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:* graphical representation, iconicity, analogy, symbol, communication, emerging literacy

*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

See googledocs for drafts of the intro and discussion. There needs to be at least one citation in order for this document to knit, so consider that preschoolers invent and comprehend iconic gestures spontaneously (Bohn, Kachel, & Tomasello, 2019).

# 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 at testing 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 afterschoolcare 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 an assisted with a fixed set of verbal prompts when necessary. 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.

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 after a few seconds. 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.

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 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. Children were now 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, each hiding places, for example, showed either a solid blue circle or square and the paper displayed a simple outline drawing of either shape. Here, the drawing was a direct representation of the target shape. 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 are submitted to analyses. For an overview of the average number of trials participants received in each condition, see tables 1, 2 and 3 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 the 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 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 shares a feature with one of the referents and is thereby referring to it while it remains distinct with regard to other surface features. In any trial 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 referents, and the identity of cues 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 95% CrI for a particular trajectory did no longer overlap with a midline demarcating 50% chance level. All analyses were preregistered prior to data collection. Analyses deviate from the preregistered analyses when comparing models using ELPD differences. 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 the 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 1, 2 and 3).

# Study 1

## 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 figure 5 and 6.

## Stimuli

Basic shapes that children know ()

Gestalt Principles, Form Completion ()

Round vs edgy (bouba kiki)

Simple and Complex Form analogy - simpler stimulus vs more information

For all stimuli in Study 1, see figure 9 in appendix B.

## Analysis

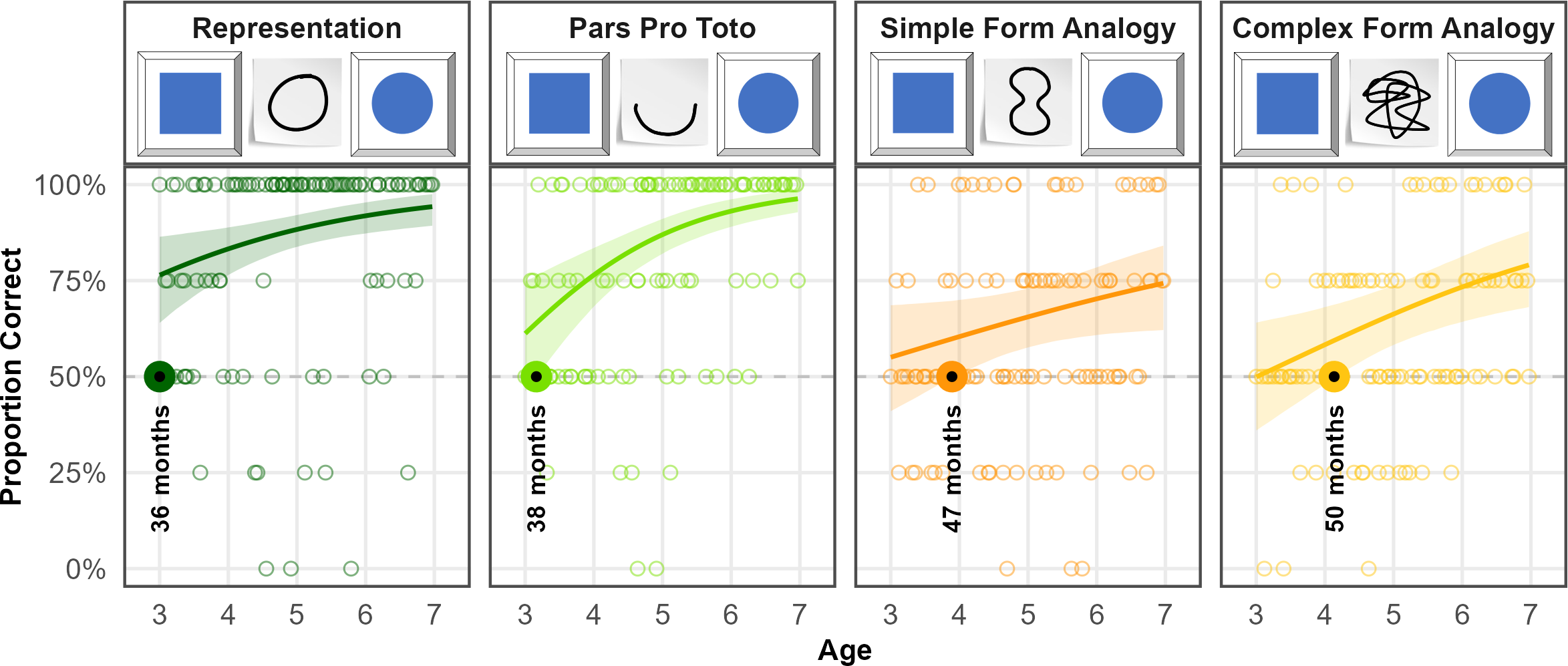
A total of 1688 trials (mean per condition = 422, range: 420 - 424) from 106 participants were submitted for analysis. The full model notation was correct ~ condition\*z.age + z.trial + sex + (z.trial | subid). In addition, a null model lacking the interaction of condition and age was fitted.

