

Young Children's Spontaneous Comprehension of 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 pre-registered. For data, analyses materials, and preregistrations see: https://github.com/GregorKachel/symlit_rep_manuscript.

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

The developmental literature testifies to young children's remarkable communicative skills in language and gesture, but much less is known about the domain of graphical communication. In three coordinated studies with identical procedures, young children's ability to spontaneously comprehend unfamiliar graphic symbols was tested in a simple object-choice task across a wide variety of symbol-referent relationships. In four conditions per study, reference was established via representation, part-whole relationships, analogies in size, number and form but also gestalt-like principles such as orientation, alignment or figure-ground relationships. The samples comprise German-speaking children continuously spread from the third to the seventh birthday (Study 1, N = 106; Study 2, N = 99; Study 3, N = 99). While children are generally able to use direct representations to identify the correct target item already prior to their third birthday, group-level success in all other conditions was observed later, but well before school age. Converging results from conditions employing several abstract and conceptual symbol-referent relationships point to a qualitative shift in symbolic competence around the fourth birthday. Together, this set of studies explores the conceptual continuum from iconic to abstract graphical representations and, therein, provides one of the most comprehensive, robust and systematic investigations of young children's ability to find meaning in unfamiliar graphic symbols - a largely unexplored component of emerging literacy.

Keywords: Symbolic Literacy, Graphic Communication, Communication, Pragmatics, Emerging Literacy, Symbol-Referent-Relationships, Symbol Grounding, Iconicity, Representation, Analogical Reasoning, Gestalt Principles, Shape, Size, Number, Form, Orientation, Spatial Relation

Word count: 12.420

Young Children's Spontaneous Comprehension of Symbol-Referent Relationships in the Graphic Domain

Introduction

Humans are cultural beings. There is almost no aspect of our psychology that is not influenced, shaped or even entirely created by and through social interactions (J. Henrich & Muthukrishna, 2021; N. Henrich & Henrich, 2007; Heyes, 2018; Legare, 2017; Moll & Tomasello, 2007; Tomasello, 2014b; Vygotsky, 1980). To a large extent, such interactions are symbolically mediated (Bruner, 1996; Donald, 1991, 2001; Tomasello, 2009). Children's developing understanding of symbols in language, gesture or symbolic artefacts is central to their enculturation and demonstrates the growth of cognitive capacities that define the human mind (Tomasello, 2014a; Vygotsky, 1980). The acquisition of language and – much later – literacy (Bruner, 1996; Dehaene, 2009) are arguably two of the most significant achievements in cultural learning that individuals undergo with dramatic consequences for cognitive abilities (Karmiloff-Smith, 1994).

The human pointing gesture provides an entry point into the sphere of symbolic communication and is one of the first means of intentional communication that children acquire at the end of the first year (Liszkowski, Brown, Callaghan, Takada, & De Vos, 2012; Tomasello, Carpenter, & Liszkowski, 2007). Following this achievement, children soon become able to use iconic gestures, learn their first words and begin to form propositions over the course of the second year of life (Tomasello, 2009, 2010). While at first lacking an understanding of the dual nature of symbolic means of communication, they soon come to grasp the standing-for-relationship of symbols in the third year of life. In order to better understand this process,

researchers have worked on gesture, words, photographs, representative drawings, maps, models, simulacra, figurines and sound symbolism (Callaghan & Corbit, 2015; Tomasello, Call, & Gluckman, 1997; Tomasello, Striano, & Rochat, 1999). Notably, symbolic behavior per se does not require symbolic insight (Namy & Waxman, 2017), which is to say that children may be successful communicators and grasp their interlocutors communicative acts or intentions but still lack insight into the standing-for relationship of symbols: the idea that something stands for something else. Therefore, the work of Judy DeLoache and colleagues has been of seminal importance in sketching children's ability to exploit the informative value of symbolic artefacts. In various studies, they have prompted children to find the location of a hidden toy by providing them with cues in the form of scale models, pictures, videos and maps representing the search space (DeLoache, 2004). Whereas even 30-month-old children would fail to retrieve hidden items by using instructions based on replicas like a model-room, 36-month-olds excel. This pattern of results in symbolic object-choice-tasks has been replicated many times by her group and other labs (DeLoache, 2011) and demonstrates that children's symbolic understanding undergoes a rapid development in the first and second year of life (Callaghan & Corbit, 2015; DeLoache, 2004; Tomasello et al., 1999).

A similar developmental trajectory has been observed in the understanding of graphic representations. Iconicity provides an entry point in this domain and even in the first year, infants are sensitive to the similarity between a picture and its 3D referent (DeLoache, Pierroustakos, Uttal, Rosengren, & Gottlieb, 1998). While pictures can incite manual exploration before the first birthday (Pierroustakos & DeLoache, 2003) this is replaced with pointing to the depiction in the second year (DeLoache et al., 1998). Around the same age, toddlers understand that pictures refer to concepts (Allen Preissler & Carey, 2004) in learning words: 15- and 18-month-olds can extend newly learned labels both from pictures to objects and from objects to pictures (Ganea, Pickard,

& DeLoache, 2008). From 2,5 to 4 children use the shape of drawings to interpret them as long as the drawings are presented to them as having been produced intentionally (Gelman & Ebeling, 1998). Yet, Callaghan (1999) found that two-year-olds were still at chance selecting an item depicted in representative drawings whereas three- and four-year-olds succeeded. One of the few studies sampling children's age continuously due to the arguably rapid changes between two and four years of age (DeLoache, 2004) found gradual improvement between 15 and 30 months of age using age in months as a predictor in standard linear regression (Ganea et al., 2008). Children's comprehension of graphic depictions remains fragile until the end of the third year. In another series of studies, Callaghan (2000) presented children either with a realistic pencil drawing, a stylized drawing or a model replica and then prompted children to choose what was depicted from a set of two realistic presentations. 30-month-olds did not benefit from iconicity and failed to use pictures and replicas as symbols. 36-month-olds succeeded in all conditions and benefitted from iconicity and the availability of verbal labels during the task (Callaghan, 2000; DeLoache, 1995). Matching graphic representations with each other is much less demanding or even so trivial that it is rarely employed in experimental contexts. Finding a graphical item in a drawn scene as is often done in picture-book reading routines or visual search paradigms and is generally accessible for children at the end of the third year (Gerhardstein & Rovee-Collier, 2002). Matching identical items in match-to-sample tasks is occasionally used as a training in research on analogical reasoning and generally provides no challenge for children older than three (Kroupin & Carey, 2022b).

Direct investigations into children's ability to use conceptual dimensions other than similarity or iconicity in the graphic domain are - to the best of our knowledge - relatively rare. Due to the graphical representations being static in time and limited to 2D-space, the conceptual dimensions along which information can be encoded is finite. Apart from representations

capturing the actual shape of what they are depicting, discussions of graphic communication generally identify primitives or “visual variables” as fundamental elements for encoding information. These variables are position, size, number, orientation, relational and gestalt-principles, as well as continuous visual properties such as brightness, color (hue) or texture (Cleveland & McGill, 1984; Monmonier, 1985; Stevenson, Alberto, Boom, & Boeck, 2014). The former group takes precedence in conceptual implementation due to their discreteness. After all, a conventional sign or representative drawing can evoke the same concept regardless of the color or brightness it is implemented with. On the perceptual level, object, color, quantity, size are easier to process than relational properties or figural analogies such as orientation and position (Donnelly et al., 2007; Stevenson et al., 2014). Similar successions have been documented throughout preschool development. Children are first able to understand iconic signs or depictions around the third birthday, whereas more conceptual or analogical dimensions are appreciated only at four to five years of age (Callaghan, 2000). Defining the precise ages and relative developmental trajectories for children’s sensitivity to using these graphic primitives as conveying meaning and disambiguating reference provides crucial insights into children’s developing symbolic competence but also provides orientation for investigations into the origins and development of graphic systems of communication. Below we discuss previous research on young children’s sensitivity for aspects such as part-whole relationships, form, orientation, number and size in the graphic domain.

Graphical representations naturally abstract from the reference they intend to depict by leaving out detail as it is generally sufficient for conveying the idea of - for example - a person without drawing their fingers. Reducing representation to a certain compositional element can also be evocative of a concept but is already typical for more abstract symbolic communication, for example, drawing a crown on a map may indicate the position of a royal residence. Using an

aspect or part of something in reference to a whole is common in rhetorical figures such as metonymy or ellipsis, and requires an inductive pragmatic reference on the side of the listener that is mastered by children at the age of three (Falkum, Recasens, & Clark, 2017). Omissions in depiction are an important aspect of developing conventions and streamlining communication (Goldin-Meadow & Feldman, 1977) as well as they are an interesting minimal step towards abstraction as they are still basically iconic, yet only showing a part or aspect of the referent. At least for basic geometrical shapes, we know that while the production and completion of shapes in drawing (Cox, 2005; Dağlı & Halat, 2016) develops throughout the preschool years, shape recognition and naming are robust in the third and fourth year (Verdine, Bunger, Athanasopoulou, Golinkoff, & Hirsh-Pasek, 2017; Verdine, Lucca, Golinkoff, Hirsh-Pasek, & Newcombe, 2016; Zambrzycka, Kotsopoulos, Lee, & Makosz, 2017) and the ability to mentally complete canonical geometric shapes is already nascent in early infancy (Kellman & Spelke, 1983; Valenza, Leo, Gava, & Simion, 2006), making part-whole relationships on the level of geometric figures an ideal target for an investigation into slightly abstract symbol-reference relationships that young children may be able to read at the beginning of the preschool years.

