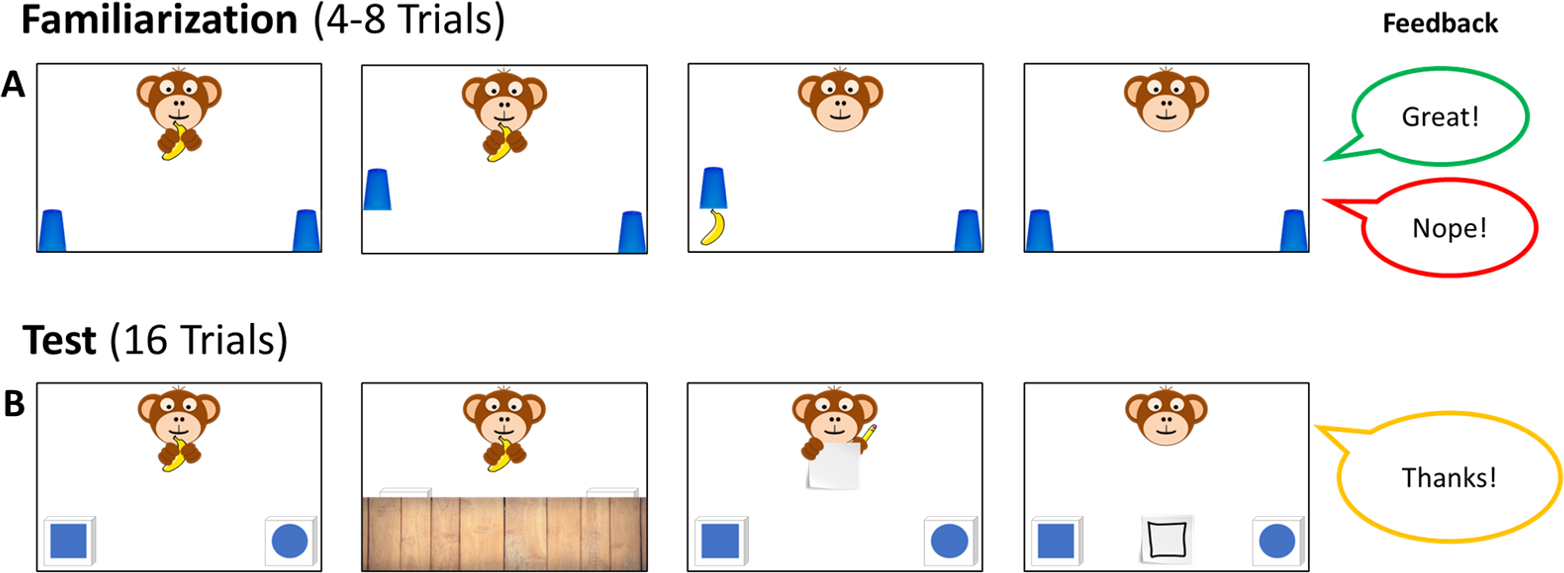


*Figure* 5. Distribution of Participants across the age-range in all three studies. Dots represent individuals. Colors indicate their respective sex.

