Early-emerging combinatorial thought

Pomiechowska, Barbara; Brody, Gabor; Teglas, Erno; Kovacs, Agnes Melinda

DOI:
10.1073/pnas.2315149121

PNAS | RESEARCH ARTICLE | PSYCHOLOGICAL AND COGNITIVE SCIENCES

Early-emerging combinatorial thought: Human infants flexibly combine kind and quantity concepts

Barbara Pomiechowska[a,b,c,1], Gábor Bródy[d], Ernő Téglás[c], and Ágnes Melinda Kovács[c]

Edited by Richard Aslin, Haskins Laboratories Inc., New Haven, CT; received August 31, 2023; accepted April 1, 2024

Combinatorial thought, or the ability to combine a finite set of concepts into a myriad of complex ideas and knowledge structures, is the key to the productivity of the human mind and underlies communication, science, technology, and art. Despite the importance of combinatorial thought for human cognition and culture, its developmental origins remain unknown. To address this, we tested whether 12-mo-old infants (N = 60), who cannot yet speak and only understand a handful of words, can combine quantity and kind concepts activated by verbal input. We proceeded in two steps: first, we taught infants two novel labels denoting quantity (e.g., "mize" for 1 item; "padu" for 2 items, Experiment 1). Then, we assessed whether they could combine quantity and kind concepts upon hearing complex expressions comprising their labels (e.g., "*padu* duck", Experiments 2-3). At test, infants viewed four different sets of objects (e.g., 1 duck, 2 ducks, 1 ball, 2 balls) while being presented with the target phrase (e.g., "padu duck") naming one of them (e.g., 2 ducks). They successfully retrieved and combined on-line the labeled concepts, as evidenced by increased looking to the named sets but not to distractor sets. Our results suggest that combinatorial processes for building complex representations are available by the end of the first year of life. The infant mind seems geared to integrate concepts in novel productive ways. This ability may be a precondition for deciphering the ambient language(s) and building abstract models of experience that enable fast and flexible learning.

cognitive development | infancy | compositionality | concepts | combinatorial thought

Significance

Human creativity is unparalleled among other species. One aspect of human cognition that contributes to this feat is combinatorial thought, or the ability to assemble infinitely many complex ideas from a finite number of simple concepts. Combinatorial thought seems to be tightly linked to language use, which facilitates building and sharing complex ideas with others. Here, we show that human infants can combine quantity and kind concepts evoked by words embedded in multielement expressions (e.g., "padu duck"; "padu" was a made-up word for the quantity 2 taught during the experiment). Therefore, combinatorial processes for developing complex ideas begin to operate during the first year and may be not a consequence of language use but a perquisite for learning in general.

Human intelligence and creativity rest on combinatorial thought, or our ability to form brand-new thoughts by combining existing concepts and ideas. Combinatorial thought spans various domains of human activity, from everyday conversation to art and science, and is often supported by external symbolic systems. Some of these systems, such as mathematical notation or programming languages, are recent, require formal education, and are only available to groups of specialists. Others, such as natural languages, are ancient, spontaneously acquired by human children, and universal across human societies. Most of the time, language use appears completely effortless, which masks the fact that understanding even the simplest linguistic messages (e.g., "Coffee is ready.", "The cat is on your pillow.") requires a systematic combination of the meanings carried by their constituents and is believed to involve complex syntactic, logicosemantic, and conceptual computations (1).

Although combinatorial thought encoded in language pervades our thinking and communication, its developmental origins remain a mystery. One possibility is that extensive experience with natural language is necessary to develop cognitive tools able to support compositional processes (2–5). Although influential, this idea has received mixed experimental support. When taken at face value, evidence from early speech production appears to corroborate it. Children start uttering isolated words around their first birthday, but they take several months to a year before stringing those words into multiword utterances (6), which suggests that combining words together into meaningful sequences poses an additional computational challenge that can be overcome through exposure to language. This argument, however, rests on the disputed assumption that early complex speech is supplied by underlying combinatorial processes, which may not be the case. Instead, the early production of multiword utterances could result from memorizing certain phrases as units (e.g., "all gone!") without appreciating the meaning of their individual components (7), with the delayed production reflecting that long units may be harder to learn and produce than short ones. Hence, speech production is not an optimal test case for early combinatorial abilities.

Recent studies using nonverbal stimuli to probe combinatorial thinking in young children have yielded conflicting results. On one hand, infants and toddlers struggle to compose functions (8) and set up complex predicate structures (4). On the other hand, they

Author affiliations: [a]Centre for Developmental Science, School of Psychology, University of Birmingham, Birmingham B15 2TT, United Kingdom; [b]Centre for Human Brain Health, School of Psychology, University of Birmingham, Birmingham B15 2TT, United Kingdom; [c]Department of Cognitive Science, Central European University, Wien 1100, Austria; and [d]Department of Cognitive, Linguistic, and Psychological Sciences, Brown University, Providence, RI 02912

Author contributions: B.P., A.K., E.T., and G.B. designed research; B.P. performed the formal analysis; A.K. and E.T. acquired funding; B.P. performed project administration; B.P. created the software; B.P. visualized data; B.P., A.K., G.B., and E.T. reviewed and edited the paper; and B.P. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: b.pomiechowska@bham.ac.uk or pomiechowskab@ceu.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2315149121/-/DCSupplemental.

Published July 9, 2024.

have access to multielement structured action schemas and recruit them to represent action roles within specific events such as chasing [e.g., *chaser*, *chasee*, (9)], giving [*giver*, *givee*, (10)], or launching [*launcher*, *launchee*, (11, 12)]. Further, by the end of the first year of life, infants deploy inferences that may be an early manifestation of logical reasoning by exclusion (13, 14). These two strands of evidence make plausible the idea that at least certain combinatorial abilities, whether dedicated to specific content domains or not, may emerge early and independently of language acquisition.

Here, we seek evidence for the early availability of general combinatorial thought. In contrast to researchers who claim that compositional skills begin to be forged during the second year of life by virtue of acquiring more sophisticated language skills (2–5), we posit that infants are naturally endowed with a cognitive apparatus for building complex representations from atomic concepts. A precondition for using this apparatus is a successful deployment of conceptual representations, that is thinking of perceived entities under discrete conceptual descriptions (e.g., AGENT, DOG, BALL). Therefore, under this proposal, the limiting factor on combinatorial thought is not the lack of dedicated cognitive mechanisms, but the difficulty to deploy the concepts that could be combined. While adults spontaneously and very rapidly filter sensory inputs through the lenses of their conceptual knowledge, recognizing things around them as computers, books, or courgettes (15), infants struggle to retrieve conceptual information but can be induced to do so (16–19). Exposure to familiar words works particularly well as a trigger of conceptual access early in life, as evidenced by neurophysiological (18) and behavioral results (19). From this, an exciting possibility follows. If the current proposal is correct, infants may be able to use their combinatorial skills to work out the meaning of multiword utterances (e.g., "small dog") at the earliest stages of language acquisition. As soon as they understand the constituent words ("dog", "small"), they could retrieve and combine the concepts linked to them (e.g., DOG, SMALL). On this view, the lack of complex utterances in early production may simply reflect higher cognitive and motor demands on speech production over comprehension, a phenomenon observed also for single words (20). Relatedly, the null results in nonverbal tasks targeting infants' combinatorial abilities (4, 8) may not be due to the lack of cognitive mechanisms for conceptual combination, but to difficulties with spontaneously deploying conceptual descriptions of the task-relevant objects [also observed in other studies, (16)] which is a precondition thereof.

Expanding on evidence from core cognition (21–23) as well as language and communicative development (24–27), we put the above hypothesis to an experimental test by investigating whether 12-mo-old infants have access to combinatorial skills for building complex conceptual representations. More specifically, we examined whether they could combine concepts activated by subsequent words in previously unheard phrases. We targeted their comprehension of minimally combinatorial linguistic expressions (28) in the form of complex noun phrases (e.g., "padu duck") consisting of newly learned quantity labels (e.g., pseudoword "padu" for two) and category labels that infants at this age tend to be familiar with (e.g., "duck"). Using novel words ensured that a potential success would be evidence of on-line combinatorial processing and not an outcome of retrieving previously formed phrase-referent mappings. We proceeded in two steps: first, we established that 12-mo-olds could learn two new abstract quantity labels denoting sets of one and two (Experiment 1). Second, we examined whether they could combine them with familiar category labels upon hearing quantified noun phrases and investigated the cognitive mechanisms underlying their performance (Experiments 2 to 3). To assess infants' performance, we created an adaptation of a looking-while-listening task (29, 30). We monitored their looking behavior using eye tracking.

