The Development of Conceptual Compositionality in Children

Stephanie Alderete (salderete@berkeley.edu)[1], Wenqing Cao (wenqing_297@berkeley.edu)[1], Steven Piantadosi (stp@berkeley.edu)[1], Fei Xu (fei_xu@berkeley.edu)[1]
[1]Department of Psychology, 2121 Berkeley
Berkeley, CA 94704 USA

Abstract

One of the core properties of human language is compositionality: the meaning of a sentence can be understood by the meaning of individual words and the rules for combining them (Szabó, 2020). We investigate the development of conceptual compositionality (the combination of concepts). In our study, 6- to 9-year-old children (N = 40) were shown a card with two objects (e.g., a car and a star). Participants were introduced to two characters (a robot and a wizard) that used their powers to change the objects in different ways (e.g., turning one object pink). In the test trials, participants were asked to predict what a card would look like after both characters used their powers on the same card. All participants successfully learned the characters' powers, but only participants 7.5 years and older succeeded in the compositionality test trials. Our findings suggest that by age 7.5 children can successfully compose functions.

Keywords: Compositionality, Conceptual Development, Language of Thought

Introduction

One fundamental principle of meaning in language is compositionality, that the meaning of a complex expression can be understood by its structure and its parts (Partee, 1984; Szabó, 2020). Compositionality allows individuals to combine words into an infinite number of novel sentences and phrases. For example, you may have no trouble understanding what a "robotic barista" is, even if you have never heard of it before. This is because you understand the meaning of "barista" and "robotic" and can combine these two concepts to infer that a "robotic barista" is a robot that makes coffee drinks and beverages. This ability to construct and deconstruct expressions for interpretation is a fundamental part of compositional thought. Within cognitive science, one important ongoing debate is concerned with whether different types of computational models exhibit compositionality (e.g., Lake & Baroni, 2023).

Until recently, the development of compositionality has primarily been studied in language. By the preschool age, several studies have shown that children demonstrate some understanding of compositionality (Barner & Snedeker, 2008; Hamburger & Crain, 1984; Hendriks, 2019; Matthei, 1982). Two studies examined how children interpret prenominal modifiers in phrases such as "the second green ball" (Hamburger & Crain, 1984; Matthei 1982). This is a strong test for compositionality because children may erroneously interpret the two modifiers as a conjunction of features, i.e., something is both green and in second position, whereas the correct interpretation requires 'second' to take scope over 'green ball", i.e., the second of all green balls. More formally these two interpretations illustrate the difference between f(x)+g(x) vs. g(f(x)). The latter captures the compositional interpretation. In a study by Matthei (1982) children were presented with an array of balls in e.g., the following order: Blue, Green, Green, Blue. Given this array of balls, the sentence "Point to the second green ball", and "Point to the green ball that is second" refer to different balls in the set. The first refers to the second of the green balls (i.e., the third ball in the set), and the latter refers to the green ball that is second in the entire set (i.e., the second ball in the set). Matthei (1982) found that 3- to 6-year-old children struggle to make the distinction between these two phrases. Children will often misinterpret complex expressions like "the second green ball" and will simplify the expression by interpreting "second" to only modify "ball", ignoring the hierarchical structure of the phrase. When asked to "point to the second green ball", children will misinterpret the phrase to mean a ball that is green and second in the sequence, not understanding that the phrase is referring to the second ball within the set of green balls.

Hamburger and Crain (1984), however, presented evidence that 4- to 6-year-olds succeed in the second-green-ball task if they had their eyes closed during the experiment, preventing them from applying "second" then "green" one at a time as they heard the phrase. This subtle manipulation improved the children's performance, suggesting that the competence is present in preschoolers but other extraneous factors (e.g., processing demands) may cause confusion and misinterpretation.

Additionally, Barner and Snedeker (2008) found that preschoolers interpret adjectives compositionally. In the study, they examined how 4-year-old children interpret the words "tall" and "short". Children had to identify novel objects called "pimwits" as either short or tall. They found that 4-year-olds can correctly identify novel objects in an array as short and tall and can adjust their standard of comparison when presented with other objects that are either taller or shorter than the objects in the original array. For example, if a child had labeled a "pimwit" as tall because it was the tallest in the array, they will adjust their labeling if the "pimwit" is placed next to a novel taller object. Their results suggest that 4-year-old children can compose adjective-noun combinations (e.g., "tall pimwit") and can adjust their comparison standard when presented with new statistical information.