## 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 has 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 [-1.70, -1.09) and *Complex Form Analogy* (beta = -1.36, 95% CrI [-1.67, -1.05) 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 [-0.48, 0.21). 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 [0.00, 0.60), 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 [0.18, 0.67). In contrast, trial number has no clear effect on performance (beta = -0.01, 95% CrI [-0.13, 0.10), 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 option. 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. For an additional conventional analysis binning participants according to their age in years, please see table 1 in appendix C.



*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.

# Study 2

## 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).

## Materials

For all stimuli in Study 2, see figure 10 in appendix B.

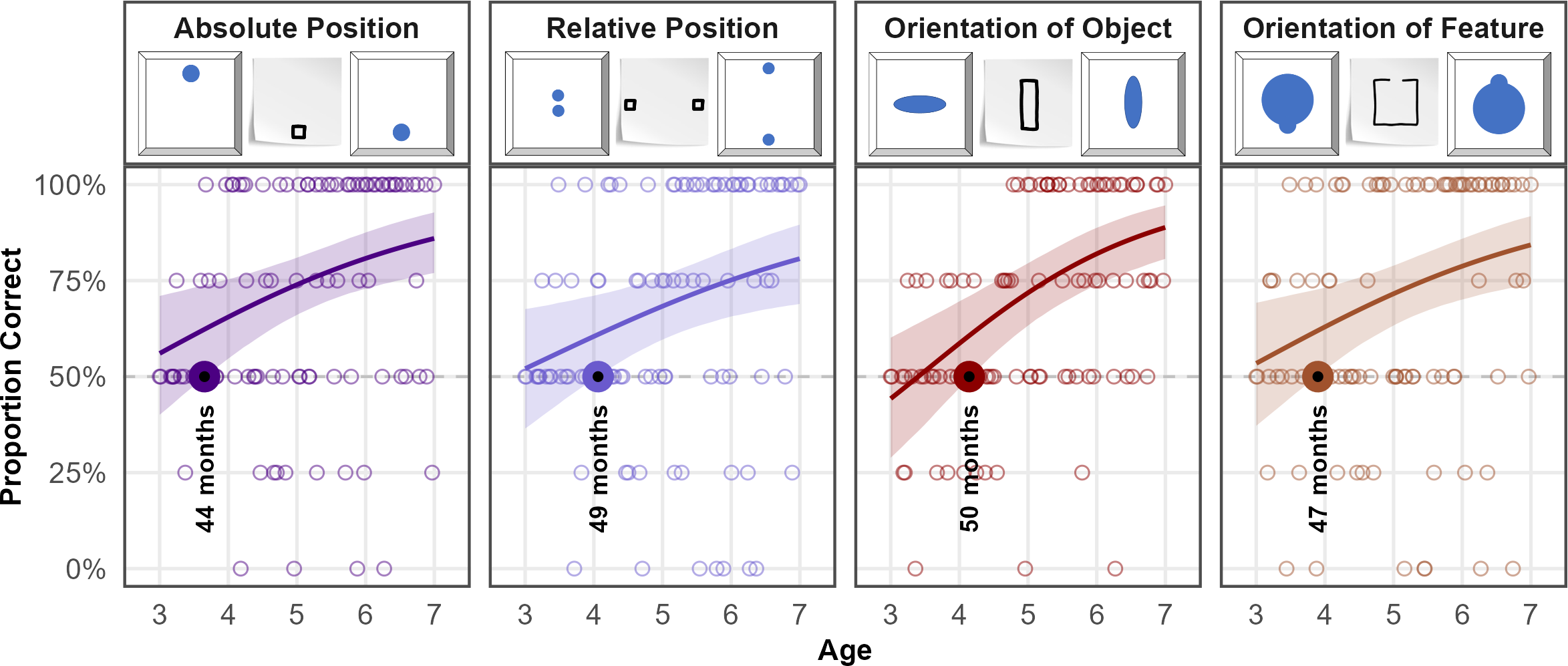
## Analysis

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 [0.18, 0.63) and slightly with trial number (beta = 0.26, 95% CrI [0.12, 0.40). Relative to *Absolute Position*, children were generally less likely to correctly solve *Relative Position* (beta = -0.28, 95% CrI [-0.58, 0.03). Performance in *Orientation of Object* (beta = -0.10, 95% CrI [-0.40, 0.22) and *Orientation of Feature* (beta = -0.12, 95% CrI [-0.43, 0.19) 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 2 in appendix C.



*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.

# Study 3

General note on the aim of the investigation

## 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).

## Materials

For all stimuli in Study 3, see figure 11 in appendix B.