We tested 12-mo-olds because they have a small receptive vocabulary of common nouns denoting object categories [e.g., "apple", "ball", "duck", (16, 25)] and are currently the youngest age group shown to rapidly learn novel words in lab settings (26, 27). Their receptive vocabulary does not yet contain numerals or quantifiers (31), but previous experimental evidence indicates that they readily attach new words to preverbal categories and concepts formed in the absence of language [e.g., for artifacts, (24); for action roles, (9)]. It is also widely established that numerical concepts are an important part of the core conceptual repertoire available to human infants (32) and were shown to be operational already at birth (33). Therefore, we reasoned that 12-mo-olds may be able to learn quantity labels for small set representations (i.e., one and two). That a mapping between novel words and numerically relevant representations might be possible at such an early age gains plausibility from recent findings suggesting that 14-mo-olds already consider verbal counting routines as indicative of set sizes (34).

Experiment 1: Infants Rapidly Learn Abstract Quantity Labels

Experiment 1 examined whether 12-mo-olds can link new words to abstract numerical content. Using a word-learning task we tested whether infants (N = 20) could associate two novel pseudowords (*"mize"*, *"padu"*; conforming to the phonotactics of their native language – Hungarian) with two distinct set sizes, one and two, respectively. The task had two parts: a *word training* (6 trials) was followed without interruption by a *word generalization test* (4 trials).

Each *training* trial consisted of an animation depicting two sets of different cardinalities (1 vs. 2). Both sets contained identical familiar-category objects likely to be recognized by name by 12-mo-olds [for Hungarian, (16, 18)]. The two sets of objects were presented on the opposite sides of the display, separated by a screen (e.g., one apple on the *right* and two apples on the *left*, Fig. 1). On each trial infants first saw a hand pointing at one set (e.g., one apple) and heard a phrase containing a new quantity word (e.g., "There is *mize* apple. *Mize*! *Mize*."). Subsequently, the same labeling procedure targeted the other set (e.g., two apples) using the other quantity label (order counterbalanced). Then, within the same trial but without pointing, infants heard a question about one of the sets (e.g., "Look, where is *mize* apple?"), to prompt them to engage in referent search and familiarize them with the structure of the upcoming generalization test. Four aspects of the training design were crucial. First, while the cardinalities of the two sets were kept constant across trials (i.e., each trial featured one set of one item and one set of two items), the set members varied in category (e.g., apples on trial 1, cars on trial 2, dogs on trial 3, etc.). Therefore, infants could leverage across-trial variation in object categories (35) to infer that the novel labels do not refer to categories but to abstract numerical properties of the sets. When object categories and numerals covary even older children have difficulties interpreting novel number words (36). Second, the items within the sets were perceptually identical, which facilitates numerosity judgments in older children (37). Third, we employed sets of one and two that infants can easily represent (21, 22) and presented them in a contrastive manner within the same scene to highlight the numerical contrast (2). Finally, we used familiar object categories and their labels to narrow down the hypothesis space for novel word meaning by inviting the inference that novel words indicate something else than category membership described by familiar category labels. Note also that we repeated the quantity label twice in isolation to increase its saliency.

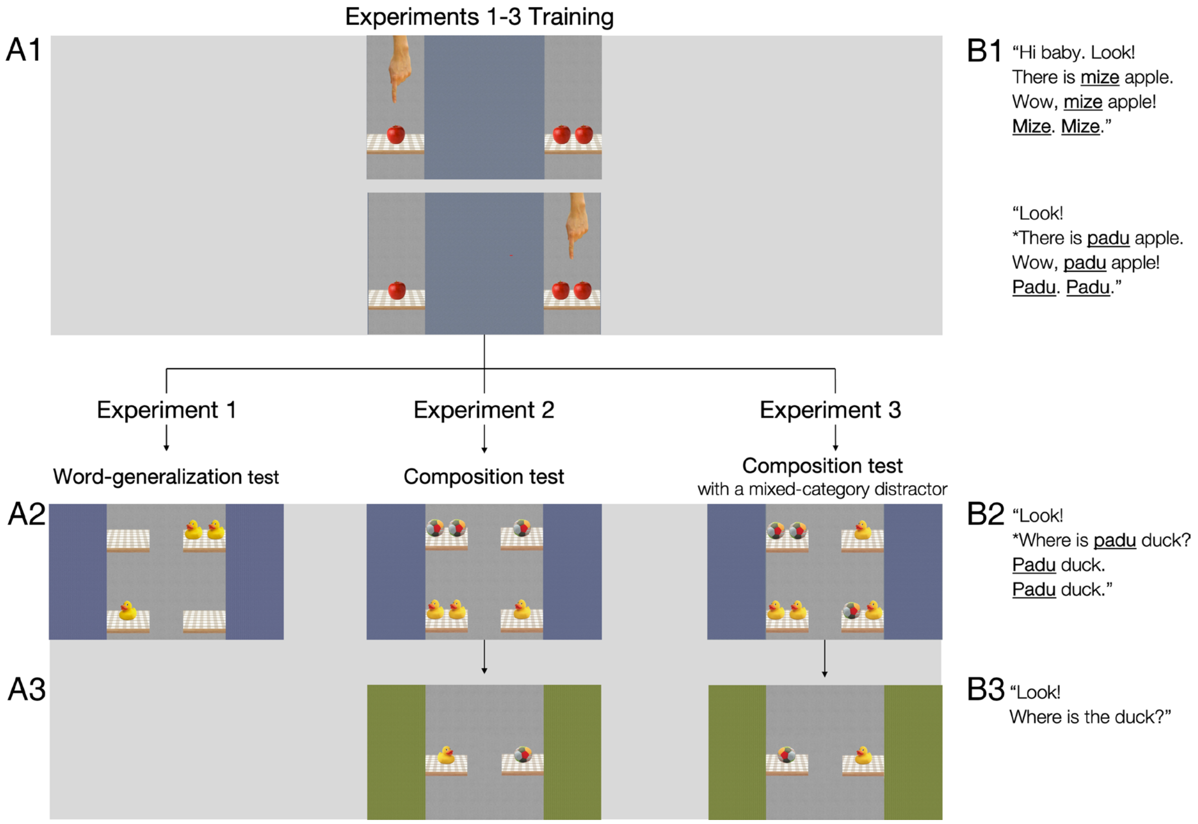

Fig. 1. Schematic of the experimental design depicting examples of (*A*) visual and (*B*) speech stimuli. (*A1* and *B1*) *Training* (6 trials). The training of novel quantity labels for sets of one and two was kept constant across experiments. Each trial comprised two sets of identical objects selected among three different familiar categories (e.g., apple, car, dog; counterbalanced across participants), with each category presented on two trials. The sets were individuated via a pointing hand and labeled sequentially using two distinct pseudowords ("mize", "padu") followed by familiar category labels. For the details of trial design and timing, please see *Materials and Methods.* (*A2* and *B2*) *Test* (4 trials). The test phase used familiar object categories that were not included in the training set (e.g., ball, duck; counterbalanced across participants). The test structure differed across experiments. Experiment 1 tested generalization of quantity labels, with one target and one distractor set. Experiments 2 to 3 tested comprehension of complex noun phrases and the test display contained 1 target and 3 distractor sets, but the structure of the distractors differed to investigate the mechanisms underlying infants' performance (Experiment 2: kind distractor, number distractor, unrelated distractor; Experiment 3: kind distractor, number distractor, mixed distractor). For the details of trial design and timing, please see *Materials and Methods.* (*A3* and *B3*) *Category-recognition test* (4 trials). In Experiments 2 to 3, the composition test was immediately followed by a category-recognition test comprising two objects, one from each category used in the composition test.