Most studies on compositionality, like the ones above, have investigated different aspects of language development. Here we ask whether compositionality exists outside of the language domain. That is, is compositionality a general principle that applies across many domains? And is our

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

ability to understand language compositionally due to our ability to compose conceptual representations in a similar way?

There has been very little research on the development of conceptual compositionality. Two recent studies have examined infants' and young children's ability to compose concepts (Piantadosi & Aslin, 2016; Piantadosi et al., 2018). Piantadosi & Aslin, (2016) found that 3.5- to 4.5-year-olds can predictively compose two functions. In their task, children had to predict, in a forced choice design, what a car would look like after it went behind a screen (i.e., function). Each car had a pattern on it, and the screens changed the pattern of the car in one of two ways; either by changing its shape (i.e., function 1), or color (i.e., function 2). If the screen had shapes on it, it would change the pattern of the car to that shape (e.g., a screen with stars would change the car's pattern to stars). If the screen had a color on it, it would change the color of the car's pattern (e.g., a red screen turns the car's pattern red). In the critical trials, children had to combine both functions (e.g., a screen with stars, and a red screen) to accurately predict what the car would like after it passed both screens (e.g., a car with red stars). 3.5- to 4.5-year-old children succeeded above chance at predicting the outcome of the car (Piantadosi & Aslin, 2016). Using a similar design, Piantadosi et al. (2018) found that 9-month-old infants learned the individual functions but had difficulties composing them in a looking time study.

While the previous study suggests that young children may be able to compose concepts, there were some conceptual and methodological issues in their experimental design. The most important conceptual issue is that success in their study did not require children to *compose* the functions. As Matthei (1982) and Hamburger and Crain (1984) demonstrated in their studies, the order of words matters in generating the meaning of linguistic expressions such as *the second green ball*. In the Piantadosi and Aslin (2016) study, the order in which children applied the functions did not matter, that is, applying a red screen (function 1) and then a screen with stars (function 2) produced the same outcome as applying a screen with stars (function 2) and then a red screen (function 1). In other words, children may have *combined* the two functions – f(x)+g(x) – but they may not have *composed* them - (g(f(x)). In addition, there is a methodological issue as well. Since the screens (e.g., the screen with stars and the red screen) show the individual functions, children did not have to compose the functions mentally. They could have simply applied the functions using the screens as prompts. A strong test for conceptual compositionality requires the learner to be able to functionally compose concepts in a specific order of operation (i.e., g(f(x)).

The present study builds on these previous studies on conceptual compositionality by testing whether children can compose two functions (i.e., g(f(x))). In our study, 6- to 9-year-old children were shown a card with two objects on it (e.g., a car and a star). They were introduced to two different characters (Earl the robot and Wally the wizard) that used their powers to change the card in different ways, representing two individual functions (f(x) and g(x)). Earl the robot changed the second of a pair of objects on a card pink (f(x)), and Wally the wizard switched the position of the two objects (g(x)) (Figure 1). Children were first trained and tested on the individual functions to see if they could learn and apply the function to predict what a card would look like after one of the characters used their powers to change the objects on the card. In the compositionality test trials, children had to predict what a card would look like after Earl the robot (Function 1) and Wally the wizard (Function 2) both used their powers to change the objects on the same card (representing g(f(x)). If children can compose two functions, then they should be able to accurately apply Earl and Wally's powers to the same card. Importantly, the order in which to apply the functions matters. Applying Earl's power and then Wally's produces a different outcome than applying Wally's powers and then Earl's. We initially piloted with 4–5-year-old children since that was the age group tested in Piantadosi & Aslin (2016). However, pilot testing suggests that this is a very difficult task for young children. Thus, in our final sample, we tested a group of 6- to 9-year-olds.