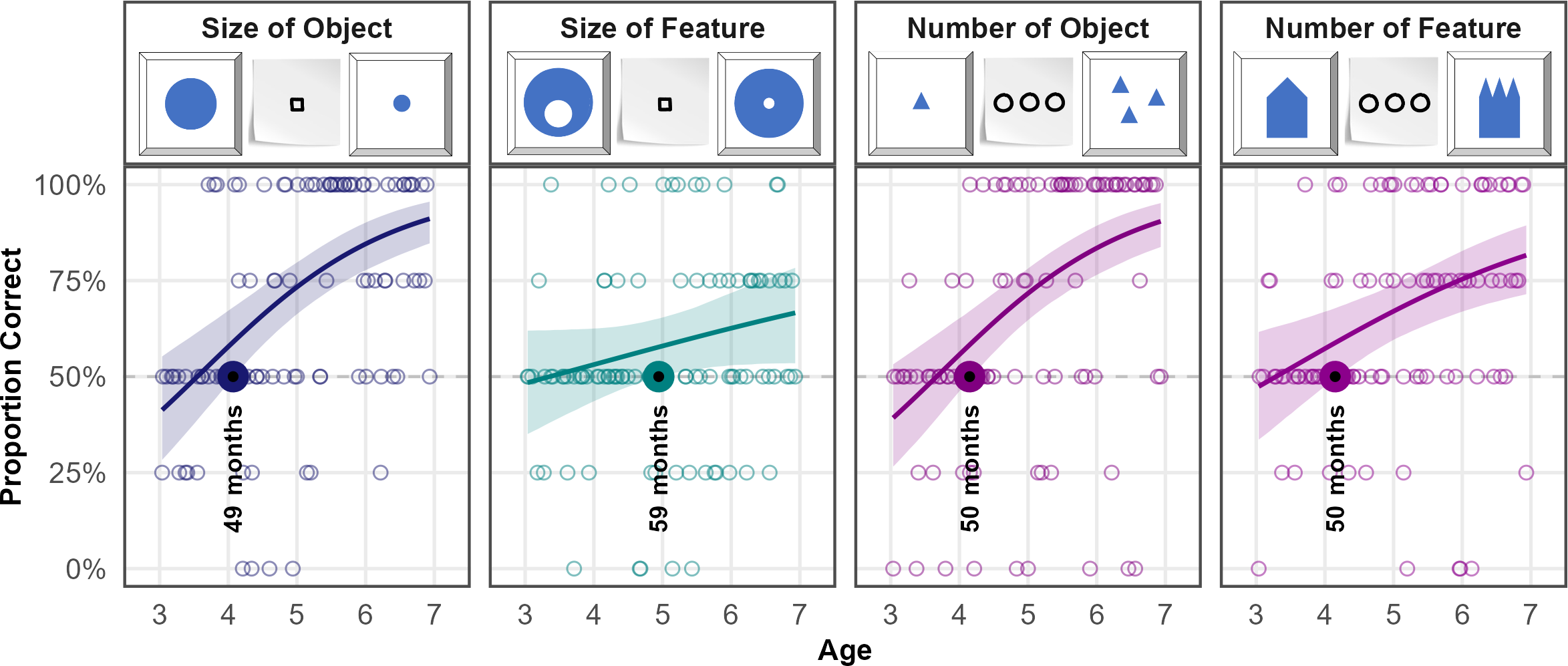
## Analysis

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 [0.48, 0.93) and with trial number (beta = 0.21, 95% CrI [0.09, 0.34), 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 [-1.01, -0.42). A smaller, marginal effect was observed in *Number of Feature* (beta = -0.31, 95% CrI [-0.62, -0.03), while *Number of Object* (beta = -0.08, 95% CrI [-0.38, 0.22) did not differ reliably from the *Size of Object*. Age moderated performance less strongly in the *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 with *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 alternative conventional analyses binning children by year of age, please see table 3 in appendix C.



*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.

# Additional Analyses

# Discussion

See googledocs for drafts of the intro and discussion

# References

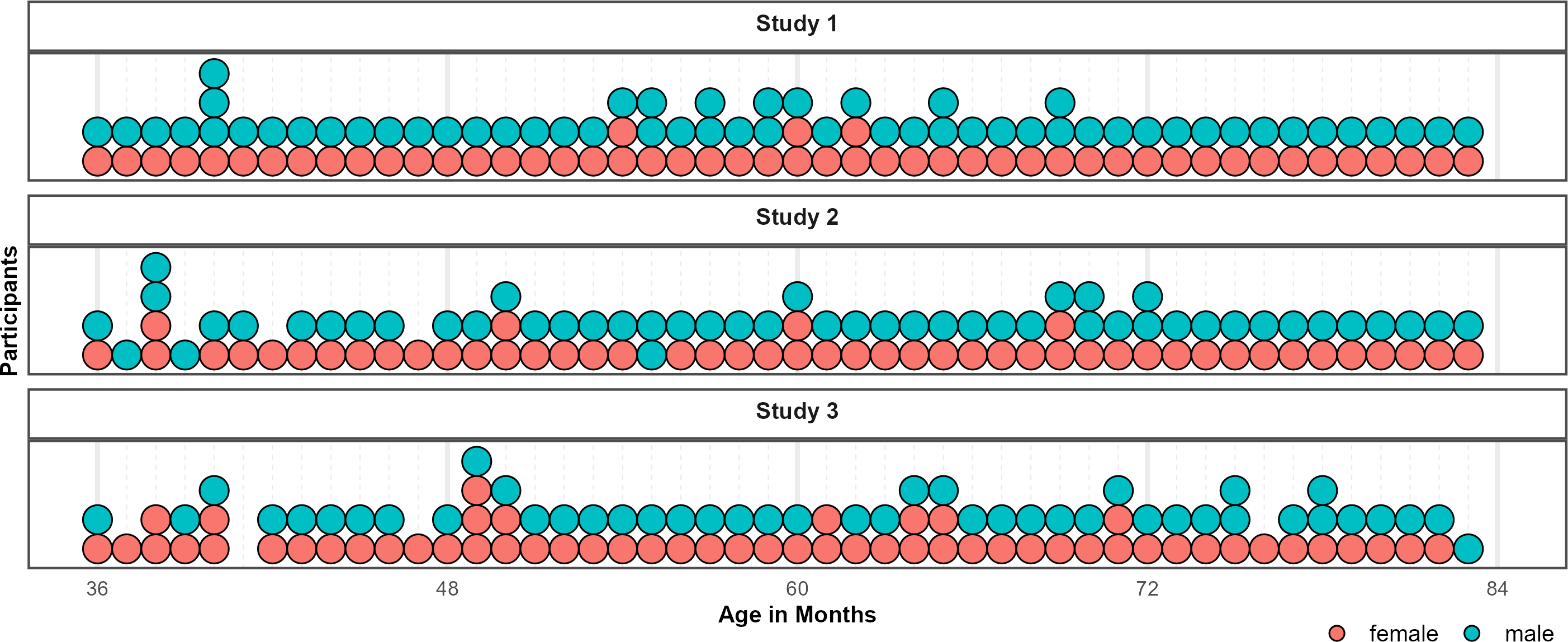
Bohn, M., Kachel, G., & Tomasello, M. (2019). Young children spontaneously recreate core properties of language in a new modality. *Proceedings of the National Academy of Sciences*, *116*(51), 26072–26077.