A highly abstract means of reference that is still rooted in iconicity may be the metaphorical use of form or shape, for example, by referring to something pointy, sharp or spiky with an edgy or more clear cut rectangular drawing (Knoeferle, Li, Maggioni, & Spence, 2017). Again, this is common in language and became a focus of research studying sound symbolism in the lexicon (Bermúdez, 2020; Sidhu, 2025) writing systems (Porto et al., 2024), as well as actions and gesture (Bohn, Call, & Tomasello, 2019; Margiotoudi & Pulvermüller, 2020). The phenomenon is robust across cultures (Ćwiek et al., 2022) and children appear sensitive to sound symbolism matching words and shapes by about 30 months of age (Fort et al., 2018; Imai, Kita, Nagumo, & Okada, 2008; Maurer, Pathman, & Mondloch, 2006) and for matching actions with

gestures or vocalizations at 36 months (Bohn, Call, et al., 2019). There is currently no data on whether children are sensitive to an agreement in the overall properties of a shape or the impression that its form conveys for grounding reference.

As with the other dimensions reported above, there is evidence that children are sensitive to Gestalt principles like closure and spatial dimensions of composition position, orientation and alignment in graphic displays but there is virtually no data on how likely children are to perceive such aspects of graphical stimuli to convey meaning. Closure appears to be a basic principle for graphic production, as children rarely draw overlapping figures (Cox, 2005; Piaget, Inhelder, & Bovet, 1997). Relative position is intuitively used by children even in their earliest figural drawings when arranging items in scenes, such as children next to an ice-cream van, but also when achieving “intellectual realism” in depicting complex objects (Luquet & Costall, 2001). That is, while children’s drawings may not look realistic per se, they are true to the invariant organisation of a depicted item via the relative position of the defining elements: when drawing a car, the wheels will be added at the bottom rather than on top (Cox, 2005). In both cases, relative position and orientation on the level of the individual object or scene may be discussed as iconic. A more abstract or conceptual use of relative position for conveying information is present in children’s developing ability to read maps. An intuitive understanding of basic maps allowing children to extract relational directions such as “behind me” or “in front of me” is accessible by four years of age and even so in the absence of training (Landau, 1986) but iconicity is again very important for grasping the conceptual space depicted in maps (Liben & Downs, 2013). Likewise, 4-year-olds can use geometric information from maps to navigate real space (Shusterman, Ah Lee, & Spelke, 2008), but more demanding transfer from maps to real-world settings with larger spatial arrays is unstable prior to six to seven years of age (Uttal & Wellman, 1989). For

example, performance in 4-year-olds decreases if mental rotation is necessary (Blades & Spencer, 1994).

Size is a relative property of graphical content and requires a comparison between items, for example when drawing a big and small dog, or an appreciation of an item's relation to the frame of reference it occurs in, as when a child may try to convey the concept of a big dog by making it occupy the entire piece of paper it is drawing on (Cox, 2005). Children are able to make spontaneous or productive size comparisons at three years of age and begin to use size words (size words (*big*), size superlatives (*biggest*), or quantifiers (more, most) (Ferry et al., 2025), yet young children are most accurate with *bigger* and only gradually gain competence with words like smaller, shorter or longer (Frausel et al., 2020). When presented with an array of graphic stimuli, preschoolers can reliably point at, for example, the smallest or largest item in a set (Haun & Tomasello, 2011). While three and four year olds can identify larger or smaller graphic items, the canonical size of the depicted objects can interfere leading to stroop-like errors (Long, Moher, Carey, & Konkle, 2019). When presented with three different geometric shapes in matching or non-matching dimensions (e.g. large, large, small) in match-to-sample tasks, children are quite unlikely to spontaneously infer size as the basis for matching even around the fourth birthday (Kroupin & Carey, 2022b). One of the few studies using size as a way for grounding reference, investigated the role of the producers' intentions in identifying concepts in ambiguous scribble drawings that were presented to participants as aiming to depict, for example, a small spider next to a big house. Children here made use of spatial analogies beginning at three years of age but appeared to benefit more as they turned four (Bloom & Markson, 1998).

Number, is another relational concept, and an absolute primitive in graphic communication (Cooperrider & Gentner, 2019; Ifrah, 2000) likely to be the conceptual space that

the very first historic systems of writing were trying to capture. Judgments of quantity have been studied extensively in developmental psychology and identified both implicit and explicit number systems (Carey & Barner, 2019). Infants already have an implicit proximate grasp of differences in number (Starr, Libertus, & Brannon, 2013) with one versus three being one of the most early comparisons possible, which is also frequently used as the most simple tasks in comparative work (Carey, 2000; Eckert, Bohn, & Spaethe, 2022; Xu & Arriaga, 2007). Preschoolers growing up in industrialized environments are constantly surrounded by math tasks in the form of various games (Daucourt, Napoli, Quinn, Wood, & Hart, 2021) as well as preschool curricula incentivizing a playful engagement with quantities, number symbols and shapes (Kucharz, 2012; Schmidt, Sauerbrey, & Smidt, 2021). By four years of age, children have a basic grasp of the cardinal principle, that is basic understanding of the relation of number words and fixed quantities (Carey & Barner, 2019; Silver & Libertus, 2022) and they are able to use that knowledge in practical everyday contexts, for example when comparing numbers with magnitudes smaller than five (Wynn, 1990). In match-to-sample tasks, children spontaneously infer number as a basis of comparison at 48–60 months (Kroupin & Carey, 2022b).

From this review of children’s ability to appreciate various basic dimensions of graphic depictions and their sensitivity to these as means for conveying information and disambiguating reference, we can identify a couple of research desiderata. Most striking is the literature’s focus on similarity and iconicity (Callaghan & Corbit, 2015). Mappings between symbol and referent can take place along a number of conceptual dimensions and few studies explicitly address these. Against this background, a systematic and comprehensive investigation of the conceptual dimensions along which graphic symbols can be imbued with meaning, will help to evaluate the extent to which children’s competence is specific to certain types of symbol-referent-relationships or generalizes across mappings based on analogy in form, size, number, or Gestalt

principles. The strong focus on iconicity often goes together with the graphic stimulus depicting actual real world objects. Here familiarity with the depicted object, its canonical features and even conventional ways of drawing them may interfere with the aim of investigating cognitive capacities rather than knowledge of culturally imbued graphic conventions, script or icons. By contrast, our work investigates young children's spontaneous comprehension of unfamiliar graphic symbols that gain meaning from being produced in a specific and highly restricted context where symbol and referent align in certain conceptual dimensions with decreasing iconicity. Therefore, our work focuses on children's pragmatic competencies and their ability to identify shared features across graphic displays used as target and referent. To the extent that children are highly unlikely to have encountered the specific graphic cues used here, our findings showcase their ability to grasp a novel symbolic practice and acquire novel symbols. This may limit the influence of children's knowledge of conventional symbols on performance and make our results more robust and informative for research about the acquisition or emergence of communicative systems despite spanning an age-range where children become competent users of script, numerals or drawings. Notably, the producers' cooperative intentions are key in individuating the symbols' meaning within the practical context of the coordination game employed in the studies below (Namy & Waxman, 2017), which has been shown to be key for success in other symbolic tasks such as the scale-model task (DeLoache, 2011), maps (Liben & Downs, 2013), and drawings (Callaghan, 2000).

General Methods

All three studies presented below share methodological and analytic approaches. 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). All stimuli, experiments, data and analyses are available at: https://github.com/GregorKachel/symnit_rep_manuscript.

Data Collection and Setup

In order to continuously trace the development of symbolic competences across the preschool years, for each study we aimed to test two children per month of age between the third and the seventh birthday for a total of 96 participants 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. All participants were recruited in Leipzig, 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 Max-Planck-Institute for Evolutionary Anthropology, Leipzig. 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, which provides a very intuitive, highly