Overall, three familiar object categories were used during training (e.g., apple, car, dog), each on two separate trials. Two more familiar categories were used in the generalization test (e.g., if apples, cars, and dogs were used in the training set, ducks and balls were used in the test set, counterbalanced across participants), each on two separate trials. *Test* trials had the following structure: a set of one and a set of two identical objects were displayed on two separate shelves (Fig. 1). After a silent baseline (1.5 s), a central attention getter attracted infants' attention to the center of the screen and the test question was delivered (e.g., "Look! Where is mize duck"?). At the offset of the test question, the attention getter disappeared, and the test response period began (6.5 s). The target noun phrase was repeated two more times (e.g., "Mize duck! Mize duck.") as a reminder and a prompt to search for the referent. Two test trials involved one novel quantity label (e.g., mize) and two the other one (e.g., padu, counterbalanced). To assess infants' interpretation of the test phrase we followed a procedure established in the literature on infant word recognition (25): we measured whether the proportion of time they spent looking at the target (i.e., the named set) increased after they heard the test question relative to the proportion of time they spent looking at the target during the silent baseline. Our measure of interest was baseline-corrected proportion of looking at the target ($corrPL_{TARGET}$). To derive $corrPL_{TARGET}$, we first calculated the proportion of looking at the target during both baseline and test periods ($PL_{TARGET} = LOOK_{TARGET}/(LOOK_{TARGET} + LOOK_{DISTRACTOR})$, wherein $LOOK_{TARGET}$ corresponds to the total dwell time on the target and $LOOK_{DISTRACTOR}$ corresponds to the total dwell time on the distractor). Then, we subtracted the baseline values from test values ($corrPL_{TARGET} = PL_{TEST} - PL_{BASELINE}$; for proportional and raw looking times, see Fig. 2 *E* and *F*; for additional details, see *SI Appendix*). This measure ranges from −1 to 1 and corrects for intrinsic preferences infants might have for certain stimuli and that can be captured before exposure to speech. Positive values indicate an increase in looking at the named set, consistent with referent identification, while negative values indicate an increase in looking at the distractor. All tests reported below are two-tailed. Although our data are interval-bounded, they did not violate the assumptions of common linear models. The same patterns of results were observed using nonparametric tests; see *SI Appendix, Non-Parametric Statistics*.

Infants generalized the trained quantity labels to sets of objects that belonged to familiar categories but were not part of the training set, $M = 0.15$, $SD = 0.20$, $t(19) = 3.400$, $P = 0.003$, $d = 0.76$, 95% CI = [0.06, 0.24]; 15 out of the 20 participants increased their target looking relative to the silent baseline, binomial: $p = 0.041$. The pattern of results was similar across the two quantity labels, $t(19) = 0.751$, $P = 0.462$, $d = 0.25$, 95% CI = [−0.29, 0.14], both eliciting an increase in target looking ("one": $M = 0.17$, $SD = 0.25$; "two": $M = 0.10$, $SD = 0.34$); see Fig. 2*C*. This generalization success indicates that infants in this task formed abstract label-quantity

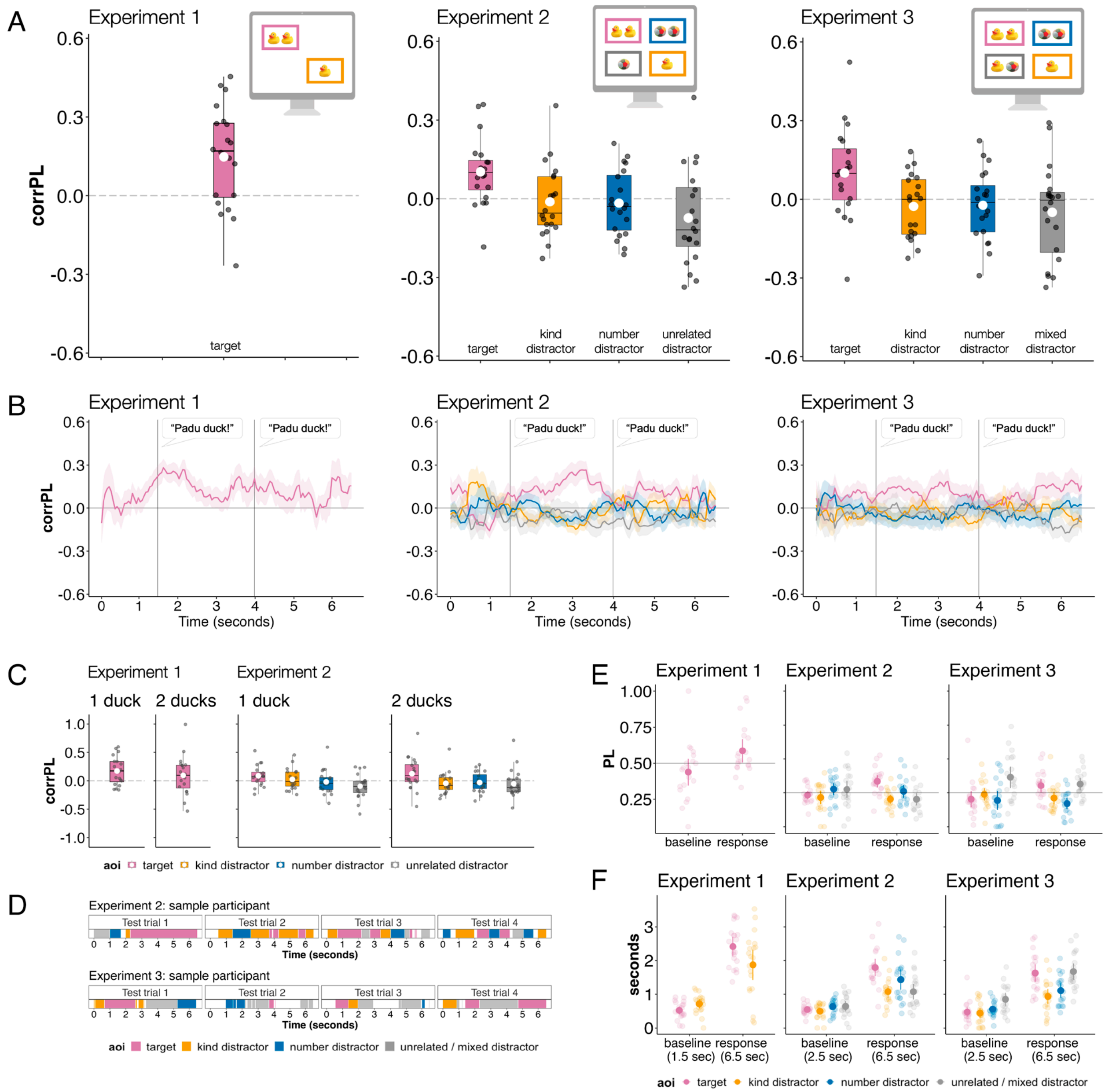

Fig. 2. Results of Experiments 1 to 3. (*A*) Mean baseline-corrected proportion of looking (*corrPL)* at potential referents. This measure captures changes in looking at the referent sets during the test response period relative to the silent baseline. Positive values indicate an increase and negative values indicate a decrease in looking at the test set. (*B*) The time course of *corrPL* during the test response period. Zero on the *x* axis marks the beginning of the response period and corresponds to the offset of the test question (e.g., "Where is *padu duck?*"). Vertical lines correspond to the onsets of two repetitions of the target noun phrase (e.g., *"Padu duck."*) administered during the test response period, at 1.5 and 4 s. Shadowing indicates SEMs, with means computed over participants for each time point. Descriptively, at the beginning of the test response period, infants seemed to display comparable looking at different referents, just as observed during baseline. Then, around the time when the test phrase was repeated their target looking rose above the baseline level. (*C*) Mean *corr*PL split by cardinality across Experiments 1 to 2. (*D*) Exploration of the referents during the test response period in individual participants. Horizontal bars represent AOI hits registered in different referent AOIs across trial time (*Materials and Methods, Areas of interest*). Colors correspond to different referents. Gaze points registered on-screen but outside of the referent AOIs are not represented in this figure. Two sample participants are depicted here. Overall, infants seemed to move their gaze around the display rather than permanently settle on the target. Please see *SI Appendix, Scene Exploration during Test* for further details and *SI Appendix*, Fig. *S2* for visualizations of datasets from all participants and further discussion. (*E*) Proportion of looking (PL) at the potential referents (*Materials and Methods, Data Analysis*) during baseline and test response periods. (*F*) Raw looking time at the potential referents during baseline and test response periods. Note that the durations of baseline and test response periods were different across experiments (Experiment 1: 1.5 s baseline, 6.5 s test response; Experiments 2 to 3: 2.5 s baseline, 6.5 s test response). (*E* and *F*) Mixed distractor in Experiment 3 seemed to attract more attention than other referents. For further details on baseline looking, see *SI Appendix, Baseline Looking during Test*. *Boxplots* (*A* and *C*): White circles indicate means; dots indicate mean values from individual infants, horizontal lines within boxes indicate medians, bottom and the top of the boxes represent the first and the third quartiles. *Scatterplots* (*E* and *F*): Solid circles indicate means; dots indicate mean values from individual infants. Bars indicate 95% CI.

mappings. The learning process was fast, occurring over the course of a 6-trial training. Infants isolated novel words from the speech stream, interpreted them as relevant to numerical properties of the stimuli and retained this information in memory.

