Journal of Experimental Child Psychology 252 (2025) 106130

Contents lists available at ScienceDirect

Journal of Experimental Child Psychology

journal homepage: www.elsevier.com/locate/jecp

Successful comparisons in novel word generalization: Executive functions or semantic knowledge?

Yannick Lagarrigue *, Jean-Pierre Thibaut *

LEAD - CNRS UMR 5022, Université de Bourgogne, 21065 Dijon, France

ARTICLE INFO

Article history:
Received 1 May 2024
Revised 24 October 2024
Available online 5 December 2024

Keywords:
Comparison
Executive functions
Vocabulary
Conceptual development
Categorization
Word generalization

ABSTRACT

Recent studies indicate that the opportunity to compare several stimuli associated with the same novel object noun, in contrast to a single stimulus design, promotes generalization along conceptually unifying dimensions. In two experiments (N = 240 4- and 5-year-olds), we assessed the link between executive functions and vocabulary (EVIP, a French version [Canadian norms] of the Peabody Picture Vocabulary Test), on the one hand, and children's novel word generalization performance in a comparison design, on the other. The experiments used two types of materials: unfamiliar objects in Experiment 1 and familiar objects in Experiment 2. In both experiments, results revealed a significant association between generalization performance and flexibility, whereas no significant links were observed with inhibition, working memory, or vocabulary. For familiar objects, we anticipated that vocabulary would play a more significant role, which was not what was observed. We interpret these results in terms of children's capacity to shift to other dimensions or to re-describe stimuli. Working memory (i.e., keeping track of dimensions) and inhibition (e.g., inhibiting irrelevant salient dimensions) did not reach significance. We also discuss the absence of correlation between vocabulary and the generalization task.

* Corresponding authors.
E-mail addresses: yannicklagarrigue@outlook.fr (Y. Lagarrigue), jean-pierre.thibaut@u-bourgogne.fr (J.-P. Thibaut).

https://doi.org/10.1016/j.jecp.2024.106130
0022-0965/

Introduction

The rapidity and efficiency of children's novel word learning have been marveled by both parents and scholars. This efficiency suggests that children successfully focus on conceptually relevant dimensions. In the case of objects, this means finding objects' dimensions that support generalization of the noun to unfamiliar entities that language refers to as instances of the same category (Bloom, 2002; Jones & Smith, 1993; Markman, 1989). Finding relevant dimensions might be challenging, with one reason being that reference is not transparent in many learning contexts (Bloom, 2002; Markman, 1990; Quine, 2013; Smith et al., 1996). A foundational hypothesis is that children would use lexical constraints that bias them toward specific properties of the stimuli (Diesendruck & Bloom, 2003; Jones & Smith, 1993; Landau et al., 1988; Markman, 1989, 1990). For example, young children follow the shape bias (Imai et al., 1994; Landau et al., 1988), according to which they generalize novel object nouns on the basis of object shape rather than texture or color. In most studies targeting these biases, the experimental design features one learning item that is introduced together with its noun. Then, children are asked which stimulus, among a set of options, would share the same noun as the learning item. It is widely accepted that even young children can learn novel words and extend them with minimal input (Bloom & Markson, 1998; Carey, 2010; Carey & Bartlett, 1978; Heibeck & Markman, 1987; Kovack-Lesh et al., 2008; Woodward et al., 1994).

In the current study, we capitalized on previous works demonstrating the benefits of comparing several objects in learning situations, particularly when the unifying dimensions of the category do not correspond to a priori salient object properties. This learning design, in which two or more learning instances are associated with the same noun, has received much attention (e.g., Augier & Thibaut, 2013; Gentner & Namy, 1999; Graham et al., 2010; Namy & Clepper, 2010; Namy & Gentner, 2002; Namy et al., 2007; Price & Sandhofer, 2021; Thibaut & Witt, 2023; Twomey et al., 2014). Overall, the hypothesis is that information redundancy might direct children toward less obvious properties. This has been shown with familiar objects (Gentner & Namy, 1999; Namy & Gentner, 2002; Thibaut & Witt, 2023), unfamiliar objects (Augier & Thibaut, 2013; Graham et al., 2010; Lagarrigue & Thibaut, 2021; Price & Sandhofer, 2021), relational terms such as nouns (Gentner et al., 2011; Thibaut & Witt, 2015), verbs (Childers, 2011, 2020; Childers et al., 2016), or adjectives (Waxman & Klibanoff, 2000) and beyond word learning, in categorization (Andrews et al., 2011; Corral et al., 2018), and in mathematical learning (Rittle-Johnson & Star, 2007) (see Alfieri et al., 2013, for a general overview of comparison and learning).

In the case of novel word learning, the idea is that introducing several instances simultaneously under a common label initiates a process of comparison that highlights common less salient properties. The general hypothesis of the current study was that when comparisons target nonobvious dimensions, in the presence of salient irrelevant dimensions, they rely on both executive functions (EFs) and vocabulary knowledge (see below). In an individual differences approach, we investigated the contribution of these two complementary explanatory mechanisms: EFs and lexical knowledge. Thibaut and colleagues (Augier & Thibaut, 2013; Thibaut & Witt, 2015) hypothesized that comparisons involve cognitive costs given that they require an active monitoring of the available information—that is, the learning exemplars and the response options—in order to uncover hidden dimensions and promote generalizations along less salient unifying dimensions. Another, complementary hypothesis is based on lexical knowledge. According to many authors, lexical knowledge is the driving force behind conceptual and novel noun development (e.g., Gentner, 1988; Gentner & Hoyos, 2017). Indeed, a richer lexical knowledge base provides children with more descriptors available to encode novel stimuli. In two experiments, we addressed these hypotheses and assessed the extent to which the benefits of comparison for generalization stem from lexical knowledge, from EFs, or from a combination of them.

Comparison in word extension tasks

As mentioned above, in many novel word extension (generalization) studies, children are first presented with a single example of a category associated with its name, for example, an apple and a novel word such as "dax." This mimics daily situations when an adult points to an object and gives its name (e.g., Diesendruck & Bloom, 2003; Gentner & Namy, 1999; Landau et al., 1988; Markman, 1989; Waxman, 1998). Then, they are asked which one of (most often) two novel objects share the same name; for example, "Show me which one is also a dax" (generalization phase). This design has been extensively used in many studies, including the aforementioned lexical biases (Kucker et al., 2019; Markman, 1990; Waxman, 1998). Basically, each bias minimizes the number of dimensions children consider as a basis for novel word generalization (Diesendruck & Bloom, 2003; Diesendruck et al., 1998; Imai & Haryu, 2004; Jones & Smith, 1993; Kucker et al., 2019; Landau et al., 1988; Markman, 1989). This is the case for object shape, which is central for the shape bias (see Kucker et al., 2019, for an extensive discussion of the literature), or for the basic level of categorization in the case of the basic level bias (Imai & Haryu, 2004; Landau et al., 1988; Markman, 1990; Waxman, 1990). These biases influence children's lexical development (e.g., Smith et al., 2002).

In the comparison design, two (or more) examples of a target category are introduced simultaneously with a common novel name ("This is a *dajo.* This is also a *dajo.*"). Then, children are introduced with at least two test objects, one taxonomically related but perceptually unrelated and the other taxonomically unrelated but perceptually similar to the learning example. For example, Gentner and Namy (1999) contrasted the single and the comparison designs. In the single design, 4-year-olds were introduced with a picture of an apple associated with a novel name ("Look, this is a *dajo*"), and then they were asked which one between a banana and a balloon is also a *dajo.* A majority of children chose the perceptual non-taxonomic choice (i.e., the balloon). In the comparison condition, two learning examples of the target category were introduced (e.g., two round fruits; "Look, this is a *dajo.* Look, this one is also a *dajo*"), and a majority of children chose the taxonomic choice, a banana. Note that a large number of word learning experiments target both children's capacity to learn and memorize the association between the learning (training) stimulus, the novel word and the underlying representation children build and/or use in these tasks involving new (generalization) stimuli. Here, word learning refers to the generalization of novel words rather than to their memorization. Indeed, there is no memory component involved in the single or comparison design given that the generalization stimuli are introduced immediately after the learning stimuli which remain in view (see Horst & Samuelson, 2008; Twomey et al., 2014).

How comparison facilitates generalizations based on unifying dimensions is an important question. According to Namy and Gentner (2002), following the structural alignment theory (Falkenhainer et al., 1989; Gentner, 1983; Gentner et al., 2003), learning stimuli properties are aligned and compared one by one, from superficial salient ones to the deepest conceptual ones (Gentner, 1983; Gentner & Gunn, 2001). Salient features such as shape and color are first aligned, and then comparisons later facilitate the alignment of deeper features. For example, comparing an apple with an orange consists in aligning them on dimensions such as shape (both are round), color (the apple is red and the orange is orange), and edibility (both are edible), which might contribute to uncover their "common taxonomic category." As an output of the alignment process, less salient properties would be focused on more easily and stored as commonalities or differences. In this design, dimensions that are not shared by the learning examples are not diagnostic given that they do not unify the target category (see Hammer et al., 2008, 2009, for discussion).

Despite many results showing benefits of comparison, depending on the comparison conditions or the structure of the stimuli that are compared, the comparison design might not always lead to the same conceptual benefits (Augier & Thibaut, 2013; Thibaut & Witt, 2015, 2023). Indeed, in most studies, the comparison condition does not elicit perfect performance (i.e., 100% taxonomic choices). Our hypothesis was that comparisons have cognitive costs, which can decrease the observed performance in a comparison design. Previous experiments suggest several factors that might be at play. For example, Thibaut and Witt (2015) manipulated the number of objects to be compared (two, three, or four learning objects) in a relational noun learning experiment and obtained the best generalization performance in the three-learning-stimuli condition. They interpreted this result as a trade-off between informativity (more learning objects means more converging evidence toward the common dimension) and cognitive costs associated with the number of comparisons between stimuli to perform and integrate. In a similar way, Augier and Thibaut (2013) showed that 4-year-olds did not benefit from additional convergent information (i.e., four exemplars rather than two or the presence of between-category information; e.g., this one is not a *dax*), whereas 6-year-olds' performance increased linearly when more converging relevant information was provided (i.e., four learning items rather than two). On more conceptual grounds, Thibaut and Witt (2023) demonstrated that the semantic distance (i.e., learning items belonging to the same basic level category or the same superordinate category) between learning objects or between learning and test objects also matters. Thibaut and colleagues (2015, 2023) interpreted their results in terms of EFs and the necessity to monitor the information and all the comparisons to be performed in the comparison conditions. However, they did not test the association between EFs and comparison format, and to the best of our knowledge no previous study has done so to date.

Here, we hypothesized that two main components of cognitive development, EFs and lexical knowledge (vocabulary), would influence the way in which comparisons are carried out and, thus, generalization performance. We tested these hypotheses with familiar objects (as in Gentner & Namy, 1999) and unfamiliar objects (as in Augier & Thibaut, 2013). We now examine how vocabulary and EFs may contribute to the mechanisms engaged in the process of comparison.

EFs and novel word generalization in comparison designs

As suggested by Thibaut and colleagues, comparison situations are complex tasks that involve successive and systematic alignments, which in turn generate cognitive costs. Indeed, alignments reveal systematic temporal patterns during the trial, as shown by a recent eye-tracking study of the comparison task (Stansbury et al., 2024). For example, the analyses revealed that some search patterns in 5- and 6-year-old children were more likely to lead to correct answers than others (e.g., an early focus on the learning items was associated with more correct answers than less early comparisons of these learning items). Thus, not only are comparisons important, but so is the way in which they are organized. Thus, alignments must be monitored during the trial, which calls for a role of EFs (e.g., Miyake et al., 2000; Zelazo et al., 1997). This central role of EFs has also been illustrated for other comparison tasks such as analogical reasoning tasks (Richland et al., 2006; Simms et al., 2018; Thibaut & French, 2016) (see below). In all comparison tasks, a larger number of stimuli increases the number of comparisons to be performed—that is, the number of alignments—in order to find one (or several) unifying feature(s). It also increases the amount of information to be manipulated in working memory or the number of irrelevant perceptual and semantic attributes (e.g., shape) children need to inhibit or the number of shifts toward new dimensions when a salient one is irrelevant for categorization.