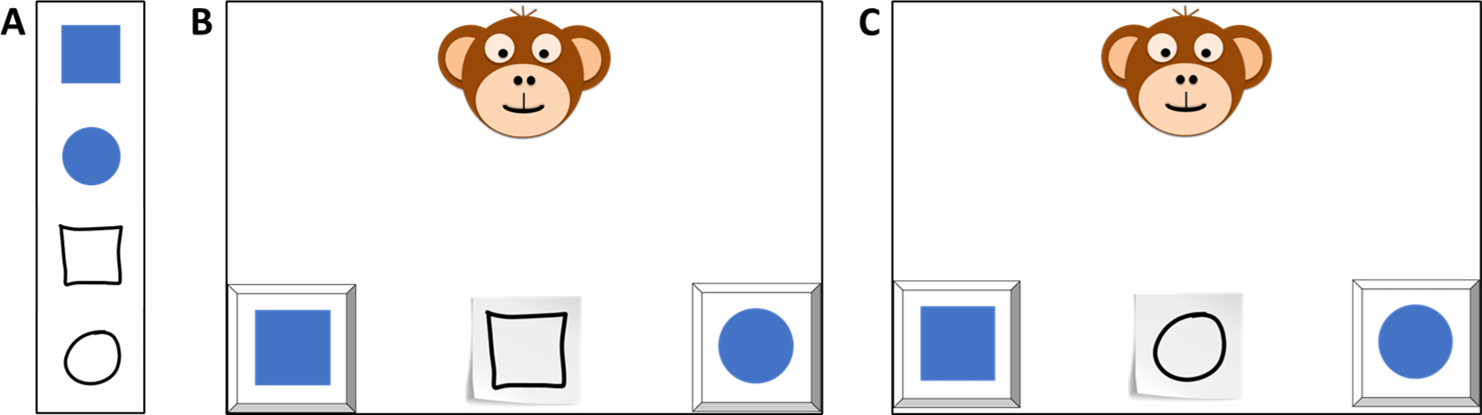


*Figure* 6. Distribution of Exclusions across the age-range in all three studies. Dots represent individuals. Colors indicate why children where not submitted to analyses.

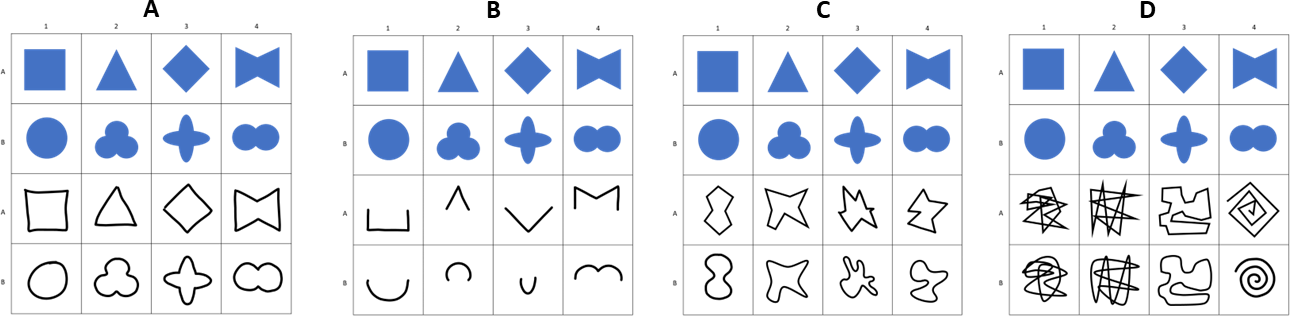
# Appendix B - Stimuli



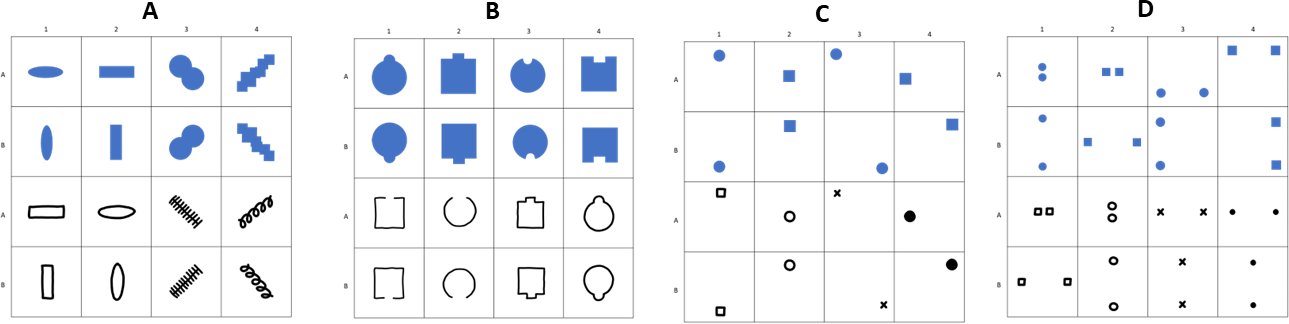
*Figure* 7. *Familiarization and Test Trials.* During Familiarization (A), children are presented with four to eight trials, in which the monkey is hiding the banana in plain sight. During familiarization children receive feedback to familiarize with the hiding game. At Test (B), children cannot know where the banana is hidden but the agent provides a cue. Here, children receive no feedback but hear a brief courtesy phrase”.