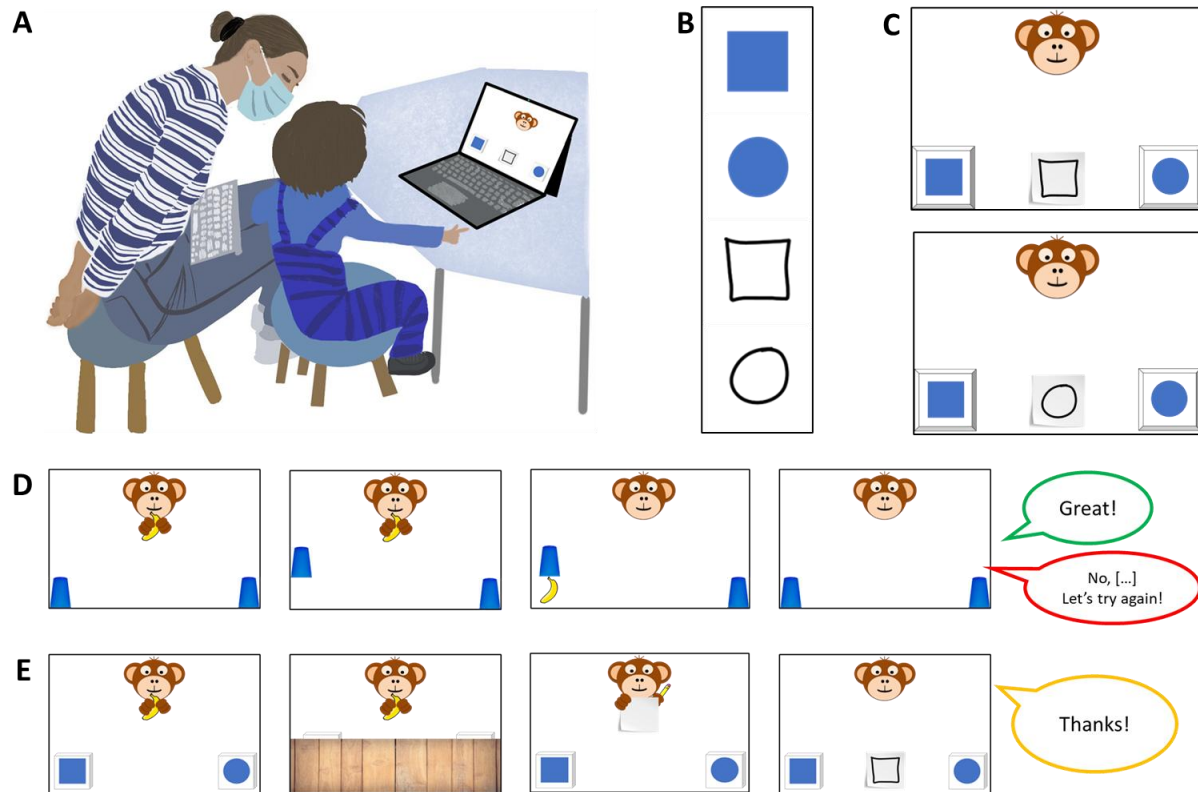


Figure 1. Setup, Stimuli and Trial Structure. Panel A shows the setup during testing. Experimenters were sitting behind the children in order to not distract them while supervising data collection. Panel B shows all elements contributing to an item of a possible test trial consisting of two target and one of two possible cues. Panel C exemplifies how two targets and one of the cues can be arranged in a specific test trial. Panel D shows a familiarization trial and Panel E shows a test trial.

controlled and efficient testing environment (Frank et al., 2016; Zack et al., 2009). 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 A for an illustration of the setup.

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. figure 1 D). 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. figure 1 E).

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. Unique graphic displays were used on every trial. 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 B. 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 1 B and C. 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 2. The first box in Panel B 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 children's binary choices (0 = incorrect, 1 = correct) to estimate the probability of interpreting cues correctly and how this probability varied as a function of age and condition. We fitted Bayesian logistic generalized linear mixed models (GLMMs) with fixed effects for condition, standardized age, their interaction, as well as trial number, and sex. To account for subject-level variability, we included random intercepts and random slopes for trial number by participant. Age and trial number were standardized prior to modeling. 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 (Stan Development Team, n.d.) 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 followed the approach of House and colleagues (House, 2017; House et al., 2013) and plotted the model estimates graphically as a smoothed curve surrounded by the respective confidence intervals. We 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 B 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 of information provided (*Pars Pro Toto*), or the amount of similarity between symbol and referent (*Simple Form Analogy*, *Complex Form Analogy*). We hypothesized (and preregistered) that as a group children will first succeed with *Representation*, then *Pars Pro Toto*, *Simple Form Analogy* and finally *Complex Form Analogy*.

Stimuli

To make the four conditions in study 1 as comparable as possible, they are all employing the same target stimuli (cf. figure 2, panels A and 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 two items of *Pars pro Toto* due to the canonical shapes (square, circle) it is less obvious with the less canonical shapes albeit they are vertically and horizontally symmetrical on purpose (cf. 2 B, box 2). 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, O'Madagain, Haun, & Bohn, 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 2 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 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. An additional exploratory analysis included an item level random effect (Model notation `correct ~ condition*z.age + z.trial + z.sex + (z.trial|id) + (z.age|item)`).

Results

Posterior predictive checks (PPC) for both full and null model indicated excellent fit of observed data and model predictions. 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 = 6222, range 4323 - 10098) and the null model (Bulk ESS, mean = 5777, range 4351 - 8307) were > 1000 , indicating reliable posterior estimations.

Comparing the models using weights based on the Widely Applicable Information Criterion (WAIC) yielded 74.52% of the model weight for the full model, and 25.48% 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 = -901.11; null model = -903.53). The standard error of the difference in predictive accuracy (SE = 3.14) is not much larger than the difference itself (ELPD diff = -2.42). Hence, 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, -1.03]) 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 ($\beta = -0.14$, 95% CrI [-0.54, 0.28]). 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 ($\beta = 0.30$, 95% CrI [-0.05, 0.67]), 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.

Figure 2. Stimuli and Developmental Trajectories for Study 1. Panels illustrate (A) an example stimulus combination (distractor, cue, target), (B) all items of the respective conditions and (C) results. 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 ensured by the 75% criterion of correct choices employed in the familiarization phase. Exclusions due to this familiarization criterion were comparably high at the lower end of the age-range (cf. Appendix A, 6) suggesting that testing younger children would require adjustments. The setup presented here worked perfectly despite providing the same task across an age-range of four years from the end of toddlerhood up into school age. The results for *Representation* further illustrate that even the youngest children in the sample understand a representational symbol-referent relationship as operationalized here - with full-color target shapes and loose outline drawings as cues - in the context of the picture-book-style object-choice-task at hand. Finally, the CrI around the estimated mean in the graphic presentation is getting narrower, showing that - as children get older - they are behaving more uniformly, providing a baseline for the uncertainty that can be expected with the current operationalization. In summary, *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 remain demanding for the oldest children in the sample, and may require a second to solve or describe verbally - even for adults. They are based 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 stimuli that are easier to grasp in line with the literature on analogical reasoning (Gentner, 1988; Richland, Morrison, & Holyoak, 2006). Finally, the 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 trials. 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 align on the same axis. 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 ellipses 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 3.

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 as well as an exploratory model including an item-level random effect (Model notation `correct ~ condition*z.age + z.trial + z.sex + (z.trial|id) + (z.age|item)`).

Results

For both full and null model, PPCs indicate excellent fit of observed data and model predictions. 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 = 6149, range 3485 - 7545) and the null model (Bulk ESS, mean = 6877, range 3839 - 10040) were > 1000, indicating reliable posterior estimations.

To compare model performance, we evaluated the WAIC estimates. The null model showed a slightly better predictive performance (ELPD = -888.15) compared to the full model (ELPD = -889.99). WAIC values also indicate better performance of the null model (WAIC = 1,776.30) over the full model (WAIC = 1,779.99). However, the difference between models (ELPD Diff = -1.84) falls within the bounds of uncertainty (SE = 1.76), 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, 0.08]). Performance in *Orientation of Object* ($\beta = -0.10$, 95% CrI [-0.46, 0.27]) and *Orientation of Feature* ($\beta = -0.12$, 95% CrI [-0.48, 0.23]) was not substantially different from the reference category when considering the full age range. Interaction terms between age and condition 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 45 months, making it the easiest task in study 2. Then in quick succession, children succeed in *Orientation of Feature* at 46 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 B. See appendix B table 3 and table 4 for an overview of model coefficients.

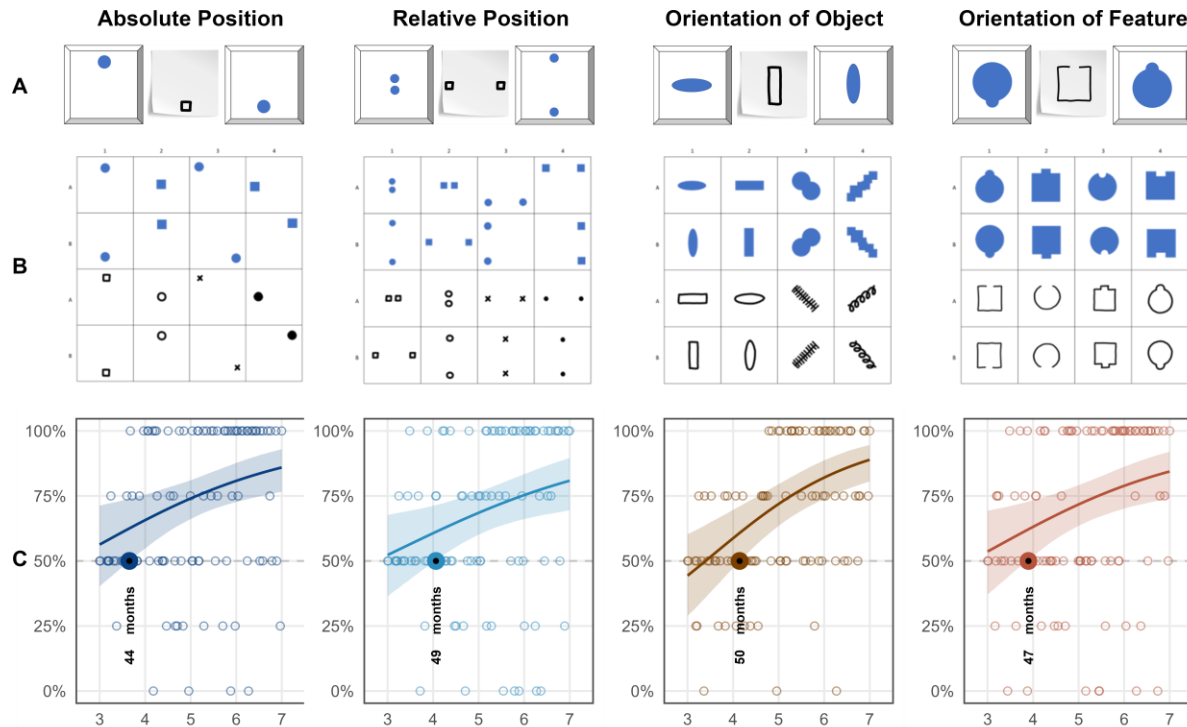


Figure 3. *Stimuli and Developmental Trajectories for Study 1.* Panels illustrate (A) an example stimulus combination (distractor, cue, target), (B) all items of the respective conditions and (C) results. Coloured lines indicate smoothed mean performance by age. Shaded areas represent 95% CIs. The dashed line demarcates chance level and the dots represent individual means. The coloured dots and annotation indicate when children's 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 reference.