EFs encompass various cognitive abilities that are central when behaviors go beyond more than simple routines, that is, when they involve coordinating multiple sources of information in a problem solving situation (Miyake et al., 2000; Zelazo & Müller, 2011). In their landmark article, Miyake et al. (2000) isolated three distinguishable yet interconnected components—inhibition, cognitive flexibility, and working memory (see Kovack-Lesh et al., 2008)—and we included these three components in our study. Note, however, that other authors have proposed different views. Numerous studies have shown positive correlations between lexical knowledge and EFs in both elementary school and preschool children (Gathercole & Pickering, 2000; St Clair-Thompson & Gathercole, 2006; Welsh et al., 2010; Yoo & Yim, 2018). These findings consistently indicate that higher performance in EF tests is correlated with higher language abilities such as vocabulary knowledge. As mentioned above, given the nature of comparison tasks, EFs might be involved in comparison designs whenever comparisons require monitoring of multiple sources of information (see Stansbury et al., 2024).

Studies investigating correlations between conceptual learning in novel name generalization tasks and EFs remain scarce. In a related domain, and following the same individual differences approach, Simms et al. (2018) showed, with a scene analogy task, that EFs and working memory are associated with the development of analogical reasoning. This is of interest in the current context because analogies are generalization tasks that require comparisons and alignments between domains (of relations). Simms et al. assessed 5- to 11-year olds' inhibitory control (Flanker task), working memory (list sorting), and cognitive flexibility (Dimensional Change Card Sort [DCCS]) with the NIH (National Institutes of Health) Toolbox (Zelazo et al., 2013). Children's working memory performance was the best predictor of their scene analogy performance even after controlling for age. Similarly, we investigated the link between executive functioning and lexical generalization in a lexical learning and generalization by comparison design. Indeed, in comparison situations, the three core EFs might be involved.

Inhibition refers to the ability to resist prepotent answers and/or to keep irrelevant dimensions out of working memory in order to produce appropriate answers (Diamond, 2006). In comparison situations, inhibition could be at play to override irrelevant salient cues. Indeed, as mentioned above, children usually rely on shape to generalize novel object words (Baldwin, 1989; Graham & Poulin-Dubois, 1999; Imai et al., 1994; Landau et al., 1988). When the salient shape is not relevant for the target category, as was the case in our experiments, discovering nonobvious conceptually relevant dimensions might require first inhibiting salient dimensions.

Cognitive flexibility is the ability to switch between dimensions of a task or between rules (Diamond, 2006). In our comparison design, cognitive flexibility might enable switching between dimensions of the objects that are compared. Cognitive flexibility also allows generating different hypotheses about potential relationships between objects, fostering creative thinking, and contributes to generalization when it requires switching to the relevant dimension (Blaye & Bonthoux, 2001; Blaye & Jacques, 2009; Deak, 2000).

Finally, working memory is the ability to hold information in mind, mentally manipulate it (Diamond, 2006), and keep in mind the dimensions that have been tested and might be conceptually relevant (without feedback). As new dimensions are focused on during the comparison process, working memory might allow for their storage and update for real-time comparisons. Previous studies have shown that comparing several instances that have been introduced simultaneously promotes more correct generalizations than viewing the same number of instances individually in immediate succession (e.g., Kovack-Lesh et al., 2008; Oakes & Ribar, 2005). However, they were targeting different variables and did not assess the link between working memory and word generalization performance.

Vocabulary and novel word learning

Why and how would vocabulary favor conceptually based generalization in comparison? One reason is that finding less salient relevant dimensions in a set of "potential" descriptors will rely on the lexical database (Murphy, 2002). The richer the lexical knowledge (i.e., more concepts, finer differences, numerous relevant dimensions), the easier it should be to find relevant descriptors for the new stimuli (Gelman, 2003; Murphy & Medin, 1985). Indeed, a broader vocabulary means, in part, a broader set of available descriptors (the dimensions that have been accumulated while learning one's lexicon). Both our experiments required finding a less salient dimension. Thus, it is reasonable to expect that a large set of available dimensions would favor the discovery of these dimensions. Vocabulary also directs children's focus during learning (Carmichael & Hayes, 2001). For novel word learning, vocabulary directs attention toward stimulus dimensions that have been specifically associated with a conceptual domain. For example, the presence of eyes means "living creature," and young children's generalizations of a novel word are driven by this information and rely more on texture that on shape (Graham & Poulin-Dubois, 1999; Jones & Smith, 1998; Jones et al., 1991; Smith et al., 2002). Macario (1991) illustrated how children might use their knowledge selectively in the case of ambiguous stimuli. Children aged 3 or 4 years were more likely to group novel items according to color when items were introduced as food and according to shape when the items were introduced as toys (see Hall et al., 1993; Mandler, 2004; Markman, 1989; Waxman, 1990).

Gentner and Hoyos (2017) argued that the richer children's linguistic knowledge, the higher the cognitive gains in a comparison situation because the existing conceptual knowledge about the compared objects provides a richer "conceptual" vocabulary. Moreover, in comparisons, a richer vocabulary enables individuals to label and articulate the objects' attributes, hence contributing to more precise descriptions of these objects, especially when comparing unfamiliar or complex objects. This view predicts that less "lexically knowledgeable" children might find fewer conceptual dimensions in a task requiring finding novel dimensions than more knowledgeable children (see Jenkins et al., 2015). If children interpret a common name as a cue that the objects share important commonalities, a richer database will benefit the comparisons (Gentner & Namy, 1999; Namy & Gentner, 2002; Waxman & Markow, 1995).

The current study

The role of EFs in a comparison design has been previously investigated through manipulation of task demands (e.g., Augier & Thibaut, 2013; Thibaut & Witt, 2023). However, to the best of our knowledge, no study has assessed whether individual differences in children's EFs and/or vocabulary affect their performance in a novel noun generalization task. Here, we assessed children's vocabulary, working memory, cognitive flexibility, and inhibition as explanatory factors of their performance in a lexical generalization by comparison task. We selected 4- and 5-year-olds for different reasons. First, previous studies suggest that younger children have difficulties in finding the less salient dimension in these tasks, and the task becomes simple for most children beyond 6 years of age (see above). Here, we used two-dimensional (2D) stimuli. Experiment 1 used 2D stimuli from Augier and Thibaut (2013, 2014), which were built around a distinctive shape and a distinctive 2D texture. Importantly, this previous experiment showed that children selected a shape match over a texture match in a majority of cases in a no-comparison condition. Thus, the shape was overwhelmingly more salient than the so-called texture. The important point is that texture was less salient. Other dimensions, such as color and size, might have played a similar role given that they are also less salient than shape (Gelman & Meyer, 2011; Landau et al., 1988). Second, this age range is subjected to EF developments that are involved in categorization tasks, as argued by Blaye and Jacques (2009). Thus, our models would seek influences that add to the influence of age. Third, children in our age range have already accumulated a rich lexical and conceptual database, including superordinate categories (e.g., Bloom, 2002; Markman & Hutchinson, 1984; Rosch et al., 1976), which would be available to encode taxonomic relations in Experiment 2 but which still undergo developments in our age range (Bloom, 2002).

Children with a broader lexical knowledge were expected to possess a richer repertoire of conceptual dimensions to encode novel conceptual similarities among learning objects leading to improved generalization performance (Schyns et al., 1998). In this study, we used a vocabulary test (EVIP, a French version [Canadian norms] of the Peabody Picture Vocabulary Test); Dunn et al., 1993). The test assesses vocabulary knowledge with items coming from various conceptual domains, increasing in complexity. Children must select a picture among multiple options depicting objects, qualities, or situations. A correct choice necessitates knowledge of either the stimulus the word describes or the elimination of options that do not correspond to the word. This means knowing the targeted word and/or knowing the other options and their names in order to decide that they are not the targeted one.

The current literature suggests three main hypotheses:

1. Variations in children's EF capacities might be a stronger predictor of their generalization performance than vocabulary.
2. Each EF might play a distinct role during different stages of comparison. Children with weaker inhibition skills might favor the perceptual lure, and those with a limited level of cognitive flexibility might struggle to identify or switch to new relevant dimensions.
3. Both vocabulary and executive control contribute to comparisons as they are implemented here. In this scenario, both the EF assessment and vocabulary assessment scores will be significant predictors of generalization performance.

To summarize, Experiment 1, with unfamiliar objects, assessed the role of inhibition, working memory, and cognitive flexibility, on the one hand, and lexical knowledge, on the other, in children's generalization of novel words, that is, in the absence of prior knowledge about the stimuli. In contrast, Experiment 2 involved familiar objects and investigated the potential influence of prior knowledge on the generalization of novel nouns. This two-experiment approach sought to enhance our understanding of how vocabulary and EFs underpin word generalization in these two "knowledge" contexts.

Experiment 1

Method

Participants

A total of 121 4- and 5-year-old preschoolers (66 female) were tested individually in a quiet room in their school (M_{age} = 53.80 months, SD = 3.78, range = 46–60). All children were native French speakers who were recruited in a city of 200,000 inhabitants in France. Eleven children were excluded from the analyses because they did not complete the task (n = 2), they were absent (n = 4), or it was found that French was not their native language (n = 5). Participants were predominantly Caucasian and came from middle-class to upper-middle-class urban areas. There were also 10% to 15% of Maghrébin children and 10% of Black children. The schools were mainly located in the center of the city and areas around it, which are middle class and upper middle class (for sociodemographic information regarding the city, see https://www.insee.fr/fr/statistiques/2011101?geo = COM-21231). The number of participants was based on recent previous studies using the same cognitive tests (Foinant et al., 2022). The procedure was in accordance with the declaration of Helsinki and was ethically reviewed and approved within an Official agreement (convention n°: 2019-0679 and endorsement n°: 2020-0566) between the Academia Inspection of the French National Education Ministry, the University of Bourgogne ("Inspection Académique de Côte d'Or"), the Université de Bourgogne and our laboratory (LEAD).

Materials

Based on previous studies using the same procedure (Augier & Thibaut, 2013; Gentner & Namy, 1999; Graham et al., 2010), we used nine sets of four unfamiliar artificial 2D gray-scale objects that were presented on a computer screen. All stimuli can be found on the Open Science Framework (https://osf.io/4rbtq/?view_only=bc4ed0214bdb4aaa981cb452cbe40555). We borrowed novel objects with distinctive shapes and distinctive 2D textures from Augier and Thibaut (2013) and Lagarrigue and Thibaut (2021). What we called "texture" are 2D perceived textures that are, by construction, "poorer" than real-world textures which are perceived as multidimensional dimensions (e.g., materials, haptic perception). The key point here is that the 2D so-called texture dimension was much less salient than the 2D shape and would not be chosen as a relevant generalization dimension by participants in a no-comparison experiment.

Each set was composed of two learning objects and two test objects (see Fig. 1). In each set, the two learning objects had the same 2D texture but different shapes. The first test object, the shape match, had the same shape as one of the two learning objects but had a different 2D texture. The other test object, the texture match, had the same texture as both learning objects but had a different shape (see Fig. 1).

The size of each object was approximately 6.0 by 6.0 cm. Textures and shapes used in one set differed from all the textures used in the other sets. The order of presentation was pseudo-randomized across participants. Each set was associated with one of nine two-syllable novel names: Youma, Buxi, Dajo, Zatu, Sepon, Xanto, Vira, Loupo, or Rodon.

Procedure

We used the same forced-choice generalization task as in previous studies (see Introduction; e.g., Augier & Thibaut, 2013; Gentner & Namy, 1999; Graham et al., 2010; Namy & Gentner, 2002). Children needed to decide which of two test objects had the same name as the two learning objects. Children were tested in French. Each learning object was introduced with a novel count noun (e.g., "This is a buxi" ["Ceci est un buxi"] while pointing to the first learning object and "This is a buxi TOO" ["Ceci est AUSSI un bux"] while pointing to the other learning object). The learning objects were presented sequentially and left in view (see Fig. 1). Then, the two test objects (i.e., the shape and 2D texture matches) were introduced, and children were asked to point to the one with the same name as the learning object (e.g., "Show me which one of these two is also a buxi" ["Montre moi lequel de ces deux-ci est aussi un buxi"]). After two practice trials, each participant saw seven experimental trials

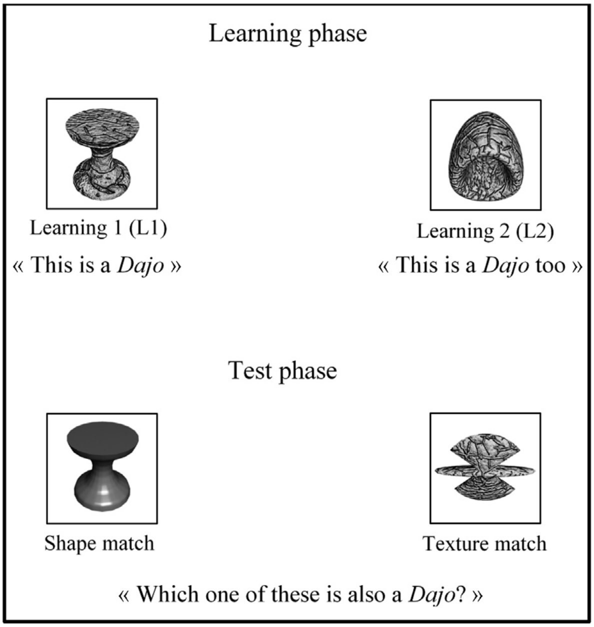

Fig. 1. Example of trial of the generalization task. The two learning objects share a common texture. Among the test objects, one possesses the identical shape as one of the learning objects, whereas the other displays the same texture as both learning objects.

in a random order. The two practice trials were chosen randomly among the set of trials. They were the same as the experimental trials in all respects except that the experimenter checked children's understanding of the game by asking them whether the second option could also be named after the same nonword as the first object. When children answered "yes," the experimenter explained that only one of the test objects could go with the learning object.

