Title

From two to many: The role of executive functions in young children's generalization of novel object names in a comparison design

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 42(0)

Authors

Lagarrigue, Yannick

Thibaut, Jean-Pierre

Publication Date

2020

Peer reviewed

From two to many: The role of executive functions in young children's generalization of novel object names in a comparison design

Yannick Lagarrigue (yannick_lagarrigue@etu.u-bourgogne.fr)
Jean-Pierre Thibaut (jean-pierre.thibaut@u-bourgogne.fr)
LEAD-CNRS, UMR 5022, Université de Bourgogne Franche-Comté,
Pôle AAFE –Esplanade Erasme, 21065 Dijon, France

Abstract

In this study, 4-year-old children were tested in an object name generalization task with a stimulus comparison design. Performance in the generalization task was correlated with performance in a vocabulary test and three executive function tasks assessing inhibition, flexibility, and working memory. Correlational analyses revealed a significant association with flexibility but not with inhibition, working memory or vocabulary test. We interpret the results in terms of a capacity to flexibly generate novel dimensions rather than inhibiting irrelevant dimensions. Individual differences in working memory and inhibition did not significantly influence performance in the word extension task. Moreover, the absence of correlation with the vocabulary performance supports the idea that children did not rely on existing knowledge to find out the relevant dimension.

Keywords: Comparison, Executive Functions, Distinctiveness, Conceptual Development.

Introduction

Learning novel words and the associated concepts relies on the ability to analyze the corresponding stimuli. It requires to figure out which properties are central for the target category. In easy situations, the salient properties are central to defining concepts (Murphy, 2002). However, in many cases, irrelevant superficial and salient similarities or differences can be more cognitively prominent than variations along more relevant dimensions, which has been shown to be challenging in learning conditions with one training stimulus or in situations in which several stimuli are introduced sequentially (Augier & Thibaut, 2013; Lawson, 2017; Son, Goldstone et Smith, 2011). Recent studies have suggested that the opportunity to compare two or more learning exemplars of a target category that are introduced simultaneously lead to better generalization performance than single presentations. Studies with both children and adults suggest that comparisons contribute to highlight non salient common and relevant properties (Gentner & Colhoun, 2010). In preschoolers, the benefits of comparison have been described for various types of words such as object names (e.g., Augier & Thibaut, 2013; Gentner & Namy, 1999), names for parts (Gentner, Anggoro and Klibanoff., 2007), action verbs (e.g., Childers & Paik, 2009), relational nouns (Gentner et al., 2011; Thibaut & Witt, 2015) or adjectives (Waxman & Klibanoff, 2000). In Gentner & Namy's study (1999), pictures of familiar objects were used to test novel names extensions in four-year old. In a single standard (no-comparison) condition (e.g. a bicycle called blicket), children extended the novel name to a perceptually similar object (e.g. a pair of glasses) more frequently than to a perceptually dissimilar taxonomically related object (e.g. a skateboard). When the experimenter introduced two taxonomically related standards that were both perceptually similar one to the other and with the perceptually similar lure (e.g. a bicycle and a tricycle, both labelled blicket) children mostly selected the taxonomically related item (i.e., the skateboard).

Most of the available evidence regarding the positive effects of comparison has been obtained with familiar objects in tasks in which perceptual similarities were pitted against taxonomic similarities. However, Graham et al. (2010) also observed a positive effect of comparison, in contrast with a no-comparison condition, with unfamiliar stimuli. They pitted a perceptually non-salient but designed to be conceptually relevant dimension (texture) against a perceptually salient but conceptually irrelevant one (shape). Four-year-old were tested either in a no-comparison condition or in a comparison condition. In the no-comparison case, the standard shared its texture but not its shape with one of two test objects, and its shape but not its texture with the other test object. As expected, a majority of children extended the new label to the same-shape test object. In the comparison condition, two standards were introduced with the same label. They had the same texture but differed along the shape dimension. The two transfer objects were the same as in the no-comparison condition. A majority of children extended the new label to the same-texture match. Hence, preschoolers were able to extract the unifying non salient dimension (texture) to guide their categorization in the comparison case.