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

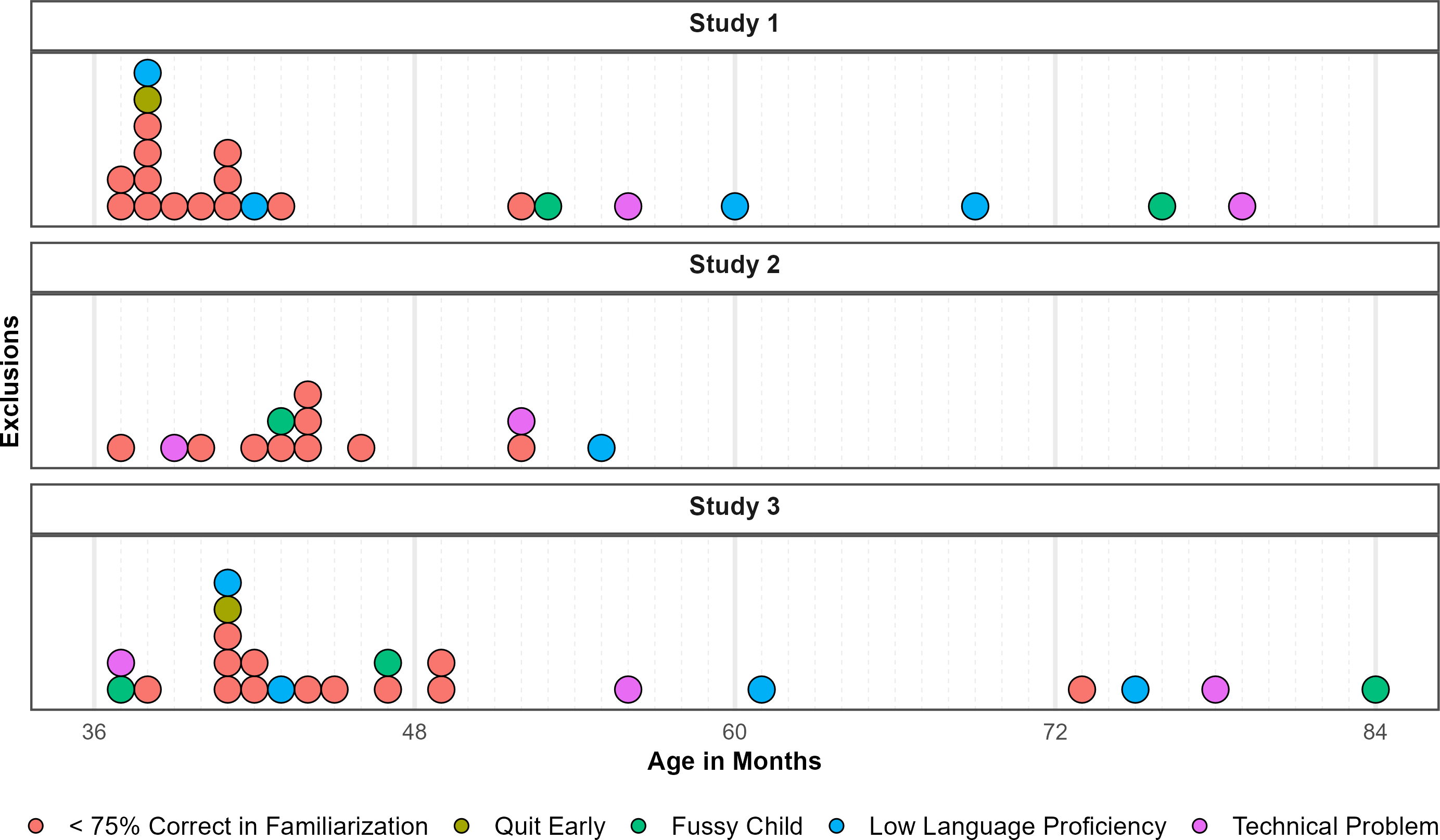
McElreath, R. (2018). *Statistical rethinking: A bayesian course with examples in r and stan*. Chapman; Hall/CRC.