All the stimuli were displayed on the computer screen when children were asked to select their responses by touching the screen. Thus, by design, children could also select one of the learning objects. On rare occasions in the practice trials, some children picked one of the learning objects. The experimenter repeated that the two learning objects were called by the same name (e.g., "This one is a Dajo and this one too; they both are Dajo" ["Celui-ci est un Dajo et celui-là aussi. Ce sont tous les deux des Dajo"]) and that children were looking for another one that might also be called Dajo. These irrelevant responses were rare (<5%). In any case, the next trial appeared only after a valid response (shape or texture) was selected.

For the cognitive assessment, vocabulary and three EFs (working memory, flexibility, and inhibition) were assessed. For the working memory and flexibility tasks, we implemented the corresponding tasks from the NIH Toolbox battery (Zelazo et al., 2013) on OpenSesame (Mathôt et al., 2012), following the NIH protocol as described by the NIH. The instructions were translated into French. We assessed participants with a touchscreen laptop.

The computerized *working memory task, the list sorting task* (see Fig. 2) was composed of a sequence of colored pictures, each depicting an item (e.g., an animal) introduced with its name. Each item was displayed for 2 s. At the end of each sequence, children were instructed to remember and to verbally report all the items from the smallest to the biggest. The list started with two items. Every two trials, an item was added to progressively increase the working memory load. The task was stopped after two consecutive errors within the same number of items. After the "1-list" version, children were presented with a "2-list" version in which two kinds of stimuli were presented (i.e., animals and food pictures). In this version, children were asked to organize stimuli from one category (i.e., food), from the

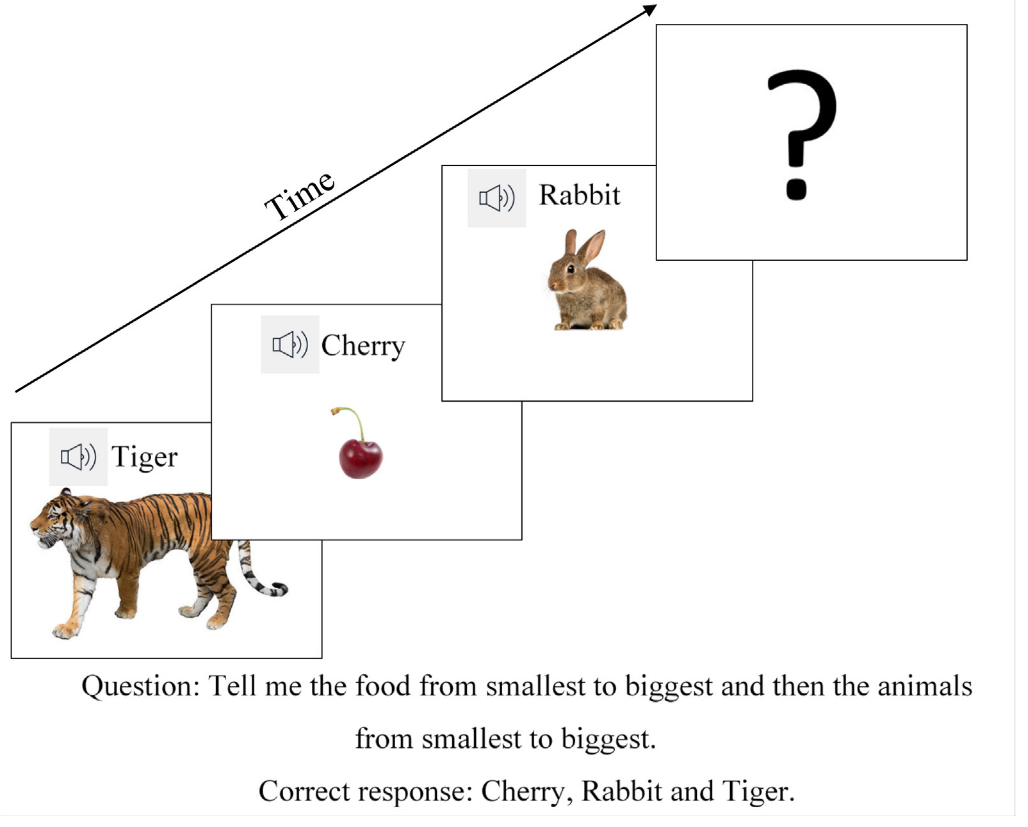

Fig. 2. Example of trial in the "2-list" sorting task with three items. The list sorting task requires children to remember all the items and to rank items from one category (i.e., food) and then from the second category (i.e., animals). The list sorting task measures working memory.

smallest to the biggest, and then from the other category (i.e., animals), also from the smallest to the biggest. The working memory score was the sum of correct trials in both lists. A correct trial was a sequence reproduced in the given order.

The *flexibility task* was a computerized version (on OpenSesame) of the DCCS of the NIH Toolbox battery. It was composed of four phases: familiarization, pre-switch, post-switch, and mixed. Before the experiment, children were taught that they were going to play two different games: the color game and the shape game. In the color game they were asked to put together two objects with the same color, and in the shape game they would need to put together the two objects with the same shape (Zelazo, 2006).

In the familiarization phase, two target stimuli (e.g., a red rabbit and a blue boat) were displayed in the lower part of the screen. With these two pictures left on the screen, a new 2D picture (a blue rabbit or a red boat), the example, was presented in the upper center of the screen (see Fig. 3). Children were instructed to match the example with one of the two target stimuli (i.e., the red rabbit or the blue boat) either according to its color or according to its shape (the order was counterbalanced between children). The two target pictures were left in view during the whole phase. The next stimulus appeared after a correct response. There was no time limit. The goal of this phase was to ensure that children clearly understood the shape or color game. After two correct trials, the second game was introduced according to the same rules. There were four familiarization trials with feedback.

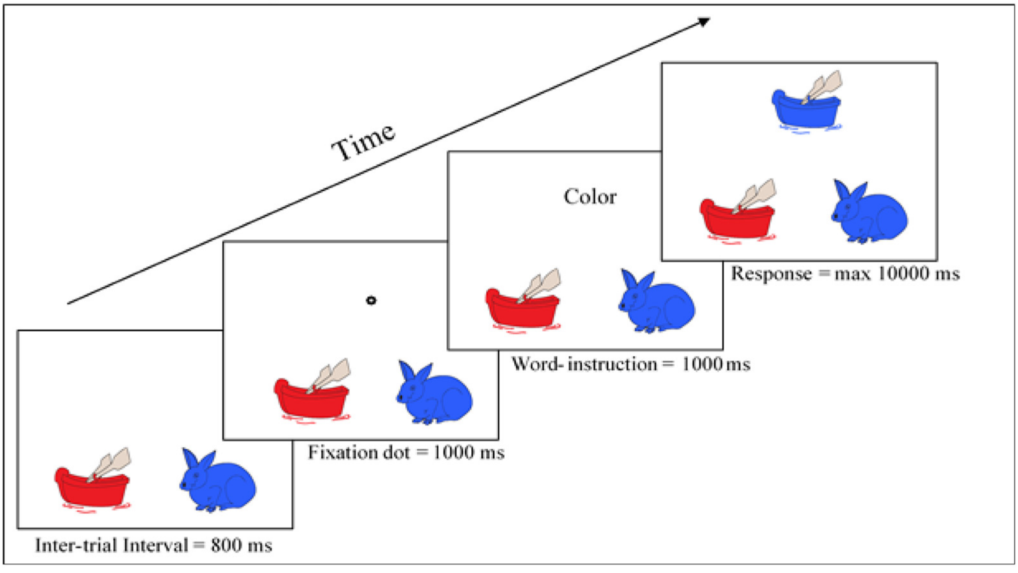

Fig. 3. Example of trial in the Dimensional Change Card Sort task (DCCS). Children needed to press one target based on the rule of categorization indicated by the word instruction printed on the screen.

In the three next phases (i.e., pre-switch, post-switch, and mixed), the two targets remained in the lower part of the screen. However, before the presentation of a stimulus, a fixation dot was displayed for 1000 ms, and the "word instruction" appeared for 1000 ms. This word instruction was read aloud by the experimenter (e.g., the word "color" in the color game). The computer's touchscreen was used to record participants' responses. The next stimulus was introduced after children provided a response (touching one of the target items) or after an absence of response of 10,000 ms. For the pre-switch phase one rule (e.g., color) was used for five trials, and for the post-switch phase the second rule (e.g., shape) was followed for five trials. The transition between the two phases was explained with instructions to switch. Children were given feedback after each trial.

The mixed phase consisted of 30 trials, including 24 "frequent" and 6 "infrequent" trials presented in a pseudo-random order (with two to five frequent trials preceding each infrequent trial). In this phase, no feedback was given.

The *inhibition task* was a Stroop-type test, the Real Animal Size Test (RAST; Catale & Meulemans, 2009) that we implemented on OpenSesame (see Fig. 4). This task was composed of three phases: control, training, and test. In each phase, children were presented with an animal picture on the computer screen and were asked to press one button for big animals and another button for small animals. Two big animals, elephant and horse, and two small animals, butterfly and bird, were used. Before the task, we checked children's knowledge of the four animals and of their size (i.e., big for horse and elephant and small for butterfly and bird).

Two big rectangles were presented on each side of the bottom part of the screen, and the animal was displayed in the center of the upper part of the screen. Children were instructed to press one rectangle (e.g., the one on the right) for big animals and the other one (e.g., the one on the left) for small animals. In the control phase, which was composed of 12 trials, all pictures were presented with the same size (medium). Before the training phase, children were told that the size of the image would change but that no matter the size of the image, they needed to say whether it was a big animal or a small animal "in real life." Children had unlimited time to respond, and feedback was provided after each trial. The goal of this phase was to ensure that children managed to sort the animals correctly because the size congruency manipulation on the RAST Stroop task is dependent on animal size knowledge.

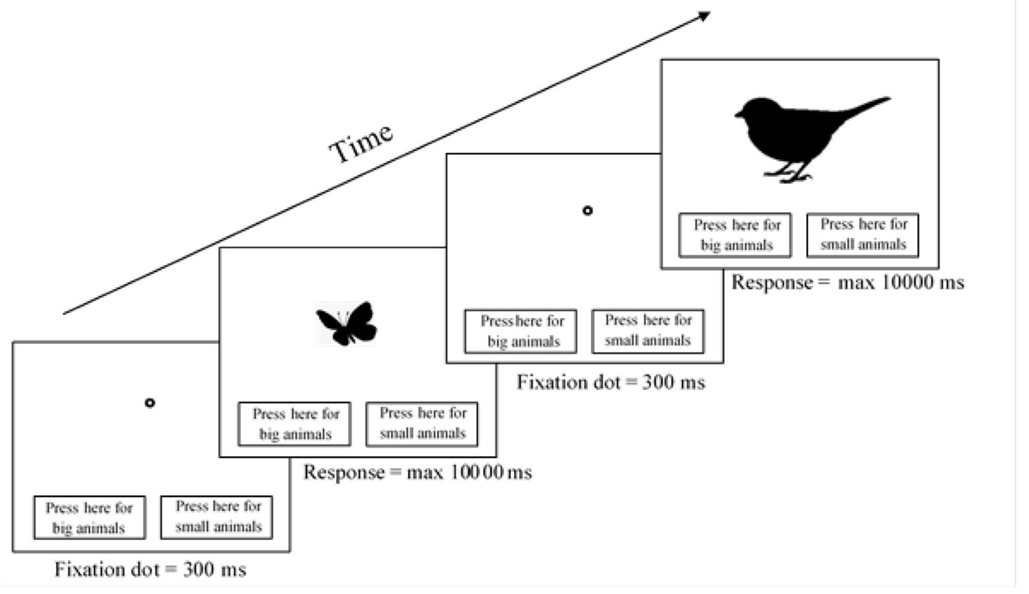

Fig. 4. Example of two trials for the Real Animal Size Test Stroop (RAST, inhibition task). Children needed to press one rectangle for small animals and another rectangle for big animals. They needed to *not* respond to the actual size of the picture but rather to the real size of the animal in congruent and incongruent trials.