Augier & Thibaut (2013) used items similar to Graham et al. (2010). They manipulated the number of items to-be-compared. They found that while four- and- six- year-old children benefited from comparison, compared to no-comparison, only six-year-old took advantage of a larger number of standards (four rather than two training standards). In an executive function context, the authors hypothesized that more training items in favor of the target dimension (i.e., texture) generated more comparisons to perform and thus increased the executive costs. Indeed, it can be argued that discovering nonobvious conceptually relevant dimensions requires the ability to inhibit immediate irrelevant and salient superficial dimensions. Once a dimension has been discovered to be conceptually irrelevant, one must flexibly rerepresent the stimuli along other dimensions or, in other words, find new less salient dimensions that might not come immediately to mind. Also, children must keep in mind which dimensions have already been tested and are irrelevant but also which new tested dimension did not unify the stimuli and which one was useful in previous trials. According to the executive functions point of view children have to inhibit salient irrelevant dimensions, flexibly find new relevant dimensions, and update their representation of conceptually relevant or irrelevant dimensions in working memory (see Richland, Morrison, & Holyoak, 2006 for discussions).

Despite a growing body of research on the benefits of comparison for novel name generalization, the respective role of individual differences in control processes (executive functions) and world knowledge in comparison word-learning situations remains unclear. In the present study, we address this question with an individual differences correlational approach in which we assess, in a comparison design, the relation between word-learning generalization performance and cognitive processes such as inhibition, cognitive flexibility and working memory (WM), and conceptual knowledge. In the analogy domain, Simms, Frausel, and Richland (2018) investigated whether executive functions (inhibition, flexibility, WM) were associated with analogical development with this individual differences approach. They first assessed 5- to 11-year olds' inhibitory control, working memory, and cognitive flexibility and found that individual differences in children's working memory were the best predictor of their performance in the scene analogy task they used. Even after controlling for age, the same relationships remained significant, suggesting a strong interrelationship between analogical reasoning and working memory development. With semantic (e.g., Richland et al., 2006) or perceptual (e.g., Thibaut et al., 2010a, b) analogical reasoning tasks, results support the idea that comparison processes, and thus generalization, are influenced by executive costs. In these contributions, the presence of irrelevant perceptual features or semantic distractors predicted children's performances, arguably in that it required inhibition and flexibility to solve the task.

The overarching purpose of the present experiment was to assess the extent to which novel name generalization is influenced by executive functions and world knowledge (see Simms et al. 2018, for executive functions). We followed a correlational approach in which performance in various executive functions, world knowledge and scores in a novel name learning task were measured. First, we assessed vocabulary knowledge because it reflects world and conceptual knowledge seen as crystallized intelligence (e.g., Ashton et al., 2000). It has been argued that world knowledge is a key factor for conceptual abstraction and understanding in the sense that the more children know about the world the more likely they will discover new conceptually relevant dimensions (e.g., Gentner & Hoyos, 2017). A positive correlation with vocabulary considered as an index of world knowledge can be interpreted as a sign that participants with better vocabulary competence used their background knowledge to make sense of the unfamiliar stimuli we used. Indeed, Gentner and colleagues (Gentner & Christie 2010; Gentner & Hoyos 2017, Gentner & Namy 1999) have suggested that hearing common label for two exemplars is an invitation to compare these exemplars. Moreover, as Gentner et Hoyos (2017) put it, the richer children's linguistic knowledge is, the higher the cognitive gains of comparison. They claimed that knowledge of the domain of the objects presented drives the comparisons benefits. Indeed, knowledge of a technical vocabulary is a direct cue of an expertise in a given area.