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

# Appendix A - Participants and Exclusions

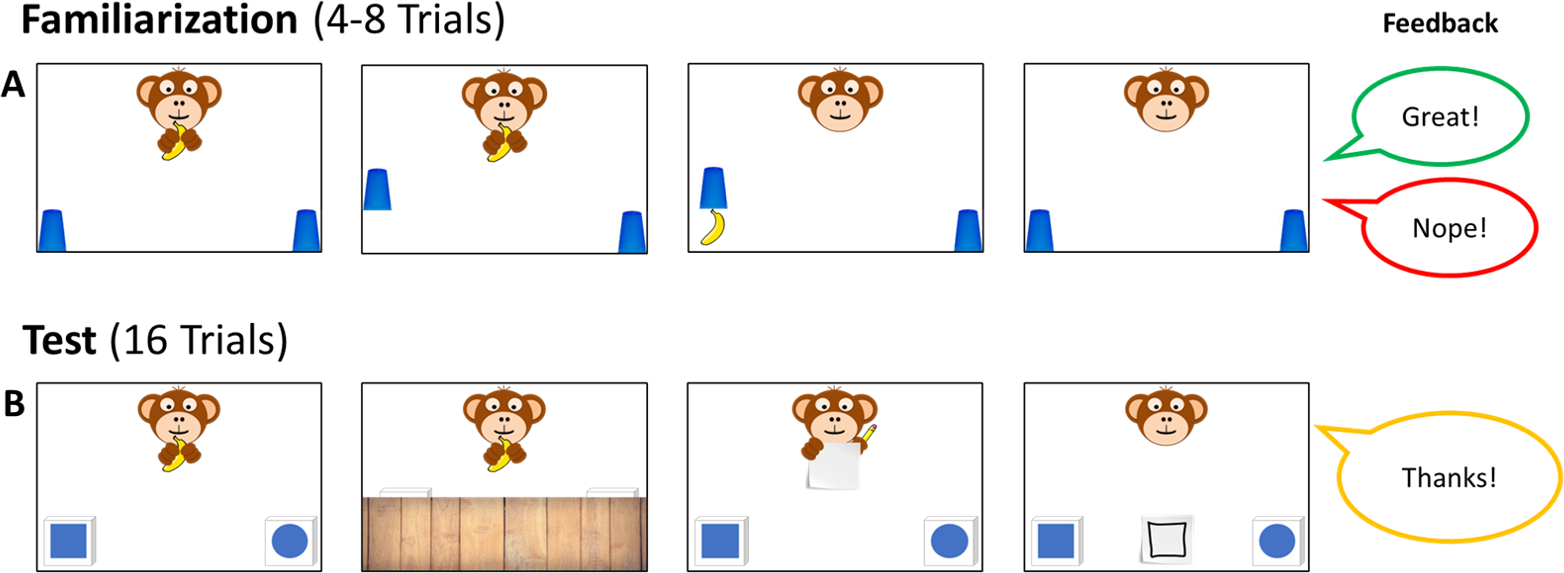


*Figure* 5. Distribution of Participants across the age-range in all three studies. Dots represent individuals. Colors