In the training phase, pictures of two different sizes were used. Large and small animals were displayed with either a large size or a small size, respectively. In the congruent trials the size of an animal in the real world was congruent with its size in the picture, whereas in the incongruent trials the size of the real animal was not congruent with its size in the picture; thus, participants had to inhibit a response to the pictorial size and to give a response related to the size of the real animal. All the animals were presented twice with each size of pictures for a total of 16 trials. A fixation dot was presented for 300 ms before every stimulus. After children's response (i.e., touching one of the target items or an absence of response within a 10,000-ms delay), the next stimulus was presented. Children received feedback after each trial. The test phase was identical to the training phase except that children completed 32 trials (four animals presented with the two sizes four times each) and feedback was no longer provided.

The vocabulary test was the EVIP (Dunn et al., 1993). In this test, children needed to point to the one picture among four that they associated with the noun given by the experimenter. Responses were recorded on a sheet of paper. The standard score was computed.

Data processing

In all the analyses, the dependent variable was the percentage of texture choices. The working memory score (list sorting) and the vocabulary score (EVIP) were computed manually. For the categorization, inhibition (RAST Stroop) and flexibility (DCCS) tasks, responses and reaction times (RTs) were recorded. Incorrect responses were discarded from the RT analysis except in the generalization task. All the RTs below 100 ms and beyond 10,000 ms or 2 standard deviations from the mean were considered as outliers and discarded from further analyses.

For the flexibility score, we followed Zelazo et al.'s (2013) procedure. The mean accuracy across all participants was 80% (*SD* = 17%). We used a two-vector method incorporating both accuracy and, for participants who maintained a high level of accuracy (at least 80% correct), RT. This two-vector method was composed of an accuracy score and an RT score. An accuracy score was computed for each participant. The sum of the correct responses in the pre-switch (5 trials), the post-switch (5 trials,) and the mixed (30 trials) phase was computed and multiplied by 0.125 to obtain a score that ranged from 0 to 5. For children (n = 47) whose accuracy was less than 80% (i.e., accuracy score < 4), the accuracy score was the corresponding score. For children (n = 74) whose accuracy was 80% or higher, an RT score was also calculated based on the median RT for correct infrequent trials from the mixed block. Following the NIH procedure (Zelazo et al., 2013), a log (base 10) transformation was applied to this median RT score, and all the median RTs between 100 and 500 ms were replaced by 500 ms and the median RTs between 3000 and 10,000 ms were replaced by 3000 ms. Like the accuracy scores, the RT scores ranged from 0 to 5. Log values were algebraically rescaled with the following formula such that smaller RT log values were at the upper end of the 0-to-5 range, whereas larger RT log values were at the lower end. Then, the rescaled RT scores were added to the accuracy scores for participants whose accuracy was 80% and beyond. For the inhibition task, we calculated the accuracy score as the proportion of correct responses in the practice phase. All participants' data are presented in Table 1.

$$\text{RT score} = 5 - \left(5 * \left[\frac{\log RT - \log(500)}{\log(3000) - \log(500)}\right]\right).$$

Results

Correlational analyses

All correlational analyses were performed with jamovi (Version 2.2). We first checked for correlations (Pearson) between age, generalization score, vocabulary score, working memory score, flexibility score, and inhibition score. These correlations revealed that older children performed better than younger children in the working memory and generalization tasks, confirming previous studies (Anderson, 2002; Best & Miller, 2010) and studies showing that performance in novel noun generalization tasks also improves with age (Augier & Thibaut, 2013; Gentner & Namy, 1999).

Because age was positively correlated with working memory (r = .227, p = .012) and generalization score (r = .227, p = .012), we also computed partial correlations controlling for age (see Table 2). Flexibility was significantly correlated with working memory (r = .290, p = .001) and inhibition (r = .193, p = .035). The most interesting result, however, was the significant positive correlation between generalization (i.e., proportion of texture choices) and flexibility (r = .182, p = .047). The remaining correlations yielded nonsignificant results. Notably, the correlation between vocabulary and the generalization task was not significant, whereas vocabulary was significantly correlated with the three EF measures: working memory (r = .322, p < .001), flexibility (r = .271, p = .002), and inhibition (r = .268, p = .003).

Linear model analysis

To test our hypotheses, we used RStudio (Version 2024.04.2) and the "MuMIn" package to estimate R^2 values of the models [function *r.squaredLR (Model)*]. We fit linear models with the novel word generalization score as the outcome measure, with children as a random factor to account for shared

Table 1
Participants' characteristics and scores for each task: Experiment 1.

Number of participants	121
Sex ratio (female/male)	66/55
Age (in months)	53.80 (*SD* = 3.78)
Generalization score (overall texture choice %)	63.16 (*SD* = 38.05)
Vocabulary score (EVIP)	113.59 (*SD* = 21.29)
Working memory score (list sorting)	5.63 (*SD* = 2.13)
Flexibility score (DCCS)	4.32 (*SD* = 1.18)
Inhibition score (RAST Stroop)	85.87 (*SD* = 13.78)

Note. EVIP, French version [Canadian norms] of the Peabody Picture Vocabulary Test; DCCS, Dimensional Change Card Sort; RAST, Real Animal Size Test.

Table 2
Correlation matrix controlled for age between categorization, vocabulary, and executive functions (working memory, inhibition, and cognitive flexibility).

	Generalization task	Vocabulary	Working memory	Inhibition
Generalization score (overall texture choice %)	–			
Vocabulary	r = .024 p = .798	–		
Working memory	r = .009 p = .921	r = .322 p < .001	–	
Inhibition	r = − .037 p = .689	r = .268 p = .003	r = .163 p = .075	–
Cognitive flexibility	r = .182 p = .047	r = .271 p = .003	r = .290 p = .001	r = .193 p = .035

Table 3
Goodness of fit of the linear model.

Model	*df*	AIC	Residuals	Pseudo-R^2	*p* Value
M0 Generalization Score ~ Age	1	108.15	2.550	.0873	
M1a … + Vocabulary	2	110.08	0.066	.0882	.798
M1b … + Flexibility	**2**	**106.09**	**4.020**	**.1400**	**.047**
M1c … + Inhibition	2	109.98	0.161	.0895	.688
M1d … + Working Memory	2	110.14	0.010	.0874	.921
M2a … + Flexibility + Inhibition	3	107.42	0.149	.1486	.418
M2b … + Flexibility + Working Memory	3	107.83	0.250	.1433	.618
M2c … + Flexibility + Vocabulary	3	108.00	0.086	.1412	.773
M3 … + Flexibility * Vocabulary	4	109.90	0.094	.1425	.911

Note. The model in bold is the best model. "…" indicates that the predictors included in the previous model were also included in the current model. AIC, Akaike information criterion.

variances within participants. The model was constructed by iteratively adding predictive variables to the null model (M0, including age only). Each variable was tested individually and added to the null model, ranked by predictive power from the best predictor to the to the worst (see Table 3 for each predictor's power). We kept variables that led to a significant decrease of the Akaike information criterion (AIC; Akaike, 1974; Wagenmakers & Farrell, 2004), as shown by chi-square tests. Each model was compared with the model of the previous level (see last column of Table 3). For example, each M1 model was compared with M0, and each M2 model was compared with the best M1 (here M1b). The model including flexibility (M1) provides a better fit to the data (i.e., has a significantly lower AIC) than the M0 model (see Table 3). Given that vocabulary correlated with the EF measures, we also tested a model including the interaction between flexibility and vocabulary, which might increase the explanatory power of the best model itself (see M3 model). Table 3 shows that the M1 model had the lowest AIC, with a within-participants fixed effect, flexibility (continuous factor). This model explained 14% of variance, as demonstrated by the pseudo-R^2, and showed that better flexibility scores predicted better generalization scores, β = .178, $t(118)$ = 2.01, p = .047. The model revealed no significant contribution of inhibition, working memory, or vocabulary alone or when they were included in a model with flexibility. In addition, the model testing the interaction between flexibility and vocabulary was not a better model than flexibility. To test how much the variance of the coefficient estimate was being inflated by multicollinearity, we computed a variance inflation factor (VIF) with the "car" package (age = 1.017, flexibility = 1.017). All values were inferior to 2.5, which is recommended to limit the confounding effect (Johnston et al., 2018).

Discussion

Our results show that only flexibility played a significant role. These interesting results suggest, first, that the best performance was associated with the ability to switch to less salient dimensions and, second, that lexical knowledge did not significantly contribute to children's generalization performance. However, we hypothesized that better vocabulary would provide a richer description of the stimuli to encode the stimuli and/or would be associated with a more differentiated database of attributes to describe unfamiliar objects, which is critical for our task in which low-saliency conceptual similarities must be uncovered (Christie & Gentner, 2010; Gentner & Hoyos, 2017).

However, it could be argued that the absence of any significant association between vocabulary and generalization resulted from the absence of familiarity of the stimuli, so that vocabulary knowledge did not help participants in finding unifying descriptors. Hence, familiar stimuli might provide a better test of our hypothesis regarding the link between vocabulary and the encoding of the stimuli. Indeed, children with a richer vocabulary could generate a richer, more diverse set of conceptual dimensions that would be used during comparisons. Our second experiment was based on former studies with familiar materials (Gentner & Namy, 1999; Graham et al., 2010; Namy & Clepper, 2010). We assessed the same variables as in Experiment 1.

In Experiment 2, we also manipulated the taxonomic distance between the learning and generalization items given that former studies on word generalization with familiar materials have documented its role (see Thibaut & Witt, 2023). Thibaut and Witt (2023) showed that taxonomic generalization to items from the same immediate superordinate level of categorization (e.g., banana or orange in the case of learning stimuli such as two apples) was higher than generalization to items from a distant same superordinate level of category (e.g., meat or eggs when the learning stimuli were two apples). The authors interpreted this result in terms of cognitive control, which was hypothesized to be more involved for longer semantic distances than for shorter semantic distance. Thus, in Experiment 2, we investigated the role of vocabulary and EFs as a function of the semantic distance between the learning items and generalization items.

Experiment 2

Method

Participants

A total of 119 4- and 5-year-old preschoolers (72 female) were tested individually in a quiet room in their school (M_{age} = 55.65 months, *SD* = 4.04, range = 48–63). As in Experiment 1, they were predominantly Caucasian and came from middle-class urban areas, with approximately 20% of Maghrébin and Sub-Saharan children). The children were native French speakers recruited from the same French city as in Experiment 1 (and its wider area) in a population from lower middle class to upper middle class. We did not have the opportunity to collect more sociodemographic data for our sample. However, population data can be found here at https://www.insee.fr/fr/statistiques/2011101?geo=COM-21231#chiffre-cle-9 and https://www.insee.fr/fr/statistiques/2011101?geo=DEP-21#tableau-SAL_G1. Nine children who did not complete all the experimental tasks were excluded from the analysis as well as 5 children whose native language, after checking, was not French. Written informed consent was obtained from their school and their parents/caregivers. The procedure followed institutional ethics board guidelines for research on humans. Because we used the same tools as in Experiment 1, we recruited the same number of participants.

Materials

Sixteen sets (plus two sets for warm-up trials) were constructed, with each set being composed of six objects. They were borrowed from Stansbury et al. (2019) and Thibaut et al. (2018; see also Thibaut & Witt, 2023) and were displayed on a computer screen. All stimuli can be found on the Open Science Framework (https://osf.io/4rbtq/?view_only=bc4ed0214bdb4aaa981cb452cbe40555). The sets came from various semantic categories such as clothing, food, tools, accessories, and animals. Each set was composed of six pictures. However, each trial was composed of five images only, depending on the condition (i.e., taxonomic distance; see below). We manipulated the semantic distance between the learning items and the transfer items: near or distant. Children saw 18 trials, 2 practice trials, and 16 experimental trials (8 trials per distance condition: near or distant). For each trial, the set, either near or distant, was randomly selected. Each set appeared only once. Fig. 5 provides an example of the object set used in the two taxonomic distance conditions. Following Thibaut and Witt (2023), taxonomic distance is defined in terms of taxonomic levels within a taxonomy, indicating the count of intermediate taxonomic steps required to reach the next common superordinate category. For example, an orange and a banana are at distance 1 because they belong to the same immediate superordinate category, fruits; an orange and a yoghurt are at distance 2 because they belong to the same superordinate category, food, for which fruits and dairies are intermediate superordinate categories. Importantly, within each set (e.g., food), the two learning objects were also perceptually similar (in terms of overall shape) *and* belonged to the same superordinate category (L1 and L2; e.g., endive and potato in Fig. 5). This latter feature is interesting because taxonomic choices, which were perceptually dissimilar, meant that children ignored perceptual similarities. The assessment of both taxonomic and perceptual distances came from ratings provided by adults and confirmed our formal definition (Thibaut & Witt, 2023, provides these ratings).