Method

Participants. Forty-three 6– to 9-year-old children were tested. Three children were excluded from data analysis (one due to parental interference, one due to inattention, and one for failing the training trials, see below). The final sample consists of 40 participants (*Age Range* = 6.08- 9.77, *Mean* = 7.6, *SD* = 1.01, 21 female and 19 male). All participants were recruited and tested at a local science museum. Participants were compensated with a $5 Amazon gift card or received a small prize (e.g., a book or toy).

Materials. Materials consist of animated images of cards with pairs of black objects on them, and two animated characters (Earl the robot and Wally the wizard). Stimuli were created using Microsoft PowerPoint. All stimuli were presented on a laptop.

Procedure. The study was divided into three phases: Function 1 training and test trials (i.e., Earl the robot's power, or f(x)), Function 2 training and test trials (i.e., Wally the wizard's power, or g(x)), and the Compositionality test trials (the combination of the Earl and Wally's powers, or g(f(x))).

In the initial set-up, participants were told that they were going to play a game to find treasure. Then Wally the wizard and Earl the robot appeared on the screen. The experimenter introduced both characters and told participants that each character had special powers, and to get the treasure, children must learn each character's power.

Function 1 Training (Earl the robot's power). The purpose of the training phase was to teach children that Earl's power is to turn the second of a pair of objects on a card pink. Participants were shown a card with two black objects (e.g., an alligator and a diamond) (Figure 1, 1a). Earl the robot

appeared beside the card holding a magic wand. He waved his wand over the card and produced a new card next to him. This card was identical to the first card except the second object was pink (e.g., an alligator and a pink diamond). Earl repeated this process with two other cards that had different objects. On each trial, after Earl has changed the card, the experimenter said, "Look, [child's name], Earl can change the last object pink! He changed this card here [pointing to the initial card] to this card here [pointing to the new card that Earl made]".

Function 1 Test Trials (Earl the robot's power). After the training phase, participants saw two test trials where they had to predict what a card would look like after Earl used his powers to change the objects on the card. Participants were shown a card with two black objects (e.g., a cat and a triangle) (Figure 1, 1b). Earl the robot appeared beside the card holding a magic wand. He waved his wand over the card and produced a new card that was blank. Participants had to select from two cards that were displayed next to the blank card. Both cards had the same objects as the initial card (e.g., a cat and a triangle), however, one card had the first object turned pink (e.g., the cat), and the other had the second object turned pink (e.g., the triangle) (Figure 1, 1b). The experimenter asked the child, "What will this card [pointing to the blank card] look like? The card on the top or the card on the bottom [pointing to the two choice cards]?". Participants either responded verbally or by pointing. Once the card was selected, the unselected card disappeared, and the blank card flipped over to reveal the correct outcome (e.g., a black cat and a pink triangle). If participants failed the trial, they repeated the trial with a different card. Since there were two test trials, participants could see as few as two trials (if they got both correct on the first try), and as many as four trials (if they failed each trial on the first try). Participants were excluded if they failed two trials in a row.

Function 2 Training (Wally the wizard's power). Function 2 training was analogous to function 1 and consisted of two parts: the training phase and two test trials. The purpose of the training phase is to teach children that Wally's power is to switch the position of two objects on a card. Participants were shown a card with two black objects (e.g., a star and an elephant) (Figure 1, 2a). Wally the wizard appeared beside the card, holding a magic wand. He waved his wand over the card and produced a new card next to him. This new card was identical to the first card except the position of the objects was switched (e.g., an elephant and a star). Wally repeated this process with two other cards that had different objects. On each trial, after Wally has changed the card, the experimenter said to the child, "Look, [child's name], Wally can switch the two objects! He changed this card here [pointing to the initial card] to this card here [pointing to the new card that Wally made]".

Function 2 Test Trials (Wally the wizard's power). After the training phase, participants saw two test trials where they had to predict what a card would look like after Wally used his powers to change the objects on the card. Participants were shown a card with two black objects (e.g., a crab and a triangle) (Figure 1, 2b). Wally the wizard appeared beside the card, holding a magic wand. He waved his wand over the card and produced a new card that was blank. Participants had to select from two cards that were displayed next to the blank card. Both cards showed the same objects as the initial card (e.g., a crab and a triangle), however, one card had the position of the objects switched (e.g., a triangle, and then a crab), and the other card was identical to the initial card (e.g., a crab and a triangle) (Figure 1, 2b). The experimenter asked the child, "What will this card [pointing to the blank card] look like? The card on the top or the card on the bottom [Pointing to the choice cards]?". Participants responded verbally or by pointing. Once the card was selected, the unselected card disappeared, and the blank card flipped over to reveal the correct outcome (e.g., a card with a triangle, and a crab). If participants failed the trial, they repeated the trial with a different card. Since there were two test trials, participants could see as few as two trials (if they got both correct on the first try), and as many as four trials (if they failed each trial on the first try). Participants were excluded if they failed two trials in a row.