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 4 B). 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 1 versus 3, and 2 versus 4. Already in infancy children can discriminate 1 versus 3 (Feigenson, Carey, & Hauser, 2002). Three- and four-year-olds can reliably distinguish these magnitudes explicitly and implicitly without counting (Halberda & Feigenson, 2008; Huntley-Fenner & Cannon, 2000; Mix, Huttenlocher, & Levine, 2002). 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 4.

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)), a null model lacking the interaction of condition and age and an additional exploratory model including an item-level random effect (Model notation correct ~ condition*z.age +z.trial +z.sex +(z.trial|id) +(z.age|item)).

Results

PPCs indicated excellent fit of observed data and model predictions in both models. 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 = 6552, range 4556 - 8586) and the null model (Bulk ESS, mean = 7253, range 5531 - 8885) were > 1000, indicating reliable posterior estimations.

When comparing performance, the full model showed a better fit (ELPD = -941.75) relative to the null model (ELPD = -945.94). The WAIC values also favored the full model (WAIC = 1,883.49) over the null model (WAIC = 1,891.88). Despite the slightly lower WAIC and higher ELPD of the full model, the difference in predictive accuracy (ELPD Diff = -4.19) remains almost within the range of sampling uncertainty (SE = 3.97). 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 [0.06, 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 [-1.06, -0.36]). A smaller, marginal effect was observed in *Number of Feature* ($\beta = -0.31$, 95% CrI [-0.67, 0.05]). *Number of Object* ($\beta = -0.08$, 95% CrI [-0.43, 0.28]) did not differ reliably from *Size of Object*. Age moderated performance less strongly in *Size of Feature* ($\beta = -0.50$, 95% CrI [-0.82, -0.20]) and *Number of Feature* ($\beta = -0.29$, 95% CrI [-0.60, 0.03]), 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 5, appendix B. For alternative analyses binning children by year of age, please see table 6 (ibid).

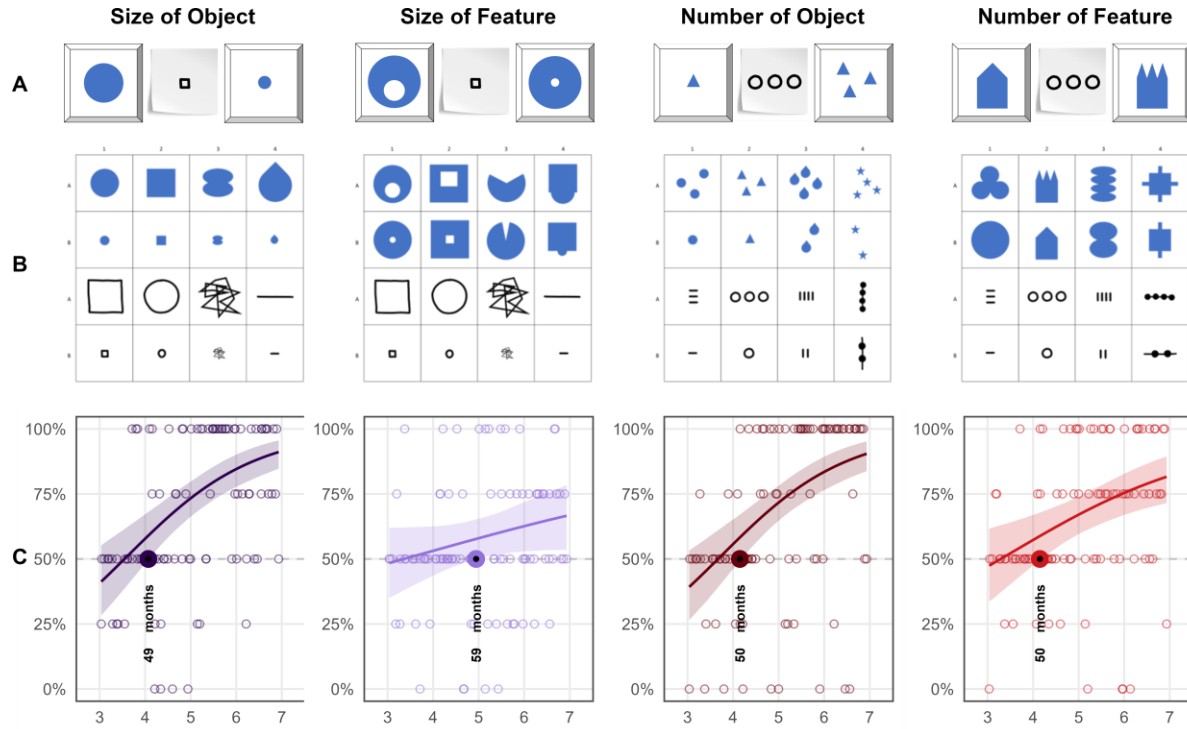


Figure 4. *Developmental Trajectories for all Conditions in Study 3.* Panels illustrate example stimulus combinations (distractor, cue, target) and results. Coloured lines indicate smoothed mean performance by age. Shaded areas represent 95% CrIs. The dashed line demarcates chance level and the dots 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. *Size of Feature* stands out as the most difficult condition in the

series of studies presented here with group-level success at 59 months. It is arguably 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. *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. Taken together, results further indicate 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.

Discussion

We presented a concerted three-part research project tracing young children's ability to spontaneously comprehend unfamiliar graphic symbols in an object-choice task across a wide variety of symbol-referent relationships. In three studies with four conditions each, we collected data of two children per month of age continuously from the third to the seventh birthday for a final sample of 304 children (Study 1, $N = 106$; Study 2, $N = 99$; Study 3, $N = 99$). Using Bayesian GLMMs, we were able to determine the month of age at which children as a group perform above chance with each symbol-referent relationship. Across the various conditions, symbolic reference between cue and target was established via direct representation, a part-whole relationship, analogies in size, number and shape, as well as gestalt-like principles like figure-ground relationships, orientation and alignment. While previous work on graphic communication

has focused almost exclusively on iconicity as a way of creating meaning and compared performance across binned age-groups, the three studies presented here map development continuously using a highly simple and streamlined procedure across various conceptual dimensions.

In Study 1, we found that - with cues that are direct outline drawings of the items they refer to - children are robustly better than chance at the lower boundary of our age-window. Representation was the most basic condition in the set and yields above-chance performance at 36 months. This demonstrates that the general task design and setup are sufficiently clear even for the youngest children in the sample, and that at the latest by three years of age children understand the pragmatics of a representational graphic cue in the context of the picture-book-style object-choice-task at hand. A second condition, Pars Pro Toto, used the exact same stimuli, but the cues provided only half of the respective figures to guide children's choices. While cues are highly similar and targets even identical, children solve Pars Pro Toto no sooner than at 38 months of age. This demonstrates that by three years of age, children's symbolic competence is still fragile if additional steps such as the straightforward completion of canonic shape is necessary for referential inference. Two further conditions aimed at abstracting from the cue representations while still retaining characteristic features of the target shapes with regard to an analogy in form, namely whether they were more round or edgy. Simple Form Analogy provided relatively simple shapes as cues with success at 47 months, and Complex Form Analogy more elaborate patterns with success at 50 months. To the best of our knowledge, this is the first time a developmental setup features abstract shape as a means of disambiguation in graphic communication and does so with an analysis open to detecting the advent of the competence necessary to solve this problem. Taken together, study 1 provides a within-subjects comparison of young children's ability to read representational, metonymical and metaphorical uses of form

in graphic communication, showing that while iconicity marks the entry point for the comprehension of graphic cues, abstract representations are only accessible by the fourth year even in the highly restricted referential space provided by a two-option object-choice coordination problem.

Study 2 explored whether children can infer symbol-referent relationships based on orientation and position. In all of the four conditions, cue and target had different shapes, which strictly limited the role of iconicity on the level of the individual cues used in the stimulus compositions. In Absolute Position, cue and target placement were aligned in their frame of presentation (e.g. the paper that the symbol was drawn on) and the symbol-referent relation was established by the respective positions in their immediate visual context. Here children succeeded at 44 months of age. In Relative Position, the target and cue were composed of two figures each that were either closer together or further apart, or aligned on a vertical or horizontal axis. Here children perform above chance at 49 months of age. Success even prior to the fourth birthday in these conditions is remarkable as it requires children to identify and compare relations (Absolute Position – figure/ground; Relative Position – figure/figure) which is demanding even later throughout the preschool years in other contexts (Shivaram, Shao, Simms, Hespos, & Gentner, 2023). Orientation of Object featured elongated shapes that were oriented either horizontally or vertically while cue and target shared the same direction. Previous research employing a same-different task indicates that children are sensitive to spatial alignment along a common axis in juxtaposed graphic displays no sooner than six to eight years of age (Zheng, Matlen, & Gentner, 2022). In the straightforward context of our study, children succeeded at 50 months. Contrary to our preregistered prediction, children succeed even earlier - at 47 months - in the complementary condition Orientation of Feature. In this condition, the cues and targets had a feature like a bump or an opening that is oriented either up- or downward. That the orientation of a feature of a

graphic display (bump, opening, indentation) appears more salient than its overall orientation is surprising but aligns with recent work on children's spontaneous ability to read asymmetric shapes as unconventional arrow cues, already prior to 48 months (Kachel et al., in prep.). In summary, Study 2 addresses an interesting caveat in previous research by presenting stimulus sets where symbol and referent feature different shapes but are composed in similar ways. While this requires uncovering abstract relations, these relations are by design still based on resemblance on the level of the overall composition, and hence, cover a middle ground between iconic and abstract representations that yet has not been operationalized in developmental work on graphic communication.