There were three test objects in both the near and distant generalization conditions:

1. A non-related choice (NR; e.g., a birdhouse in Fig. 5) that was rated as neither perceptually nor taxonomically related to the learning objects. It was the same in both generalization distance conditions.
2. A perceptual choice (P; e.g., a rugby balloon in Fig. 5) that was perceptually similar to the two learning objects but taxonomically unrelated to them. The average perceptual distance with learning objects was equivalent in both distance conditions. The distractor was perceptually similar to

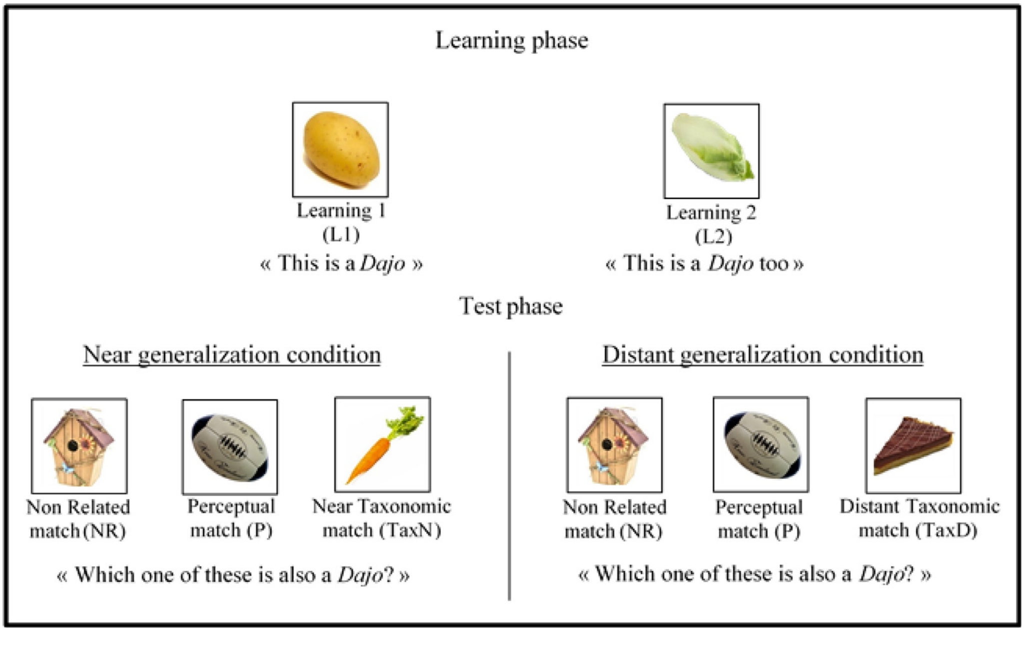

Fig. 5. Example of a trial of the generalization task (see text). The two learning objects are from the same superordinate taxonomic category (vegetables). The non-related match (NR) is not taxonomically or perceptually related to the learning objects. The perceptual match (P) is perceptually similar to the learning objects but is taxonomically unrelated. The taxonomic match (either TaxN or TaxD) is from the same taxonomic category as the learning objects.

keep the same logic as in Experiment 1, and as in most previous studies perceptual similarity was a strong explanatory factor in many theories (e.g., Jones & Smith, 1993). Another reason is that the preferred level of generalization for young children is the basic level of categorization, and this basic level of categorization is characterized by a strong perceptual similarity between items (Rosch & Lloyd, 1978; Rosch et al., 1976). This perceptual distractor ensured that children would need to inhibit a perceptual similarity to uncover the taxonomic relation between the learning items.

3. A taxonomic choice, either near or distant, as a function of the generalization condition (TaxN or TaxD; e.g., a carrot or a piece of pie, respectively, in Fig. 5). In the near generalization condition (TaxN), the taxonomic choice belonged to the same immediate superordinate category as the learning objects (e.g., vegetables). In the distant generalization condition (TaxD), the taxonomic choice came from a more distant superordinate category (e.g., food). In both generalization conditions, the taxonomic choice was perceptually dissimilar to the two learning objects.

In sum, a trial was composed of five pictures: L1, L2, Tax (TaxN or TaxD), P, and NR.

Procedure

The procedure and data processing were the same as in Experiment 1. All participants' data are presented in Table 4.

Results

Correlational analyses

We ran the same analyses as in Experiment 1. Pearson's correlations were used to explore the relations between age, generalization score (in each taxonomic distance condition separately or combined), vocabulary, working memory, flexibility, and inhibition. Quite expectedly, age was positively correlated with working memory ($r = .381$, $p < .001$), and so were flexibility ($r = .306$, $p < .001$) and vocabulary ($r = .314$, $p < .001$). For this reason, we computed partial correlations controlling for age. The partial correlation matrix can be seen in Table 5. Vocabulary was significantly correlated with working memory ($r = .376$, $p < .001$), flexibility ($r = .331$, $p < .001$), and inhibition ($r = .198$, $p = .031$). Flexibility was significantly correlated with working memory ($r = .269$, $p = .003$) and inhibition ($r = .439$, $p < .001$) (Miyake et al., 2000).

However, we were looking for correlations between generalization scores and our cognitive factors. Importantly, there were positive and significant correlations between generalization (overall score) and flexibility ($r = .197$, $p = .033$) and inhibition ($r = .182$, $p = .049$), but no significant correlation with vocabulary. In other words, higher flexibility and inhibition scores meant more taxonomic responses. Given that the near generalization condition ($M = 48.00$, $SD = 27.18$) led to significantly more taxonomic responses than the distant generalization condition ($M = 38.34$, $SD = 26.68$, $t(4.37)$, $p < .001$),

Table 4
Participants' characteristics and scores for each task: Experiment 2.

Number of participants	119
Sex ratio (female/male)	72/47
Age (in months)	55.65 (*SD* = 4.04)
Generalization score: All	43.18 (*SD* = 24.08)
Generalization score: Near	48.00 (*SD* = 27.18)
Generalization score: Distant	38.34 (*SD* = 26.68)
Vocabulary score (EVIP)	111.34 (*SD* = 20.03)
Working memory score (list sorting)	4.93 (*SD* = 1.91)
Flexibility score (DCCS)	4.21 (*SD* = 1.22)
Inhibition score (RAST Stroop)	86.06 (*SD* = 14.62)

Note. EVIP, French version [Canadian norms] of the Peabody Picture Vocabulary Test; DCCS, Dimensional Change Card Sort; RAST, Real Animal Size Test.

Table 5
Percentage of each response type for each generalization distance (near or distant) and comparison with chance level (33%, *t* tests).

Distance	Taxonomic		Perceptual		Non-related	
	Mean %	*p* Value	Mean %	*p* Value	Mean %	*p* Value
Near	48.00 (*SD* = 27.18)	<.001	37.18 (*SD* = 26.61)	.089	14.81 (*SD* = 19.32)	<.001
Distant	38.34 (*SD* = 26.68)	.031	39.08 (*SD* = 26.47)	<.001	14.71 (*SD* = 19.23)	<.001
All	43.17 (*SD* = 27.31)	<.001	38.13 (*SD* = 26.50)	<.001	14.76 (*SD* = 19.24)	<.001

we decided to compute separate correlations between our cognitive factors and taxonomic choices for the two generalization distances separately. These correlations revealed that only the near generalization distance was correlated with children's flexibility (r = .218, p = .018) and inhibition (r = .224, p = .015) scores. The correlation with working memory was marginally significant (r = .173, p = .062). All the other correlations were not significant. In the distant generalization case, there was no significant correlation between generalization scores and any of the cognitive factors.

Linear regression analysis

As in Experiment 1, we fit mixed linear models with the generalization score as the outcome measure and with children as a random factor to account for shared variances within participants. We used RStudio (Version 2024.04.2) and the "MuMIn" package to estimate R^2 of the models [function *r.squaredGLMM(Model)*]. The model was constructed by iteratively adding predictive variables to the null model (M0, containing just the intercept, age, and the random effect). Each variable was tested individually and added to the null model in order of predictive power, from the best predictor to the worst (see Table 6 for each predictor's power).The predictive variables were kept in the model when their addition led to a significant decrease of the AIC (Akaike, 1974; Wagenmakers & Farrell, 2004), as shown by chi-square tests. Each model was compared with the model of the previous level (see last column of Table 7). M2 was the best model, with the lower AIC, and the predictive variables were a random effect (participant) and two within-participants fixed effects: flexibility (continuous factor) and distance (near or distant). As in Experiment 1, adding the other factors (vocabulary, working memory, and inhibition) to flexibility never increased the power of the model (M4). This model explained 24.46% of variance (pseudo-R^2). To test how much the variance of the coefficient estimate was being inflated by multicollinearity, we computed a VIF with the package "car" (age = 1.10, distance = 1.00, flexibility = 1.10). All values are inferior to 2.5, which is recommended to limit the confounding effect (Johnston et al., 2018).

First, we observed a significant effect of the distance on the generalization score, $F(1, 115) = 4.37$, $p < .001$, $d = 0.81$), revealing that children selected the taxonomic response more often in the near generalization condition ($M = 48.00$, $SD = 27.18$) than in the distant generalization condition ($M = 38.34$, $SD = 26.68$) which is consistent with Thibaut and Witt (2023). Second, a significant effect of flexibility was found, $F(1, 115) = 2.16$, $p = .033$, $d = 0.40$. Children with higher flexibility scores selected more taxonomic choices than the other children. Finally, these analyses revealed no significant effect of inhibition, working memory, or vocabulary.

Discussion

Experiment 2 investigated the role of vocabulary and EFs in a similar learning by comparison design as in Experiment 1 with familiar items. We manipulated the semantic distance between the learning and generalization items. It was hypothesized that vocabulary knowledge would have a larger explanatory power with familiar stimuli than with unfamiliar stimuli (E1). Results revealed that vocabulary knowledge did not significantly explain the percentage of taxonomic choices. As in Experiment 1, the most important explanatory factor was cognitive flexibility as assessed by the DCCS, espe-

Table 6
Correlation matrix controlling for age between generalization score (in all trials, only near condition, and only distant condition), vocabulary, and executive functions (working memory, inhibition, and cognitive flexibility).

	Generalization scores (all conditions)	Generalization scores (near condition)	Generalization scores (distant condition)	Vocabulary	Working memory	Inhibition	Cognitive flexibility
Generalization score (all conditions)	–						
Generalization score (near condition)	r = .895 p < .001	–					
Generalization score (distant condition)	r = .892 p < .001	r = .599 p < .001	–				
Vocabulary	r = .000 p = .999	r = .009 p = .922	r = .009 p = .920	–			
Working memory	r = .093 p = .315	r = .173 p = .062	r = − .007 p = .942	r = .376 p < .001	–		
Inhibition	r = .182 p = .049	r = .224 p = .015	r = .100 p = .281	r = .198 p = .031	r = .193 p = .036	–	
Cognitive flexibility	r = .197 p = .033	r = .218 p = .018	r = .134 p = .150	r = .331 p < .001	r = .269 p = .003	r = .439 p < .001	–

cially in the case of the near generalization items. Working memory (+list sorting task) and inhibition (RAST Stroop) did not contribute to individual differences.

Thus, our data did not confirm the hypothesis that a broader vocabulary would provide a better lexico-semantic encoding of the stimulus. In contrast, our results confirm Experiment 1, showing that the ability to switch was positively associated with taxonomic choices. We hypothesized that the absence of any significant contribution of our measured factors in the distant case resulted from the difficulty of the task given that participants were below chance in this condition (whereas they were not below chance in the close condition; see Table 5).

General discussion

How do children generalize novel words to novel stimuli given the limited available evidence during learning? Here, we capitalized on former studies based on comparison designs, which have been shown to help children find unifying less salient dimensions. This design has been hypothesized to involve systematic explorations of the learning stimuli in order to progressively align them according to one unifying dimension (Childers et al., 2016; Graham et al., 2010; Namy & Gentner, 2002; Namy et al., 2007; Thibaut & Witt, 2023; Waxman & Klibanoff, 2000). We hypothesized that successful alignments require cognitive control (e.g., inhibition of the salient non-unifying dimension and/or the ability to re-represent stimuli when salient dimensions do not unify the stimuli). Finally, lexical knowledge was also hypothesized to ground children's search, providing additional descriptors when the most salient dimensions were not the unifying ones. Our main purpose was to assess whether EFs, vocabulary, or both would explain children's generalization performance with both unfamiliar and familiar materials. The findings revealed that cognitive flexibility was the most important explanatory factor of generalization performance for both familiar and unfamiliar objects, whereas inhibition and working memory revealed no significant association with performance. Quite surprisingly, vocabulary did not contribute statistically to generalization performance, whereas it was correlated with the EF measures.