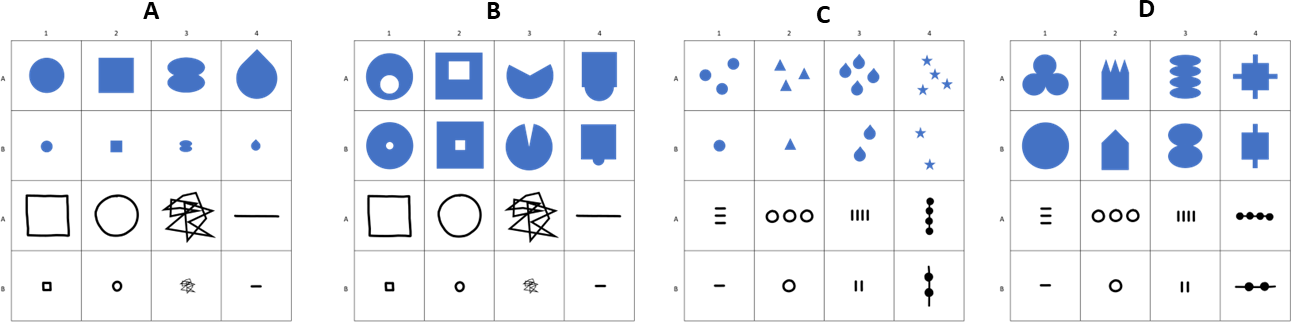
*Figure* 8. *Trial Composition.* Each trial was composed of a cue and two possible referents serving as target and distractor. In all conditions, the referents exemplified opposites of a concept (cf. Panel A: form - square and round). For each pair there were two cues that occured in counterbalanced order across participants (cf. Panel B and C).



*Figure* 9. *Stimuli for Study 1.* Panels illustrate all items used in the conditions (A) Representation, (B) Pars Pro Toto, (C) Simple Form Analogy and (D) Complex Form Analogy.



*Figure* 10. *Stimuli for Study 2.* Panels illustrate all items used in the conditions (A) Absolute Position, (B) Relative Position, (C) Orientation of Object and (D) Orientation of Feature.



*Figure* 11. *Stimuli for Study 3.* Panels illustrate all items used in the conditions (A) Size of Object, (B) Size of Feature, (C) Number of Object and (D) Number of Feature.

# Appendix C - Results

Table 1: Study 1 - Posterior Estimates for the Full Model.

