Young Children and Adults Use Reasoning by Exclusion Rather Than Attraction to Novelty to Disambiguate Novel Word Meanings

Natalie Bleijlevens and Tanya Behne

Department of Developmental Psychology, University of Göttingen, Germany

Leibniz ScienceCampus "Primate Cognition", Göttingen, Germany

Author Note

Natalie Bleijlevens https://orcid.org/0000-0002-4421-4153

Tanya Behne https://orcid.org/0000-0001-7649-4949

This study was preregistered on OSF; see https://osf.io/x9aej. The complete data, analysis code and experimental stimuli are accessible there as well (https://osf.io/8dpw4/). There is no known conflict of interest to disclose. This research project has been approved by the ethics committee of the Institute for Psychology, University of Göttingen (project numbers 300 & 301). There are no conflicts of interest to disclose. The research was funded by the Deutsche Forschungsgemeinschaft DFG (254142454/GRK 2070). We want to thank Roger Mundry for statistical support, as well as Liudmila Gromova, Larissa Benedetti for their help with the study implementation, Leonie-Zoe Schulz and Jana Tatian for their help with coding, and Marlene Meyer, Lia Künnemann, and Sofya Nechukhaeva for their help with the stimuli creation.

Correspondence concerning this article should be addressed to Natalie Bleijlevens, University of Göttingen, Waldweg 26, 37073 Göttingen, Germany. Email: natalie.bleijlevens@uni-goettingen.de

Abstract

Upon hearing a novel label, listeners tend to assume that it refers to a novel, rather than a familiar object. While this disambiguation or mutual exclusivity (ME) effect has been robustly shown across development, it is unclear what it involves. Do listeners use their pragmatic and lexical knowledge to exclude the familiar object and thus select the novel one? Or is the effect, at least in early childhood, simply based on an attraction to novelty and a direct mapping of the novel label to a novel object? In a pre-registered online-study with 2-to-3-year-olds (n=75) and adults (n=112), we examined i) whether relative object novelty alone (without pragmatic or lexical information) could account for participants' disambiguation and ii) whether participants' decision processes involved reasoning-by-exclusion. Participants encountered either a known and an unknown object (classic ME condition) or two unknown objects, one completely novel and one pre-exposed (novelty condition) as potential referents of a novel label. Reasoning-by-exclusion was assessed by children's looking patterns and adults' explanations. In the classic ME condition, children and adults significantly chose the novel object and both used reasoning-by-exclusion. In contrast, in the novelty condition, children and adults chose randomly. Across conditions, a retention test revealed that adults remembered their prior selections, while children's performance was fragile. These results suggest that referent disambiguation is not based on relative object novelty alone. Instead, to resolve referential ambiguity, both young children and adults seem to make use of pragmatic and/or lexical sources of information and to engage in reasoning-by-exclusion strategies.

Keywords: word learning, disambiguation, mutual exclusivity, disjunctive syllogism, reasoning by exclusion, novelty

Public significance statement: From early on, children seem to find the referents of novel words with relative ease. This study found that young toddlers' success is not driven by a simple attraction to the novelty of certain referents. Instead, they seem to use their pragmatic and lexical knowledge, in combination with their logical reasoning abilities, to exclude unlikely referents of novel words.

Young Children and Adults Use Reasoning by Exclusion Rather Than Attraction to Novelty to Disambiguate Novel Word Meanings

Young word learners face many challenges. Ambiguity is a constant part of their learning environment – whenever they are confronted with a new label, their surroundings offer multiple possible referents (Quine, 1960). In a world full of environmental impressions and unknown factors, they need to learn which kind of information is relevant in a given situation, how to make inferences based on their experiences and how to store the important information for later use.

Despite all the ambiguity and uncertainty young children encounter, they seem to link novel words to their correct referents with relative ease. For example, when asked to find the "dax" while facing a well-known object (e.g., car) and an unknown one, children consistently choose the unknown object as the correct referent (Markman & Wachtel, 1988). This behavioral response, also known as the *disambiguation* or *mutual exclusivity (ME) effect*, can be robustly found in children from 17 months of age (Lewis et al., 2020; see Pomiechowska et al., 2021 for evidence in even younger infants) as well as in adults (e.g., Halberda, 2006).

While the disambiguation effect itself is quite uncontroversial, the explanations for children's behavioral responses are not. Three theoretical proposals describe children as using different sources of information for identifying the correct referent. Two of them assume children to reason by exclusion: either (1) based on lexical constraints like the mutual exclusivity bias ("The car is already named car; thus, the dax must be the other one"; Markman & Wachtel, 1988) or (2) based on pragmatic context information regarding the speaker's intentions ("If she meant the car, she would have said so; thus, she must be referring to the other one"; Clark, 2015; Diesendruck & Markson, 2001). In contrast, the third proposal assumes children to directly map the novel word to the most novel object, based on their attraction to object novelty (Horst et al., 2011; Mather & Plunkett, 2012; Merriman et al., 1995).

Thus, while the first two proposals presume that children's word learning relies on some form of reasoning by exclusion, implying deductive inferences, the latter suggests a positive link between label and object. How can we distinguish between children's use of reasoning by exclusion vs. the direct positive mapping of words to objects? In the classic ME set-up, the same response would be predicted by all three proposals. However, there are two approaches to discern the underlying strategy: 1) manipulating the sources of information available for disambiguation and 2) analyzing participants' responses with regard to indications of reasoning by exclusion. In the current project, we combined both approaches to get a complete picture of children's behavior in the ME task.

Sources of Information Available for Disambiguation

The most parsimonious explanation for children's disambiguation was offered by accounts focusing on cues of object novelty: Children might select the novel object in the disambiguation task not because they rule out the familiar distractor (based on either social-cognitive inferences or the distractor's nameability), but because of the target's novelty (Horst et al., 2011; Mather & Plunkett, 2012; Merriman et al., 1995). But is object novelty alone sufficient to drive the disambiguation effect? Previous research has revealed a mixed pattern regarding this question.

Merriman and colleagues first demonstrated the role of object novelty for word disambiguation: Upon pre-exposing the unfamiliar (target) object, the disambiguation effect in 2-year-olds (but less so in older children) disappeared or even reversed (Merriman et al., 1989; Merriman & Schuster, 1991). These researchers argued that children's behavior was guided by a "feeling of novelty" principle, i.e., the "expectation that new names will map onto physical entities that feel new" (Merriman et al.,1995). They stressed that this was specific to word disambiguation, since children selected the unfamiliar object more often when asked for a label than in a no-label control condition (Merriman & Schuster, 1991).

Since then, a couple of studies have examined the role of novelty, using a more stringent manipulation, i.e., the presentation of several unknown objects that differed in their relative (token) novelty. For example, Horst et al. (2011), used an object selection task and manipulated the relative novelty of several unfamiliar objects by pre-exposing some of them before test. Twenty-four-month-olds selected a completely novel object over some unfamiliar, but pre-exposed ones when being asked for a novel word (see Dysart et al., 2016 for similar results in 3-year-olds). However, as Mather & Plunkett (2012) have stressed, it is difficult to distinguish whether children's selections in these studies reflect an intrinsic bias to link novel words to novel objects or whether they merely reflect children's interest in exploring novel objects (see Graham et al., 2005; Markman, 1991 for similar arguments, and see Kucker et al., 2020 for findings showing that children may even pick the novel object when asked for familiar one).

Thus, to account for such a baseline attraction to novelty, Mather and Plunkett (2012) ran two eye-tracking studies that examined children's attentional shifts to the novel object after label onset. The first of these studies found that, upon hearing a novel word, 22-month-olds shifted their gaze to the most novel object on the screen (versus a well-known object and an unknown, but pre-exposed object; experiment 1), beyond baseline preferences. However, in their second, more stringent test of the effect of object novelty, in which the novel and the pre-exposed object were in direct competition (and no familiar object was shown), they could not replicate this overall effect[1] (see also Bleijlevens et al., 2023; Graham et al., 2005; Marno, 2021 for findings that pragmatic factors, rather than novelty, drive young children's disambiguation). Based on such mixed patterns of results[2], and especially considering the

[1] In this more stringent test, Mather and Plunkett (2012) found a significant effect only in their third trial, but not in the first two trials nor in the overall analysis across trials.

[2] Note also that studies which presented the same speaker/voice to introduce the pre-exposed object and later label the novel one (e.g., Dysart et al., 2016; Mather & Plunkett, 2012) cannot exclude that children based their decisions on pragmatic inferences such as "If she meant the pre-exposed object, she would have named it earlier, so she must mean the novel one" instead of being guided by their attraction to object novelty alone.

current replication crisis in our field, we need more research to determine the source of information used for referent disambiguation.

Signs of Reasoning by Exclusion

Explanations of the disambiguation effect also differ with regard to the underlying decision process or mental computations they propose: Do listeners identify the intended referent by applying a process of elimination (not A, thus B) or do they link the novel word directly to the novel target object? Lexical and pragmatic accounts both suggest that listeners settle on the novel object by first excluding the familiar object (e.g., car) as a potential referent, though offering different reasons for this elimination ("It cannot be the car, because the car already has a name" vs. "If the speaker meant the car, she would have said so"). In contrast, accounts that focus on object novelty may not necessarily expect children to exclude a referent: Merriman and colleagues (1995), e.g., argued that their proposed "feeling of novelty" does not involve negation, but assumes children to make an association between the felt novelty of the stimulus and its name.

Halberda (2006) found a way to assess the process of elimination by measuring listeners' looking patterns. He showed that children (and adults) "double check" the distractor object before looking at the target object in the disambiguation task. Specifically, 3-to-4-year-olds and adults who happened to look at the target object while hearing the novel label switched their gaze to the distractor object before switching back to the target object. These findings suggest that children and adults may solve the disambiguation task by using a process of elimination. They leave open, however, the specific kind of deductive reasoning ability underlying this process, as well as its emergence in development.

Halberda (2006) interpreted 3-to-4-year-olds' use of double-checks as evidence for their application of the *disjunctive syllogism*: "A or B. Not A. Therefore B." Research using object search tasks suggests that it is not until 2.5 to 3 years of age that children engage in this kind of deductive inferences (Gautam et al., 2021; Mody & Carey, 2016). However, there are ongoing debates about whether specific task demands mask earlier competence, and whether tasks that already reveal competence in infants (Cesana-Arlotti et al., 2018) or non-human animals (Call, 2004) are open to alternative explanations (Feiman et al., 2022; Leahy & Carey, 2020). Similarly, it is debatable whether double checks reflect the same reasoning processes as the more complex object search tasks that were developed to measure this kind of deductive inferences (Gautam et al., 2021; Mody & Carey, 2016).

Due to the limited research (with rather small samples) on children's deductive reasoning abilities in the ME task and their emergence, it may be insightful to extend the approach by Halberda (2006) to a sample of 2-to-3-year-olds to see at which age children start using double-checks. Their patterns would thus be of interest for both investigating at which age children start to resolve referential ambiguity based on reasoning by exclusion, and the early emergence of deductive reasoning more generally.

When Do Children Learn Words After Successful Disambiguation?

While the main aim of the current project, as well as its sample size calculation, focused on listeners' disambiguation strategies, we acknowledge the importance of addressing a more recent debate: the question how children's disambiguation relates to the long-term learning of the novel label-object-link. A large number of studies showed that after successful disambiguation, children aged 2-3 years could remember a new word-object-mapping after different time delays (Bleijlevens et al., 2023; Carey & Bartlett, 1978; Jaswal & Markman, 2003; Markson & Bloom, 1997; Spiegel & Halberda, 2011; Vlach & Sandhofer, 2012; Zosh et al., 2013). However, findings from some other studies indicated that, at least at 24 months of age, children did not retain the label-object links, which they had successfully identified in disambiguation tasks, for even 5 minutes (but see Kalashnikova et al., 2018 for contradictory findings for 24-month-olds raised monolingually). Consequently, this led to the claim that children's disambiguation indicates a cognitive process that allows for a successful communication in situation time, while being more or less independent from word learning

(Kucker et al., 2015; McMurray et al., 2012; *c.f.*, Zosh et al., 2013). More data is needed to find consistent patterns across studies and disentangle which factors may contribute to children's word learning.