Table 7
Goodness of fit of the linear mixed model.

Model	N_{par}	AIC	Residual deviance	Pseudo-R^2	*p* Value
M0 Generalization Score ∼ (1 | Participant) + Age	4	20.35	12.35	.5549	
M1a … + Distance	5	4.48	−5.52	.6144	<.001
M1b … + Vocabulary	5	22.35	12.35	. 5578	.999
M1c … + Flexibility	5	17.65	7.66	.5576	.302
M1d … + Inhibition	5	18.36	8.35	.5576	.046
M1e … + Working Memory	5	21.31	11.31	.5578	.308
M2a … + Distance + Vocabulary	6	6.48	−5.52	.6169	>.999
M2b … + Distance + Flexibility	**6**	**1.78**	**−10.22**	**.6167**	**.030**
M2c … + Distance + Inhibition	6	20.19	8.19	.5604	>.999
M2d … + Distance + Working Memory	6	23.14	11.14	.5607	.308
M3 … + Distance * Flexibility	7	8.23	−5.77	.6151	.618
M4a … + Distance + Flexibility + Vocabulary	7	3.19	−10.81	.6191	.442
M4b … + Distance + Flexibility + Inhibition	7	2.38	−11.62	.6191	.236
M4c … + Distance + Flexibility + Working Memory	7	3.56	−10.44	.6191	.641

Note. The model in bold is the best model. "…" indicates that the predictors included in the previous model were also included in the current model. N_{par}, number of parameters.

Vocabulary knowledge does not explain novel word generalization in comparison designs

The absence of association between vocabulary and word generalization is an important result because it means that vocabulary did not limit or facilitate learning novel words when the task was a matter of aligning/comparing stimuli, which is particularly surprising in the case of familiar objects. Indeed, we hypothesized that a richer lexical knowledge base would provide a richer set of encoding dimensions, a richer set of conceptual primitives (Gentner & Hoyos, 2017; Mandler, 1992; Schyns et al., 1998) that would contribute to the encoding of the stimuli. The absence of correlation was not due to a ceiling effect given that performance was around 50% correct. Given the level of performance and the vocabulary score range, we hypothesized that the more knowledgeable participants would take advantage of their knowledge background. Indeed, in the context of analogies, it has been argued that vocabulary is the major determinant of successful alignments. Our task and analogical reasoning tasks both are comparison tasks that involve alignments of the stimuli (our task) or of domains (analogy tasks). In both cases, the reasoning was that a richer database would provide a more differentiated vocabulary for better abstraction of commonalities (see Childers, 2020; Gentner & Hoyos, 2017). Our results do not show that lexical knowledge does not play any role here. Indeed, in the familiar case at least, children must know the superordinate categories to which learning stimuli belong. However, what the data suggest is that how children use this knowledge—that is, how control processes are involved—is more important than lexical knowledge per se.

Flexibility matters for comparison but not for inhibition or working memory

We hypothesized that the comparison process means aligning the stimuli on salient perceptual dimensions and progressively aligning them on unifying ones (e.g., Namy & Gentner, 2002). In this context, each EF might, in principle, contribute at any stage of the comparison process—inhibition to discard the irrelevant salient shape distractor, flexibility to flexibly switch to dimensions that might not come immediately to mind or re-represent the stimuli, and working memory to update dimensions that have already been tested or keep the representation of both stimuli in mind. Data show that cognitive flexibility best explained novel word extension performance.

Salient perceptual distractors were expected to be attractive for children because shape similarities are salient (Landau, 1994; Rattermann & Gentner, 1998). Being unable to inhibit shape similarities was hypothesized to interfere with the search of unifying dimensions. However, the absence of relation between generalization performance and inhibition suggests that differences in performance were not due to differences in inhibition. This suggests that children were able to inhibit shape-based alignments in the generalization phase when they noticed that these commonalities did not unify the category or that a failure to inhibit the salient similarities did not separate participants who failed to do it from those who succeeded. In the case of unfamiliar stimuli, in Experiment 1 the two learning stimuli did not share this salient shape similarity. For familiar learning stimuli, which shared both conceptual features (same taxonomy) and perceptual similarities, inhibition did not explain differences in performance either, that is, ignoring the shape similarities in favor of taxonomic similarities.

Thus, how can differences in flexibility explain the two conditions? The significant association between cognitive flexibility and the percentage of correct answers suggests that switching toward novel dimensions is the most important factor, that is, switching to a less salient descriptor or recoding the stimuli. In the case of unfamiliar stimuli, this means that once shape differences are perceived (shape salience makes shape differences easy to detect), the main explanatory factor was to generate a new unifying feature. In the generalization phase, children who lacked cognitive flexibility chose the shape distractor because it was the only stimulus for which they perceived a similarity with the learning stimuli even though it was with only one stimulus. In the familiar stimuli case, participants could use both perceptual (shape) similarities and taxonomic similarities as unifying dimensions. Because the stimuli were familiar, children could refer to their knowledge that taxonomic similarities are more important in the case of lexical generalization than other similarities (see Markman & Hutchinson, 1984, for evidence). Thus, most likely perceptual similarities were the most easily accessed information, but inhibiting them was not what differentiated participants. The next step, switching to the taxonomic alignment between the learning stimuli (at some point), was the most predictive step. Because it was the case for unfamiliar stimuli, when children did not switch to a taxonomic description of the learning stimuli, they selected the perceptual distractor (which was also similar to both learning items) in the generalization phase (Namy & Gentner, 2002; Price & Sandhofer, 2021).

What our results seem to suggest is that flexibly switching between dimensions, or being able to generate less obvious dimensions, contributed to children's performance. These switching difficulties might be compatible with two different complementary interpretations. First, they could result from a difficulty to align the two learning stimuli in terms of their common unifying features. In Experiment 2, aligning the two learning stimuli along the common taxonomic dimension might have proved difficult because of the perceptual commonality (i.e., same shape) between the two learning stimuli. Second, an equivalent difficulty might arise in the generalization phase when children tried to align the two learning stimuli with the options along the same unifying "learning" feature. Of these two complementary interpretations, we tend to favor the alignment of learning stimuli, which we think is crucial. Indeed, once children succeed at switching to the less salient dimension (either the texture or taxonomic relation) for the learning stimuli, they can refer to this dimension in the generalization phase. This dimension then belongs to their conceptual vocabulary. We think this is also consistent with Stansbury et al. (2024) and Thibaut et al. (2018), who showed that the depth of processing of learning stimuli (as assessed by the time spent on the learning items at the beginning of the trial and the number of early transitions between them) was a significant predictor of taxonomic performance, which the authors interpreted as a necessary step to find and encode the dimension unifying them. Their stimuli were similar to our Experiment 2 stimuli. The importance of cognitive flexibility in categorization was also demonstrated by Blaye and Jacques (2009), who used a cross-categorization task in which they introduced a target stimulus (e.g., a dog) and then two semantically related options, either a taxonomic choice or a thematic choice (i.e., a snail or a kennel) together with an unrelated choice. Children were asked to select one stimulus that goes with the target and then another one that would also go with the target. The authors argued that a failure to switch between the two semantic choices means insufficient executive control. Half of the 4-year-old children could perform this sequential selection correctly. The authors interpreted this discrepancy in terms of cognitive flexibility. Thus, there is converging evidence that cognitive flexibility is an important aspect of tasks that require different encoding of the stimuli either because this is requested by the task (Blaye & Jacques, 2009) or because the most accessible encoding does not provide a common description of the learning stimuli or the most task relevant encoding of these stimuli. Recent data by Foinant et al. (2024) assessed this link between categorization and EFs and showed that cognitive flexibility was the most important explanatory factor and that cognitive flexibility was a significant mediator between a psychological condition (child food neophobia) and several categorization scores.

Our data show that the capacity to inhibit a prepotent dimension did not explain significant differences between children in our learning by comparison task. This does not mean that inhibiting local perceptual similarities did play any role but shows that differences between children in doing so did not explain differences between them. In our experiments, working memory did not explain the generalization performance either. This is compatible with the idea that comparisons generate low memory load. One reason is that all the stimuli remained in view during the entire trial, decreasing the necessity to mentally manipulate, or combine, several dimensions. Working memory loads might increase in a sequential presentation of the learning stimuli. Children would need to compare stored mental representations of the stimuli rather than two "available" representations of the stimuli. In addition, in a sequential design, it would be more difficult to progressively update the representations of both stimuli because only one would be available at a time. Along these lines, results by Lawson (2017) and Vlach et al. (2012) revealed lower performance when learning stimuli were not introduced simultaneously, which can be explained by the larger working memory load that decreased the comparison process efficiency.

Beyond language learning

Our study did not have a no-comparison condition, in contrast to previous studies with the same age groups. These no-comparison studies showed that children mostly chose the perceptual choice when they had the choice between a perceptual same-shape lure and another choice (see Augier & Thibaut, 2013; Gentner & Namy, 1999; Graham et al., 2010; Landau et al., 1988; Waxman, 1990). Many explanations have been proposed for this result. Going for comparison designs meant introducing the complexities of cognitive manipulations. Indeed, if one defines cognitive complexity as the number of sources of variation to be related and processed in parallel (Andrews & Halford, 2002; Frye et al., 1998), comparison conditions are more cognitively demanding than no-comparison conditions, which are better suited to study constraints that children put on novel word reference. One interesting feature of our comparison design, however, is that it is "cognitively simple" given that the available information (the number of constitutive dimensions) remains small (mainly shape and texture in Experiment 1) and the number of types of comparisons is also small (i.e., between learning objects, between test objects, and between learning and test objects). This simplicity is an interesting feature of our paradigm compared with paradigms such as analogical reasoning tasks in which the number of dimensions potentially associated with the domains that are compared is larger (see Thibaut et al., 2022, for a discussion). An interesting consequence is that the mapping between each EF and the dimensions defining the task is more straightforward than in more multidimensional tasks (e.g., the relation between mathematical competence and EFs). In this context, we could figure out the most plausible role of cognitive flexibility in the entire comparison scenario during the learning stage.

In this respect, it would be interesting to test whether EFs are at play in a richer comparison context. Indeed, the *meta*-analysis of Alfieri et al. (2013) shows that comparison situations are implemented not only in language learning but also in many different learnings such as mathematics procedures (Rittle-Johnson & Star, 2007, 2009), definitions (Christie & Gentner, 2010), and analogy (Chen & Daehler, 1989). In our case, we could progressively increase the cognitive load of the task by increasing the number of stimuli, the number of stimuli dimensions, or their saliency and assess when other cognitive dimensions might progressively contribute to learning as the cognitive constraints of the task increase.

Limitations

One limitation is that we did not manipulate the number of stimuli involved in the comparison (three or four), and this number interacts with performance (Thibaut & Witt, 2015). Indeed, more stimuli to integrate might have involved working memory. The same is true for the temporal format of comparison. In daily situations, children are often shown stimuli sequentially, so that the comparison is between a real object and representations of the learning objects. This format might involve working memory more centrally but also inhibition. More learning stimuli might also constrain the scope of generalization more efficiently (e.g., Spencer et al., 2011; Xu & Tenenbaum, 2007). Along similar lines, one could argue that our two tasks differed slightly in their structure given that Experiment 1 had only two options versus three in Experiment 2, which might preclude direct comparisons between them. However, Experiment 2 results showed that the unrelated option was chosen far less often than the other two options. Hence, ultimately the two experiments remain very similar in this respect. In addition, these small differences between the experiments (beyond the main difference of type of stimuli) give similar results, that is, the same correlations and the same explanatory factor of flexibility.

We acknowledge that the absence of relation between generalization and inhibition, working memory, and vocabulary does not mean that they are not at all and/or never involved in the comparison process or in concept learning. Our results only show that at 4 and 5 years of age, the difference in performance in our comparison task depends more specifically on cognitive flexibility skills. For example, regarding inhibition, the absence of effect of inhibition as an explanatory factor does not rule out the possibility that children might struggle against irrelevant perceptual similarities. As mentioned above, more distractors, unfamiliar *real* objects rather than familiar objects, might elicit a more important role of inhibition or of vocabulary.