Since all four conditions in study 2 have an iconic aspect to them. Study 3 set out to further reduce the amount of iconicity by drawing abstract analogies in size and number. In the final study, the stimuli used as cue and target/distractor had nothing in common except for their relative magnitude in size and number. Number of Object featured graphic displays with one or three items per target or referent. Number of Feature reuses the same cues but employs targets varying in number with regard to salient features (e.g., spikes or dents). Children succeed in both conditions at 50 months. A crucial advantage of the study design and analytical approach employed here is not only that it allows to determine when children as a group are able to solve the respective conditions, but it also provides a comparison of the respective developmental trajectories. For example, performance rates on Number of Feature and Number of Object share the same cut-off for group-level success, but the curve for the number of features items has a more gradual growth, suggesting that it may still be slightly more demanding. A different pattern emerges for the two conditions featuring size relationships. Children succeed at 49 months of age in Size of Object, but ten months later in Size of Feature in which target and distractor figures have a small or large opening, bump or hole. This aligns with research showing that while

preschoolers are generally sensitive to the canonical size of objects depicted in graphical displays (Bloom & Markson, 1998; Cox, 2005; Long et al., 2019) they are much less likely to spontaneously extract size relations among abstract shapes (Kroupin & Carey, 2022b).

A set of major findings emerges from the studies presented here. First, our results robustly confirm that symbol-referent relationships are easiest for children when they are based in iconicity and become gradually more difficult as visual similarity decreases (Ganea et al., 2008). The findings presented here perfectly illustrate how children are generally able to grasp and benefit from graphic means of communication but that access to conceptual information appears to develop later in the preschool years (Callaghan, 2000). These findings are of practical relevance for early childhood education as they showcase children's ability to use various expressive dimensions in the graphic domain and can guide the design of pedagogical materials by showing specifically how accessible various ways of conveying meaning in graphic design may be at various ages.

Second, the reduced accuracy in analogy-based and feature-referenced tasks are both consistent with evidence that younger children tend to focus on surface features rather than relations (Gentner & Hoyos, 2017; Rattermann & Gentner, 1998), making tasks that require attentional shifting or nested comparison structures particularly demanding (Halford, Wilson, & Phillips, 1998). Relational or abstract as opposed to feature-based processing taxes working memory and inhibitory control more strongly (Gick & Holyoak, 1980; Goswami, 1991). This, in turn, points to aspects of cognitive maturation that are likely to foster children's performance in the tasks outlined above, as seen in work on the development of executive functions — particularly the early emergence and developmental trajectories of working memory, inhibitory

control, and cognitive flexibility in childhood [De Luca and Leventer (2010); Diamond (2013); reilly2022developmental].

Third, just around their fourth birthday at 50 months of age, children are generally able to identify various types of analogies spontaneously in unfamiliar symbolic displays. Previous findings have challenged assumptions of domain specificity in symbol development modalities such as graphic, gestural or verbal communication and rather highlighted the flexibility of symbolic and pragmatic abilities (Callaghan, 1999). Our findings champion a similar perspective with regard to various conceptual or basic graphic dimensions and underline the flexibility of symbolic cognition across various symbol-referent relationships within the graphic domain. For conditions that relied strongly on children to abstract from surface level features and focus on analogies in form, orientation, size and number, there is no strong evidence for sequences in the ages of success. Rather, children appear to master all symbol-referent relationships almost at the same point in time, which is highly indicative of a qualitative shift in pragmatic and analogical reasoning skills in the fourth year that drive children's ability for solving communicative coordination problems across various symbol-referent relationships.

To the extent that the studies presented above highlight the fourth birthday as a milestone in symbolic development, this connects to a rich literature on major developmental shifts in this important period of middle-childhood. As change at this age is again rapid and foundational across a variety of cognitive processes (Karmiloff-Smith, 1994; Rakoczy, 2022), several factors are likely to contribute to the convergence of above-chance performance across the experimental conditions used here. What has been described as the four-year-revolution in social cognition endows children with a full-fledged metarepresentational theory of mind and firmly developed belief-desire psychology (Gopnik & Astington, 1988; Rakoczy, 2022; Wellman, Cross, &

Watson, 2001). Theory of mind performance, in turn, was shown to be correlated with representational insight into the standing-for-relationship of pictorial representations (Callaghan, Rochat, & Corbit, 2012), as well as insight into communicative intent (Callaghan & Rochat, 2008), and qualitative shifts in perspective-taking abilities (Flavell, Everett, Croft, & Flavell, 1981). All of which are necessary or conducive for the flexible and spontaneous interpretation of graphic displays as well as inference-based or pragmatically challenging communicative behavior broadly construed. Specifically helpful for the tasks employed here may be developmental gains typically occurring around the fourth birthday in the domain of analogical reasoning (Christie & Gentner, 2014; Gentner, 1977; Shivaram et al., 2023), which are again fostered by parallelly growing capacities in working memory (Richland et al., 2006; Simms, Frausel, & Richland, 2018), executive functions (Bohn, Tessler, Kordt, Hausmann, & Frank, 2023; Thibaut, French, & Vezneva, 2010) and inhibitory control (Carlson & Wang, 2007; Jablonski, 2014).

While work on analogical reasoning abilities has highlighted that preschoolers are generally able to identify agreement in graphic displays and specifically involving size, number, and shape also at the age of four (Kroupin & Carey, 2022b, 2022a), this has mostly been established in match-to-sample paradigms where children are conventionally presented with three items and prompted with a phrase such as “which ones go together” directly inducing the cognitive operation tested. In our task, children make matches and comparisons across various dimensions without any verbal scaffolding addressing the match-making process or the dimension of comparison and, hence, showcases abilities for implicit judgements across a wide variety of basic conceptual and visual features. Therefore, our work provides a crucial extension to the literature on analogical reasoning by showing that children can actually put their developing skills to work in the practical context of a versatile communicative coordination game requiring both relational reasoning and pragmatic inference.

Taken together, our work provides one of the most comprehensive, robust and systematic investigations of young children's ability to find meaning in unfamiliar graphic symbols and may provide a benchmark data set for the development of preschoolers symbolic competence. Yet by design, the project is exploratory and while it traces symbolic development across the preschool years, it has limitations that future research needs to address. A crucial problem is that the project aims to establish robust developmental trajectories but does so only in a single, industrialized population (J. Henrich, Heine, & Norenzayan, 2010) that is spending their days immersed in symbolic artefacts and being educated almost daily by preschool curricula geared to prepare them for script-centered formal education. Future work needs to address the role of enculturation and specifically the influence of permanent exposure to symbolic and pictorial illustrations, writing and numbers or depictions featuring canonical shapes - ideally doing so both by investigating children in various cultural settings, as well as less industrialized communities, with a much more limited access and exposure to symbolic artefacts in their daily routines (Callaghan et al., 2012; Dehaene, Izard, Pica, & Spelke, 2006; Martlew & Connolly, 1996; Zhu, Nduku, et al., 2025; Zhu, Pitchik, et al., 2025). Second, having established several milestones for symbolic development across various symbol-referent relationships in a cross-sectional design, we can only explain patterns or shifts by reference to developments occurring at similar developmental stages. Our work needs to be complemented with studies employing an individual-differences perspective to untangle the relative contributions of knowledge about various conventional systems of reference, pragmatic skills and the growth of basic level cognitive capacities. Finally, it would be highly valuable to complement the current results on the comprehension of unfamiliar symbols by establishing at what ages children would be able to use their skills in the production of graphic communication and, thereby, inform theories on the development of communication systems broadly construed. Currently, our work highlights that iconicity is key in establishing new

communication systems (Fay, Ellison, & Garrod, 2014) and that more abstract systems of reference may then build on these (Cooperrider & Gentner, 2019; Gentner & Asmuth, 2019). Finally, findings support experimental work on the development of communication systems by showing children at four years of age are already extremely flexible and skilled communicators ready to spontaneously decode novel means and systems of communication without explicit instruction or training (Bohn, Kachel, & Tomasello, 2019; Raviv & Arnon, 2018). It would be extremely interesting to see at what ages and under what specific set of circumstances (Nölle, Fusaroli, Mills, & Tylén, 2020; Nölle & Spranger, 2022) young children may begin to use various dimensions of graphic communication for disambiguating reference (Nölle, Staib, Fusaroli, & Tylén, 2018) or establishing a even compositional communication systems (Bohn, Kachel, et al., 2019) when producing signs themselves. Our work here demonstrates that at least by the age of four, they would be apt receivers of symbolic creations whose reference is based on iconicity, metonymy, alignment, orientation, as well as analogies in form, size and number.

The work presented here provides a clear and methodologically rigorous account of early symbolic development. Its straightforward streamlined design, conceptual breadth, robust sample, and open but fine-grained statistical approach offer a reliable foundation for precisely the future work outlined here and already address an important caveat in our understanding of children's growing aptitude in graphic communication: the conceptual space spanning from iconic to abstract graphical representation.