indicate their respective sex.

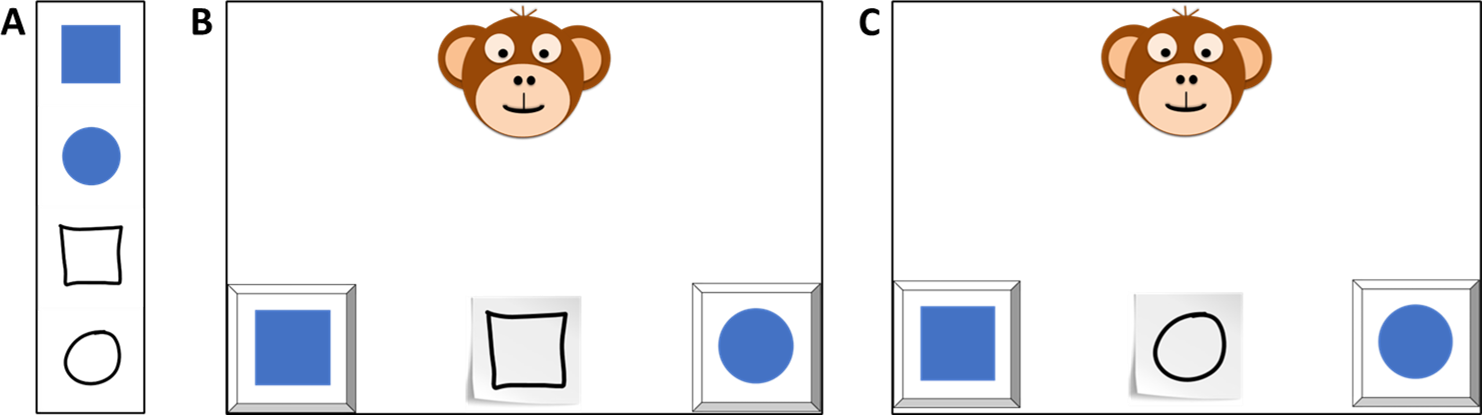


*Figure* 6. Distribution of Exclusions across the age-range in all three studies. Dots represent individuals. Colors indicate why children where not submitted to analyses.