Another very general limitation arises from the way in which EFs were generally assessed, with instruments that purport to target specific aspects of cognitive functioning but in fact lack specificity. This was due to the lack of consensus about their nature and the way in which EFs have been modeled (Miyake et al., 2000). Here, they were correlated with each other, as is generally the case. In general, it is recommended to use several tools to capture more stable converging information for each EF. In addition, the diverse nature of each EF does not simplify the matter. Nevertheless, in both experiments, the pattern of correlations was very consistent. Gaining precision on the specificity of our correlations is also a goal for future work.

Future research should collect specific data on socioeconomic status and language background (e.g., monolingual vs. bilingual) given their influence on language development and EF skills in children. These demographic data might help to better understand how these variables interact with the constructs being studied, thereby enhancing the reliability and broader applicability of the results.

Conclusion

Our results show a correlation between a novel conceptual learning task and EFs. In many other studies aimed at finding correlations between academic competences such as mathematics and EFs, the relation between the competence and EFs was often more obscure because the competence was multidimensional, with both declarative knowledge and processing. Our task was much simpler and showed that flexibility is at play, meaning that children are required to re-describe the stimuli. In sum, our data provide evidence that EFs, but not vocabulary, might contribute to learning and generalizing a novel name.

CRediT authorship contribution statement

Yannick Lagarrigue: Writing – review & editing, Writing – original draft, Validation, Supervision, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Jean-Pierre Thibaut:** Writing – review & editing, Writing – original draft, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis.

Acknowledgments

Part of these data have already been presented in a Cognitive Science conference paper: Lagarrigue, Y., & Thibaut, J-P. (2020). From two to many: The role of executive functions in young children's generalization of novel object names in a comparison design. In S. Denison, M. Mack, Y. Xu, and B.C. Armstrong (Eds.), Proceedings of the 42nd Annual Conference of the Cognitive Science Society, pp. 1573-1579. Austin, TX: Cognitive Science Society. The authors wish to thank the National Agency of Research (ANR) for their financial support (COMPARE project ANR-18-CE28). They would like to thank

Damien Foinant, Eleanor Stansbury, Yoris Gouas, Lilou Ranjard, Rachel Mattei, and Marine Grota for their help in running the experiment.

Data availability

Our data and stimuli can be found at https://osf.io/4rbtq/?view_only=5ef96d58e2bb45a3a6ceb726d2a55e18.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, 19*(6), 716–723. https://doi.org/10.1109/TAC.1974.1100705.
Alfieri, L., Nokes-Malach, T. J., & Schunn, C. D. (2013). Learning through case comparisons: A meta-analytic review. *Educational Psychologist, 48*(2), 87–113. https://doi.org/10.1080/00461520.2013.775712.
Anderson, P. (2002). Assessment and Development of Executive Function (EF) During Childhood. *Child Neuropsychology, 8*(2), 71–82. https://doi.org/10.1076/chin.8.2.71.8724.
Andrews, G., & Halford, G. S. (2002). A cognitive complexity metric applied to cognitive development. *Cognitive Psychology, 45*(2), 153–219. https://doi.org/10.1016/s0010-0285(02)00002-6.
Andrews, J. K., Livingston, K. R., & Kurtz, K. J. (2011). Category learning in the context of co-presented items. *Cognitive Processing, 12*(2), 161–175. https://doi.org/10.1007/s10339-010-0377-5.
Augier, L., & Thibaut, J.-P. (2013). The benefits and costs of comparisons in a novel object categorization task: Interactions with development. *Psychonomic Bulletin & Review, 20*(6), 1126–1132.
Augier, L., & Thibaut, J-P. (2014). Dimensional distinctiveness constraints on comparison processing across development. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 36, pp. 1874–1879). https://escholarship.org/content/qt9vt8c7q8/qt9vt8c7q8.pdf.
Baldwin, D. A. (1989). Priorities in children's expectations about object label reference: Form over color. *Child Development, 60*(6), 1291–1306. https://doi.org/10.2307/1130921.
Best, J. R., & Miller, P. H. (2010). A Developmental Perspective on Executive Function. *Child Development, 81*(6), 1641–1660. https://doi.org/10.1111/j.1467-8624.2010.01499.x.
Blaye, A., & Bonthoux, F. (2001). Thematic and taxonomic relations in preschoolers: The development of flexibility in categorization choices. *British Journal of Developmental Psychology, 19*(3), 395–412. https://doi.org/10.1348/026151001166173.
Blaye, A., & Jacques, S. (2009). Categorical flexibility in preschoolers: Contributions of conceptual knowledge and executive control. *Developmental Science, 12*(6), 863–873. https://doi.org/10.1111/j.1467-7687.2009.00832.x.
Bloom, P. (2002). *How children learn the meanings of words*. MIT Press.
Bloom, P., & Markson, L. (1998). Intention and analogy in children's naming of pictorial representations. *Psychological Science, 9*(3), 200–204. https://doi.org/10.1111/1467-9280.00038.
Carey, S. (2010). Beyond fast mapping. *Language Learning and Development, 6*(3), 184–205.
Carey, S., & Bartlett, E. (1978). Acquiring a single new word. In *Proceedings of the Stanford Child Language Conference* (Vol. 15, pp. 17–29).
Carmichael, C. A., & Hayes, B. K. (2001). Prior knowledge and exemplar encoding in children's concept acquisition. *Child Development, 72*(4), 1071–1090.
Catale, C., & Meulemans, T. (2009). The Real Animal Size Test (RAST): A new measure of inhibitory control for young children. *European Journal of Psychological Assessment, 25*(2), 83–91.
Chen, Z., & Daehler, M. W. (1989). Positive and negative transfer in analogical problem solving by 6-year-old children. *Cognitive Development, 4*(4), 327–344. https://doi.org/10.1016/S0885-2014(89)90031-2.
Childers, J. B. (2011). Attention to multiple events helps two-and-a-half-year-olds extend new verbs. *First Language, 31*(1), 3–22. https://doi.org/10.1177/0142723710361825.
Childers, J. B. (Éd.). (2020). *Language and concept acquisition from infancy through childhood: Learning from multiple exemplars.* Springer International. https://doi.org/10.1007/978-3-030-35594-4.
Childers, J. B., Parrish, R., Olson, C. V., Burch, C., Fung, G., & McIntyre, K. (2016). Early verb learning : How do children learn how to compare events? *Journal of Cognition and Development, 17*(1), 41–66. https://doi.org/10.1080/15248372.2015.1042580.
Christie, S., & Gentner, D. (2010). Where hypotheses come from: Learning new relations by structural alignment. *Journal of Cognition and Development, 11*(3), 356–373. https://doi.org/10.1080/15248371003700015.
Corral, D., Kurtz, K. J., & Jones, M. (2018). Learning relational concepts from within- versus between-category comparisons. *Journal of Experimental Psychology: General, 147*(11), 1571–1596. https://doi.org/10.1037/xge0000517.
Deak, G. O. (2000). The growth of flexible problem solving: Preschool children use changing verbal cues to infer multiple word meanings. *Journal of Cognition and Development, 1*(2), 157–191. https://doi.org/10.1207/S15327647JCD010202.
Diamond, A. (2006). The early development of executive functions. In E. Bialystok & F. I. M. Craik (Eds.), *Lifespan cognition: Mechanisms of change* (pp. 70–95). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195169539.003.0006.
Diesendruck, G., & Bloom, P. (2003). How specific is the shape bias? *Child Development, 74*(1), 168–178.
Diesendruck, G., Gelman, S. A., & Lebowitz, K. (1998). Conceptual and linguistic biases in children's word learning. *Developmental Psychology, 34*(5), 823–839.
Dunn, L. M., Dunn, L. M., & Thériault-Whalen, C. M. (1993). *Échelle de vocabulaire en images Peabody: EVIP.* PsyCan.
Falkenhainer, B., Forbus, K. D., & Gentner, D. (1989). The structure-mapping engine: Algorithm and examples. *Artificial Intelligence, 41*(1), 1–63. https://doi.org/10.1016/0004-3702(89)90077-5.

Foinant, D., Lafraire, J., & Thibaut, J.-P. (2022). Relationships between executive functions and food rejection dispositions in young children. *Appetite*. https://doi.org/10.1016/j.appet.2022.106102 106102.
Foinant, D., Lafraire, J., & Thibaut, J.-P. (2024). Individual differences in categorization development: The mediation of executive functions and factual knowledge, the case of food. *Developmental Psychology, 60*(10), 1785–1800. https://doi.org/10.1037/dev0001785.
Frye, D., Zelazo, P. D., & Burack, J. A. (1998). Cognitive complexity and control: I. Theory of mind in typical and atypical development. *Current Directions in Psychological Science, 7*(4), 116–121. https://doi.org/10.1111/1467-8721.ep10774754.
Gathercole, S. E., & Pickering, S. J. (2000). Working memory deficits in children with low achievements in the national curriculum at 7 years of age. *British Journal of Educational Psychology, 70*(2), 177–194. https://doi.org/10.1348/000709900158047.
Gelman, S. A. (2003). *The essential child: Origins of essentialism in everyday thought*. Oxford University Press.
Gelman, S. A., & Meyer, M. (2011). Child categorization. *Wiley Interdisciplinary Reviews: Cognitive Science, 2*(1), 95–105. https://doi.org/10.1002/wcs.96.
Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science, 7*(2), 155–170. https://doi.org/10.1016/S0364-0213(83)80009-3.
Gentner, D. (1988). Metaphor as structure mapping: The relational shift. *Child Development, 59*(1), 47–59. https://doi.org/10.2307/1130388.
Gentner, D., Anggoro, F. K., & Klibanoff, R. S. (2011). Structure mapping and relational language support children's learning of relational categories: Structure mapping and relational language. *Child Development, 82*(4), 1173–1188. https://doi.org/10.1111/j.1467-8624.2011.01599.x.
Gentner, D., & Gunn, V. (2001). Structural alignment facilitates the noticing of differences. *Memory & Cognition, 29*(4), 565–577. https://doi.org/10.3758/BF03200458.
Gentner, D., & Hoyos, C. (2017). Analogy and abstraction. *Topics in Cognitive Science, 9*(3), 672–693.
Gentner, D., Loewenstein, J., & Thompson, L. (2003). Learning and transfer: A general role for analogical encoding. *Journal of Educational Psychology, 95*(2), 393–408. https://doi.org/10.1037/0022-0663.95.2.393.
Gentner, D., & Namy, L. L. (1999). Comparison in the development of categories. *Cognitive Development, 14*(4), 487–513. https://doi.org/10.1016/S0885-2014(99)00016-7.
Graham, S. A., Namy, L. L., Gentner, D., & Meagher, K. (2010). The role of comparison in preschoolers' novel object categorization. *Journal of Experimental Child Psychology, 107*(3), 280–290. https://doi.org/10.1016/j.jecp.2010.04.017.
Graham, S. A., & Poulin-Dubois, D. (1999). Infants' reliance on shape to generalize novel labels to animate and inanimate objects. *Journal of Child Language, 26*(2), 295–320. https://doi.org/10.1017/S0305000999003815.
Hall, D. G., Waxman, S. R., & Hurwitz, W. M. (1993). How two- and four-year-old children interpret adjectives and count nouns. *Child Development, 64*(6), 1651–1664. https://doi.org/10.1111/j.1467-8624.1993.tb04205.x.
Hammer, R., Bar-Hillel, A., Hertz, T., Weinshall, D., & Hochstein, S. (2008). Comparison processes in category learning: From theory to behavior. *Brain Research, 1225*, 102–118. https://doi.org/10.1016/j.brainres.2008.04.079.
Hammer, R., Diesendruck, G., Weinshall, D., & Hochstein, S. (2009). The development of category learning strategies: What makes the difference? *Cognition, 112*(1), 105–119. https://doi.org/10.1016/j.cognition.2009.03.012.
Heibeck, T. H., & Markman, E. M. (1987). Word learning in children: An examination of fast mapping. *Child Development, 58*(4), 1021–1034.
Horst, J. S., & Samuelson, L. K. (2008). Fast mapping but poor retention by 24-month-old infants. *Infancy, 13*(2), 128–157. https://doi.org/10.1080/15250000701795598.
Imai, M., Gentner, D., & Uchida, N. (1994). Children's theories of word meaning: The role of shape similarity in early acquisition. *Cognitive Development, 9*(1), 45–75. https://doi.org/10.1016/0885-2014(94)90019-1.
Imai, M., & Haryu, E. (2004). The nature of word learning biases and their roles for lexical development: From a crosslinguistic perspective. In D. G. Hall & S. R. Waxman (Eds.), *Weaving a lexicon* (pp. 411–444). Boston Review.
Jenkins, G. W., Samuelson, L. K., Smith, J. R., & Spencer, J. P. (2015). Non-Bayesian noun generalization in 3- to 5-year-old children: Probing the role of prior knowledge in the suspicious coincidence effect. *Cognitive Science, 39*(2), 268–306.
Johnston, R., Jones, K., & Manley, D. (2018). Confounding and collinearity in regression analysis: A cautionary tale and an alternative procedure, illustrated by studies of British voting behaviour. *Quality & Quantity, 52*(4), 1957–1976. https://doi.org/10.1007/s11135-017-0584-6.
Jones, S. S., & Smith, L. B. (1993). The place of perception in children's concepts. *Cognitive Development, 8*(2), 113–139.
Jones, S. S., & Smith, L. B. (1998). How children name objects with shoes. *Cognitive Development, 13*(3), 323–334. https://doi.org/10.1016/S0885-2014(98)90014-4.
Jones, S. S., Smith, L. B., & Landau, B. (1991). Object properties and knowledge in early lexical learning. *Child Development, 62*(3), 499–516.
Kovack-Lesh, K. A., Horst, J. S., & Oakes, L. M. (2008). The cat is out of the bag: The joint influence of previous experience and looking behavior on infant categorization. *Infancy, 13*(4), 285–307. https://doi.org/10.1080/15250000802189428.
Kucker, S. C., Samuelson, L. K., Perry, L. K., Yoshida, H., Colunga, E., Lorenz, M. G., & Smith, L. B. (2019). Reproducibility and a unifying explanation: Lessons from the shape bias. *Infant Behavior and Development, 54*, 156–165. https://doi.org/10.1016/j.infbeh.2018.09.011.
Lagarrigue, Y., & Thibaut, J.-P. (2021). Solid ground makes solid understandings: Does simple comparison pave the way for more complex comparisons? In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 43, pp. 2726–2732). https://escholarship.org/uc/item/3635f1zx.
Landau, B. (1994). Object shape, object name, and object kind: Representation and development. In D. Medin (Ed.). *Psychology of learning and motivation: Advances in research and theory* (Vol. 31, pp. 253–304). Academic Press.
Landau, B., Smith, L. B., & Jones, S. S. (1988). The importance of shape in early lexical learning. *Cognitive Development, 3*(3), 299–321.
Lawson, C. A. (2017). The influence of task dynamics on inductive generalizations: How sequential and simultaneous presentation of evidence impacts the strength and scope of property projections. *Journal of Cognition and Development, 18*(4), 493–513. https://doi.org/10.1080/15248372.2017.1339707.