| Predictor | Estimate | SD | MAD | 95% CrI | Bulk ESS | Tail ESS |
| --- | --- | --- | --- | --- | --- | --- |
| Intercept | 2.03 | 0.21 | 0.20 | [1.63, 2.44] | 4,762 | 5,635 |
| Pars Pro Toto | -0.13 | 0.21 | 0.21 | [-0.54, 0.28] | 7,175 | 7,679 |
| Simple Form Analogy | -1.39 | 0.19 | 0.19 | [-1.75, -1.04] | 6,931 | 7,148 |
| Complex Form Analogy | -1.36 | 0.19 | 0.19 | [-1.73, -1.00] | 6,978 | 7,206 |
| Age\* | 0.42 | 0.15 | 0.14 | [0.13, 0.72] | 4,412 | 5,591 |
| Trial\* | -0.01 | 0.07 | 0.07 | [-0.15, 0.12] | 8,206 | 6,164 |
| Sex (Male) | -0.30 | 0.20 | 0.20 | [-0.70, 0.09] | 4,433 | 6,217 |
| Pars Pro Toto × Age\* | 0.31 | 0.18 | 0.18 | [-0.05, 0.66] | 6,302 | 7,377 |
| Simple Form Analogy × Age\* | -0.20 | 0.16 | 0.16 | [-0.52, 0.11] | 5,815 | 6,755 |
| Complex Form Analogy × Age\* | -0.08 | 0.16 | 0.16 | [-0.40, 0.24] | 5,710 | 6,711 |
| Note. Estimates represent posterior means with 95% equal-tailed credible intervals (CrIs). MAD indicates Median Absolute Deviation. ESS refers to effective sample size. R̂ values omitted as they are all ~1 indicating convergence. \* = variables were standardized. | | | | | | |

Table 2: Study 1 - Descriptives and Conventional Analyses.

| Condition | Age | N | Trials | Trials/N | M | SD | p | df | t(N-1) | d |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Representation | 3-year-olds | 26 | 103 | 3.96 | 75.00 | 22.36 | <0.001 | 25.00 | 5.70 | 1.12 |
|  | 4-year-olds | 28 | 109 | 3.89 | 81.25 | 33.07 | <0.001 | 27.00 | 5.00 | 0.94 |
|  | 5-year-olds | 28 | 112 | 4.00 | 87.50 | 28.46 | <0.001 | 27.00 | 6.97 | 1.32 |
|  | 6-year-olds | 24 | 96 | 4.00 | 87.50 | 20.85 | <0.001 | 23.00 | 8.81 | 1.80 |
| Pars Pro Toto | 3-year-olds | 26 | 104 | 4.00 | 66.35 | 22.30 | <0.001 | 25.00 | 3.74 | 0.73 |
|  | 4-year-olds | 28 | 112 | 4.00 | 75.00 | 31.91 | <0.001 | 27.00 | 4.15 | 0.78 |
|  | 5-year-olds | 28 | 112 | 4.00 | 85.71 | 21.97 | <0.001 | 27.00 | 8.60 | 1.63 |
|  | 6-year-olds | 24 | 96 | 4.00 | 91.67 | 15.93 | <0.001 | 23.00 | 12.82 | 2.62 |
| Simple Form Analogy | 3-year-olds | 26 | 104 | 4.00 | 53.85 | 23.12 | 0.404 | 25.00 | 0.85 | 0.17 |
|  | 4-year-olds | 28 | 112 | 4.00 | 59.82 | 28.33 | 0.078 | 27.00 | 1.83 | 0.35 |
|  | 5-year-olds | 28 | 112 | 4.00 | 58.93 | 28.23 | 0.106 | 27.00 | 1.67 | 0.32 |
|  | 6-year-olds | 24 | 96 | 4.00 | 69.79 | 23.29 | <0.001 | 23.00 | 4.16 | 0.85 |
| Complex Form Analogy | 3-year-olds | 25 | 100 | 4.00 | 53.00 | 25.33 | 0.559 | 24.00 | 0.59 | 0.12 |
|  | 4-year-olds | 28 | 112 | 4.00 | 54.46 | 23.62 | 0.326 | 27.00 | 1.00 | 0.19 |
|  | 5-year-olds | 28 | 112 | 4.00 | 59.82 | 25.77 | 0.054 | 27.00 | 2.02 | 0.38 |
|  | 6-year-olds | 24 | 96 | 4.00 | 76.04 | 18.77 | <0.001 | 23.00 | 6.80 | 1.39 |
| Note. Descriptive and inferential statistics by age group and condition. Performance (\*M\*, \*SD\*) is percent correct responses.\*p\*, \*t\*, and \*d\* reflect one-sample t-tests vs. chance level (0.5). \*Trials\* = total trials across participants; \*Trials/N\* = average per participant. | | | | | | | | | | |