The Current Study

We ran a pre-registered, unmoderated online study (due to the Covid-19 pandemic) with 2-to-3-year-olds and adults. Our main study aim was to assess the mechanisms underlying referent disambiguation: Do children select referents by excluding distractor objects (based on lexical constraints or pragmatic inferences) or by a direct positive mapping of the novel label to the novel object (based on their attraction to novelty)?

To answer this question, we used two approaches. First, we manipulated the sources of information available for disambiguation. We presented children with a disambiguation task in which a speaker uses a novel label to ambiguously refer to one of two objects. Across conditions, we manipulated whether the two objects differed from each other in both novelty and nameability (Classic ME condition) or only in relative novelty (Novelty condition). In the Novelty condition, both objects were unknown, but one object was completely novel, while the other one was pre-exposed in a previous scene (without naming it) in a way that did not allow for a pragmatic interpretation of the subsequent labeling event.[3] If object novelty is sufficient to guide referent disambiguation, we expected participants, upon hearing a novel label, to select the novel object above chance in both conditions. If the additional (lexical/pragmatic) cues in the Classic ME condition support disambiguation beyond the pure attraction to novelty, we expected better performances and higher certainty (indicated by reaction times and/or social referencing) in the Classic ME than the Novelty condition. Additionally, we predicted the performance to increase with age in the Classic ME task (see,

[3] In some previous studies that stress the role of object novelty in disambiguation, the same speaker/voice introduced the pre-exposed object and later labeled the novel one (e.g., Dysart et al., 2016; Mather & Plunkett, 2012), allowing for the pragmatic inference "If she meant the pre-exposed object, she would have named it earlier".

e.g., Lewis et al., 2020), but to possibly decrease with age in the Novelty condition (see, e.g., Merriman et al., 1989).

The second approach was to assess whether participants show signatures of reasoning by exclusion during disambiguation in different contexts. Therefore, we planned to conceptually replicate children's "double checks" gaze patterns (Halberda, 2006) and, for the first time, extend it to a younger age range, as well as to different disambiguation contexts. Additionally, we measured signatures of reasoning-by-exclusion in adults' strategy descriptions. We expected reasoning-by-exclusion strategies in the Classic ME condition, and assessed whether the same applies to situations in which objects only differed in their situational novelty.

Our second (subordinate) aim was to assess if and in which contexts children and adults remembered the new word-referent-links after a 5-minute-delay. We predicted better learning performance in adults than in children, and (across ages) better performances in the Classic ME than the Novelty condition, based on findings questioning word learning after disambiguation based on novelty cues alone (e.g., Kucker et al., 2020).

Method

We preregistered the experimental design, procedure, sample sizes, and statistical analyses on OSF (https://osf.io/x9aej). The complete study materials, data, analysis scripts, and details regarding the sample size calculation, the counterbalancing/randomization plan, and results are accessible on OSF as well (Bleijlevens & Behne, 2021; https://osf.io/8dpw4/). This project has been approved by the ethics committee of the Institute for Psychology, University of Göttingen (project numbers 300 & 301).

Participants

The final sample included 75 typically developing 2-to-3-year-old German-speaking children (24-47 months, $M = 32.9$, $SD = 7.2$; 40 female, 35 male, all monolingual) and 112 adults (18-66 years, $M = 30.1$, $SD = 10.4$; 42 female, 67 male, 3 diverse; 28 bilingual). We used data simulation to determine the required number of participants *a priori* with the goal to obtain .8 power. The calculation was based on the test of children's and adults' performance in referent selection trials in both conditions against chance. Children were invited via the department's database of children whose parents previously agreed to be contacted for our studies. Adults were recruited via an online recruitment platform (www.prolific.com) and paid £1.95 (= £9/hour) for their successful participation.

The participating children lived in a German university city and its surrounding. Further demographic data (race, income, education etc.) was not collected due to the data protection rules of the university. For adults, Prolific provided ethnical background data: 104 White, 2 Black, 4 Mixed, 1 other, 1 unknown.

Based on our pre-registered exclusion criteria, ten additional children were excluded due to: technical issues (1), parental interference with impact on the whole experiment (1)[4], uncooperative behavior (1), at least one mistake in the familiar-label trials (2), and multilingual language acquisition (5). Further, we excluded 11 single trials due to: children's lack of attention[5] (2 disambiguation and their corresponding retention trials and 1 single retention trial), parental interference (2 disambiguation and corresponding retention trials), and technical issues (1 disambiguation and its corresponding retention trial).

Design

We used a 2 (condition: "Classic ME" vs. "Novelty") x 2 (age group: children vs. adults) factorial repeated-measures design with two trials per participant. We randomly assigned participants to conditions.

[4] One parent named one of the unknown objects during the object pre-exposure phase.

[5] A trained, blinded coder used the webcam videos of children's referent selection and retention trials and pre-screened their overall attention. In case of partial distraction, the experimenter decided whether the child was distracted during crucial parts of the trials (i.e., before or during the speaker's request). Unclear cases were discussed with an additional coder.

Stimuli

We used and extended[6] the experimental stimuli from Bleijlevens et al. (2023; see Bohn et al., 2022 and https://github.com/manuelbohn/mcc for the original source of visual stimuli). Ten pictures of known objects and four of unknown objects were included. Three female German native speakers, one for each speaker in the experimental videos, were recorded for the auditory stimuli. Two non-words ("modi", "toma") served as novel labels for objects in referent selection and six known words as labels in practice and familiar-label trials (apple, bus, house, dog, ball, shoe). Based on the Wordbank (Frank, Braginsky, et al., 2017), each of these German known words is produced by at least 70% of 24-month-old German-speaking children. The labels of the familiar distractor objects from the Classic ME task (car and flower, see below) were understood by 100% of the children whose parents provided an answer to this question at the end of the experiment (67 out of 75; 8 parents concluded the experiment before reaching this question). Videos were created via PsychoPy (Peirce et al., 2019) and Powerpoint. A non-verbal video cartoon for children was used as a time delay prior to the retention test.[7]

Procedure

The experiment was conducted as an unmoderated remote online study (see Rhodes et al., 2020) via participants' computers/laptops at home. Participants were instructed via videos and instruction texts. We video-recorded children's, but not adults', testing sessions via their webcams. After providing informed consent, each participant took part in the following study phases: practice (4 trials), pre-exposure-I, familiar-label test (2 trials), pre-exposure-II, referent disambiguation (2 trials) and retention (2 trials). On each trial (except pre-exposure videos), an animal speaker, located at center stage and looking straight ahead, asked for a referent. Participants could select between several objects that were presented on a row of

[7] "Elefant und Hase gehen auf Reisen" from the German WDR production "Die Sendung mit der Maus" (see Figure 1)

tables below (Figure 1). After 11 seconds of no selection, parents encouraged their children to answer by reading out the text presented on the screen “What do you think is the right one? Just choose what you think is right”. No further spontaneous help was permitted (except for Practice trials). Adult participants were asked about their selection strategy at the end of the study.

Practice

Frog introduced herself. On the first two practice trials, two known objects fell down on the two empty tables in front of frog, with fixed order and target locations (Figure 1). Frog named one of these objects and asked participants to show it to her: “Oh! There is a [known label]! Look at the [known label]! Can you show me the [known label]?”

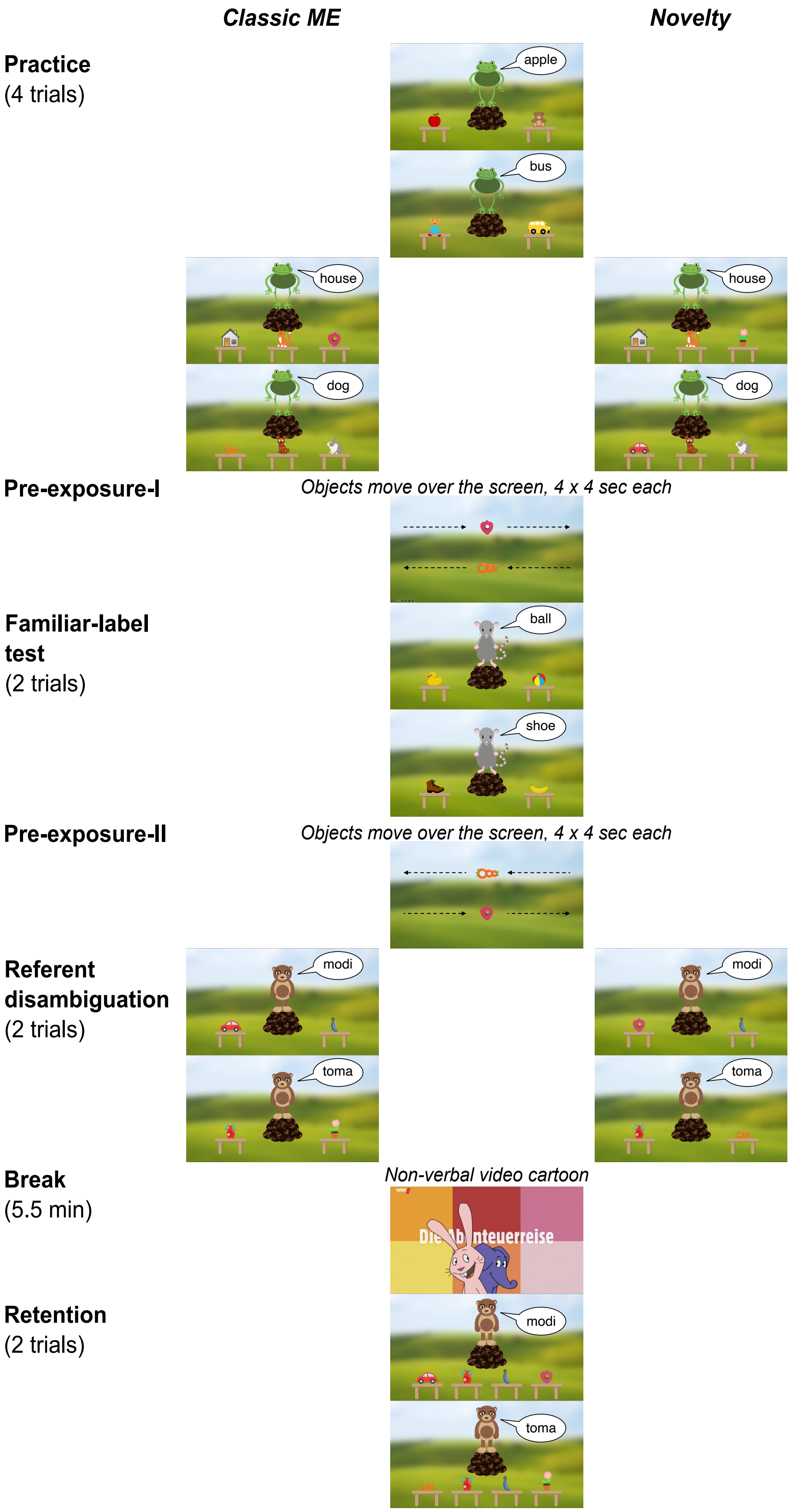
The following two practice trials were similar, except that there were three objects on the row of tables. Two of these, a known target and a known distractor object (e.g. house & cat, respectively) were the same across conditions; however, the identity of the third differed across conditions: in the Classic condition this third object was one of the pre-exposed novel objects, and in the Novelty condition a known object (see Appendix B). We created 3 options for object locations in both trials, across which each object was overall presented equally often at each location and changed its location between trials. We randomized which of the options a participant received, as well as the trial/label order. Participants’ time to react was unlimited. Only in practice trials, parents were asked to help their children until they provided the correct answer. Children received positive feedback from frog.

Pre-exposure-I

Two unknown objects subsequently moved over the screen one at a time, each of them four times à four seconds. Background music was played to facilitate children’s attention. The presented objects would later serve as distractor objects in referent selection trials in the Novelty condition. However, participants in both conditions were presented with the same pre-exposure videos to make the visual input as similar as possible across conditions.