Note that although the current test involved complex phrases, it was not designed to test combinatorial processing. This is because both target and distractor sets contained objects from the same category (e.g., 1 duck vs. 2 duck). Therefore, attending solely to the quantity label while disregarding the category label would have been sufficient to find the target. Infants' success at this task was a precondition to test their combinatorial abilities in Experiments 2 to 3.

Experiments 2 to 3: Infants Combine Quantity and Kind Concepts

Experiment 2 tested whether 12-mo-olds could combine quantity and kind concepts activated by newly learned quantity labels and familiar category labels, respectively. As before, the task had two parts: a *word training* (identical to Experiment 1) and a *composition test*. The composition test involved the same speech prompt as the word generalization test in Experiment 1 (e.g., "Where is mize duck"?) and a different visual display (i.e., four sets of potential referents instead of two). There was one target set (i.e., the set labeled by the test phrase, e.g., one duck) and three distinct distractor sets: 1) the kind distractor matching the target in kind and corresponding to the category label (e.g., two ducks); 2) the number distractor matching the target in number and corresponding to the quantity label (e.g., one ball); 3) the unrelated distractor (e.g., two balls) having no overlap with the target set, matching neither of the labels in the test phrase, and used to equalize the number of objects per category (Fig. 1).

A composition test structured this way could only be solved by considering both constituents of the test phrase (e.g., "mize duck"). If infants combine the concepts linked to both constituents into a single referent, they should selectively increase their looking at the target. Alternatively, if they selectively focus on one of the constituents, either the new quantity label or the familiar category label, they should increase their looking to two referents: the target and the distractor corresponding to the attended constituent (e.g., the target and the number distractor, if their attention was captured by the new quantity label, or the target and the kind distractor, if their attention was captured by the familiar category label). Because the composition test involved more objects than the generalization test in Experiment 1, we extended the duration of the baseline to 2.5 s to give infants more time to explore the display. Object categories used in the composition test were different from those used during training to ensure that infants combine the heard words on-line and cannot rely on memory traces of phrases heard before. Finally, it is important to note that performance at the composition test critically depends not only on infants' compositional abilities and learning novel quantity labels, but also on their knowledge of the presented category labels, i.e., it is impossible to compute the complex meaning of the phrase "mize duck" without being able to retrieve the concept linked to the word "duck". Although we selected category labels that infants at this age are on average able to recognize, there is significant interindividual variability in their word-recognition performance (38). Therefore, to assess recognition of the familiar labels used in the composition test, we administered a category-recognition test that followed without interruption. To be included in the composition analysis infants had to pass the category-recognition test: that is, they had to display understanding of the familiar category labels from the composition test (for details, see *SI Appendix, Category-Recognition Test*).

To evaluate comprehension of complex noun phrases, we computed baseline-corrected proportion of looking at each of the four referents present during the composition test (target, kind distractor, number distractor, unrelated distractor; see *Materials and Methods, Experiment 2*). Infants (N = 20) increased their looking to the target significantly above the baseline level, M = 0.10; SD = 0.13; $t(19)$ = 3.570, P = 0.002, d = 0.80, 95% CI = [0.04, 0.16], 16/20 infants, binomial: P = 0.012, but failed to do so for any of the distractors (kind distractor: M = −0.01; SD = 0.14; $t(19)$ = 0.387, P = 0.703, d = 0.08, 95% CI = [−0.08, 0.05]; number distractor: M = −0.02; SD = 0.13; t(19) = 0.612, P = 0.548, d = 0.14, 95% CI = [−0.07, 0.04]; unrelated distractor: M = −0.07; SD = 0.19; t(19) = 1.770, P = 0.093, d = 0.40, 95% CI = [−0.16, 0.01]). The increase in target looking was significantly higher than the changes in looking at the distractors (target vs. kind distractor: $t(19)$ = 2.502, P = 0.022, d = 0.56, 95% CI = [0.02, 0.21]; target vs. number distractor: $t(19)$ = 2.969, P = 0.008, d = 0.66, 95% CI = [0.04, 0.20]; target vs. unrelated distractor: $t(19)$ = 2.830, P = 0.011, d = 0.63, 95% CI = [0.05, 0.31]).

This pattern of results supports two conclusions. First, in line with Experiment 1, 12-mo-olds rapidly mapped novel words onto abstract concepts with numerical content. Second, and importantly for our understanding of their combinatorial abilities, they used this newly acquired knowledge to determine the composite meaning of complex noun phrases by considering both the newly trained quantity labels and familiar nouns. To succeed at the composition test, infants must have accessed the concepts linked to both constituent words and used them jointly to find the referent set that satisfied both. Moreover, the selective increase in looking at the target but not the distractors matching either the quantity or category labels suggests that infants combined the activated concepts into a unified referent representation (e.g., of a single duck, or of a pair of ducks) rather than independently focusing on and processing single words (e.g., "mize" at time t, "duck" at time t+1). The latter strategy would have led to an increase in looking at all sets that contain a token of at least one of the labeled concepts (e.g., in the case of mize duck ~ *1 duck*, these would be the set comprising one duck along with the sets comprising one ball and two ducks).

What kind of compositional processes do infants recruit to interpret quantified noun phrases? There are three main possibilities. Infants may simply coactivate multiple concepts, concatenate coactivated concepts in a list-like representation, or just as adults, properly combine them in a single function. The first two processes, activating two concepts and setting up composite representations by concatenating them (e.g., [2, DUCK]) predict that referent search should be satisfied by all sets that individually satisfy both concepts. For example, in this case the expression "padu duck" would be applicable to any set that contains two items, at least one of which is a duck, taking even mixed-category sets, e.g., {ball duck}, as valid referents along with single-category sets, e.g., {duck duck}. In contrast, if infants use a quantification-like strategy to track the number of items within sets defined through category labels by applying functions with variables whose value is assigned based on the category label, they should not accept mixed sets as referents. For example, if [x] is linked to the word "mize" and [x x] to the word "padu" and the interpretation rule takes the value of x from the category label, then the expression "padu duck" maps exclusively to the homogenous, single-category set {duck duck}, and not to the mixed-category set {ball duck}.

Experiment 3 was designed to disentangle these strategies: Infants watched the same training of quantity labels as in Experiments 1 to 2 and a modified composition test. As before, the composition test contained four referents, 1 target and 3 distractors, but the unrelated distractor was now replaced with a mixed distractor (Fig. 1), and the test question always featured the quantity label paired with the set of two. The mixed distractor was a set of two different objects, one token of the labeled kind and one token of a different kind (e.g., a ball and a duck). If infants simply coactivate the two concepts or recruit a concept concatenation strategy, they should increase their looking not only to the target set (e.g., a set of two ducks) but also to the mixed distractor. In contrast, if they recruit a quantification-like process, they should selectively increase their looking to the target. As in Experiment 2, the composition test was followed by a category-recognition test aimed to determine which participants displayed recognition of the presented kind labels and could be included in the composition analysis. To evaluate infants' comprehension during the composition test, we computed baseline-corrected proportion of looking at each of the four referents.