References

- Allen Preissler, M., & Carey, S. (2004). Do both pictures and words function as symbols for 18- and 24-month-old children? *Journal of Cognition and Development*, 5(2), 185–212.
- Bermúdez, N. (2020). Sound symbolism. *The International Encyclopedia of Linguistic Anthropology*, 1–3.
- Blades, M., & Spencer, C. (1994). The development of children’s ability to use spatial representations. *Advances in Child Development and Behavior*, 25, 157–199.
- Bloom, P., & Markson, L. (1998). Intention and analogy in children’s naming of pictorial representations. *Psychological Science*, 9(3), 200–204.
- Bohn, M., Call, J., & Tomasello, M. (2019). Natural reference: A phylo-and ontogenetic perspective on the comprehension of iconic gestures and vocalizations. *Developmental Science*, 22(2), e12757.
- 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.
- Bohn, M., Tessler, M. H., Kordt, C., Hausmann, T., & Frank, M. C. (2023). An individual differences perspective on pragmatic abilities in the preschool years. *Developmental Science*, 26(6), e13401.
- Bruner, J. S. (1996). *The culture of education*. Harvard University Press.

- Bürkner, P.-C. (2017). Brms: An r package for bayesian multilevel models using stan. *Journal of Statistical Software*, 80, 1–28.
- Callaghan, T. (1999). Early understanding and production of graphic symbols. *Child Development*, 70(6), 1314–1324.
- Callaghan, T. (2000). Factors affecting children's graphic symbol use in the third year: Language, similarity, and iconicity. *Cognitive Development*, 15(2), 185–214.
- Callaghan, T., & Corbit, J. (2015). The development of symbolic representation. In R. M. Lerner (Ed.), *Handbook of child psychology and developmental science*. Wiley.
<https://doi.org/10.1002/9781118963418.childpsy207>
- Callaghan, T., & Rochat, P. (2008). Children's understanding of artist-picture relations: Implications for their theories of pictures. In C. Milbrath & H. M. Trautner (Eds.), *Children's understanding and production of pictures, drawings, and art: Theoretical and empirical approaches* (pp. 187–205). Hogrefe & Huber Publishers.
- Callaghan, T., Rochat, P., & Corbit, J. (2012). Young children's knowledge of the representational function of pictorial symbols: Development across the preschool years in three cultures. *Journal of Cognition and Development*, 13(3), 320–353.
- Carey, S. (2000). The origin of concepts. *Journal of Cognition and Development*, 1(1), 37–41.
- Carey, S., & Barner, D. (2019). Ontogenetic origins of human integer representations. *Trends in Cognitive Sciences*, 23(10), 823–835.
- Carlson, S. M., & Wang, T. S. (2007). Inhibitory control and emotion regulation in preschool children. *Cognitive Development*, 22(4), 489–510.

Christie, S., & Gentner, D. (2014). Language helps children succeed on a classic analogy task.

Cognitive Science, 38(2), 383–397.

Cleveland, W. S., & McGill, R. (1984). Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American Statistical Association*, 79(387), 531–554.

Cooperrider, K., & Gentner, D. (2019). The career of measurement. *Cognition*, 191, 103942.

Cox, M. (2005). *The pictorial world of the child*. Cambridge University Press.

Ćwiek, A., Fuchs, S., Draxler, C., Asu, E. L., Dediu, D., Hiovain, K., et al.others. (2022). The bouba/kiki effect is robust across cultures and writing systems. *Philosophical Transactions of the Royal Society B*, 377(1841), 20200390.

Dağlı, Ü. Y., & Halat, E. (2016). Young children's conceptual understanding of triangle. *Eurasia Journal of Mathematics, Science and Technology Education*, 12(2), 189–202.

Daucourt, M. C., Napoli, A. R., Quinn, J. M., Wood, S. G., & Hart, S. A. (2021). The home math environment and math achievement: A meta-analysis. *Psychological Bulletin*, 147(6), 565.

De Luca, C. R., & Leventer, R. J. (2010). Developmental trajectories of executive functions across the lifespan. In *Executive functions and the frontal lobes* (pp. 57–90). Psychology Press.

Dehaene, S. (2009). *Reading in the brain: The new science of how we read*. New York: Penguin.

Dehaene, S., Izard, V., Pica, P., & Spelke, E. (2006). Core knowledge of geometry in an amazonian indigene group. *Science*, 311(5759), 381–384.

- DeLoache, J. S. (1995). Early understanding and use of symbols: The model model. *Current Directions in Cognitive Science*, 4, 109–113.
- DeLoache, J. S. (2004). Becoming symbol-minded. *Trends in Cognitive Sciences*, 8(2), 66–70.
- DeLoache, J. S. (2011). Early development of the understanding and use of symbolic artifacts. *The Wiley-Blackwell Handbook of Childhood Cognitive Development*, 2, 312–336.
- DeLoache, J. S., & Marzolf, D. (1992). When a picture is not worth a thousand words: Young children's understanding of pictures and models. *Cognitive Development*, 7(3), 317–329.
- DeLoache, J. S., Pierroutsakos, S. L., Uttal, D. H., Rosengren, K. S., & Gottlieb, A. (1998). Grasping the nature of pictures. *Psychological Science*, 9(3), 205–210.
- Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, 64(1), 135–168.
- Donald, M. (1991). *Origins of the modern mind: Three stages in the evolution of culture and cognition*. Harvard University Press.
- Donald, M. (2001). *A mind so rare: The evolution of human consciousness*. WW Norton & Company.
- Donnelly, N., Cave, K., Greenway, R., Hadwin, J. A., Stevenson, J., & Sonuga-Barke, E. (2007). Visual search in children and adults: Top-down and bottom-up mechanisms. *Quarterly Journal of Experimental Psychology*, 60(1), 120–136.
- Eckert, J., Bohn, M., & Spaethe, J. (2022). Does quantity matter to a stingless bee? *Animal Cognition*, 25(3), 617–629.

- Falkum, I. L., Recasens, M., & Clark, E. V. (2017). “The moustache sits down first”: On the acquisition of metonymy. *Journal of Child Language*, 44(1), 87–119.
- Fay, N., Ellison, M., & Garrod, S. (2014). Iconicity: From sign to system in human communication and language. *Pragmatics & Cognition*, 22(2), 244–263.
- Feigenson, L., Carey, S., & Hauser, M. (2002). The representations underlying infants’ choice of more: Object files versus analog magnitudes. *Psychological Science*, 13(2), 150–156.
- Ferry, A. L., Corcoran, M. G., Williams, E., Curtis, S. M., Gale, C. J., & Twomey, K. E. (2025). Bigger versus smaller: Children’s understanding of size comparison words becomes more precise with age. *Child Development*, 96(2), 492–507.
- Flavell, J. H., Everett, B. A., Croft, K., & Flavell, E. R. (1981). Young children’s knowledge about visual perception: Further evidence for the level 1–level 2 distinction. *Developmental Psychology*, 17(1), 99.
- Fort, M., Lammertink, I., Peperkamp, S., Guevara-Rukoz, A., Fikkert, P., & Tsuji, S. (2018). Symbouki: A meta-analysis on the emergence of sound symbolism in early language acquisition. *Developmental Science*, 21(5), e12659.
- Frausel, R. R., Silvey, C., Freeman, C., Dowling, N., Richland, L. E., Levine, S. C., ... Goldin-Meadow, S. (2020). The origins of higher-order thinking lie in children’s spontaneous talk across the pre-school years. *Cognition*, 200, 104274.
- Ganea, P. A., Pickard, M. B., & DeLoache, J. S. (2008). Transfer between picture books and the real world by very young children. *Journal of Cognition and Development*, 9(1), 46–66.
<https://doi.org/10.1080/15248370701836592>

- Gelman, S. A., & Ebeling, K. S. (1998). Shape and representational status in children's early naming. *Cognition*, 66(2), B35–B47.
- Gentner, D. (1977). Children's performance on a spatial analogies task. *Child Development*, 1034–1039.
- Gentner, D. (1988). Metaphor as structure mapping: The relational shift. *Child Development*, 47–59.
- Gentner, D., & Asmuth, J. (2019). Metaphoric extension, relational categories, and abstraction. *Language, Cognition and Neuroscience*, 34(10), 1298–1307.
- Gentner, D., & Hoyos, C. (2017). Analogy and abstraction. *Topics in Cognitive Science*, 9(3), 672–693.
- Gerhardstein, P., & Rovee-Collier, C. (2002). The development of visual search in infants and very young children. *Journal of Experimental Child Psychology*, 81(2), 194–215.
- Gick, M. L., & Holyoak, K. J. (1980). Analogical problem solving. *Cognitive Psychology*, 12(3), 306–355.
- Goldin-Meadow, S., & Feldman, H. (1977). The development of language-like communication without a language model. *Science*, 197(4301), 401–403.
- Gopnik, A., & Astington, J. (1988). Children's understanding of representational change and its relation to the understanding of false belief and the appearance-reality distinction. *Child Development*, 26–37.

Goswami, U. (1991). Analogical reasoning: What develops? A review of research and theory.

Child Development, 62(1), 1–22.

Halberda, J., & Feigenson, L. (2008). Developmental change in the acuity of the "number sense":

The approximate number system in 3-, 4-, 5-, and 6-year-olds and adults. *Developmental Psychology*, 44(5), 1457.

Halford, G. S., Wilson, W. H., & Phillips, S. (1998). Processing capacity defined by relational complexity: Implications for comparative, developmental, and cognitive psychology.

Behavioral and Brain Sciences, 21(6), 803–831.

Haun, D. B., & Tomasello, M. (2011). Conformity to peer pressure in preschool children. *Child*

Development, 82(6), 1759–1767.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral*

and Brain Sciences, 33(2-3), 61–83.

Henrich, J., & Muthukrishna, M. (2021). The origins and psychology of human cooperation.

Annual Review of Psychology, 72(1), 207–240.