Figure 1

Experimental Procedure: Example Trials for Each Phase in both conditions



Familiar-label test

Mouse introduced herself. In each of the two familiar-label trials, two known objects fell down on the two empty tables in front of her, followed by mouse's request to show one of them to her. In contrast to the practice trials, the response time was now limited to 33 seconds after mouse finished her request and children did not receive any feedback. Further, parents were instructed to not help their children spontaneously, but only to read out the text displayed on the screen (see above). We randomized the trial order and target location for the first trial and used the opposite location for the second trial. Participants who made at least one mistake in familiar-label trials were excluded.

Pre-exposure-II

As in Pre-exposure-I, the same two objects subsequently moved over the screen, each of them four times à four seconds, while background music was playing. Across both pre-exposure phases, each of the objects was presented moving on the upper or lower part of the screen and moving in each direction equally often. After the second pre-exposure, each object was presented for 32 seconds in total. A blinded coding revealed that, on average, children spent 77.4 sec (out of 82 sec total duration of both videos) looking at the screen, with little difference between conditions (Classic ME: 75.5 sec, Novelty: 79.4 sec).

Referent Disambiguation

Participants were presented with two disambiguation trials in their assigned condition. Bear introduced herself while standing behind two empty tables. Two objects appeared on top of the screen and descended until they rested on the tables. Both conditions only differed in the objects presented: In the Classic ME condition, we presented one well-known object (e.g., car) and one novel object. In the Novelty condition, both objects were unknown, but one of them was completely novel while the other one had already been presented in the pre-exposure videos. Afterwards, without changing her frontal gaze direction, bear said excitedly: "Oh cool, there is a [novel label] on the table! How nice! A [novel label] on the table! Can you show me the [novel label]?". Response times were limited to 33 seconds after bear finished her request. We randomized the trial order and target location for the first trial and used the opposite location for the second trial. We counterbalanced between participants which objects were presented during the pre-exposure and which were the novel objects.

Break

A non-verbal video cartoon for young children was played, serving as a time delay of 5.5 minutes prior to retention trials.

Retention

Bear was standing behind four tables when four objects fell down on them. In each retention trial, bear made a request using one of the novel labels introduced in the disambiguation trials, saying: "Oh, there is a [novel label]! Look at the [novel label]! Can you show me the [novel label]?". The objects presented were the same across conditions. For example, in a "modi" retention trial we presented the target from the "modi" disambiguation trial, the distractor from the "modi" disambiguation in the Classic ME condition, the distractor from the "modi" disambiguation in the Novelty condition and the target from the "toma" disambiguation trial (with an analogous set-up for the "toma" retention trial). To ensure that previous exposure times to each object did not differ across conditions, the object that served as a distractor in the *other* condition's "modi" disambiguation trial, served as the third object in one of the participant's last two practice trials (see Appendix B for a detailed explanation of this rational). The presentation of these four items allowed us a) to have a stringent test of specific word-object links for which it does not suffice to select one of the objects one has previously chosen and b) to test retention even in those participants who previously selected an object we did not define as the target. Moreover, the inclusion of items from the other (not presented) condition served to keep the visual input and difficulty in both conditions as equal as possible.

We randomized the trial order and which of eight pre-defined options for object locations was presented. Across these options, each object appeared equally often on each location. Each object changed its position between trials. Participants were given 33 seconds to respond.

Measures

Object choices

We measured object choices by adults' mouse clicks and parents' clicks confirming children's pointing gestures. We were interested in "correct choices" in referent disambiguation trials, i.e., selecting the unknown/novel (vs. known/pre-exposed) object, and in "consistent choices" in retention trials, i.e., selecting the same object they had previously selected in the corresponding disambiguation trial. A blinded coder coded children's pointing directions in 25% of the disambiguation videos. They corresponded to parents' clicks in 100% of the cases.

Proportion of Target Looking (PTL, Exploratory)

A trained, blinded coder used the software ELAN version 6.0 (2020) to code the onsets and ends of children's looking directions (left object, right object, speaker) in referent disambiguation trials via their webcam videos. We then summed up the looking times to each object per participant, trial and phase (before/after first label onset). To measure the proportion of looking time to the target (vs. distractor; PTL), we divided the looking time to the target by the combined looking time to the target + distractor. We used a similar procedure to measure the proportion of looking to the subsequently *selected* (vs. unselected) object, i.e., to children's "subjective" target (PSL). Reliability coding by a second blinded coder for 25% of the trials revealed highly correlated PTLs (r = 0.85) with a mean absolute deviation of 0.06.

Double Checks (Reasoning by Exclusion)

To analyze children's use of reasoning by exclusion in disambiguation vs. familiar-label trials, we coded "double checks" in a binary manner. Based on Halberda (2006), we focused on the time frame between label onset of the first label mention and 3000 ms thereafter.[8] We only attended to object fixations and excluded shifts to the speaker in between the object fixations. Importantly, we analyzed looking patterns based on children's "subjective" targets, i.e., the objects they subsequently selected by pointing. We coded 1 for double checks in the following cases: a) Children fixated the subjective target during label onset, then shifted their gaze to the subjective distractor and back to the subjective target; b) Children fixated the speaker during label onset, then first shifted their gaze to the subjective distractor and then to the subjective target. We coded 0 if children fixated the subjective target or the speaker during label onset, but then showed another looking behavior than described above. We coded NA in all other cases, including trials in which children fixated the subjective distractor during label onset, because in these cases we cannot disentangle whether children excluded the distractor object or simply shifted their gaze directly to the target (see Halberda, 2006).

Uncertainty

To assess uncertainty, we measured participants' response times in disambiguation trials. For adults, we used the time of their mouse clicks (relative to the trial start). For children, a trained and blinded coder coded the webcam videos and measured the time from first label onset until the child's executed pointing gesture.[9] Additionally, we used these videos to measure children's social referencing, i.e., at least one look (vs. no look) to the

[8] In our preregistration, we argued that we will use the same time window as Halberda (2006). Note however, that we mistakenly defined this window there as being 2000 instead of 3000 ms long.

[9] We preregistered to measure the time between video onset and the child's response. However, since this would encompass a time frame before the speaker requested an object, i.e., potential baseline preferences for objects, we started the measurement at the onset of the first label mention. Note that none of the children responded before the label was used at least once.

parent before/during the child's response. Reliability coding by a second blinded coder for 25% of the videos revealed 95% agreement for social referencing and highly correlated response times ($r = 0.99$) with an average absolute deviation of 0.48 sec.

Adults' Selection Strategies

After the experiment, we asked adults to explain the reasoning strategies behind their selections and categorized them in two different ways. First, we assessed whether they engaged in "reasoning by exclusion". We coded 0 if adults indicated reasoning based on arguments in favor of the selected object and 1 if they reasoned based on arguments against the non-selected object (for examples see Appendix, Table C1). Second, we assessed the "specific reasoning strategy" by assigning their answers to five pre-registered categories (Table 1). We coded only one strategy per individual because we asked only one question referring to their overall reasoning approach across trials. Whenever participants' indicated strategies differed for the two novel labels ($n = 1$), we coded the strategy related to the participant's first trial. A second blinded coding for 25% of adults' strategy descriptions revealed a correspondence of 95% for reasoning by exclusion and 95% for the specific reasoning strategies.

Statistical analyses

For the data analysis, we used R (version 4.2.1; R Core Team, 2020) and RStudio (version 2023.6.0.421; Posit team, 2023) Appendix A lists all functions and packages used. The data set, R scripts, detailed model results and assumption tests are accessible on OSF (Bleijlevens & Behne, 2021; https://osf.io/8dpw4/). If not stated otherwise, we followed our preregistered analysis plan and the model assumptions were met.

Before interpreting model parameters, we tested for the overall effect of our fixed effects for each model with more than one predictor by using Likelihood Ratio Tests comparing the fit of the full model to that of a null model, lacking the predictors of interest. This way, we avoided "cryptic multiple testing" (Forstmeier & Schielzeth, 2011).

Table 1

Categories of Adults' Specific Reasoning Strategies

Strategy	Explanation	Examples
Speaker intent	Reasoning based on the speaker, her behavior/intentions	"I assumed that bear knows the words 'car' and 'flower'. Since he did not use them, I chose the other object, respectively."
Nameability/ Familiarity[10]	Reasoning based on the nameability/familiarity of an object	"I could exclude the other ones because I knew their names." "I knew the other object, thus, the unknown object must be the modi/toma."
Perceptual features	Reasoning based on objects' perceptual (visual/auditory) properties or salience	"I imagined which object would best match the sound of the word." "The modi object looked modern and futuristic, consistent with the name."
Experimental logic	Reasoning based on the pragmatics of the experiment/ the experimenters' intentions	"I saw two of them in the inter-sequence. Therefore, it seemed more likely to me that it were the previously shown objects."
Explicit guessing	Indication of own ignorance/ selection based on intuition	"I decided intuitively." "I randomly chose one of the objects."

Object Choices

To analyze children's and adults' object choices (combined), we fitted two GLMMs with binomial error distribution: one on their correct choices in disambiguation trials and one on their consistent choices in retention trials. In both models, we used condition, age (z- and log-transformed) and their interaction as predictors and added random intercepts for

[10] Note that we preregistered to assign responses to this category based on nameability only. Since we realized that many participants in the Classic ME condition (n = 20) answered in line with this category description, but referred to objects' *familiarity* instead of their labels, we slightly extended the category description.

participants. To analyze whether their choices significantly differed from chance level, we inspected the models' predicted probabilities (and their 95% confidence intervals), as depicted in a plot. In addition, we ran an exploratory GLMM with binomial error distribution predicting consistent choices by group (each combination of age group and condition), with suppressed intercept and random intercepts for participants, and tested each group's performance against the chance level of 0.25.

Children's Looking Patterns in Referent Disambiguation Trials

To analyze looking times to the target (vs. distractor), we fitted an exploratory GLMM with beta error distribution. We predicted PTLs in disambiguation trials by labeling phase (pre vs. post labeling), condition, and their interaction and added random intercepts for participants.

To analyze children's use of double checks (i.e., reasoning by exclusion), we conducted two GLMMs with binomial error distribution, one for each condition. In both models, we predicted double checks by trial type (familiar-label vs. disambiguation) and added random intercepts for participants.[11] In addition, we analyzed potential age effects by fitting two exploratory models on children's double checks (one for each condition), with trial type, age (z-transformed), and their interaction as predictors, and random intercepts for participants.

Uncertainty

To test for condition effects on response times, we ran two LMMs, one per age group (children/adults).[12] In both models, we used log-transformed response times, included

[11] In addition to the pre-registered separate analysis for both conditions, we analyzed the data in a single model (adding condition and its interaction with trial type as predictors) and revealed a similar pattern: While children in the Classic ME condition showed significantly more double checks in disambiguation than in familiar-label trials (b = 1.26, SE = 0.47, p = .008), this was not the case for children in the Novelty condition (b = 0.41, SE = 0.48, p = .387).

[12] We preregistered to analyze children's and adults' response times together in one model. However, since we measured response times for children starting from label onset (instead of video onset as it is the case for adults) until a response was made, response times were not comparable anymore between age groups and we decided to analyze them in separate models instead.

condition as a predictor and random intercepts for participants.

To compare children's social referencing in both conditions, we conducted a GLMM with binomial error distribution. We predicted social referencing by condition, including random intercepts for participants. Note that we adapted the alpha level to .025 for children's uncertainty analyses, because we measured uncertainty via response times as well as social referencing (i.e., multiple testing).

Adults' Selection Strategies

To analyze adults' described selection strategies, we ran two models. First, we ran a binomial model to predict adults' use of "reasoning by exclusion" strategies by condition. We removed the intercept in this model to test both conditions against 0.5 (i.e., equal use of both strategies).[13] Second, we predicted adults' specific selection strategies by condition in a multinomial model. We excluded the "speaker intent" category from this analysis because of its low frequency (2). Since the "experimental logic" category occurred in only one condition (Novelty), possibly leading to a problem of complete separation, we ran the model 1000 times, based on a decision in our preregistration: We interchanged the response of one participant at a time such that a participant's response that was coded as not-"experimental logic" was then coded as "experimental logic". We then used the mean of all estimates for the evaluation of our hypotheses. To interpret the model results, we used the model's predicted probabilities (and their 95% confidence intervals) for each strategy per condition, as depicted in a plot.