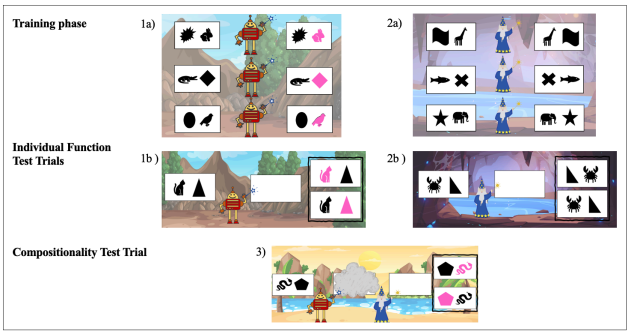

Figure 1: Training phase, Function Test Trials, and Compositionality Test Trials. Images 1a and 2a depict the training phase for Function 1 (Earl the robot) and Function 2 (Wally the Wizard). Images 1b and 2b depict the test trials for Function 1 and Function 2. Image 3 depicts the compositionality test trials.

After the training and testing on the individual functions, participants were shown one last slide to remind them of each character's power. Earl and Wally's powers were displayed one at a time. The image from Earl's training phase appeared first (Figure 1, 1a). The experimenter said to the child, "Let's see Earl's power one more time. Earl can turn these cards [pointing to the initial cards] into these cards [pointing to the transformed cards]." Then the image from Wally's training phase appeared next to the image of Earl's training phase (Figure 1, 2a). The experimenter said to the child, "Let's see Wally's power one more time. Wally can turn these cards [pointing to the initial cards] into these cards [pointing to the transformed cards]."

Compositionality Test Trials. Participants completed 6 compositionality test trials where they had to apply Earl's power (f(x)), and Wally's power (g(x)) to the same card (g(f(x)). On each trial, participants were presented with a card containing two black objects (e.g., a snake and a pentagon). Earl the robot appeared beside the card holding a magic wand. Earl waved his wand over the card and produced a new card that was covered by a cloud so participants could not see the objects. Then, Wally the wizard appeared beside the occluded card holding a magic wand. Wally waved his magic wand over the occluded card and produced a new card that was blank. Participants had to select from two cards that were displayed next to the blank card. Both cards contained the same objects as the initial card (e.g., a snake and a pentagon). One card depicted the outcome if children had applied Earl's power first and then Wally's power (e.g., a pink pentagon and a black snake). The second card depicted the card outcome if children mistakenly applied Wally's and then Earl's powers (e.g., a black pentagon and pink snake). The experimenter told participants, "Earl changed this first card here [pointing to the initial card] to this card here [pointing to the middle card occluded by the cloud], and then Wally changed this second card here [pointing to the middle card occluded by the cloud] to this card [pointing to the blank card]. What does this last card look like? Does it look like the top one or the bottom one [pointing to the two choice cards]?". Participants responded verbally or by pointing. This time participants did not receive any feedback and never saw the outcome of the final card. After each trial, the experimenter said, "Thank you, let's do another one", and proceeded to the next trial. This procedure was repeated for all 6 test trials.

All compositionality test trials had Earl the robot transform the card first, and then Wally the wizard. Participants saw one of two orders for where the correct card was located on each trial: [Top, Bottom, Bottom, Top, Bottom, Top] or [Bottom, Top, Top, Bottom, Top, Bottom], counterbalanced across participants.

Results

Function training Trials. For *Function 1 training* (Earl the robot), 34 participants passed both test trials without needing a repeat trial; 6 participants needed 1 repeat trial, completing a total of 3 trials. For *Function 2 training* (Wally the wizard), 37 participants passed both test trials without needing a repeat trial; 3 participants needed 1 repeat trial, completing a total of 3 trials (Table 1). Overall, participants of all ages (6-9.99) were able to learn and apply the individual functions.