Henrich, N., & Henrich, J. (2007). *Why humans cooperate: A cultural and evolutionary explanation*. Oxford University Press.

Heyes, C. (2018). *Cognitive gadgets: The cultural evolution of thinking*. Harvard University Press.

House, B. R. (2017). Diverse ontogenies of reciprocal and prosocial behavior: Cooperative

development in fiji and the united states. *Developmental Science*, 20(6), e12466.

- House, B. R., Silk, J. B., Henrich, J., Barrett, H. C., Scelza, B. A., Boyette, A. H., & Laurence, S. (2013). Ontogeny of prosocial behavior across diverse societies. *Proceedings of the National Academy of Sciences*, 110(36), 14586–14591.
- Huntley-Fenner, G., & Cannon, E. (2000). Preschoolers' magnitude comparisons are mediated by a preverbal analog mechanism. *Psychological Science*, 11(2), 147–152.
- Ifrah, G. (2000). *The universal history of numbers*. Harvill London.
- Imai, M., Kita, S., Nagumo, M., & Okada, H. (2008). Sound symbolism facilitates early verb learning. *Cognition*, 109(1), 54–65.
- Jablonski, S. (2014). Inhibitory control and literacy development among 3-to 5-year-old children. *L1-Educational Studies in Language and Literature*, 1–25.
- Kachel, G., O'Madagain, C., Haun, D., & Bohn, M. (in prep.). *Young children's comprehension of deixis in the graphic domain*.
- Karmiloff-Smith, B. A. (1994). Beyond modularity: A developmental perspective on cognitive science. *European Journal of Disorders of Communication*, 29(1), 95–105.
- Kellman, P. J., & Spelke, E. S. (1983). Perception of partly occluded objects in infancy. *Cognitive Psychology*, 15(4), 483–524.
- Knoeferle, K., Li, J., Maggioni, E., & Spence, C. (2017). What drives sound symbolism? Different acoustic cues underlie sound-size and sound-shape mappings. *Scientific Reports*, 7(1), 5562.

- Kroupin, I. G., & Carey, S. E. (2022a). The importance of inference in relational reasoning: Relational matching as a case study. *Journal of Experimental Psychology: General*, *151*(1), 224.
- Kroupin, I. G., & Carey, S. E. (2022b). You cannot find what you are not looking for: Population differences in relational reasoning are sometimes differences in inductive biases alone. *Cognition*, *222*, 105007.
- Kucharz, D. (2012). *Elementarbildung*. Weinheim, Basel: Beltz Verlag.
- Landau, B. (1986). Early map use as an unlearned ability. *Cognition*, *22*(3), 201–223.
- Legare, C. H. (2017). Cumulative cultural learning: Development and diversity. *Proceedings of the National Academy of Sciences*, *114*(30), 7877–7883.
- Liben, L. S., & Downs, R. M. (2013). The role of graphic representations in understanding the world. In *Visions of aesthetics, the environment & development* (pp. 139–180). Psychology Press.
- Liszkowski, U., Brown, P., Callaghan, T., Takada, A., & De Vos, C. (2012). A prelinguistic gestural universal of human communication. *Cognitive Science*, *36*(4), 698–713.
- Long, B., Moher, M., Carey, S., & Konkle, T. (2019). Real-world size is automatically encoded in preschoolers' object representations. *Journal of Experimental Psychology: Human Perception and Performance*, *45*(7), 863.
- Luquet, G.-H., & Costall, A. T. (2001). *Children's drawings (le dessin enfantin)*. International Specialized Book Services.

- Margiotoudi, K., & Pulvermüller, F. (2020). Action sound–shape congruencies explain sound symbolism. *Scientific Reports*, 10(1), 12706.
- Martlew, M., & Connolly, K. J. (1996). Human figure drawings by schooled and unschooled children in papua new guinea. *Child Development*, 67(6), 2743–2762.
- Maurer, D., Pathman, T., & Mondloch, C. J. (2006). The shape of boubas: Sound–shape correspondences in toddlers and adults. *Developmental Science*, 9(3), 316–322.
- McElreath, R. (2018). *Statistical rethinking: A bayesian course with examples in r and stan*. Chapman; Hall/CRC.
- Mix, K. S., Huttenlocher, J., & Levine, S. C. (2002). *Quantitative development in infancy and early childhood*. Oxford University Press.
- Moll, H., & Tomasello, M. (2007). Cooperation and human cognition: The vygotskian intelligence hypothesis. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 362(1480), 639–648.
- Monmonier, M. (1985). *Semiology of graphics: Diagrams, networks, maps*. JSTOR.
- Namy, L. L., & Waxman, S. R. (2017). Symbols redefined. In *Symbol use and symbolic representation* (pp. 269–278). Psychology Press.
- Nölle, J., Fusaroli, R., Mills, G. J., & Tylén, K. (2020). Language as shaped by the environment: Linguistic construal in a collaborative spatial task. *Palgrave Communications*, 6(1), 1–10.
- Nölle, J., & Spranger, M. (2022). From the field into the lab: Causal approaches to the evolution of spatial language. *Linguistics Vanguard*, 8(s1), 191–203.

- Nölle, J., Staib, M., Fusaroli, R., & Tylén, K. (2018). The emergence of systematicity: How environmental and communicative factors shape a novel communication system. *Cognition*, 181, 93–104.
- Piaget, J., Inhelder, B., & Bovet, M. (1997). *Mental imagery in the child: A study of the development of imaginal representation*. Routledge. Retrieved from <https://books.google.de/books?id=BXrHqqUYu0QC>
- Pierroutsakos, S. L., & DeLoache, J. S. (2003). Infants' manual exploration of pictorial objects varying in realism. *Infancy*, 4(1), 141–156.
- Porto, A., Basalyga, A., Huckle, M. N., Santiago, J., Fein, E., & Kranjec, A. (2024). Sound symbolism across diverse writing systems. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 46.
- Rakoczy, H. (2022). Foundations of theory of mind and its development in early childhood. *Nature Reviews Psychology*, 1(4), 223–235.
- Rattermann, M. J., & Gentner, D. (1998). More evidence for a relational shift in the development of analogy: Children's performance on a causal-mapping task. *Cognitive Development*, 13(4), 453–478.
- Raviv, L., & Arnon, I. (2018). Systematicity, but not compositionality: Examining the emergence of linguistic structure in children and adults using iterated learning. *Cognition*, 181, 160–173.

- 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.
- Schmidt, T., Sauerbrey, U., & Smidt, W. (2021). *Frühpädagogische handlungskonzepte: Eine wissenschaftliche bestandsaufnahme*. UTB.
- Shivaram, A., Shao, R., Simms, N., Hespos, S., & Gentner, D. (2023). When do children pass the relational-match-to-sample task? *Proceedings of the Annual Meeting of the Cognitive Science Society*, 45.
- Shusterman, A., Ah Lee, S., & Spelke, E. S. (2008). Young children's spontaneous use of geometry in maps. *Developmental Science*, 11(2), F1–F7.
- Sidhu, D. M. (2025). Sound symbolism in the lexicon: A review of iconic-systematicity. *Language and Linguistics Compass*, 19(1), e70006.
- Silver, A. M., & Libertus, M. E. (2022). Environmental influences on mathematics performance in early childhood. *Nature Reviews Psychology*, 1(7), 407–418.
- Simms, N. K., Frausel, R. R., & Richland, L. E. (2018). Working memory predicts children's analogical reasoning. *Journal of Experimental Child Psychology*, 166, 160–177.
- 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*.
- Stan Development Team. (n.d.). *RStan: The R interface to Stan*. Retrieved from <https://mc-stan.org/>

- Starr, A., Libertus, M. E., & Brannon, E. M. (2013). Number sense in infancy predicts mathematical abilities in childhood. *Proceedings of the National Academy of Sciences*, *110*(45), 18116–18120.
- Stevenson, C. E., Alberto, R. A., Boom, M. A. van den, & Boeck, P. A. de. (2014). Visual relations children find easy and difficult to process in figural analogies. *Frontiers in Psychology*, *5*, 827.
- Thibaut, J.-P., French, R., & Vezneva, M. (2010). The development of analogy making in children: Cognitive load and executive functions. *Journal of Experimental Child Psychology*, *106*(1), 1–19.
- Tomasello, M. (2009). *The cultural origins of human cognition*. Harvard university press.
- Tomasello, M. (2010). *Origins of human communication*. MIT press.
- Tomasello, M. (2014a). *A natural history of human thinking*. Cambridge, Massachusetts; London, England: Harvard University Press.
- Tomasello, M. (2014b). The ultra-social animal. *European Journal of Social Psychology*, *44*(3), 187–194. <https://doi.org/10.1002/ejsp.2015>
- Tomasello, M., Call, J., & Gluckman, A. (1997). Comprehension of novel communicative signs by apes and human children. *Child Development*, *68*(6), 1067–1080.
- Tomasello, M., Carpenter, M., & Liszkowski, U. (2007). A new look at infant pointing. *Child Development*, *78*(3), 705–722.

Tomasello, M., Striano, T., & Rochat, P. (1999). Do young children use objects as symbols?

British Journal of Developmental Psychology, 17(4), 563–584.

Uttal, D. H., & Wellman, H. M. (1989). Young children's representation of spatial information

acquired from maps. *Developmental Psychology*, 25(1), 128.

Valenza, E., Leo, I., Gava, L., & Simion, F. (2006). Perceptual completion in newborn human

infants. *Child Development*, 77(6), 1810–1821.

Verdine, B. N., Bunger, A., Athanasopoulou, A., Golinkoff, R. M., & Hirsh-Pasek, K. (2017).