Table 3: Study 2 - Posterior Estimates for the Full Model.

| Predictor | Estimate | SD | MAD | 95% CrI | Bulk ESS | Tail ESS |
| --- | --- | --- | --- | --- | --- | --- |
| Intercept | 1.08 | 0.20 | 0.20 | [0.69, 1.47] | 4,475 | 6,307 |
| Relative Position | -0.28 | 0.18 | 0.18 | [-0.64, 0.09] | 8,152 | 7,410 |
| Orientation of Object | -0.10 | 0.19 | 0.19 | [-0.47, 0.28] | 8,277 | 8,170 |
| Orientation of Feature | -0.12 | 0.19 | 0.19 | [-0.49, 0.24] | 8,388 | 7,492 |
| Age\* | 0.40 | 0.14 | 0.14 | [0.14, 0.68] | 3,951 | 5,701 |
| Trial\* | 0.26 | 0.08 | 0.08 | [0.10, 0.43] | 6,840 | 7,672 |
| Sex (Male) | -0.02 | 0.22 | 0.22 | [-0.45, 0.43] | 3,842 | 5,576 |
| Relative Position × Age\* | -0.05 | 0.16 | 0.16 | [-0.37, 0.26] | 6,542 | 7,410 |
| Orientation of Object × Age\* | 0.19 | 0.16 | 0.16 | [-0.13, 0.51] | 6,841 | 7,114 |
| Orientation of Feature × Age\* | -0.00 | 0.16 | 0.16 | [-0.32, 0.31] | 6,918 | 6,858 |
| Note. Estimates represent posterior means with 95% equal-tailed credible intervals (CrIs). MAD indicates Median Absolute Deviation. ESS refers to effective sample size. R̂ values omitted as they are all ~1 indicating convergence. \* = variables were standardized. | | | | | | |

Table 4: Study 2 - Descriptives and Conventional Analyses.

| Condition | Age | N | Trials | Trials/N | M | SD | p | df | t(N-1) | d |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Absolute Position | 3-year-olds | 20 | 80 | 4.00 | 58.75 | 18.63 | 0.049 | 19.00 | 2.10 | 0.47 |
|  | 4-year-olds | 25 | 100 | 4.00 | 62.00 | 33.17 | 0.083 | 24.00 | 1.81 | 0.36 |
|  | 5-year-olds | 27 | 108 | 4.00 | 73.15 | 30.95 | <0.001 | 26.00 | 3.89 | 0.75 |
|  | 6-year-olds | 25 | 100 | 4.00 | 81.00 | 29.12 | <0.001 | 24.00 | 5.32 | 1.06 |
| Relative Position | 3-year-olds | 22 | 88 | 4.00 | 54.55 | 21.32 | 0.329 | 21.00 | 1.00 | 0.21 |
|  | 4-year-olds | 24 | 96 | 4.00 | 59.38 | 26.39 | 0.095 | 23.00 | 1.74 | 0.36 |
|  | 5-year-olds | 27 | 108 | 4.00 | 68.52 | 34.39 | 0.010 | 26.00 | 2.80 | 0.54 |
|  | 6-year-olds | 25 | 100 | 4.00 | 76.00 | 34.97 | 0.001 | 24.00 | 3.72 | 0.74 |
| Orientation of Object | 3-year-olds | 22 | 85 | 3.86 | 48.86 | 19.64 | 0.789 | 21.00 | -0.27 | 0.06 |
|  | 4-year-olds | 25 | 100 | 4.00 | 57.00 | 25.54 | 0.183 | 24.00 | 1.37 | 0.27 |
|  | 5-year-olds | 27 | 108 | 4.00 | 75.93 | 24.50 | <0.001 | 26.00 | 5.50 | 1.06 |
|  | 6-year-olds | 25 | 100 | 4.00 | 81.00 | 25.29 | <0.001 | 24.00 | 6.13 | 1.23 |
| Orientation of Feature | 3-year-olds | 21 | 80 | 3.81 | 55.95 | 28.40 | 0.348 | 20.00 | 0.96 | 0.21 |
|  | 4-year-olds | 25 | 100 | 4.00 | 67.00 | 28.61 | 0.007 | 24.00 | 2.97 | 0.59 |
|  | 5-year-olds | 27 | 108 | 4.00 | 65.74 | 34.77 | 0.027 | 26.00 | 2.35 | 0.45 |
|  | 6-year-olds | 25 | 100 | 4.00 | 79.00 | 33.60 | <0.001 | 24.00 | 4.32 | 0.86 |
| Note. Descriptive and inferential statistics by age group and condition. Performance (\*M\*, \*SD\*) is percent correct responses.\*p\*, \*t\*, and \*d\* reflect one-sample t-tests vs. chance level (0.5). \*Trials\* = total trials across participants; \*Trials/N\* = average per participant. | | | | | | | | | | |