Results

Target Selection on Disambiguation Trials

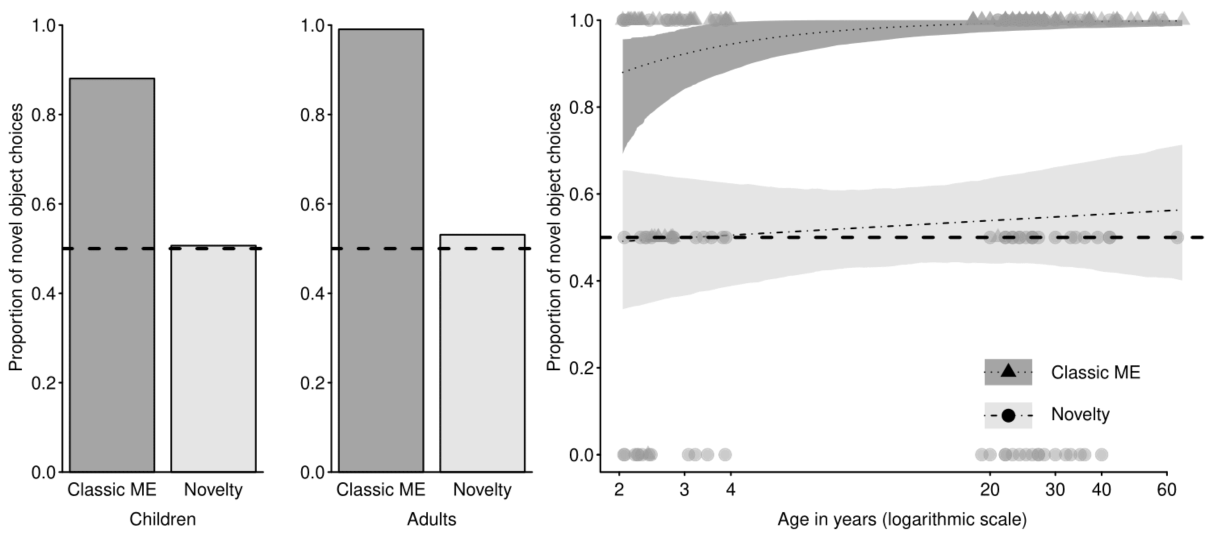
In the Classic ME condition, both children and adults chose the target significantly above chance (see Figure 2: the confidence interval does not include the value of 0.5). In

[13] By suppressing the intercept, the model calculates one estimate per condition, representing the logit of the probability of selecting the novel object in this condition. It then tests both of them against a value of 0 in the model's logit-space = a probability of 0.5 (logit(0.5) = 0).

contrast, in the Novelty condition, participants across the age range selected the target at chance levels. In fact, participants across the age range (i.e., children and adults) selected the correct object significantly more often in the Classic ME than in the Novelty condition.

Figure 2

Correct Choices in Referent Disambiguation Trials



Note. Left: proportion of correct object choices per age group and condition (descriptive data). Right: Grey triangles (Classic ME) and dots (Novelty) show the proportions of correct object choices per participant, based on trials in which any selection was made ($n_{\text{Classic ME}}$ = 174, n_{Novelty} = 188). Dotted (Classic ME) and dash-dotted (Novelty) lines represent the fitted values; and darker (Classic ME) and lighter (Novelty) polygons show the 95% confidence intervals, both revealed by the GLMM and calculated via bootstrapping with 10000 boots.

Furthermore, the model revealed a significant interaction of condition and age (b = -1.39, SE = 0.63, p = .027): Participants' performance significantly improved with age in the Classic ME condition (b = 1.49, SE = 0.60, p = .013), but not in the Novelty condition (b = 0.10, SE = 0.20, p = .615). The full model predicting correct choices by condition, age and their interaction explained the data significantly better than the null model ($\chi^2 = 86.97$, $df = 3$, $p < .001$).

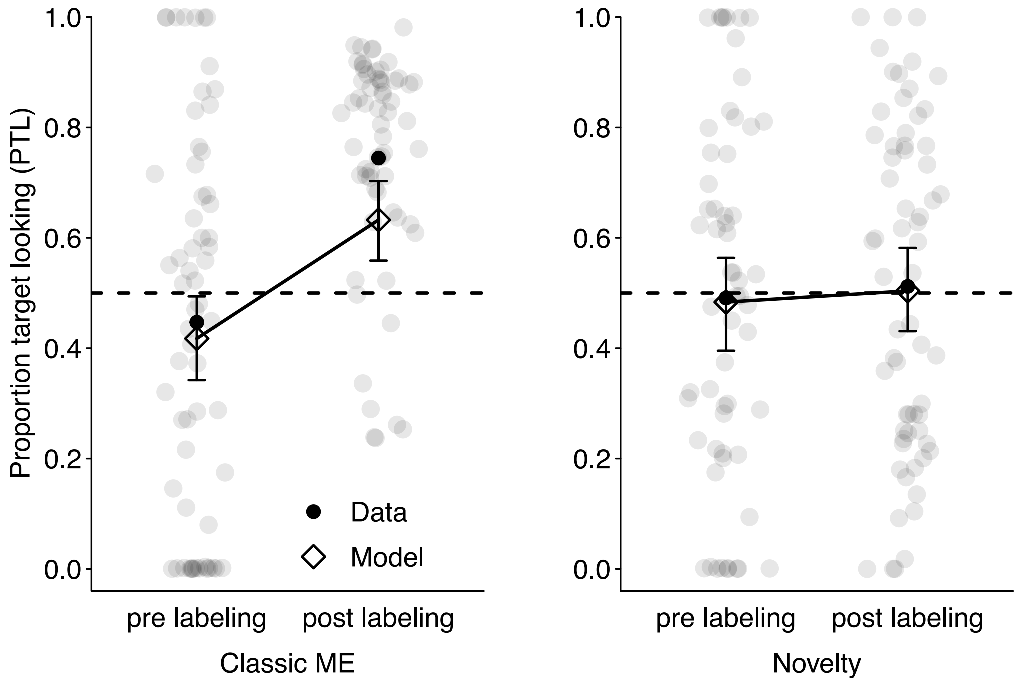
Children's Proportion of Target Looking (PTL)

Our exploratory analysis revealed that children's looking patterns exactly mirrored their behavioral responses on the disambiguation trials: There was a significant interaction between condition and labeling phase ($b = -0.79$, $SE = 0.32$, $p = .013$). That is, in the Classic ME condition, children significantly increased their PTL after label onset ($b = 0.88$, $SE = 0.23$, $p < .001$) and after label onset, they looked significantly longer to the target than expected by chance (Figure 3: the confidence interval does not include the chance level of 0.5). In contrast, in the Novelty condition, children's proportion of target looking did not differ before and after label onset ($b = 0.08$, $SE = 0.23$, $p = .721$) and was not above chance in the post-labeling phase. The model described the data significantly better than the corresponding null model ($\chi^2 = 15.21$, $df = 3$, $p = .002$).

Additionally, we found that children looked longer at the objects they subsequently selected (i.e., the "subjective" target): In both conditions, children significantly increased their proportion of looking to the subjective target after first label onset (Classic ME: $b = 0.55$, $SE = 0.22$, $p = .014$, Novelty: $b = 0.55$, $SE = 0.22$, $p = .014$), and they looked significantly longer to the subjective target vs. distractor after label onset (Figure C1). There was no significant interaction between condition and labeling phase ($b = 0.00$, $SE = 0.31$, $p = .998$). The model described the data significantly better than the corresponding null model ($\chi^2 = 12.32$, $df = 3$, $p = .006$).

Figure 3

Proportion of Target Looking Time by Condition and Labeling Phase



Note. Grey dots represent children's PTLs per codable trial ($n_{\text{ClassicME\_pre}}$ = 62, $n_{\text{ClassicME\_post}}$ = 62, $n_{Novelty\_pre}$ = 59, $n_{Novelty\_post}$ = 63) and black dots the aggregated PTLs per condition and trial type. Diamond shapes indicate the predicted probabilities; and vertical lines the 95% confidence intervals, both obtained by the GLMM and calculated via bootstrapping with 1000 boots.

Uncertainty in Referent Selection Trials

Adults responded significantly faster in the Classic ME (M = 1.20, SD = 1.23) than the Novelty condition (M = 2.07, SD = 2.24; b = 0.46, SE = 0.16, p = .004). In contrast, children's response times did not differ between conditions ($M_{Classic\ ME}$ = 11.47, $SD_{Classic\ ME}$ = 5.86; $M_{Novelty}$ = 12.11, $SD_{Novelty}$ = 7.02; b = 0.00, SE = 0.10, p = .996).

Additionally, children's frequency of social referencing did not differ significantly between the Classic ME (11% of trials) and the Novelty condition (14% of trials, b = 0.29, SE

= 1.66, p = .863). Thus, while adults seemed to experience higher certainty in disambiguation in the Classic ME (vs. Novelty) condition, we found no signs for such a difference in children.

Reasoning by Exclusion

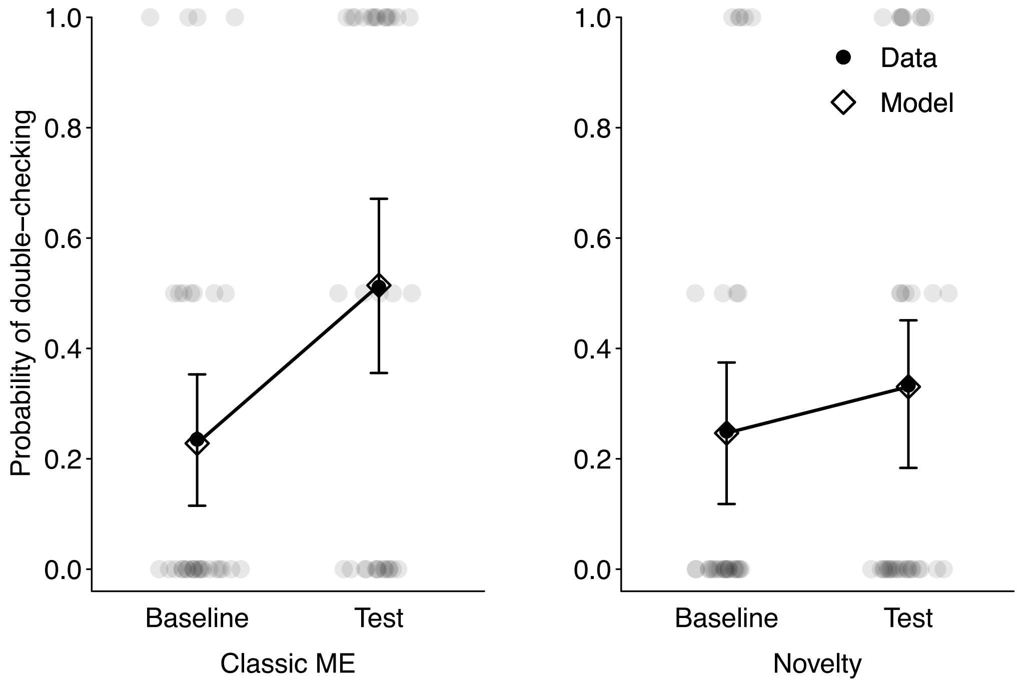
Children's Double Checks

Children in the Classic ME condition significantly increased their use of double checks in test (disambiguation) trials compared to baseline (familiar-label) trials (b = 1.28, SE = 0.50, p = .011). In the Novelty condition, in contrast, children's use of double checks did not differ between trial types (b = 0.41, SE = 0.48, p = .388), indicating that they selectively used reasoning by exclusion in the Classic ME condition (Figure 4).

The exploratory analysis of potential age effects in children's use of double checks revealed a significant interaction of age and trial type in the Classic ME condition (b = 1.02, SE = 0.52, p = .049): Children's use of double checks increased with age in Classic ME disambiguation trials (b = 0.90, SE = 0.40, p = .023), but not familiar-label trials (b = -0.12, SE = 0.33, p = .727). Children's double-checking started to differ significantly between disambiguation and familiar-label trials at the age of 2.4 years. The model for the Classic ME condition described the data significantly better than the corresponding null model (χ^2 = 14.63, df = 3, p = .002). This was not the case, however, for the Novelty condition (χ^2 = 5.66, df = 3, p = .130), in which children's use of double checks did not seem to change with age (Figure 5).