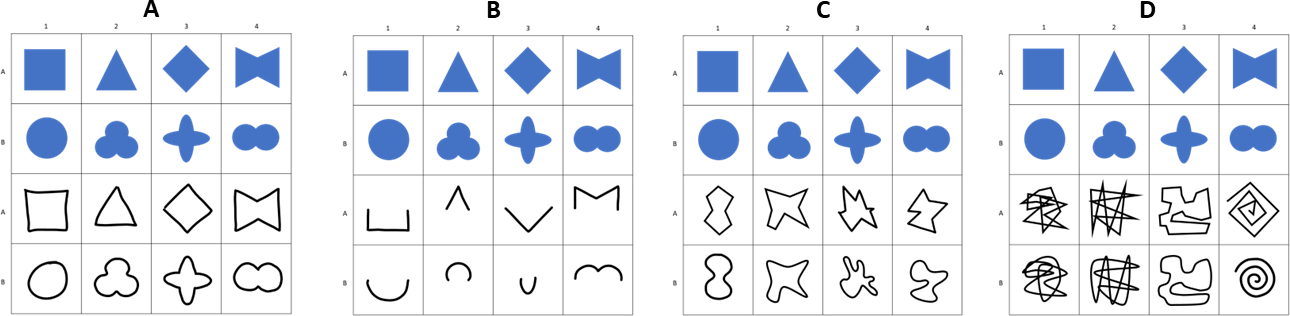
# Appendix B - Stimuli



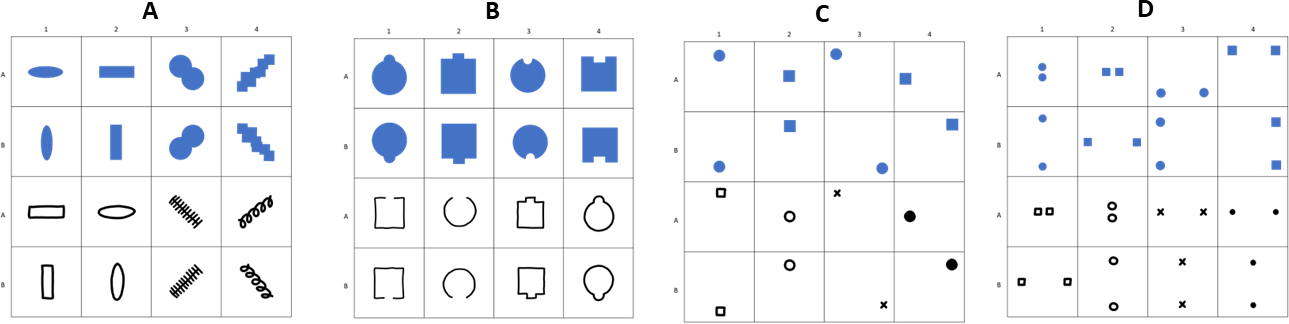
*Figure* 7. *Familiarization and Test Trials.* During Familiarization (A), children are presented with four to eight trials, in which the monkey is hiding the banana in plain sight. During familiarization children receive feedback to familiarize with the hiding game. At Test (B), children cannot know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



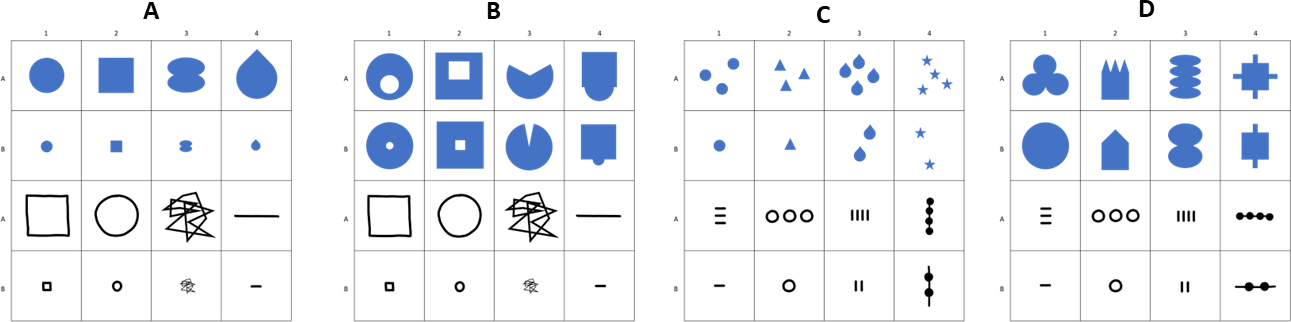
*Figure* 8. *Trial Composition.* Each trial was composed of a cue and two possible referents serving as target and distractor. In all conditions, the referents exemplified opposites of a concept (cf. Panel A: form - square and round). For each pair there were two cues that occured in counterbalanced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

Table 1: Descriptive Analyses for Study 1.

| 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 |

Table 2: Descriptive Analyses for Study 2.

| 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 |

Table 3: Descriptive Analyses for Study 3.

| 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 |

# Appendix D - Model Diagnostics

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.