Shape up: An eye-tracking study of preschoolers' shape name processing and spatial development. *Developmental Psychology*, 53(10), 1869.

Verdine, B. N., Lucca, K. R., Golinkoff, R. M., Hirsh-Pasek, K., & Newcombe, N. S. (2016). The

shape of things: The origin of young children's knowledge of the names and properties of geometric forms. *Journal of Cognition and Development*, 17(1), 142–161.

Vygotsky, L. S. (1980). *Mind in society: The development of higher psychological processes*.

Cambridge, MA: Harvard University Press.

Wellman, H., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The

truth about false belief. *Child Development*, 72(3), 655–684.

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

Wynn, K. (1990). Children's understanding of counting. *Cognition*, 36(2), 155–193.

Xu, F., & Arriaga, R. I. (2007). Number discrimination in 10-month-old infants. *British Journal*

of Developmental Psychology, 25(1), 103–108.

Zambrzycka, J., Kotsopoulos, D., Lee, J., & Makosz, S. (2017). In any way, shape, or form?

Toddlers' understanding of shapes. *Infant Behavior and Development*, 46, 144–157.

<https://doi.org/https://doi.org/10.1016/j.infbeh.2016.12.002>

Zheng, Y., Matlen, B. J., & Gentner, D. (2022). Spatial alignment facilitates visual comparison in children. *Annual Meeting of the Cognitive Science Society*.

Zhu, R., Nduku, T., Arieda, J. O., Verma, A., Fan, J. E., & Frank, M. C. (2025). Investigating children's performance on object-and picture-based vocabulary assessments in global contexts: Evidence from kisumu, kenya. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 47.

Zhu, R., Pitchik, H. O., Kilonzo, T. N., Engelmann, J., Fernald, L. C., & Gopnik, A. (2025). The development of picture comprehension across early environments: Evidence from urban and rural toddlers in western kenya. *Developmental Science*, 28(1), e13579.

Appendix

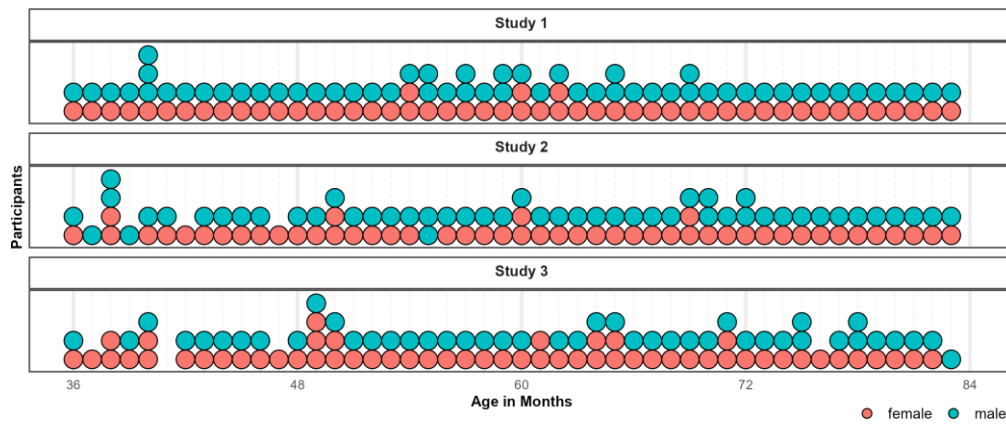


Figure 5. Distribution of Participants across the age-range in all three studies. Dots represent individuals. Colors indicate their respective sex.

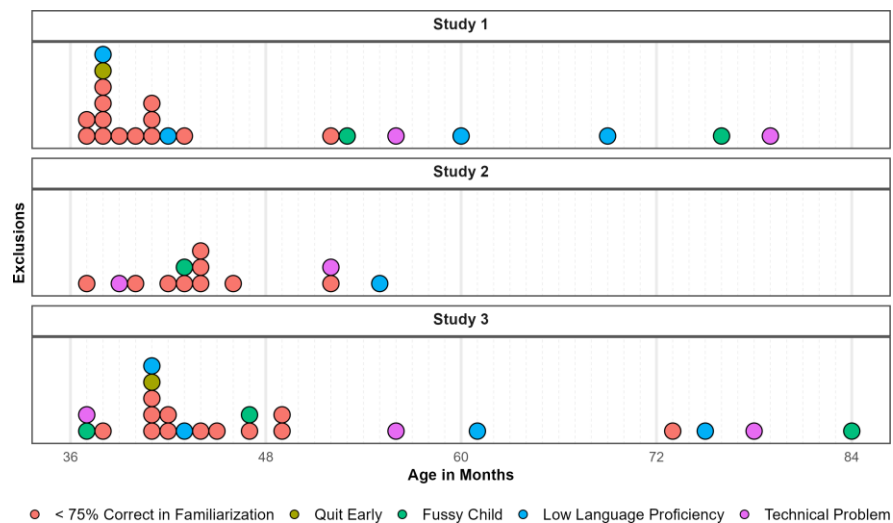


Figure 6. Distribution of Exclusions across the age-range in all three studies. Dots represent individuals. Colors indicate why children were not submitted to analyses.

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

| Predictor | Estimate | SD | MAD | 95% CrI | Bulk ESS | Tail ESS |
|------------------------------------|----------|------|------|----------------|----------|----------|
| Intercept | 2.03 | 0.20 | 0.20 | [1.64, 2.43] | 5,184 | 6,617 |
| Pars Pro Toto | -0.14 | 0.21 | 0.21 | [-0.54, 0.28] | 7,491 | 7,577 |
| Simple Form Analogy | -1.39 | 0.18 | 0.18 | [-1.75, -1.03] | 6,933 | 7,821 |
| Complex Form Analogy | -1.36 | 0.19 | 0.18 | [-1.73, -1.00] | 6,810 | 6,666 |
| Age* | 0.42 | 0.15 | 0.15 | [0.13, 0.72] | 4,323 | 5,827 |
| Trial* | -0.01 | 0.07 | 0.07 | [-0.15, 0.12] | 10,098 | 7,496 |
| Sex (Male) | -0.29 | 0.20 | 0.20 | [-0.70, 0.10] | 4,512 | 6,169 |
| Pars Pro Toto \times Age* | 0.30 | 0.18 | 0.18 | [-0.05, 0.67] | 5,856 | 7,147 |
| Simple Form Analogy \times Age* | -0.20 | 0.16 | 0.16 | [-0.52, 0.13] | 5,344 | 6,849 |
| Complex Form Analogy \times Age* | -0.08 | 0.17 | 0.17 | [-0.40, 0.25] | 5,673 | 7,198 |

Note. Estimates represent posterior means with 95% equal-tailed credible intervals (CrIs). MAD indicates Median Absolute Deviation. ESS refers to effective sample size. \hat{R} 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, 1.48] | 3,776 | 6,041 |
| Relative Position | -0.28 | 0.18 | 0.18 | [-0.64, 0.08] | 7,227 | 7,370 |
| Orientation of Object | -0.10 | 0.19 | 0.19 | [-0.46, 0.27] | 7,545 | 7,408 |
| Orientation of Feature | -0.12 | 0.18 | 0.18 | [-0.48, 0.23] | 7,402 | 7,559 |
| Age* | 0.40 | 0.14 | 0.14 | [0.14, 0.68] | 4,012 | 4,964 |
| Trial* | 0.26 | 0.08 | 0.08 | [0.10, 0.43] | 5,662 | 7,109 |
| Sex (Male) | -0.01 | 0.22 | 0.22 | [-0.45, 0.43] | 3,485 | 5,359 |
| Relative Position \times Age* | -0.06 | 0.16 | 0.16 | [-0.37, 0.25] | 7,445 | 7,754 |
| Orientation of Object \times Age* | 0.19 | 0.16 | 0.16 | [-0.12, 0.51] | 7,513 | 7,844 |
| Orientation of Feature \times Age* | -0.01 | 0.16 | 0.16 | [-0.32, 0.31] | 7,428 | 8,385 |

Note. Estimates represent posterior means with 95% equal-tailed credible intervals (CrIs). MAD indicates Median Absolute Deviation. ESS refers to effective sample size. \hat{R} 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 | 0.18 | [0.70, 1.38] | 5,074 | 6,931 |
| Size of Feature | -0.71 | 0.18 | 0.18 | [-1.06, -0.36] | 7,351 | 7,601 |
| Number of Object | -0.08 | 0.18 | 0.18 | [-0.43, 0.28] | 7,735 | 8,265 |
| Number of Feature | -0.31 | 0.18 | 0.18 | [-0.67, 0.05] | 7,429 | 7,348 |
| Age* | 0.70 | 0.14 | 0.13 | [0.44, 0.97] | 4,556 | 6,771 |
| Trial* | 0.21 | 0.08 | 0.08 | [0.06, 0.36] | 8,586 | 7,671 |
| Sex (Male) | -0.18 | 0.19 | 0.18 | [-0.55, 0.18] | 5,392 | 6,583 |
| Size of Feature \times Age* | -0.50 | 0.16 | 0.16 | [-0.82, -0.20] | 6,020 | 6,979 |
| Number of Object \times Age* | 0.00 | 0.16 | 0.16 | [-0.32, 0.32] | 6,741 | 7,976 |
| Number of Feature \times Age* | -0.29 | 0.16 | 0.16 | [-0.60, 0.03] | 6,633 | 6,930 |

Note. Estimates represent posterior means with 95% equal-tailed credible intervals (CrIs). MAD indicates Median Absolute Deviation. ESS refers to effective sample size. \hat{R} 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.