Figure 4

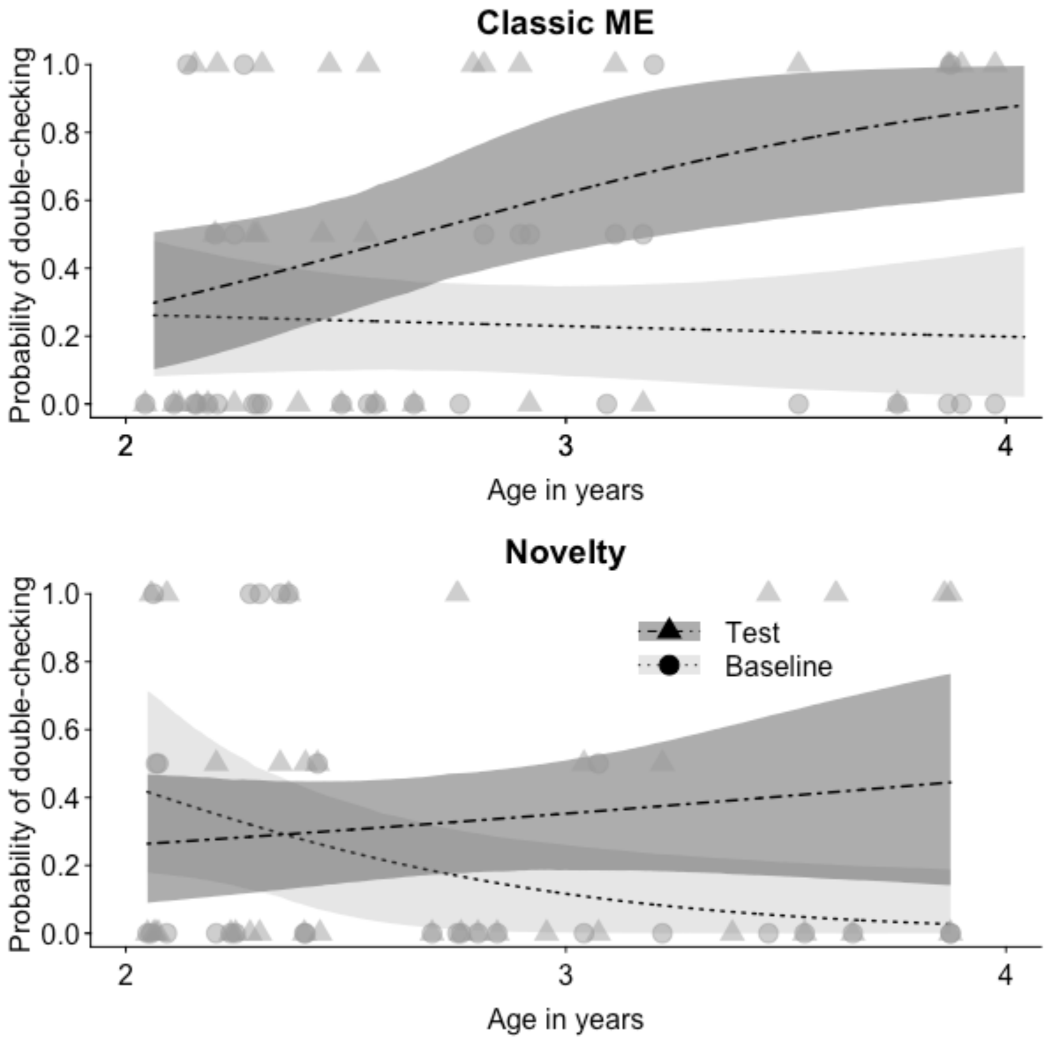
Children's Use of Double Checks (Reasoning by Exclusion) by Condition and Trial Type



Note. Grey dots represent the proportions of codable trials with double checks per child ($n_{\text{ClassicME\_Baseline}} = 51$, $n_{ClassicME\_Test} = 45$, $n_{Novelty\_Baseline} = 40$, $n_{Novelty\_Test} = 51$) and black dots the aggregated proportions per condition and trial type. Diamond shapes indicate the predicted probabilities and vertical lines the 95% confidence intervals, both obtained by the two GLMMs (one per condition) and calculated via bootstrapping with 1000 boots.

Figure 5

Age Effects in Children's Use of Double Checks (Reasoning by Exclusion)



Note. Developmental change in the probability of double checking in both conditions. Grey dots (Baseline/ familiar-label trials) and triangles (Test/ disambiguation trials) show the proportions of double-checks per participant, based on codable trials ($n_{\text{ClassicME\_Baseline}}$ = 51, $n_{\text{ClassicME\_Test}}$ = 45, $n_{\text{Novelty\_Baseline}}$ = 40, $n_{\text{Novelty\_Test}}$ = 51). Dotted (Baseline) and dash-dotted (Test) lines represent the fitted values and lighter (Baseline) and darker (Test) polygons the 95% confidence intervals, both obtained by the two GLMMs (one per condition) and calculated via bootstrapping with 10000 boots.

Adults' Selection Strategies

As predicted, we found that participants in the Classic ME condition were significantly more likely to use reasoning by exclusion (b = 2.06, SE = 0.43, p < .001) while adults in the Novelty condition were significantly more likely to search for arguments in favor of the selected object (b = -2.62, SE = 0.60, p < .001; see Table C1).

The second model revealed that adults in the Classic ME condition indicated strategies based on the objects' nameability/familiarity significantly more often than participants in the Novelty condition (Figure C2). In contrast, adults in the Novelty condition described explicit guessing and inferences based on perceptual object features significantly more often than those in the Classic ME condition.

Object Choices in Retention Trials

Our analysis on participants' consistent object choices revealed no interaction between condition and age (b = -0.73, SE = 0.54, p = .175). Therefore, we ran a reduced model excluding the interaction term (as preregistered). Both models fitted the data better than the corresponding null models (full model: χ^2 = 54.17, df = 3, p < .001; reduced model: χ^2 = 52.34, df = 2, p < .001). The results showed that participants' consistent choices did not differ between conditions (b = -0.24, SE = 0.54, p = .653), but their performance increased with age (b = 2.08, SE = 0.41, p < .001). While adults selected the consistent object far above chance in both conditions, children's pattern of performance was rather fragile and less easy to interpret (see Figure C3). An exploratory model focusing on the children's data suggests that within the group of children, there was no effect of age or condition (Figure C4, Table C2). Furthermore, an exploratory model testing each age group in each condition against the chance level of 0.25 revealed that adults (Classic ME: b = 4.75, SE = 0.93, p < .001; Novelty: b = 3.67, SE = 0.81, p < .001) as well as children (Classic ME: b = -0.21, SE = 0.50, p = .037; Novelty: b = 0.23, SE = 0.48, p = .003) performed significantly above chance in both conditions. For details on children's choices see Table C3.

Discussion

The present study aimed to contribute to current discussions concerning children's referent disambiguation and word learning by using two approaches. First, we manipulated the *source of information* available to identify the intended referent and found that, across different measures, both children and adults disambiguated in the classic ME task (offering lexical, pragmatic and novelty cues), but not in an adapted version in which relative object novelty was the only cue. Second, we assessed their use of *reasoning by exclusion* based on children's looking patterns (i.e., "double checks") and adults' written reports. We found that both children (starting from around two-and-a-half years of age) and adults engaged in reasoning by exclusion in the ME task, but not if objects only differed in their relative novelty. And finally, while adults robustly retained words after the initial selection, children's performance was more fragile.

The Sources of Information Critical for Disambiguation – All About Novelty?

Our manipulation showed that relative object novelty alone was not sufficient to guide participants' disambiguation: Neither children nor adults preferentially selected the more novel object as the referent of the novel label if no lexical or pragmatic cues were given. At first glance, this does not seem to be in line with earlier findings (e.g., Dysart et al., 2016; Horst et al., 2011; Mather & Plunkett, 2012). However, a closer look at previous findings raises some questions regarding the role of novelty and its robustness.

First, in previous studies that did not include a baseline comparison (e.g., Dysart et al., 2016; Horst et al., 2011), children's novel object choices may not reflect a mapping of the novel label to the novel object, but may instead simply be based on children's curiosity to explore the novel object (see Graham et al., 2005; Markman, 1991; Mather & Plunkett, 2012 for this argument). And second, when controlling for this baseline preference for novel objects, the effect may be rather unstable: e.g., in Mather & Plunkett (2012), there was an overall attentional shift to the most novel object in the first, but not the second experiment, which used a more stringent manipulation of novelty. Thus, just like Mather & Plunkett (2012), we found that when a pre-exposed and a completely novel object are presented in direct contrast, upon hearing a novel label, children did not consistently look towards the completely novel object.

The present findings, in combination with previous data, suggest that children's guidance by object novelty for word disambiguation does not seem to be as robust as assumed (see also, e.g., Bleijlevens et al., 2023). The analysis of children's looking data shows that the lack of novelty preference in children's selections was not due to differences between their "implicit" attraction to novelty and their explicit object selections: Children did not only select objects randomly, but also had no systematic looking preferences. Additionally, an extensive coding of children's attention to the screen revealed that the observed results are not due to children in an online study being inattentive during the object pre-exposure. Instead, children in the Novelty condition looked longer at one of the objects (be it based on individual preferences or chance) and then also selected this object.

In contrast to their lack of preference in the Novelty condition, children (and adults) did choose the target object when in addition to object novelty they could also rely on linguistic and/or pragmatic information. In our classic ME task, children both selected and looked at the target object far above chance, in line with a wealth of research (see Lewis et al., 2020 for a review). The set-up in the classic ME task cannot distinguish the influence of lexical and pragmatic factors on participants' disambiguation. However, recent work has shown that children can identify the intended referents of novel words based on pragmatic inferences even when lexical information is lacking (i.e., when both potential referents have no label; Akhtar et al., 1996; Bleijlevens et al., 2023; Bohn et al., 2022).

Importantly, children's (and adults') success in the classic ME condition in this asynchronous online study demonstrates the robustness of the disambiguation effect across different paradigms and methodologies. This is increasingly important given the current replication crisis in our field (Frank, Bergelson, et al., 2017). In addition, it emphasizes that children's behavior in the Novelty condition was not a consequence of the online setting, but rather suggests that their disambiguation in the classic ME task may not be reducible to a direct mapping based on object novelty alone. The results highlight the role of lexical (e.g., Lewis et al., 2020) and pragmatic (Bleijlevens et al., 2023; Bohn et al., 2022) information as critical sources to solve the ME task. They leave open, however, in how far the importance of different information sources may differ for younger children (e.g., Mather & Plunkett, 2010; Yurovsky & Frank, 2017).

The Mental Computations Underlying Children's Disambiguation

Our second approach was to investigate the mental computations underlying referent disambiguation. We found that participants in the classic ME task reasoned by exclusion: Instead of directly attending to the novel object, participants first excluded the known object. For adults, this was indicated by describing arguments *against* the to-be-excluded object instead of arguments *in favor of* the selected object. Similarly, children demonstrated reasoning by exclusion by their "double checks" (Halberda, 2006): Upon label onset in the disambiguation (vs. familiar-label) trials, children (from 2.4 years onwards) shortly looked to the known, to-be-excluded object before attending to and selecting the novel target object. In contrast, in the Novelty condition, participants did not indicate reasoning strategies based on excluding distractor objects: Adults described their search for arguments in favor of the selected object, and children did not show any more double checks than in the familiar-label trials. The pattern of results highlights that the disambiguation effect is probably not a result of children's attraction to novelty alone, but rather the product of lexical and/or pragmatic inferences that initiate the exclusion of the known object as the correct referent.

Thus, in the current study we could replicate and extend the results of Halberda (2006), using a younger age group of 2-to-3-year-olds and a greater sample. Like in the original study, we found children to double-check in the classic ME task, despite many methodological changes including a different format and setting (i.e., online, with only one instead of two locally separated screens and the target label embedded in several sentences), highlighting the robustness and replicability of the effect. Additionally, and extending previous results, our data suggests that this exclusion process is initiated by lexical/pragmatic information in the scene (and not present in contexts in which objects only differ in their relative novelty) and that the starting point of this reasoning ability is around the age of two-and-a-half years.