Infants (N = 20) increased their looking to the target set relative to the silent baseline, M = 0.10, SD = 0.17, $t(19)$ = 2.574, P = 0.019, d = 0.59, 95% CI = [0.02, 0.18], a pattern displayed by 15 out of 20 participants, binomial: P = 0.041. The changes in looking to individual distractors did not differ from baseline (kind distractor: M = −0.03; SD = 0.12; $t(19)$ = 1.000, P = 0.330, d = 0.22, 95% CI = [−0.08, 0.03]; number distractor: M = −0.02; SD = 0.13; $t(19)$ = 0.768, P = 0.452, d = 0.17, 95% CI = [−0.08, 0.04]; mixed distractor: M = −0.05; SD = 0.18; $t(19)$ = −1.237, P = 0.231, d = 0.28, 95% CI = [−0.13, 0.03]). The increase in target looking was larger than changes in looking at distractor sets (target vs. kind distractor: $t(19)$ = 2.279, P = 0.034, d = 0.51, 95% CI = [0.01, 0.24]; target vs. number distractor: $t(19)$ = 2.328, P = 0.031, d = 0.52, 95% CI = [0.01, 0.23]; target vs. mixed distractor: $t(19)$ = 2.141, P = 0.046, d = 0.48, 95% CI = [0.00, 0.29]).

Following naming, infants focused on the target set, consisting of two objects from the named category. The proportion of time they spent exploring the mixed distractor, consisting of two objects only one of which represented the named category, did not change relative to the silent baseline. This behavior is consistent with using kind concept to define the domain over which the abstract quantity concepts should be deployed, similarly to how adults interpret quantified noun phrases containing numerals. It is also possible that infants interpreted the quantity label denoting the set of two as more instead of two. Importantly, however, such interpretation could yield the same pattern of results as observed here only through a similar combination of the named concepts (i.e., by applying the concept MORE to the domain specified by the kind concept). Further, these results rule out that infants simply mentally coactivated or concatenated the two labeled concepts (e.g., [2, DUCK] and applied two separate search criteria (i.e., "search for DUCK", "search for 2"), as this strategy would not lead to the elimination of the mixed distractor. Only a modified version of this strategy, wherein infants apply category concepts (e.g., DUCK) directly to sets rather than individuals (e.g., "search for homogenous groups of DUCK", "search for two"), could produce a similar outcome by excluding the mixed set from consideration as a potential referent. While this possibility does not seem very likely given the current work on early categorization demonstrating that infants spontaneously parse arrays of two objects based on their categories as early as 4 mo of age (39), pinpointing the nature of the early quantity labels and operations they enter remains an exciting challenge for future work.

Discussion

The ability to build complex representations from simple ones is a landmark of human thought and language (40–43). Our findings reveal that well before being able to speak and express their own thoughts via language, young infants begin to interpret complex thoughts expressed verbally by others by spontaneously combining concepts elicited by words. In our experiments, 12-mo-olds were able to identify referents of two-element noun phrases (e.g., "padu duck") consisting of familiar category labels (e.g., duck), acquired prior to the lab visit, and new quantity labels (e.g., pseudoword "padu" for 2 items), introduced and taught during the experiment. Upon hearing a stream of speech, they segmented it into constituent words, and not only activated the associated concepts but also combined them into representations that singled out specific sets of objects in their visual environment. This composition process was systematic (43), enabling a successful combination of numerical concepts and kind concepts linked to familiar category labels that were not part of the training set. Therefore, at least some components of the cognitive apparatus that supports combinatorial thought are operational in infancy, long before children gain experience combining linguistic symbols in their own utterances. From early on, the human mind seems geared to construct complex representations, an ability that enables the unbounded creativity and flexibility of thought.

There is a considerable debate about the origins of the human capacity for complex thought. Some argued that it arises from the human-specific language faculty, and it can manifest in various combinatorial capacities that, intuitively, seem disconnected from language such as mathematics (44). Others proposed that the combinatorial processes carried out on concepts may be nonverbal and independent from natural language (41). The former proposal is corroborated by the fact that the complexity of human thought is unparalleled in the animal world and resonates with our subjective impression of relying on language while thinking. The latter proposal has recently gained support from neuroimaging work indicating that complex combinatorial operations supporting advanced mathematical judgments (45) and computer code comprehension (46) involve neural circuits that do not overlap with linguistic networks. Comparative work suggesting that certain nonhuman species form multielement expressions from simple meaning-bearing units (47) can also be seen as supporting the independence of certain combinatorial operations from language, although the nature of these processes is debated (48).

Our investigation of preverbal infants aimed to shed further light on whether combinatorial abilities can be deployed by the end of the first year of life. Contrary to theoretical proposals that combinatorial thought is underpinned by progresses in language acquisition (3), the present results indicate that neither the experience with language production nor a large mental lexicon or mastery of syntax are required to perform basic combinatorial operations, such as combining two different concepts into a single complex representation. However, given that our task used natural language stimuli, it remains open whether the presence of words is necessary, or they merely catalyze conceptual access and through that enable conceptual combination. Words may only trigger the activation of nonverbal concepts, which then serve as input to nonlinguistic combinatorial operations. Future studies should establish how complex thought comes into being during early ontogeny, addressing whether infants can form complex representations without the mediation of language, as well as in phylogeny, by investigating nonhuman animals. Further, future research should also clarify how the mediation of language affects compositional thought. One exciting avenue is to explore whether early compositional abilities support context-sensitive concept composition (e.g., as required by scalar adjectives "big mouse" vs. "big car").

Besides uncovering early emerging combinatorial abilities, our study also provides evidence for infants' early ability to link words and abstract numerical representations. These findings are in line with previous findings that 14-to-18-mo-olds interpret verbal counting routines as numerically relevant (34). The present success at generalizing new quantity labels provides three initial conclusions that may inform future efforts to better characterize the concepts that guide the early learning of numerical expressions. First, the numerical concepts infants recruited in the present task were abstract, i.e., dissociated from featural and kind information. Second, they were readily indexed under external symbols in the form of words. Third, and most importantly, they entered compositional computations that led to the formation of complex referent representations.

Regarding the origins and format of these numerical concepts, several possibilities should be considered. Infants may be recruiting one of the two number representation systems available to them [(i.e., the object tracking system, or the analog magnitude system, (23)], deploying representations such as singular/plural also documented in nonhuman primates (49), or, perhaps, using primitive representations such as singleton and pair (50, 51); or chunk, (19). Importantly, however, independently of the exact format of those numerical representations, the present data show that infants can flexibly combine numerical concepts with other concepts to form complex thoughts that guide their behavior.

Finally, the application of rapidly learned quantity labels to identify referent sets named by others bears a certain resemblance to counting. As such, it raises the question of why infants succeed in this task while older children take years to learn numerals and understand the cardinality principle underlying counting (23). We suggest that infants have access to sophisticated representational formats that may be precursors of quantification. Our Experiment 3 shows that in addition to coactivating numeral and kind concepts, they combine them in such a way that the kind concept specifies the domain of the numeral. However, as mentioned above, there are reasons to believe that these concepts may be specific to set sizes of one and two, and likely are not the concepts of integers required for counting. Further, although learning label-quantity associations in the lab task tailored for this purpose was remarkably fast, we do not know whether infants stored them in their long-term memory and would be able to apply them in other contexts. Note also that counting rests on the use of mental algorithms that specify the relationship between number symbols and the successor function (52). These counting algorithms are believed not to be part of our evolved numerical toolkit but a product of cultural learning including, albeit not limited to, the acquisition of number words.

In conclusion, the present study offers initial evidence that human infants may be naturally compositionally minded. A cognitive apparatus that gives rise to combinatorial thought becomes operational already by the end of the first year of life. Infants' propensity to combine atomic concepts into complex conceptual structures likely guides the extraction of linguistic meaning during language acquisition and, more widely, may be fundamental to building symbolic models of the world claimed to be the engine of unparalleled human learning (53).