Macario, J. F. (1991). Young children's use of color in classification: Foods and canonically colored objects. *Cognitive Development, 6*(1), 17–46. https://doi.org/10.1016/0885-2014(91)90004-W.
Mandler, J. M. (1992). How to build a baby: II. Conceptual primitives. *Psychological Review, 99*(4), 587–604.
Mandler, J. M. (2004). *The foundations of mind: Origins of conceptual thought*. Oxford University Press.
Markman, E. M. (1989). *Categorization and naming in children: Problems of induction*. MIT Press.
Markman, E. M. (1990). Constraints children place on word meanings. *Cognitive Science, 14*(1), 57–77.
Markman, E. M., & Hutchinson, J. E. (1984). Children's sensitivity to constraints on word meaning: Taxonomic versus thematic relations. *Cognitive Psychology, 16*(1), 1–27. https://doi.org/10.1016/0010-0285(84)90002-1.
Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods, 44*(2), 314–324. https://doi.org/10.3758/s13428-011-0168-7.
Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology, 41*(1), 49–100. https://doi.org/10.1006/cogp.1999.0734.
Murphy, G. L. (2002). *The big book of concepts*. MIT Press.
Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review, 92*(3), 289–316.
Namy, L. L., & Clepper, L. E. (2010). The differing roles of comparison and contrast in children's categorization. *Journal of Experimental Child Psychology, 107*(3), 291–305. https://doi.org/10.1016/j.jecp.2010.05.013.
Namy, L. L., & Gentner, D. (2002). Making a silk purse out of two sow's ears: Young children's use of comparison in category learning. *Journal of Experimental Psychology: General, 131*(1), 5–15.
Namy, L. L., Gentner, D., & Clepper, L. E. (2007). How close is too close? Alignment and perceptual similarity in children's categorization. *Cognition, Brain, Behavior, 11*(4).
Oakes, L. M., & Ribar, R. J. (2005). A comparison of infants' categorization in paired and successive presentation familiarization tasks. *Infancy, 7*(1), 85–98.
Price, G. F., & Sandhofer, C. M. (2021). One versus many: Multiple examples in word learning. *Journal of Experimental Child Psychology, 209* 105173.
Quine, W. V. O. (2013). *Word and object.* MIT Press. https://books.google.com/books?hl=fr&lr=&id=R-TxCwAAQBAJ&oi=fnd&pg=PR9&dq=quine+word+and+object&ots=YTqOIQJcJE&sig=clHU17URzoWgNdvcKa4ZPCemuk0.
Rattermann, M. J., & Gentner, D. (1998). More evidence for a relational shift in the development of analogy: Children's performance on a causal-mapping task. *Cognitive Development, 13*(4), 453–478. https://doi.org/10.1016/S0885-2014(98)90003-X.
Richland, L. E., Morrison, R. G., & Holyoak, K. J. (2006). Children's development of analogical reasoning: Insights from scene analogy problems. *Journal of Experimental Child Psychology, 94*(3), 249–273. https://doi.org/10.1016/j.jecp.2006.02.002.
Rittle-Johnson, B., & Star, J. R. (2007). Does comparing solution methods facilitate conceptual and procedural knowledge? An experimental study on learning to solve equations. *Journal of Educational Psychology, 99*(3), 561–574.
Rittle-Johnson, B., & Star, J. R. (2009). Compared with what? The effects of different comparisons on conceptual knowledge and procedural flexibility for equation solving. *Journal of Educational Psychology, 101*(3), 529–544.
Rosch, E., & Lloyd, B. B. (Éds.). (1978). *Cognition and categorization.* Lawrence Erlbaum.
Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8*(3), 382–439.
Schyns, P. G., Goldstone, R. L., & Thibaut, J.-P. (1998). The development of features in object concepts. *Behavioral and Brain Sciences, 21*(1), 1–17. https://doi.org/10.1017/S0140525X98000107.
Simms, N. K., Frausel, R. R., & Richland, L. E. (2018). Working memory predicts children's analogical reasoning. *Journal of Experimental Child Psychology, 166*, 160–177. https://doi.org/10.1016/j.jecp.2017.08.005.
Smith, L. B., Jones, S. S., & Landau, B. (1996). Naming in young children: A dumb attentional mechanism? *Cognition, 60*(2), 143–171.
Smith, L. B., Jones, S. S., Landau, B., Gershkoff-Stowe, L., & Samuelson, L. (2002). Object name learning provides on-the-job training for attention. *Psychological Science, 13*(1), 13–19.
Spencer, J. P., Perone, S., Smith, L. B., & Samuelson, L. K. (2011). Learning words in space and time: Probing the mechanisms behind the suspicious-coincidence effect. *Psychological Science, 22*(8), 1049–1057. https://doi.org/10.1177/0956797611413934.
St Clair-Thompson, H. L., & Gathercole, S. E. (2006). Executive functions and achievements in school: Shifting, updating, inhibition, and working memory. *Quarterly Journal of Experimental Psychology, 59*(4), 745–759. https://doi.org/10.1080/17470210500162854.
Stansbury, E., Witt, A., Bard, P., & Thibaut, J.-P. (2024a). How children generalize novel nouns: An eye-tracking analysis of their generalization strategies. *PLoS One, 19*(4). https://doi.org/10.1371/journal.pone.0296841 e296841.
Stansbury, E., Witt, A., & Thibaut, J.-P. (2019). Children's generalization of novel object names in comparison contexts: An eye tracking analysis. In *Proceedings of the 41th Annual Meeting of the Cognitive Science Society* (pp. 2871–2877).
Stansbury, E., Witt, A., & Thibaut, J.-P. (2024b). How children generalize novel nouns: An eye-tracking analysis of their generalization strategies. *PLoS One, 19*(4) e296841.
Thibaut, J.-P., & French, R. M. (2016). Analogical reasoning, control and executive functions: A developmental investigation with eye-tracking. *Cognitive Development, 38*, 10–26. https://doi.org/10.1016/j.cogdev.2015.12.002.
Thibaut, J.-P., Glady, Y., & French, R. M. (2022). Understanding the what and when of analogical reasoning across analogy formats: An eye-tracking and machine learning approach. *Cognitive Science, 46*(11). https://doi.org/10.1111/cogs.13208 e13208.
Thibaut, J.-P., Stansbury, E., & Witt, A. (2018). Generalization of novel names for relations in comparison settings: The role of conceptual distance during learning and at test. In *Conférence Proceedings of the 40th Annual Meeting of the Cognitive Science Society* (pp. 1114–1119).
Thibaut, J.-P., & Witt, A. (2015). Young children's learning of relational categories: Multiple comparisons and their cognitive constraints. *Frontiers in Psychology, 6*. https://doi.org/10.3389/fpsyg.2015.00643 643.

Thibaut, J.-P., & Witt, A. (2023). Children's generalization of novel names in comparison settings: The role of semantic distance during learning and at test. *Journal of Experimental Child Psychology, 234* 105704.

Twomey, K. E., Ranson, S. L., & Horst, J. S. (2014). That's more like it: Multiple exemplars facilitate word learning. *Infant and Child Development, 23*(2), 105–122. https://doi.org/10.1002/icd.1824.

Vlach, H. A., Ankowski, A. A., & Sandhofer, C. M. (2012). At the same time or apart in time? The role of presentation timing and retrieval dynamics in generalization. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*(1), 246–254.

Wagenmakers, E.-J., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin & Review, 11*(1), 192–196. https://doi.org/10.3758/BF03206482.

Waxman, S. R. (1990). Linguistic biases and the establishment of conceptual hierarchies: Evidence from preschool children. *Cognitive Development, 5*(2), 123–150.

Waxman, S. R. (1998). Linking object categorization and naming: Early expectations and the shaping role of language. In D. L. Medin (Éd.), *Psychology of learning and motivation* (Vol. 38, pp. 249–291). Academic Press. https://doi.org/10.1016/S0079-7421(08)60189-9.

Waxman, S. R., & Klibanoff, R. S. (2000). The role of comparison in the extension of novel adjectives. *Developmental Psychology, 36*(5), 571–581.

Waxman, S. R., & Markow, D. B. (1995). Words as invitations to form categories: Evidence from 12- to 13-month-old infants. *Cognitive Psychology, 29*(3), 257–302.

Welsh, J. A., Nix, R. L., Blair, C., Bierman, K. L., & Nelson, K. E. (2010). The development of cognitive skills and gains in academic school readiness for children from low-income families. *Journal of Educational Psychology, 102*(1), 43–53. https://doi.org/10.1037/a0016738.

Woodward, A. L., Markman, E. M., & Fitzsimmons, C. M. (1994). Rapid word learning in 13- and 18-month-olds. *Developmental Psychology, 30*(4), 553–566.

Xu, F., & Tenenbaum, J. B. (2007). Sensitivity to sampling in Bayesian word learning. *Developmental Science, 10*(3), 288–297.

Yoo, J., & Yim, D. (2018). Relationship among executive functions, vocabulary and reading skills in school-aged children with and without poor vocabulary. *Communication Sciences & Disorders, 23*(3), 570–583 https://doi.org/10.12963/csd.18523.

Zelazo, P. D. (2006). The Dimensional Change Card Sort (DCCS): A method of assessing executive function in children. *Nature Protocols, 1*(1), 297–301.

Zelazo, P. D., Anderson, J. E., Richler, J., Wallner-Allen, K., Beaumont, J. L., & Weintraub, S. (2013). II. NIH Toolbox Cognition Battery (CB): Measuring executive function and attention. *Monographs of the Society for Research in Child Development, 78*(4), 16–33. https://doi.org/10.1111/mono.12032.

Zelazo, P. D., Carter, A., Reznick, J. S., & Frye, D. (1997). Early development of executive function: A problem-solving framework. *Review of General Psychology, 1*(2), 198–226. https://doi.org/10.1037/1089-2680.1.2.198.

Zelazo, P. D., & Müller, U. (2011). Executive function in typical and atypical development. In U. Goswami (Ed.), *The Wiley–Blackwell handbook of childhood cognitive development* (2nd ed., pp. 574–603). Wiley-Blackwell.