While children's use of "double checks" is indicative of some kind of reasoning by exclusion, it remains unclear in how far it is evidence that children as young as 2-3 years of age engage in a form of logical inference known as the *disjunctive syllogism*: "A or B. Not A. Therefore, B." (Halberda, 2006). In the logical reasoning literature, certain tasks that were designed to measure deductive reasoning in infants or apes leave room for different explanations. For example, in the 2-cups-task (Call, 2004), a reward is hidden in one of two cups, before participants are shown that one of them (A) is empty. Participants' success in this task (choosing cup B) can be explained by three inferences with differing levels of sophistication (see Mody & Carey, 2016): 1) "avoid (empty) cup A" without representing A and B as alternatives and their dependent relationship, 2) "maybe A, maybe B" including the representation of A and B, but not their dependent relationship, or 3) the disjunctive syllogism "A or B, not A, therefore B" including both the representation of A and B as alternatives as well as their dependent relationship. Which of these inferences are applicable to children's double checks in the ME task?

Children's gaze shifts away from the correct object indicate an inference that is more sophisticated than option 1, since the simple avoidance of A (e.g., car) would not necessitate any further checks after already attending to the correct object.[14] On the other end of the

[14] It is, of course, feasible, that even though the 'avoid A' heuristic does not explain children's looking pattern, there may be other heuristic responses (not yet articulated in the literature) that may result in the looking pattern observed here. Further theoretical and empirical work is needed to explore and evaluate this.

continuum, their behavior is compatible with the application of the disjunctive syllogism (Cesana-Arlotti & Halberda, 2021; Halberda, 2006; see also Grassmann, 2013). Accordingly, the age of onset in our sample (two-and-a-half years) equals the age of success in tasks measuring the disjunctive syllogism in other domains (e.g., Gautam et al., 2021). However, since the measure of children's double checks cannot distinguish between the full-fledged logical inference "not A, therefore *necessarily* B" vs. the weaker inference "not A, therefore *probably* B", it is also compatible with option 2 ("maybe A, maybe B"[15]) that includes the representation of two options and the exclusion of one of them, but lacks the representation of their dependent relationship (see Grigoroglou & Ganea, 2022; and see Rescorla, 2009 for a Bayesian framework where the probability of B increases as the probability of A decreases).

Thus, children from around two-and-a-half years seem to represent both objects in the ME task as possible referents and exclude one of them before selecting the other one. Children younger than this, in contrast, may select the novel object based on a weaker inference: Since they do not engage in double checks, nor in a positive mapping based on novelty (as suggested by their random selections in the Novelty condition), they may exclude the known object, but instead of representing both objects as potential referents, they may be simply avoiding the known one and therefore selecting the other one. Future research is needed to elaborate the full developmental pattern and to disentangle which kind of logical reasoning capacities are underlying children's double checks.

[15] Note that there are two different readings of this option: In the logical reading, "maybe A" may be translated to "not impossibly A" and is thus less informative and probably less sophisticated than what children in our task inferred. In the probabilistic reading, "maybe A" may refer to the inference that it might be A, but with unclear probability – the uncertainty about it actually being A may then lead to further searches. This latter reading of "maybe A, maybe B" is the most parsimonious explanation for (two-and-a-half-year-old) children's behavior in our classic ME task.

Word Learning After Successful Disambiguation

In the current study, children showed an overall fragile pattern of retention performance.[16] Our exploratory analysis suggests that children made consistent choices in retention trials above chance level, without differences between conditions. Together with a wealth of previous work, this suggests that referent disambiguation does indeed contribute to children's learning of novel word meanings. However, children's performance in the present study was not robust at all. A factor that may have contributed to this is the online presentation of the task (see Strouse & Samson 2021). Further research is needed to establish which factors foster word learning after disambiguation, such as children's active involvement, the salience and relevance of objects and tasks, the difficulty of encoding, and other factors known to influence children's memory performance more generally (see, Wojcik, 2013).

Conclusion

Our aim was to investigate the mechanisms underlying children's (and adults') referent disambiguation. Across measures, we found that 2-to-3-year-olds, as well as adults, chose an unknown (vs. known) object as the referent of a novel label in the classic disambiguation task. However, after controlling for lexical and pragmatic cues, relative object novelty alone guided neither their object selections, nor their looking behavior. Adults' strategy descriptions and children's looking patterns revealed instead that, in the classic disambiguation task, participants engaged in reasoning by exclusion, starting at around two-and-a-half years.

The pattern of results contributes to the theoretical debate on which cues and mechanisms are crucial for disambiguation and word learning. It suggests that children (and adults) may resolve referential ambiguity not by simply mapping the novel label to the most

[16] The preregistered analysis, i.e., visual inspection of fitted values and confidence intervals, revealed an unstable pattern that was not easy to interpret (see Figure C3). Further, note that we probably lacked statistical power for this specific analysis, such that the results should be interpreted with caution.

novel object, contradictory to associative novelty accounts of word learning. Instead, they solve the disambiguation task by a process of elimination, highlighting the importance of lexical and/or pragmatic sources of information for the ability of young children to disambiguate novel word meanings.

References

Akhtar, N., Carpenter, M., & Tomasello, M. (1996). The role of discourse novelty in early word learning. *Child Development*, *67*(2), 635–645. https://doi.org/10.2307/1131837

Bache, S. M., & Wickham, H. (2020). *magrittr: A forward-pipe operator for R*. https://CRAN.R-project.org/package=magrittr

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bleijlevens, N., & Behne, T. (2021). *Young children and adults use reasoning by exclusion rather than attraction to novelty to disambiguate novel word meanings*. https://osf.io/8dpw4/

Bleijlevens, N., Contier, F., & Behne, T. (2023). Pragmatics aid referent disambiguation and word learning in young children and adults. *Developmental Science*, e13363. https://doi.org/10.1111/desc.13363

Bohn, M., Tessler, M. H., Merrick, M., & Frank, M. C. (2022). Predicting pragmatic cue integration in adults' and children's inferences about novel word meanings. *Journal of Experimental Psychology: General*. Advance online publication. https://doi.org/10.1037/xge0001216

Bolker, B., Warnes, G. R., & Lumley, T. (2022). *gtools: Various R programming tools*. https://CRAN.R-project.org/package=gtools

Brooks, M. E., Kristensen, K., Benthem, K. J. van, Magnusson, A., Berg, C. W., Nielsen, A., Skaug, H. J., Maechler, M., & Bolker, B. M. (2017). glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *The R Journal*, *9*(2), 378–400.

Call, J. (2004). Inferences about the location of food in the great apes (pan paniscus, pan troglodytes, gorilla gorilla, and pongo pygmaeus). *Journal of Comparative*

Psychology, *118*(2), 232–241. https://doi.org/10.1037/0735-7036.118.2.232

Carey, S., & Bartlett, E. (1978). Acquiring a single new word. *Papers and Reports on Child Language Development*, *15*, 17–29.

Cesana-Arlotti, N., & Halberda, J. (2021). *Domain-general logical inference by 2.5-year-old toddlers*. PsyArXiv. https://doi.org/10.31234/osf.io/qzxkp

Cesana-Arlotti, N., Martín, A., Téglás, E., Vorobyova, L., Cetnarski, R., & Bonatti, L. L. (2018). Precursors of logical reasoning in preverbal human infants. *Science*, *359*(6381), 1263–1266. https://doi.org/10.1126/science.aao3539

Clark, E. V. (2015). Common Ground. In B. MacWhinney & W. O'Grady (Eds.), *The handbook of language emergence* (pp. 328–353). John Wiley & Sons. https://doi.org/10.1002/9781118346136.ch15

Diesendruck, G., & Markson, L. (2001). Children's avoidance of lexical overlap: A pragmatic account. *Developmental Psychology*, *37*(5), 630–641. https://doi.org/10.1037/0012-1649.37.5.630

Dysart, E. L., Mather, E., & Riggs, K. J. (2016). Young children's referent selection is guided by novelty for both words and actions. *Journal of Experimental Child Psychology*, *146*, 231–237. https://doi.org/10.1016/j.jecp.2016.01.003

ELAN (Version 6.0) [Computer software]. (2020). Max Planck Institute for Psycholinguistics, The Language Archive. https://archive.mpi.nl/tla/elan

Feiman, R., Mody, S., & Carey, S. (2022). The development of reasoning by exclusion in infancy. *Cognitive Psychology*, *135*, 101473. https://doi.org/10.1016/j.cogpsych.2022.101473

Forstmeier, W., & Schielzeth, H. (2011). Cryptic multiple hypotheses testing in linear models: Overestimated effect sizes and the winner's curse. *Behavioral Ecology and Sociobiology*, *65*(1), 47–55. https://doi.org/10.1007/s00265-010-1038-5

Fox, J., & Weisberg, S. (2019). *An R companion to applied regression* (Third). Sage.

https://socialsciences.mcmaster.ca/jfox/Books/Companion/

Frank, M. C., Bergelson, E., Bergmann, C., Cristia, A., Floccia, C., Gervain, J., Hamlin, J. K., Hannon, E. E., Kline, M., Levelt, C., Lew-Williams, C., Nazzi, T., Panneton, R., Rabagliati, H., Soderstrom, M., Sullivan, J., Waxman, S., & Yurovsky, D. (2017). A collaborative approach to infant research: Promoting reproducibility, best practices, and theory-building. *Infancy*, *22*(4), 421–435. https://doi.org/10.1111/infa.12182

Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2017). Wordbank: An open repository for developmental vocabulary data. *Journal of Child Language*, *44*(3), 677–694. https://doi.org/10.1017/S0305000916000209

Gautam, S., Suddendorf, T., & Redshaw, J. (2021). When can young children reason about an exclusive disjunction? A follow up to Mody and Carey (2016). *Cognition*, *207*, 104507. https://doi.org/10.1016/j.cognition.2020.104507

Graham, S. A., Turner, J. N., & Henderson, A. M. E. (2005). The influence of object pre-exposure on two-year-olds' disambiguation of novel labels. *Journal of Child Language*, *32*(1), 207–222. https://doi.org/10.1017/S030500090400666X

Grassmann, S. (2013). Word learning by exclusion—Pragmatics, logic and processing. In F. Liedtke & C. Schulze, *Beyond words: Content, context, and inference*. Walter de Gruyter.

Grigoroglou, M., & Ganea, P. A. (2022). Language as a mechanism for reasoning about possibilities. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *377*(1866), 20210334. https://doi.org/10.1098/rstb.2021.0334

Grolemund, G., & Wickham, H. (2011). Dates and times made easy with lubridate. *Journal of Statistical Software*, *40*(3), 1–25.

Halberda, J. (2006). Is this a dax which I see before me? Use of the logical argument disjunctive syllogism supports word-learning in children and adults. *Cognitive Psychology*, *53*(4), 310–344. https://doi.org/10.1016/j.cogpsych.2006.04.003

Horst, J. S., Samuelson, L. K., Kucker, S. C., & McMurray, B. (2011). What's new? Children prefer novelty in referent selection. *Cognition*, *118*(2), 234–244. https://doi.org/10.1016/j.cognition.2010.10.015

Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal*, *50*(3), 346–363.