Materials and Methods

Participants. Participants were monolingual 12-mo-olds growing up in Hungarian-speaking families. All participants were typically developing and born full term. The final samples in each experiment consisted of 20 infants (*Experiment 1*: 8 females, age: M = 12 mo 2 d; *R* = 11 mo 18 d to 12 mo 20 d; *Experiment 2*: 13 females; age: M =12 mo 1 d, R = 11 mo 11 d to 12 mo 19 d, *Experiment 3*: 11 females, age: M = 12 mo 3 d, R = 11 mo 18 d to 12 mo 25 d). In Experiment 1, an additional six infants were tested but had to be excluded from the final sample ($n = 1$ cried; $n = 5$ did not provide enough data); in Experiment 2, 9 infants were excluded ($n = 5$ did not complete the task; $n = 2$ provided less than 3 valid training trials; $n = 1$ due to experimenter error; $n = 1$ was identified as bilingual after testing); in Experiment 3, *14* infants were excluded ($n = 7$ failed to complete the task; $n = 3$ due to parental interference, i.e., talking during the task; $n = 2$ failed to provide enough valid training trials; $n = 1$ failed to provide enough valid category-recognition trials; $n = 1$ due to equipment failure). Further, in Experiments 2 to 3, 17 infants did not meet the criteria for inclusion in the composition test (i.e., they failed to display familiar word recognition in the category-recognition test, that would have been necessary for composition: Experiment 2: $n = 9$; Experiment 3: $n = 8$). Please see the *Data Analysis* section for the details of the inclusion criteria. All caregivers provided written informed consent. Infants received small gifts for their participation. The sample size was determined a priori based on word-mapping studies (9, 27) which used similar methodology (i.e., eye tracking measures, training using ostensive signals and/or deictic gestures), and investigated a similar age group (12-to-14-mo-olds). The relevant experiments yielded large effect sizes [(27): Experiment 1, $d = 0.84$; (9): Experiment 1, $d = 1.47$]. Therefore, using G*Power 3.1 (49), we estimated that testing 20 participants per condition would be sufficient to provide 90% statistical power to detect a large effect size (Cohen's $d = 0.80$) in comparison against chance, applying an α of 0.05. The sample size of 20 was kept constant across all experiments.

Stimuli. We selected 5 object categories whose names are likely to be recognized by 12-mo-olds [apple, ball, car, dog, duck, for word-recognition studies in Hungarian (16, 18) and 5 colorful object images (display size: approx. 150 × 150 px per image], 1 per category. The images were used to create the experimental animations (*Design*). *Word stimuli* were 5 category labels (i.e., Hungarian common nouns: "*alma*" - apple, "*autó*" - car, "*labda*" - ball, "*kacsa*" - duck, "*kutya*" - dog) and 2 novel CVCV pseudowords that served as quantity labels (*"padu"*:, *"mize"*). They were phonologically distinct and compatible with the Hungarian phonotactics. One was used to label sets of one and the other one to label sets of two (counterbalanced across-subjects). The category and quantity labels were combined into complex noun phrases of the following structure: *quantity label* + *noun*, e.g., "mize kacsa" (English: "mize duck"). The noun phrases were embedded in carrier phrases that differed across training and test. At training, infants heard: "*Szia baba, nézd csak! Itt van* QUANTIFIER LABEL NOUN. *Hűű,* QUANTIFIER NOUN*!* QUANTIFIER*!* QUANTIFIER*!"* (English: "Hi baby, look! There is QUANTIFIER NOUN! Wow, QUANTIFIER NOUN! QUANTIFIER! QUANTIFIER!", e.g., "Hi baby, look! There is mize duck! Wow, mize duck! Mize! Mize!"). At test, they heard: "*Nézd csak! Hol van* QUANTIFIER NOUN*?* QUANTIFIER NOUN*!* QUANTIFIER NOUN*!"* (English: "Look! Where is QUANTIFIER NOUN? QUANTIFIER NOUN! QUANTIFIER NOUN!", e.g., "Look! Where is mize duck? Mize duck! Mize duck!"). In Hungarian, nouns in numeral noun phrases are not supplied with a plural marker, hence we used nouns and verbs in the singular form and the same carrier phrases for sets of one and two. The speech stimuli were recorded by a female native speaker of Hungarian using infant-directed speech. For each combination of quantifier and noun, we used a single voice recording to ensure that infants looking responses would not be affected by auditory differences between labeling phrases.

Design and Procedure. The task had two main parts administered without interruption: *training* (6 trials, identical across experiments) and *test* (4 trials, different across experiments). A *word-generalization* test was administered in Experiment 1; a *composition test* was administered in Experiments 2 to 3. Further, in Experiments 2 to 3, the composition test was followed by a category-recognition test (4 trials). One set of three object categories was used during training (e.g., dog, apple, shoe) and another set of two different categories was used at test (e.g., duck, ball, counterbalanced across subjects).

The experiment took place in a dimly lit soundproof room. Infants were seated on their caregivers' lap, approximately 60 cm away from the monitor. Caregivers wore opaque sunglasses to ensure that they could not see the screen. The task was preceded by an infant-friendly five-point calibration routine. Calibration stimuli were colorful rotating spirals whose appearance was accompanied by short jingles. The calibration sequence was repeated until the infant provided calibration data for at least four calibration points.

Training (Experiments 1-3, 6 trials). Each training trial had two phases: *labeling* and *question*. In the labeling phase, infants saw two shelves arranged symmetrically at the opposite sides of the display and separated by an opaque screen (of varying color: light blue, light green, or light purple; randomized across trials to make the stimuli more variable, and thus more engaging). There were three identical objects spatially divided into two sets. One object was placed on one shelf and two objects on the other. First, the objects were displayed for 2 s, accompanied by the phrase "Hi baby!". Then, a pointing hand appeared at the top of the display above one of the shelves, moved downward (1.3 s, sound: "Look!"), stopped above the targeted set of objects (e.g., two dogs), and the naming phrase was delivered (e.g., "There is padu dog. Wow, padu dog. Padu! Padu!", 7.6 s). Following naming, the hand moved upward and left the display (1 s). After 1 s of pause, this sequence of events was then repeated targeting the other referent set (e.g., one dog) using a different quantity label (e.g., "Look! There is mize dog. Wow, mize dog. Mize! Mize!"). Next, the objects were covered by two opaque screens that moved in horizontally from the sides of the display. Once the objects were fully occluded, the question phase began without interruption: The same sets of objects, as used in the labeling phase, were placed in new locations and a question prompt was delivered in a gaze-contingent manner. The aim of the question phase was to familiarize infants with the structure of the upcoming word-generalization test and to explore the time course of learning novel quantity labels (for details, see *SI Appendix*).

The training set contained objects from three basic-level categories (e.g., apple, car, dog), different than those used at test. Infants were shown a total of 6 training trials, 2 per category. Overall, one training trial lasted approximately 30 s, depending on how rapidly infants oriented to the within-trial attention getter. Trials were preceded by a central attention getter displayed against a uniform background.

Word-generalization test (Experiment 1, 4 trials). The word-generalization test involved two object categories that were not part of the training set. There were four shelves located in the central part of the display. Two of the shelves were empty and two others held previously unseen familiar objects, identical to each other and spatially divided into two sets: One object was placed on one shelf and two on another shelf. A 1.5 s static baseline display, administered to measure infants' spontaneous looking behavior, was followed by the appearance of a centrally located gaze-contingent attention getter. Looking at the attention getter continuously for 500 ms triggered the onset of the test question (e.g., "Look! Where is mize duck?"; in case of failure to accumulate 500 ms of cumulative looking at the attention getter, the test question was programmed to start after 4 s but all participants succeeded to trigger the question with their gaze). The disappearance of the attention getter was timed to the offset of the test question and marked the beginning of the test response period (that lasted 6.5 s). The target noun phrase was repeated two more times (e.g., "Mize duck! Mize duck"; onset at 1.5 s and 4 s relative to the beginning of the measurement period). The four-location structure of the word-generalization test display was designed to match the structure of the composition test display used in the subsequent experiments, which involved four sets of potential referents. Infants were presented with 4 word-generalization test trials. For the details of the applied counterbalancing, please see *SI Appendix*.