As for executive functions, in our concept learning-comparison context, they can play a role at various levels of the task. A first hypothesis is that discovering nonobvious conceptually relevant dimensions in the case of unfamiliar stimuli, requires the ability to inhibit immediate irrelevant superficial aspects (for example, the shape of a stimulus is a priori salient, see Landau et al., 1988). Then, a second hypothesis is that once a (salient) dimension has been tested to be conceptually irrelevant, one must flexibly rerepresent the stimuli along other potentially relevant dimensions in order to find new, less salient, dimensions that did not immediately come to mind. Under this view, the difference between the saliency of the nonobvious but relevant dimension and the saliency of the superficial but not relevant dimension, directly impact cognitive costs. Working memory is also involved, particularly updating, because children must keep in mind which dimensions have already been tested for conceptual relevance, which new tested dimension were found to be conceptually irrelevant, and which one unified the stimuli in previous trials.

In the present experiment, we used unfamiliar objects similar to the ones used by Graham et al. (2010) or Augier and Thibaut (2013). We tested our participants with various cognitive tasks assessing three executive functions (working memory, inhibition and cognitive flexibility, see Miyake et al., 2000). The reasoning was that depending on the observed correlations (if any) between performance in the novel name learning task and the executive function tasks, different control mechanisms might be at play (Anderson, 2002; Zelazo & Müller, 2007). First, we hypothesized that a correlation between the learning task and our measure of inhibition would result from participants' difficulties to inhibit the salient and irrelevant common shape. A second hypothesis was that a correlation between the novel name generalization task and flexibility would mean that flexibly redescribing the stimuli in terms of a less salient dimension (once the salient one has been recognized as irrelevant) was also more difficult for participants with lower generalization scores. Third, a correlation between naming and working memory performance would be interpreted as a sign that participants who have difficulties keeping former hypotheses or descriptions of previous stimuli in working memory have difficulties with the conceptualization task. Augier and Thibaut (2013) followed the same line of reasoning. However, the major difference between the present experiment and Augier and Thibaut (2013) is that they directly manipulated the cognitive costs of the task with the number of training items, age and the presence of a contrast. They did not assess participants' executive functions skills. Here, we will assess participants' executive functions and correlate these measures with their scores at the name extension task.

Methods

Participants

39 female and 37 male 4- to 5-year old preschoolers were tested individually in a quiet room in their school (mean age = 54.45m, SD = 3.73, range: 46-60m). Informed consent was obtained from their school and their parents. The procedure followed institutional ethics board guidelines for research on humans.

Materials

Eighteen sets of four unfamiliar artificial grey-scale objects depicted on cards were created and divided into two groups of nine sets. One group used more salient textures and the other group less salient textures. The shapes remained the same in both groups. A participant saw one of the two groups of nine stimuli. Each set was composed of four stimuli, two training standards and two transfer-test options. In each set, the two standards shared the same texture but had different shapes. The first test object, the shape match, had the same shape as one of the two standards but had a different texture. The other test object, the texture match, had the same texture as both standards but had a different shape (see Figure 1). Two sets out of the nine sets were used as practice trials.

The size of each object was approximately 6.0 cm by 6.0 cm. They were printed on a laminated card measuring 12.0 cm x 9.0 cm. Textures and shapes that were used in one set differed from all the textures that were used for the other sets. The order of presentation was pseudo-randomized within and across participants. Each set was associated with one of nine two-syllable novel names, *Youma*, *Buxi*, *Dajo*, *Zatu*, *Sepon*, *Xanto*, *Vira*, *Loupo* and *Rodon*.

Procedure

We used a forced-choice categorization task in which children had to decide which of two simultaneously presented objects had the same name as the standards. Children were tested in French. Each standard was introduced with a novel count noun (e.g. “This is a buxi. / Ceci est un buxi.” – pointing to the first standard, and “This is a buxi TOO. / Ceci est AUSSI un buxi.” – pointing to the other standard. The objects were presented sequentially and left in view. Then, the two test objects (i.e., the shape and the texture matches) were introduced and the child was asked to point to the one which had the same name which would also be given the same name (e.g., “Show me which one of these two is also a buxi / Montre-moi lequel de ces deux est aussi un buxi”). Each child started with two practice trials followed by seven test trials presented in random order.