Table 5: Study 3 - Posterior Estimates for the Full Model.

| Predictor | Estimate | SD | MAD | 95% CrI | Bulk ESS | Tail ESS |
| --- | --- | --- | --- | --- | --- | --- |
| Intercept | 1.04 | 0.18 | 0.17 | [0.70, 1.39] | 5,262 | 6,963 |
| Size of Feature | -0.71 | 0.18 | 0.18 | [-1.07, -0.37] | 7,985 | 8,161 |
| Number of Object | -0.08 | 0.18 | 0.18 | [-0.44, 0.27] | 7,687 | 7,931 |
| Number of Feature | -0.31 | 0.18 | 0.18 | [-0.67, 0.03] | 8,132 | 7,790 |
| Age\* | 0.70 | 0.14 | 0.13 | [0.44, 0.97] | 4,655 | 5,699 |
| Trial\* | 0.21 | 0.08 | 0.07 | [0.07, 0.36] | 8,714 | 7,473 |
| Sex (Male) | -0.18 | 0.19 | 0.19 | [-0.55, 0.20] | 4,875 | 6,203 |
| Size of Feature × Age\* | -0.50 | 0.16 | 0.16 | [-0.81, -0.19] | 6,132 | 7,369 |
| Number of Object × Age\* | 0.00 | 0.17 | 0.17 | [-0.32, 0.33] | 6,727 | 7,330 |
| Number of Feature × Age\* | -0.28 | 0.16 | 0.17 | [-0.61, 0.02] | 6,519 | 7,167 |
| Note. Estimates represent posterior means with 95% equal-tailed credible intervals (CrIs). MAD indicates Median Absolute Deviation. ESS refers to effective sample size. R̂ values omitted as they are all ~1 indicating convergence. \* = variables were standardized. | | | | | | |

Table 6: Study 3 - Descriptives and Conventional Analyses.