Composition test (Experiments 2-3, 4 trials). The composition test used object categories that were not part of the training set and had the same overall event structure as the word-generalization test. However, we introduced one critical change to the test display. Unlike in the word-generalization test, here, each test trial involved four sets of potential referents representing two distinct categories (kept constant across four test trials, randomized across participants) and two quantities. There were one target and three distractors, presented in four spatially distinct locations (e.g., 1 duck, 2 ducks, 1 ball, 2 balls). Distinct sets of objects were located on separate shelves, each equidistant from the center of the display. To accommodate for the higher number of referent sets (i.e., 4 sets here instead of 2 sets in Experiment 1) and provide infants with enough time to explore all sets presented, the duration of the baseline was extended to 2.5 s. The duration of the measurement period remained unchanged (6.5 s). The target set was the set labeled by the test phrase (e.g., "Look! Where is mize duck?" for 1 duck). Two distractor sets corresponded to the individual components of the noun phrases featured in the test question: The number distractor set (e.g., 1 ball) was matched in numerosity with the target, thus satisfying the meaning of the novel quantity label (e.g., "mize"); the kind distractor set (e.g., 2 ducks) was matched in kind with the target, thus satisfying the meaning of the familiar category label (e.g., "duck"). The third distractor set varied across experiments: An *unrelated* distractor in Experiment 2 (e.g., 2 balls) did not match any of the target words. It was included to equalize the number of objects per category and to ensure that test displays on the singleton and pair trials will look the same. A *mixed* distractor in Experiment 3 was a set of two different items (e.g., 1 ball and 1 duck) comprising one object falling under the category label used in the test phrase. In Experiment 2 all sets were labeled on subsequent trials (order counterbalanced), while in Experiment 3 only the sets of two items were labeled.

Category-recognition test (Experiments 2-3, 4 trials). This phase followed the composition test without interruption. Each trial involved two objects from the two categories used before in the composition test. The test display involved two shelves located in the central area of the screen, each containing a single object, as in the training. The trials comprised a short silent baseline (1.5 s), followed by a centrally located gaze-contingent attention getter present during the delivery of the test phrase ("Look, where is the [noun]?"; Hungarian: "Nézd csak, hol van a [noun]?"), and a measurement period (3.5 s), during which the target noun was repeated two more times in isolation. We counterbalanced within subjects the order of the test labels (ABAB) and the location of the target object (ABBA); and across subjects which target location was tested first (target on the left first vs. target on the left second).

Data Analysis.

Inclusion criteria in Experiment 1. To be included in the final analysis, infants had to provide a minimum of 3 valid training trials (out of 6) and 2 valid word-generalization test trials (out of 4), minimum one per condition (singleton vs. pair). A training trial was considered valid if the infant provided eye tracking data for a minimum of 50% of the total labeling phase (encompassing the initial presentation of the objects, and the two sequences consisting of the following events: pointing, naming, hand exiting the scene, see *Design*) as well as a minimum of 60% of each individual naming sequence. A word-generalization test trial was considered valid if the infant attended to the screen for a minimum of 60% of both the baseline and measurement periods (data analyses using 50% and 70% thresholds yield identical patterns of results; see *SI Appendix, Inclusion criteria* for further details). Five out of 27 participants failed to satisfy these criteria.

Inclusion criteria in Experiments 2-3. We used the same trial inclusion criteria as in Experiment 1, and we extended these criteria to the category-recognition test trials (i.e., 60% attendance during baseline and measurement periods). To be considered for the main composition analysis, each infant had to contribute a minimum of 3 valid training trials, 2 valid composition trials, and 2 valid category-recognition trials, as well as provide a positive score in the category-recognition test to ensure that they can recognize category labels presented to them in the composition test, a precondition for interpreting complex noun phrases.

Areas of interest (AOIs). To quantify infants' gaze behavior, we defined non-overlapping AOIs (size: 525 × 570 px) around the potential referents (i.e., sets of objects), as well as around the attention getter (AOI size: 150 × 120 px). We determined where infants looked by deriving AOI hits: For each sample and each AOI, the AOI hit was scored as 1 if the gaze was recorded within that AOI and as 0 if the gaze was recorded outside. The analysis script including the AOI coordinates defined for different phases of the experiment is available in the project's OSF repository: https://osf.io/cnez5/.

Main Measure.

Experiment 1. To assess learning of the novel quantity labels, we calculated the mean baseline-corrected proportion of target looking ($corrPL_{\mathrm{TARGET}}$) by 1) obtaining the proportions of looking *(PL)* at the target (PL_{TARGET}) during baseline and test of each trial; this was done by dividing the time spent looking at the target by the total time spent looking at the target and distractor for each baseline and test period ($PL_{TARGET} = LOOK_{TARGET}/(LOOK_{TARGET} + LOOK_{DISTRACTOR})$, where $LOOK_{TARGET}$ indicates the total looking at the target, i.e., sum of AOI hits in the target AOI, and $LOOK_{DISTRACTOR}$ indicates the total looking at the distractor, i.e., the sum of AOI hits in the distractor AOI; 2) subtracting the baseline PL_{TARGET} from the test PL_{TARGET} ($corrPL_{TARGET} = PL_{TARGET\_TESTRESPONSE} - PL_{TARGET\_BASELINE}$); 3) averaging $corrPL_{TARGET}$ within-participants. The $corrPL_{TARGET}$ reflects how infants' looking behavior changed in response to the test phrase relative to the silent baseline within each trial and corrects for preferences that infants might have for individual stimuli (e.g., for 2 objects over one object). It ranges from -1 to 1 and corrects for intrinsic preferences infants may have for different objects and quantities; positive values indicate an increase in looking at the target relative to baseline, while negative values indicate an increase in looking at the distractor. We used proportions and not cumulative gaze duration to ensure that infants who spent overall more time looking at the screen did not disproportionately influence the results.

Experiment 2. To assess infants' performance at the composition test, we computed the mean proportion of baseline-corrected looking at each of the four sets: target ($corrPL_{\mathrm{TARGET}}$), number distractor ($corrPL_{\mathrm{NUM}}$), kind distractor ($corrPL_{\mathrm{KIND}}$), unrelated distractor ($corrPL_{\mathrm{UNREL}}$). This computation had three steps: 1) we derived the *PL* at each referent during baseline and during tests (e.g., $PL_{TARGET} = LOOK_{TARGET}/(LOOK_{TARGET} + LOOK_{NUM} + LOOK_{KIND} + LOOK_{UNREL})$, where $LOOK_{TARGET}$, indicates the total time spent looking at the target, $LOOK_{NUM}$ indicates the total time spent looking at the number distractor, $LOOK_{KIND}$ indicates the total time spent looking at the kind distractor, $LOOK_{UNREL}$ indicates the total time spent looking at the unrelated distractor); 2) we derived the baseline-corrected proportion of looking at each referent (e.g., $corrPL_{TARGET} = PL_{TARGET\_TEST} - PL_{TARGET\,BASELINE}$; $corrPL_{NUM} = PL_{NUM\_TEST} - PL_{NUM\,BASELINE}$); 3) we averaged within-participants the *corrPL* scores for each referent.

Experiment 3. We applied the same procedure as in Experiment 2, replacing the unrelated distractor with the mixed distractor.

Exploratory analyses: time course. The exploratory time course plots were obtained for the target set (Fig. 2) to describe the dynamics of infants' responses over the course of the test response period. We plotted the evolution of the *corr-*PL_{TARGET}. The time course data were derived within-trial by subtracting the overall $PL_{TARGET\_BASELINE}$ (see above) from PL_{TARGET_TEST} computed for 50 ms bins obtained by grouping the raw data of the test response period. Then, we computed individual averages by averaging over test trials for each participant, and finally, we computed grand averages averaging over participants.

Data, Materials, and Software Availability. Anonymized raw gaze data have been deposited in OSF (https://osf.io/cnez5/) (54).

ACKNOWLEDGMENTS. We thank Agnes Volein, Mari Toth for help with participant recruitment; Krisztina Andrasi, Dorottya Meszegeto, Iulia Savos for help with data acquisition; Eszter Körtvélyesi for help with speech recordings; Oana Stanciu, Gergely Csibra, Romy Frömer, Lei Zhang for feedback. This study was partly supported by the European Union's Horizon 2020 Research and Innovation Programme ERC Starting Grants #284236, #639840, and by James S. McDonnell Foundation, Grant #220020449.