Cognitive assessment: Vocabulary and three executive functions (working memory, flexibility and inhibition) were assessed. For the working memory and the flexibility tasks, we adapted the corresponding tasks from the National Institutes of Health Toolbox battery (NIH Toolbox CB). We followed the same protocol except that we implemented the task on Open Sesame. We assessed participants’ skills with a computer and the instructions were given in French.

In the computerized working memory task, children were presented with a series of animal pictures along with their auditory name. They were instructed to remember and to verbally rank all the animals from the smallest to the biggest. The number of items in the list increased every two trials. Two list were presented. The task was stopped after two errors in two trials of the same number of items. The score was the sum of correct trials in both lists.

In the flexibility task, we adapted the Dimensional Change Card Sort (DCCS) test which is used in the NIH Toolbox battery and was implemented on Open Sesame. The instructions were provided in French. In this task, children were shown two target cards (e.g., a red rabbit and a blue boat) and asked to follow a rule, which is “to choose the one with the same color (or shape) as the example”. After a fixed number of trials, they were asked to reverse the rule and to select the one with the same shape (or color). This task assesses children’s ability to switch from one rule to another rule.

The inhibition task was the Real Animal Size Test (Catale & Meulemans, 2009) which was computerized on Open Sesame. Children were presented with an animal picture on the computer 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 presented. Two different sizes of pictures were used. Big and small animals could be displayed on the screen either with a big size, either with a small size. Thus, in the congruent trials the size of an animal in the real world was congruent with its size on the picture whereas in the incongruent trials the size of the real animal was not congruent with the size of the picture. In the latter case, children had to inhibit their tendency to respond to the size of the animal in the picture and rather to answer to the real size of the animal.

For the vocabulary test, we used the EVIP which is a French adaptation (Canadian norms) of the PPVT (Peabody Picture Vocabulary Test, Dunn & Dunn, 2007). In this test, children had to select the one out of four-images associated with a noun given by the experimenter. Responses were recorded on a paper sheet and a standard score was computed according to the age. Data analysis

For the categorization, Stroop and DCCS tasks, we measured the reaction times. Except for the categorization task (in which there were not incorrect responses), all reaction times of incorrect responses were discarded from the analysis. For the three tasks, all the reaction times inferior to 100ms and more than two deviation standards away from the mean was considered as outliers and discarded from the analysis. Then, a z-score called Score Time was calculated for each participant’s reaction time.

Given that in these tasks we were more interested in correctness than speed, the weight given to the percentage of correct responses was higher than the weight attributed to the reaction times. To do so, we added the Score Time to the percentage of correct responses. This score was used in all the analyses.

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

Number of participants	76
Sex ratio (F/M)	39/37
Age (in months)	54.44 (±3.73)
Categorization score	52.09 (±38.29)
Vocabulary score (EVIP)	109.72 (±19.16)
Working memory score	5.34 (±2.06)
Flexibility score (DCCS)	77.75 (±17.95)
Inhibition score (Stroop)	84.38 (±13.74)
First trial score	52.63 (±50.29)

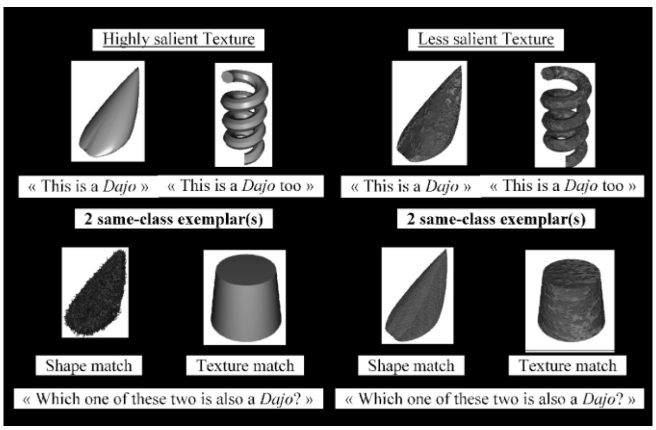

Figure 1: Example of a stimulus set with the two saliencies of Texture.