| Condition | Age | N | Trials | Trials/N | M | SD | p | df | t(N-1) | d |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Size of Object | 3-year-olds | 21 | 84 | 4.00 | 51.19 | 23.02 | 0.815 | 20.00 | 0.24 | 0.05 |
|  | 4-year-olds | 27 | 105 | 3.89 | 54.63 | 31.80 | 0.456 | 26.00 | 0.76 | 0.15 |
|  | 5-year-olds | 26 | 104 | 4.00 | 84.62 | 25.57 | <0.001 | 25.00 | 6.90 | 1.35 |
|  | 6-year-olds | 24 | 96 | 4.00 | 78.12 | 22.50 | <0.001 | 23.00 | 6.12 | 1.25 |
| Size of Feature | 3-year-olds | 21 | 84 | 4.00 | 46.43 | 19.82 | 0.419 | 20.00 | -0.83 | 0.18 |
|  | 4-year-olds | 26 | 104 | 4.00 | 51.92 | 23.37 | 0.678 | 25.00 | 0.42 | 0.08 |
|  | 5-year-olds | 26 | 104 | 4.00 | 56.73 | 32.06 | 0.295 | 25.00 | 1.07 | 0.21 |
|  | 6-year-olds | 24 | 96 | 4.00 | 62.50 | 19.50 | 0.005 | 23.00 | 3.14 | 0.64 |
| Number of Object | 3-year-olds | 21 | 84 | 4.00 | 42.86 | 21.13 | 0.137 | 20.00 | -1.55 | 0.34 |
|  | 4-year-olds | 27 | 108 | 4.00 | 61.11 | 30.49 | 0.069 | 26.00 | 1.89 | 0.36 |
|  | 5-year-olds | 26 | 104 | 4.00 | 72.12 | 34.88 | 0.003 | 25.00 | 3.23 | 0.63 |
|  | 6-year-olds | 24 | 96 | 4.00 | 83.33 | 32.69 | <0.001 | 23.00 | 4.99 | 1.02 |
| Number of Feature | 3-year-olds | 21 | 82 | 3.90 | 50.00 | 19.36 | >0.999 | 20.00 | 0.00 | 0.00 |
|  | 4-year-olds | 27 | 108 | 4.00 | 62.96 | 24.39 | 0.010 | 26.00 | 2.76 | 0.53 |
|  | 5-year-olds | 26 | 104 | 4.00 | 65.38 | 31.68 | 0.020 | 25.00 | 2.48 | 0.49 |
|  | 6-year-olds | 24 | 96 | 4.00 | 76.04 | 27.07 | <0.001 | 23.00 | 4.71 | 0.96 |
| Note. Descriptive and inferential statistics by age group and condition. Performance (\*M\*, \*SD\*) is percent correct responses.\*p\*, \*t\*, and \*d\* reflect one-sample t-tests vs. chance level (0.5). \*Trials\* = total trials across participants; \*Trials/N\* = average per participant. | | | | | | | | | | |

# Appendix D - THIS IS OBSOLETE

Additional Tables and illustrations for the convenience of the reader. Add illustrations they said; it will add value they said.

## Study 1

Model diagnostics were drawn for the full and null model in study 1. Rhat values in both models were equal to one, indicating convergence across all chains. Effective sample sizes for all fixed effects in the full model (Bulk ESS, mean = 6072, range 4412 - 8206) and the null model (Bulk ESS, mean = 6655, range 4610 - 10721) were > 1000, indicating reliable posterior estimations.

## Study 2

Model diagnostics were drawn for the full and null model in study 2. Rhat values in both models were equal to one, indicating convergence across all chains. Effective sample sizes for all fixed effects in the full model (Bulk ESS, mean = 6423, range 3842 - 8388) and the null model (Bulk ESS, mean = 6501, range 4502 - 8732) were > 1000, indicating reliable posterior estimations.

## Study 3

Model diagnostics were drawn for the full and null model in study 3. Rhat values in both models were equal to one, indicating convergence across all chains. Effective sample sizes for all fixed effects in the full model (Bulk ESS, mean = 6669, range 4655 - 8714) and the null model (Bulk ESS, mean = 7344, range 5803 - 9047) were > 1000, indicating reliable posterior estimations.

# Appendix E - Additional Analyses

Preregistration: An additional exploratory analysis will include a random effect for item level effects (Model: correct ~ task\*z.age +z.trial +z.sex +(z.trial|id) +(z.age|item)). Results will help to evaluate the equivalence of items within a task and be reported in the supplements.

\*\*correct ~ condition\*z.age +z.trial +sex +(z.trial|subid) +(z.age|cue)\*\*

Cue Level Models

## Study 1

Additional Tables and illustrations for the convenience of the reader. Add illustrations they said; it will add value they said.

## Study 2

Additional Tables and illustrations for the convenience of the reader. Add illustrations they said; it will add value they said.

## Study 3

Additional Tables and illustrations for the convenience of the reader. Add illustrations they said; it will add value they said.