Table 1: Participant's Performance on Training Trials

Training Trial	No Repeat Trials	One Repeat Trial	Two Repeat Trials
Function 1	34	6	0
Function 2	37	3	0

Compositionality Test Trials. Chance was established at 50% as there were two card options in each trial. The mean performance for participants was 62%. A Wilcoxon test found that the participants' performance was marginally different from chance (50%), ($T = 382$, $p = 0.06$, $r = 0.33$).

Next, we examined the effects of age (younger children aged 6.0-7.5, and older children aged 7.5-9.99), order (the placement of the correct card on each trial (top or bottom), gender (male vs. female), and trial order (first half of the trials vs. the second half of trials) on the compositionality test trials. For the trial order, we wanted to know if performance significantly improved in the latter half of the trials. For age, we used a median split (7.5 years) to see if there was a significant performance difference between younger and older children.

A Generalized Linear Mixed Effects Model (GLMM) was fit predicting the participant's binary responses (1 = correct, 0 = incorrect), from the fixed effects of gender, age, order, and trial order with a random intercept for participant id. There was no effect of gender, ($\hat{\beta} = 1.34$, $SE = 0.93$, $z = 1.44$, $p = 0.14$), or trial order ($\hat{\beta} = 0.53$, $SE = 0.90$, $z = 0.58$, $p = 0.56$). There was also no effect of trial order ($\hat{\beta} = 0.07$, $SE = 0.37$, $z = 1.18$, $p = 0.85$). Thus, participants' performance did not significantly improve in the latter half of the test trials. There was an effect of age ($\hat{\beta} = 2.43$, $SE = 1.0$, $z = 2.43$, $p = 0.01$), older children (range 7.5 - 9.99) selected the correct card 75% of the time while younger children (range = 6.0-7.5) selected it 50% of the time (Figure 2). Older children performed significantly better than younger children at predicting the outcome of the cards on the compositionality test trials.

Given the significant effect of age, we conducted two separate Wilcoxon tests to compare the performance of the different age groups to chance (50%). Older children's average performance (N = 20, *Range* = 7.5-9.99, *Mean performance* = 75%) was significantly different from chance ($T = 138$, $p = .02$, $r = .56$) whereas younger children's average performance (N = 20, *Range* = 6.0-7.5, *Mean performance* = 50%) was not significantly different from chance ($T = 56$, $p = 0.84$, $r = 0.03$).

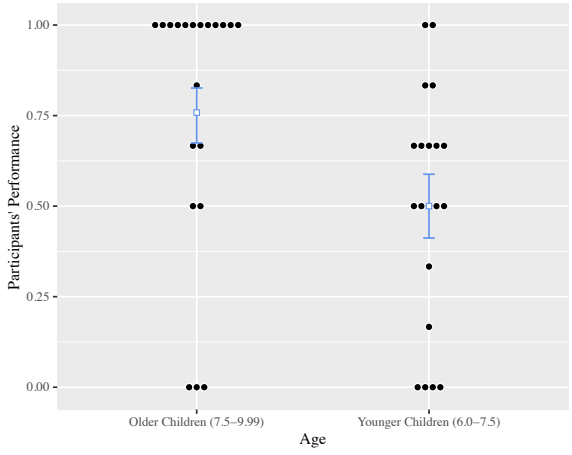

Figure 2: Participants' performance across test trials. Each dot represents a participant's average performance across all 6 test trials. The square dot represents the mean for each group, with 95% confidence intervals.

Discussion

Our study provides preliminary evidence that by 7.5 years of age, children can successfully compose functions (g(f(x)). Younger children (6-7.5) can learn and apply single functions (e.g., g(x), f(x)), but fail to successfully compose them.

Our findings contrast with the data from Piantadosi and Aslin (2016), who found success in children as young as 3.5 years of age. What could account for the different results? One possibility is that success on our task required children to mentally compose the functions (g(f(x)), since applying Earl's and then Wally's power yielded a different outcome compared to applying Wally's and then Earl's power. However, we did not manipulate the order of the two functions directly. This will be an important step for future research.