Jaswal, V., & Markman, E. (2003). The relative strengths of indirect and direct word learning. *Developmental Psychology*, *39*, 745–760. https://doi.org/10.1037/0012-1649.39.4.745

Kalashnikova, M., Escudero, P., & Kidd, E. (2018). The development of fast-mapping and novel word retention strategies in monolingual and bilingual infants. *Developmental Science*, *21*(6), e12674. https://doi.org/10.1111/desc.12674

Kucker, S. C., McMurray, B., & Samuelson, L. K. (2015). Slowing down fast mapping: Redefining the dynamics of word learning. *Child Development Perspectives*, *9*(2), 74–78. https://doi.org/10.1111/cdep.12110

Kucker, S. C., McMurray, B., & Samuelson, L. K. (2020). Sometimes it is better to know less: How known words influence referent selection and retention in 18- to 24-month-old children. *Journal of Experimental Child Psychology*, *189*, 104705. https://doi.org/10.1016/j.jecp.2019.104705

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Leahy, B. P., & Carey, S. E. (2020). The acquisition of modal concepts. *Trends in Cognitive Sciences*, *24*(1), 65–78. https://doi.org/10.1016/j.tics.2019.11.004

Lewis, M., Cristiano, V., Lake, B. M., Kwan, T., & Frank, M. C. (2020). The role of developmental change and linguistic experience in the mutual exclusivity effect. *Cognition*, *198*, 104191. https://doi.org/10.1016/j.cognition.2020.104191

Lüdecke, D., Ben-Shachar, M. S., Patil, I., Waggoner, P., & Makowski, D. (2021).

performance: An R package for assessment, comparison and testing of statistical models. *Journal of Open Source Software*, *6*(60), 3139. https://doi.org/10.21105/joss.03139

Markman, E. M. (1991). The whole-object, taxonomic, and mutual exclusivity assumptions as initial constraints on word meanings. In S. A. Gelman & J. P. Byrnes (Eds.), *Perspectives on Language and Thought* (1st ed., pp. 72–106). Cambridge University Press. https://doi.org/10.1017/CBO9780511983689.004

Markman, E. M., & Wachtel, G. F. (1988). Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive Psychology*, *20*(2), 121–157. https://doi.org/10.1016/0010-0285(88)90017-5

Markson, L., & Bloom, P. (1997). Evidence against a dedicated system for word learning in children. *Nature*, *385*, 813–815. https://doi.org/10.1038/385813a0

Marno, H. (2021). Is 2-year-old children's referential disambiguation of a novel word primarily aimed at identifying the word meaning or at understanding their communicative partner? *Developmental Psychology*, *57*(7), 1058–1066. https://doi.org/10.1037/dev0001204

Mather, E., & Plunkett, K. (2010). Novel labels support 10-month-olds' attention to novel objects. *Journal of Experimental Child Psychology*, *105*(3), 232–242. https://doi.org/10.1016/j.jecp.2009.11.004

Mather, E., & Plunkett, K. (2012). The role of novelty in early word learning. *Cognitive Science*, *36*(7), 1157–1177. https://doi.org/10.1111/j.1551-6709.2012.01239.x

McMurray, B., Horst, J. S., & Samuelson, L. K. (2012). Word learning emerges from the interaction of online referent selection and slow associative learning. *Psychological Review*, *119*(4), 831–877. https://doi.org/10.1037/a0029872

Merriman, W. E., Bowman, L. L., & MacWhinney, B. (1989). The mutual exclusivity bias in children's word learning. *Monographs of the Society for Research in Child*

Development, *54*(3/4), i. https://doi.org/10.2307/1166130

Merriman, W. E., Marazita, J., & Jarvis, L. (1995). Children's disposition to map new words onto new referents. In M. Tomasello & W. E. Merriman (Eds.), *Beyond names for things: Young children's acquisition of verbs* (pp. 147–183). Hillsdale, NJ: Lawrence Erlbaum Associates.

Merriman, W. E., & Schuster, J. M. (1991). Young children's disambiguation of object name reference. *Child Development*, *62*(6), 1288. https://doi.org/10.2307/1130807

Mody, S., & Carey, S. (2016). The emergence of reasoning by the disjunctive syllogism in early childhood. *Cognition*, *154*, 40–48. https://doi.org/10.1016/j.cognition.2016.05.012

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*, 195–203. https://doi.org/10.3758/s13428-018-01193-y

Phillips, N. (2017). *Yarrr: A companion to the e-book "YaRrr!: The pirate's guide to R."* https://CRAN.R-project.org/package=yarrr

Pomiechowska, B., Bródy, G., Csibra, G., & Gliga, T. (2021). Twelve-month-olds disambiguate new words using mutual-exclusivity inferences. *Cognition*, *213*, 104691. https://doi.org/10.1016/j.cognition.2021.104691

Posit team. (2023). *RStudio: Integrated Development Environment for R*. Posit Software, PBC. http://www.posit.co/

Quine, W. V. O. (1960). *Word and object: An inquiry into the linguistic mechanisms of objective reference* (pp. xv, 294). MIT Press.

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rescorla, M. (2009). Chrysippus' dog as a case study in non-linguistic cognition. In R. W. Lurz (Ed.), *The Philosophy of Animal Minds* (pp. 52–71). Cambridge University Press.

Rhodes, M., Rizzo, M. T., Foster-Hanson, E., Moty, K., Leshin, R. A., Wang, M., Benitez, J., & Ocampo, J. D. (2020). Advancing developmental science via unmoderated remote research with children. *Journal of Cognition and Development*, *21*(4), 477–493. https://doi.org/10.1080/15248372.2020.1797751

Spiegel, C., & Halberda, J. (2011). Rapid fast-mapping abilities in 2-year-olds. *Journal of Experimental Child Psychology*, *109*(1), 132–140. https://doi.org/10.1016/j.jecp.2010.10.013

Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S* (Fourth). Springer. https://www.stats.ox.ac.uk/pub/MASS4/

Vlach, H. A., & Sandhofer, C. M. (2012). Fast mapping across time: Memory processes support children's retention of learned words. *Frontiers in Psychology*, *3*. https://www.frontiersin.org/articles/10.3389/fpsyg.2012.00046

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., … Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Wojcik, E. (2013). Remembering new words: Integrating early memory development into word learning. *Frontiers in Psychology*, *4*. https://www.frontiersin.org/articles/10.3389/fpsyg.2013.00151

Yurovsky, D., & Frank, M. C. (2017). Beyond naïve cue combination: Salience and social cues in early word learning. *Developmental Science*, *20*(2), e12349. https://doi.org/10.1111/desc.12349

Zosh, J. M., Brinster, M., & Halberda, J. (2013). Optimal contrast: Competition between two referents improves word learning. *Applied Developmental Science*, *17*(1), 20–28. https://doi.org/10.1080/10888691.2013.748420

Appendix A

Packages and Functions

We used the following packages for data handling, preparation, and visualization: tidyverse version 2.0.0 (Wickham et al., 2019), magrittr version 2.0.3 (Bache & Wickham, 2020), and lubridate version 1.9.2 (Grolemund & Wickham, 2011).

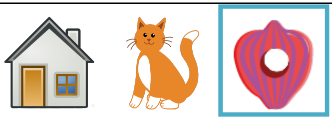
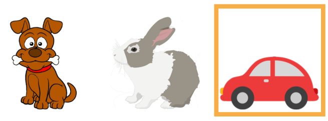
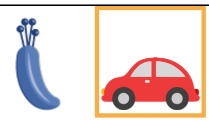
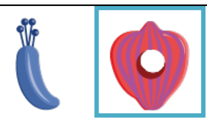
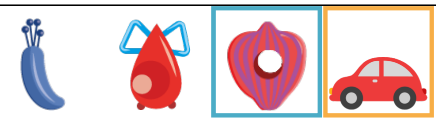
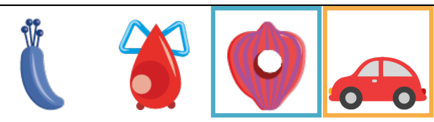
For data analysis, we used the following packages and functions: glmer() from the package lme4 version 1.1-32 (Bates et al., 2015) for GLMMs with binomial error distribution, glht() from the package multcomp version 1.4-23 (Hothorn et al., 2008) to test choices against a chance level other than 0.5, lmer() from the package lmerTest version 3.1-3 (Kuznetsova et al., 2017) for LMMs, glmmTMB() from the package glmmTMB version 1.1.7 (Brooks et al., 2017) for GLMMs with beta error distribution, multinom() from the package nnet version 7.3-18 (Venables & Ripley, 2002) for multinomial models, logit() from the package gtools version 3.9.4 (Bolker et al., 2022), vif() from the package car version 3.1-2 (Fox & Weisberg, 2019) and check_collinearity() from the package performance version 0.10.3 (Lüdecke et al., 2021) to calculate variance inflation factors, and pirateplot() from the package yarrr version 0.1.5 (Phillips, 2017) for data visualization.

Appendix B

Methodological Details

Table B1

Rational Behind Object Presentations in Practice and Retention Trials

	Classic ME	Novelty
Practice "house"		
Practice "dog"		
Disambiguation "modi"		
Disambiguation "toma"		
Retention "modi"		
Retention "toma"		

Note. Example objects presented in practice, disambiguation and retention trials. The left object in each line represents the target object, respectively.

To keep the task demands in retention trials as similar as possible across conditions, we presented the very same objects in both conditions. Take for example, the "modi" retention trial. Here we presented the supernovel object from the "modi" disambiguation trial (i.e., the blue target object in our example above), the distractors that had been presented in the "modi" disambiguation trial in each condition (i.e., the car in the Classic ME condition and the pinkish pre-exposed object in the Novelty condition) as well as the supernovel object from the "toma" disambiguation trial. However, depending on the condition, without any further adjustments, participants would have had different exposures to some of these objects prior to this retention trial. Participants in the Novelty condition would not have encountered the car previously and participants in the classic ME condition would not have encountered the pre-exposed pinkish object in a selection trial previously (in other words, this applied to the distractor object from the disambiguation trial in the other condition; see colored frames). To ensure that participants had seen all the objects prior to the retention trials, we thus presented these objects already in the practice phase as an additional (irrelevant) distractor object. Due to the logic of our conditions, this concerned either an unfamiliar (Classic ME) or familiar object (Novelty).

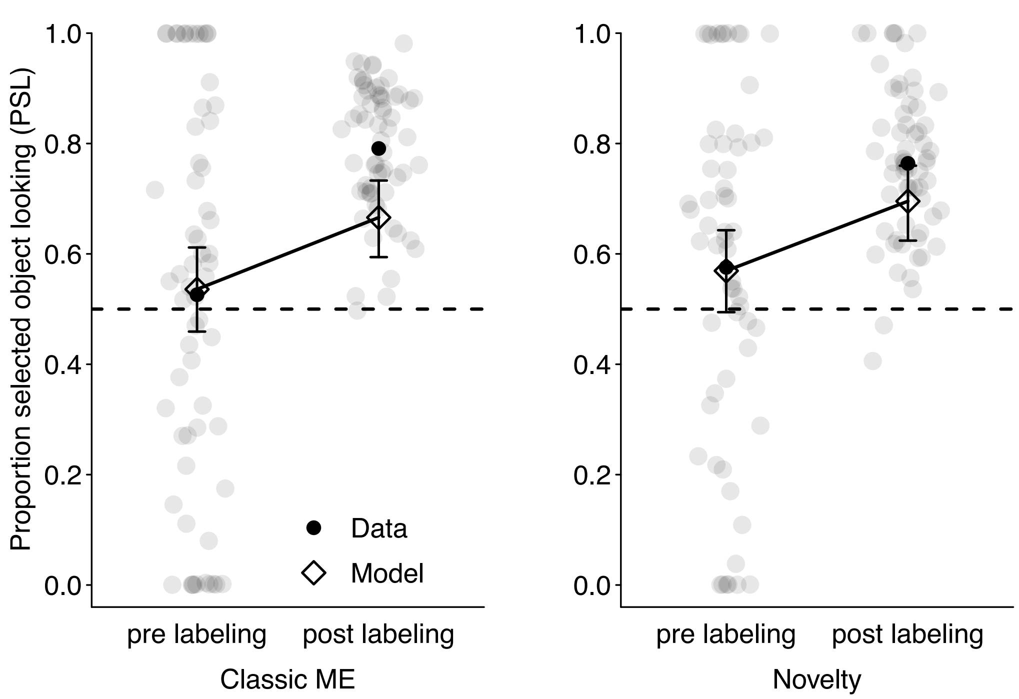
Appendix C

Analysis Details

Children's Proportion of Looking to the Selected vs. Unselected Object

Figure C1

Proportion of Looking Time to the Selected Object by Condition and Labeling Phase



Note. Grey transparent dots represent children's proportion of looking to their subsequently selected object (PSL) per trial, based on trials in which the gaze patterns were codable (Classic ME: $n_{\text{pre}} = 62$, $n_{\text{post}} = 62$; Novelty: $n_{pre} = 59$, $n_{post} = 63$). Filled black dots indicate the aggregated PSL per condition and trial type and diamond shapes the predicted probabilities obtained by the GLMM with beta error distribution predicting PSL by condition, labeling phase and their interaction. Vertical lines indicate the 95% confidence intervals. Predicted values and their confidence intervals have been obtained via bootstrapping with 1000 boots.