1. L. Pylkkänen, The neural basis of combinatory syntax and semantics. *Science* **366**, 62–66 (2019).
2. E. S. Spelke, Core knowledge, language, and number. *Lang. Learn. Dev.* **13**, 147–170 (2017).
3. E. S. Spelke, *What Babies Know: Core Knowledge and Composition* (Oxford University Press, 2022), **vol. 1**.
4. M. Shukla, J. de Villiers, The role of language in building abstract, generalized conceptual representations of one-and two-place predicates: A comparison between adults and infants. *Cognition* **213**, 104705 (2021).
5. W. Hinzen, Reference across pathologies: A new linguistic lens on disorders of thought. *Theor. Linguist.* **43**, 169–232 (2017).
6. E. V. Clark, *First Language Acquisition* (Cambridge University Press, 2009).
7. M. Tomasello, The item-based nature of children's early syntactic development. *TiCS* **4**, 156–163 (2000).
8. S. Piantadosi, H. Palmeri, R. Aslin, Limits on composition of conceptual operations in 9-month-olds. *Infancy* **23**, 310–324 (2018).
9. J. Yin, G. Csibra, Concept-based word learning in human infants. *Psychol. Sci.* **26**, 1316–1324 (2015).
10. D. Tatone, A. Geraci, G. Csibra, Giving and taking: Representational building blocks of active resource-transfer events in human infants. *Cognition* **137**, 47–62 (2015).
11. L. B. Cohen, G. Amsel, Precursors to infants' perception of the causality of a simple event. *Infant Behav. Dev.* **21**, 713–731 (1998).
12. G. E. Newman, H. Choi, K. Wynn, B. J. Scholl, The origins of causal perception: Evidence from postdictive processing in infancy. *Cogn. Psychol.* **57**, 262–291 (2008).
13. N. Cesana-Arlotti *et al.*, Precursors of logical reasoning in preverbal human infants. *Science* **359**, 1263–1266 (2018).
14. N. Cesana-Arlotti, A. M. Kovács, E. Téglás, Infants recruit logic to learn about the social world. *Nat. Commun.* **11**, 5999 (2020).
15. K. Grill-Spector, N. Kanwisher, Visual recognition: As soon as you know it is there, you know what it is. *Psychol. Sci.* **2**, 152–160 (2005).
16. B. Pomiechowska, G. Brody, G. Csibra, T. Gliga, Twelve-month-olds disambiguate new words using mutual-exclusivity inferences. *Cognition* **213**, 104691 (2021).
17. M. Kibbe, A. M. Leslie, Conceptually rich, perceptually sparse: Object representations in 6-month-old infants' working memory. *Psychol. Sci.* **30**, 362–375 (2019).
18. E. Parise, G. Csibra, Electrophysiological evidence for the understanding of maternal speech by 9-month-old infants. *Psychol. Sci.* **23**, 728–733 (2012).
19. L. Feigenson, J. Halberda, Conceptual knowledge increases infants' memory capacity. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 9926–9930 (2008).
20. C. Houston-Price, E. Mather, E. Sakkalou, Discrepancy between parental reports of infants' receptive vocabulary and infants' behaviour in a preferential looking task. *J. Child Lang.* **34**, 701–724 (2007).
21. L. Feigenson, S. Carey, On the limits of infants' quantification of small object arrays. *Cognition* **2005**, 295–313 (2005).
22. L. Feigenson, S. Carey, M. Hauser, The representations underlying infants' choice of more: Object files versus analog magnitudes. *Psychol. Sci.* **2002**, 150–156 (2002).
23. S. Carey, *The Origins of Concepts* (Oxford University Press, 2009).
24. B. Pomiechowska, G. Csibra, Nonverbal action interpretation guides novel word disambiguation in 12-month-olds. *Open Mind* **6**, 51–76 (2022).
25. E. Bergelson, D. Swingley, At 6–9 months, human infants know the meanings of many common nouns. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 3253–3258 (2012).
26. Sh. Tsuji *et al.*, Communicative cues in the absence of a human interaction partner enhance 12-month-old infants' word learning. *J. Exp. Child Psychol.* **191**, 104740 (2020).
27. B. Pomiechowska, T. Gliga, Lexical acquisition through category matching: 12-month-old infants associate words to visual categories. *Psychol. Sci.* **30**, 288–299 (2019).
28. D. K. Bemis, L. Pylkkänen, Simple composition: A magnetoencephalography investigation into the comprehension of minimal linguistic phrases. *J. Neurosci.* **31**, 2801–2817 (2011).
29. A. Fernald, R. Zangl, A. L. Portillo, V. Marchman, Looking while listening: Using eye movements to monitor spoken language. *Dev. Psycholinguist.* **44**, 97 (2008).
30. R. M. Golinkoff, W. Ma, L. Song, K. Hirsh-Pasek, Twenty-five years using the intermodal preferential looking paradigm to study language acquisition: What have we learned? *Perspect. Psychol. Sci.* **8**, 316–339 (2013).
31. M. Frank *et al.*, *Variability and Consistency in Early Language Learning: The Wordbank Project* (MIT Press, 2021).
32. L. Feigenson, S. Dehaene, E. Spelke, Core systems of number. *TiCs* **8**, 307–314 (2004).
33. V. Izard *et al.*, Newborn infants perceive abstract numbers. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 10382–10385 (2009).
34. J. Wang, L. Feigenson, Infants recognize counting as numerically relevant. *Dev. Sci.* **22**, e12805 (2019).
35. J. M. Siskind, A computational study of cross-situational techniques for learning word-to-meaning mappings. *Cognition* **61**, 39–91 (1996).
36. Y. T. Huang, E. Spelke, J. Snedeker, When is four far more than three? Children's generalization of newly acquired number words *Psychol. Sci.* **21**, 600–606 (2010).
37. T. Posid, S. Cordes, Verbal counting moderates perceptual biases found in children's cardinality judgments. *J. Cogn. Dev.* **16**, 621–637 (2015).
38. E. Bergelson, The comprehension boost in early word learning: Older infants are better learners. *CDP* **14**, 142–149 (2020).
39. C. Spriet, E. Abassi, J. R. Hochmann, L. Papeo, Visual object categorization in infancy. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2105866119 (2022).
40. N. Chomsky, *Syntactic Structures* (De Gruyter Mouton, 2009).
41. J. A. Fodor, *The Language of Thought* (Harvard University Press, 1975).
42. B. Lake *et al.*, Building machines that learn and think like people. *Behav. Brain Sci.* **40**, e253 (2017).
43. J. A. Fodor, E. Lepore, *The Compositionality Papers* (Oxford University Press, 2002).
44. N. Chomsky, *Language and Mind* (Cambridge University Press, 2006).
45. M. Amalric, S. Dehaene, Origins of the brain networks for advanced mathematics in expert mathematicians. *Proc. Natl. Acad. Sci. U.S.A.* **2016**, 4909–4917 (2016).
46. A. Ivanova *et al.*, Comprehension of computer code relies primarily on domain-general executive brain regions. *Elife* **9**, e58906 (2020).
47. K. Zuberbühler, Syntax and compositionality in animal communication. *Philos. Trans. R. Soc. Lond. B* **375**, 201900062 (2020).
48. J. Bolhuis *et al.*, Meaningful syntactic structure in songbird vocalizations? *PLoS Biol.* **16**, e2005157 (2018).
49. D. Barner *et al.*, Evidence for a non-linguistic distinction between singular and plural sets in rhesus monkeys. *Cognition* **107**, 603–622 (2008).
50. S. Piantadosi, J. Tenenbaum, N. Goodman, Bootstrapping in a language of thought: A formal model of numerical concept learning. *Cognition* **123**, 199–217 (2012).
51. A. Leslie, M. Chen, Individuation of pairs of objects in infancy. *Dev. Sci.* **10**, 423–430 (2007).
52. S. Carey, D. Barner, Ontogenetic origins of human integer representations. *TiCS* **23**, 823–835 (2019).
53. S. Dehaene, *How We Learn: The New Science of Education and the Brain* (Penguin, 2020).
54. B. Pomiechowska, G. Brody, E. Téglás, Á. M. Kovács, EYE TRACKING [composition]. OSF. https://osf.io/cnez5. Deposited 22 August 2022.