Another possibility is that our study did not provide children with any visual cues for how the characters changed the cards. Children had to remember each character's power and mentally apply them to the same card. In Piantadosi and Aslin (2016), the pattern/color on the screens served as cues to how the screen would change the object. Thus, children may have simply applied the pattern and color on the screens to the test object.

Alternatively, there may be task demands in our study that prevent younger children from successfully composing the two functions. For example, our task required children to mentally switch the position of two objects when applying Wally's powers. This form of mental rotation may be difficult for some children. Yet the results from the function training and testing trials suggest that all children successfully learned and applied each individual function in the training phase (Table 1). One explanation for the apparent success in learning function 2 (Wally's power) is that there is a flaw in our experimental design. In the function 2 test trials, children were presented with two cards when determining which card accurately depicted Wally's power: one card showed the initial objects unchanged, and the other showed the initial objects with their position switched. Since one of the options was identical to the initial card, some children could have passed the training by using a heuristic such as "the card should look different from the initial card" to identify the correct card without having learned Wally's powers. If children passed the Function 2 test trials by using a heuristic rather than by applying Function 2 (Wally's power), then they would be unable to accurately apply Function 2 in the test trials. This issue was not present in the Function 1 (Earl the robot) test trials, as both card options were different from the initial card. Thus, it is possible that some of the younger children did not successfully learn both functions, which may explain their failure in the compositionality test trials. It is possible that younger children can compose two functions but were not able to demonstrate their ability in our study because they failed to learn function 2 (Wally the wizard's power).

Another explanation for the poor performance found with younger children is that they may find it difficult to keep track of two objects on a card while also keeping track of the character's powers in the compositionality test trials. Tracking two objects and two characters' powers may exceed children's working memory capacity and make it difficult for them to accurately compose the two functions.

In future research, we plan to simplify our experimental design by using one object instead of two, and to test the effects of the order of the two functions directly.

Conclusion

The results of our study provide some preliminary evidence that by 7.5 years of age, children can successfully learn two functions and compose them. Younger children (6-7.5) may be able to learn and apply individual functions but struggle to compose them. Future research should investigate whether there are simpler ways to test children's conceptual compositionality. Perhaps with a simpler design, younger children can succeed at composing functions. We hope this study provides the first step to understanding the developmental origins of conceptual compositionality.

Acknowledgments

We would like to thank the members of the Berkeley Early Learning Lab for helpful discussion.

References

Barner, D., & Snedeker, J. (2008). Compositionality and Statistics in Adjective Acquisition: 4-Year-Olds Interpret *Tall* and *Short* Based on the Size Distributions of Novel Noun Referents. *Child Development*, *79*(3), 594–608. https://doi.org/10.1111/j.1467-8624.2008.01145.x

Hamburger, H., & Crain, S. (1984). Acquisition of cognitive compiling. *Cognition*, *17*(2), 85–136. https://doi.org/10.1016/0010-0277(84)90015-5

Hendriks, P. (2020). The acquisition of compositional

meaning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *375*(1791), 20190312. https://doi.org/10.1098/rstb.2019.0312

Lake, B. M., & Baroni, M. (2023). Human-like systematic generalization through a meta-learning neural network. *Nature*, *623*(7985), 115–121. https://doi.org/10.1038/s41586-023-06668-3

Matthei E. H. (1982). The acquisition of prenominal modifier sequences. *Cognition*, *11*(3), 301–332. https://doi.org/10.1016/0010-0277(82)90018-x

Partee, B. H. (1984). *Compositionality in formal semantics.* Blackwell Publishing.

Piantadosi, S., & Aslin, R. (2016). Compositional Reasoning in Early Childhood. *PLOS ONE*, *11*(9), e0147734. https://doi.org/10.1371/journal.pone.0147734

Piantadosi, S. T., Palmeri, H., & Aslin, R. (2018). Limits on Composition of Conceptual Operations in 9-Month-Olds. *Infancy*,*23*(3),310–324. https://doi.org/10.1111/infa.12225

Szabó, Zoltán Gendler (2022). Compositionality. *The Stanford Encyclopedia of Philosophy*