Adults' Reasoning Strategies

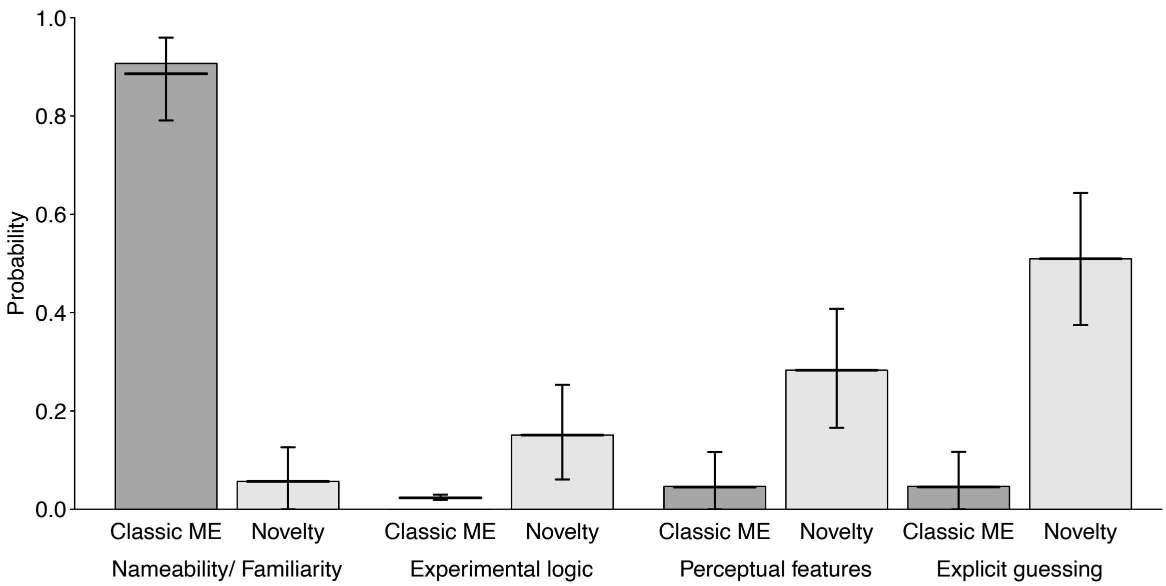
Table C1

Adults' Indication of Reasoning by Exclusion

Strategy	Example	Classic ME	Novelty
Reasoning by exclusion	"Since I knew the word for the other object for sure, I knew the other object must be the modi/toma."	47 (85.5%)	3 (5.3%)
Direct mapping	"I associated 'toma' with tower, so I chose the higher, red object"	6 (10.9%)	41 (71.9%)
Other	"I don't know"	2 (3.6%)	13 (22.8%)

Figure C2

Adults' Selection Strategies by Condition

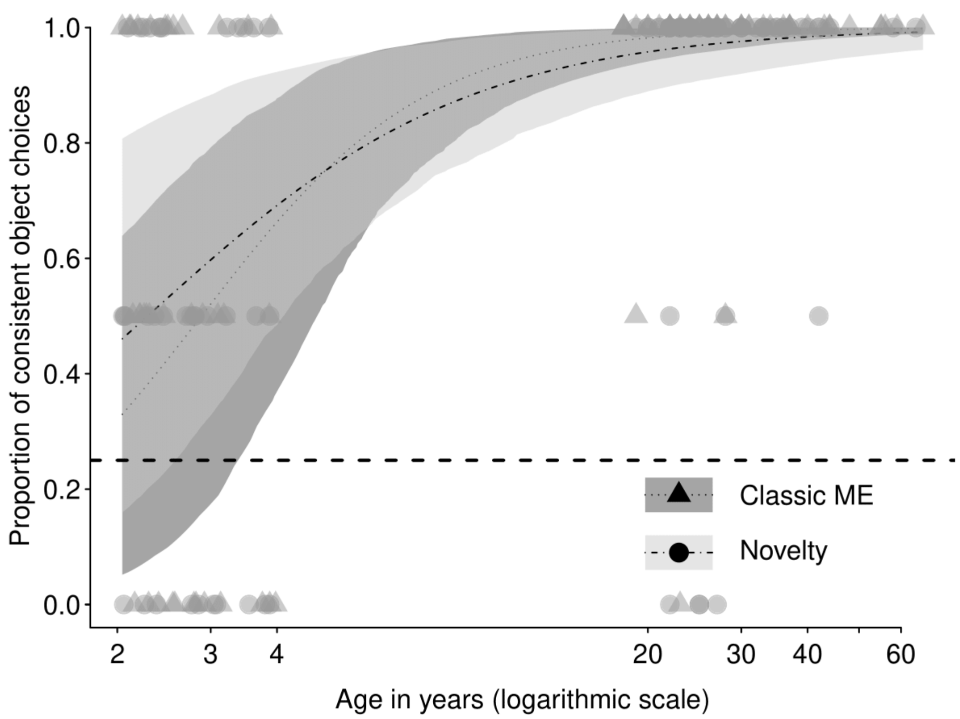


Note. The bars show the proportions of selection strategies based on adults who indicated a strategy matching any of the four depicted categories (n_{Classic} = 43, n_{Novelty} = 53). The "speaker intent" category was excluded because it comprised only 2 observations overall. Horizontal lines show the predicted probabilities of each category by the multinomial model and vertical lines the 95% confidence intervals, both obtained via bootstrapping with 1000 boots. The figure shows that conditions differed significantly in the occurrence of each strategy (confidence intervals of one condition do not include fitted values of the respective other condition).

Children's and Adults' Retention Performance

Figure C3

Consistent Choices in Retention Trials by Condition and Age



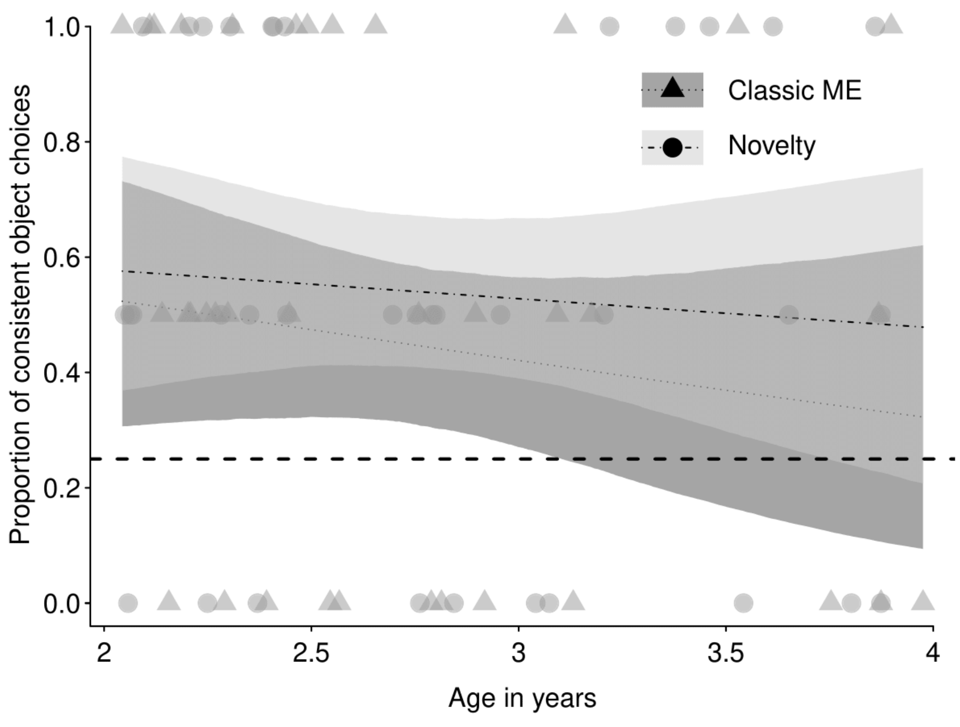
Note. Grey transparent triangles (Classic ME) and dots (Novelty) show the proportions of consistent object choices per participant, based on trials in which any selection was made (children: $n_{\text{Classic ME}} = 66$, $n_{\text{Novelty}} = 71$; adults: $n_{\text{Classic ME}} = 105$, $n_{\text{Novelty}} = 113$). Dotted (Classic ME) and dash-dotted (Novelty) lines represent the fitted values revealed by the binomial mixed effects model predicting participants' consistent choices by age (log- and z-transformed), condition, and their interaction. Darker (Classic ME) and lighter (Novelty) polygons show the 95% confidence intervals. Chance level = 25%. Fitted values and confidence intervals were calculated via bootstrapping with 10000 boots. They show that adults selected the consistent object above chance in both conditions (confidence intervals do not include the 0.25 chance value), while children's confidence intervals partly overlap with the 0.25 chance level, especially in the Classic ME condition.

Age Effects on Children's Retention Performance

We ran additional exploratory models to further investigate age effects within *children's* (independent of adults') object choices in retention trials. We fitted a GLMM with binomial error distribution predicting only children's consistent choices by condition, age, and their interaction, including random intercepts for children. The model did not explain the data better than the null model including only random intercepts for children (χ^2 = 1.88, *df* = 3, *p* .597). Figure C4 further supports that there was no indication for effects of age or condition on children's retention performance. The full results are shown in Table C2.

Figure C4

Effect of Age and Condition on Children's Consistent Choices in Retention Trials



Note. Grey transparent triangles (Classic ME) and dots (Novelty) show the proportions of consistent object choices per child, based on trials in which any selection was made ($n_{\text{Classic ME}}$ = 66, n_{Novelty} = 71). Dotted (Classic ME) and dash-dotted (Novelty) lines represent the fitted values revealed by the binomial mixed effects model predicting children's consistent choices by age (z-transformed), condition, and their interaction. Darker (Classic ME) and lighter (Novelty) polygons show the 95% confidence intervals. Fitted values were calculated via bootstrapping with 10000 boots. Chance level = 25%.

Table C2

Results of the Exploratory Model Predicting Children's Consistent Choices in Retention Trials by Age, Condition and Their Interaction

	Estimate	*SE*	*p*	95% *CI*
Reference group for condition = Classic ME				
Intercept	-0.230	0.287	.423	-0.829, 0.313
Condition	0.384	0.395	.331	-0.371, 1.193
Z-age	-0.256	0.297	.388	-0.901, 0.297
Condition x z-log-age	0.136	0.400	.734	-0.654, 0.973
SD random intercepts for subjects	0.617			
Reference group for condition = Novelty				
Intercept	0.154	0.271	.571	-0.372, 0.644
Condition	-0.384	0.395	.331	-1.160, 0.375
Z-age	-0.120	0.272	.660	-0.734, 0.383
Condition x z-log-age	-0.136	0.400	.734	-0.910, 0.615
SD random intercepts for subjects	0.617			

Note. Binomial mixed effects model on children's consistent choices in retention trials with condition, z-transformed age (in years, continuous, M = 0, SD = 1) and their interaction as predictors and random intercepts for participants. $N_{\text{observations}}$= 134. N_{groups} = 72. The 95% confidence intervals were obtained via bootstrapping with 1000 boots.

Table C3

Children's Object Selections in Retention Trials

Object	Classic ME condition	Novelty condition
Previously selected object from the correct trial (modi/ toma)[a]	28 (40.6%)	38 (50.0%)
Previously selected object from the other trial (toma/ modi)[b]	19 (27.5%)	10 (13.2%)
Previously unselected object from the correct trial (modi/ toma)[c]	6 (8.7%)	15 (19.7%)
Previously unselected object from the other trial (toma/ modi)[d]	2 (2.9%)	4 (5.2%)
Unrelated object	11 (15.9%)	4 (5.2%)
No choice	3 (4.4%)	5 (6.5%)

[a] e.g., in a "modi"-retention trial, this would be the object the child previously selected in the "modi"-referent disambiguation trial.

[b] e.g., in a "modi"-retention trial, this would be the object the child previously selected in the "*toma*"-referent disambiguation trial.

[a] e.g., in a "modi"-retention trial, this would be the object the child previously *did not* select in the "modi"-referent disambiguation trial.

[a] e.g., in a "modi"-retention trial, this would be the object the child previously *did not* select in the "*toma*"-referent disambiguation trial.