Results

We performed Pearson's correlation analyses to investigate the link between age, vocabulary, working memory, inhibition, flexibility and children's performance in the generalization task, which was the main target of the present paper. We expected positive correlation so we computed one-tail correlations. The correlation matrix is reported in Table 2. Age was found to be significantly correlated with working memory (*r* (76) = .280, *p* = .007), inhibition (*r* (76) = .386, *p* <.001) and the categorization score (*r* (76) = .285, *p* = .006). Older children performed better than younger children in the working memory, the inhibition and the categorization tasks. This is in line with many studies showings that executive functions develop with age (Anderson, 2002; Best & Miller, 2010) and studies showing that performance in generalization tasks also improves with age (Augier & Thibaut 2013; Gentner, 1999). Working memory was also significantly correlated with inhibition (*r* (76) = .196, *p* = .045) and with vocabulary (*r* (76) = .279, *p* = .007), which was less expected.

The most interesting result was the correlation between the global categorization performance and flexibility score (DCCS) was significant (*r* (76) = .190, *p* = .05). In other words, higher flexibility scores meant more texture matches choices. We also introduced an unusual analysis, a correlation between the result in the first trial (correct or error). This was motivated by recent results (Lagarrigue & Thibaut, submitted) showing in a similar task that the result for the first trial was a good predictor of a later performance in a task that relied on the same relevant feature as the one relevant in the first trial. We interpret this result in terms of inhibition and cognitive flexibility: participants who were able to inhibit the irrelevant salient in the first trial or were able to rerepresent the first stimulus with another dimension might also be those who would reach the best performance level. Following this logic, we computed correlations which revealed a significant association between age and first trial score (*r* (76) = .264, *p* = .010), mean categorization score (*r* (76) = .656, *p* <. 001) and inhibition (*r* (76) = .198, *p* = .043). The other correlations were not significant. In our context, the correlation between inhibition and first trial score suggests that children with a higher level of inhibition were also the ones who were more prone to succeed in the first trial and in later trials.

To identify the variables that were most predictive of the generalization score among the children, we carried out a stepwise procedure using the AIC as our criterion for model selection (Hu, 2007). We used a backward selection procedure which starts with all predictors in the model, iteratively removes the least contributive predictors, and stops when all predictors are statistically significant. Then the model with the lowest AIC criterion is selected (see Table 3). Our predictive variables were age, performance in the first trial, scores for flexibility, vocabulary, inhibition and working memory, which were continuous factors. Results show that the best model is Model 5 which includes both flexibility score and first trial score as significant predictive variables. This model significantly predicted generalization score variation across our sample (*p* < .001) and explained 45% of this variation, as demonstrated by the adjusted R^2. It confirmed the previous results by revealing a significant effect of flexibility (*F* = 2.101, *p* = .039) and first trial score (*F* = 7.62, *p* < .001).

Table 2: correlation matrix (one-tail). Significant results are written in bold.

	Categorization score	Age	Working memory score	Flexibility score (DCCS)	Inhibition score (DCCS)	Vocabulary score (EVIP)	First trial
Categorization score		***r* = .285** ***p* = .006**	*r* = .013 *p* = .457	*r* = .190 *p* = .051	*r* = .108 *p* = .176	*r* = -.070 *p* = .725	***r* = .656** ***p* < .001**

Age	**r = .285** **p = .006**		**r = .280** **p = .007**	r = .054 p = .322	**r = .386** **p < .001**	r = -.044 p = .649	**r = .264** **p = .010**
Working memory score	r = .013 p = .457	**r = .280** **p = .007**		r = .067 p = .284	**r = .196** **p = .045**	**r = .279** **p = .007**	r = -.010 p = .534
Flexibility score (DCCS)	r = .190 p = .051	r = .054 p = .322	r = .067 p = .284		r = 090 p = .220	r = .059 p = .306	r = .014 p = .452
Inhibition score (Stroop)	r = .108 p = .176	**r = .386** **p < .001**	**r = .196** **p = .045**	r = 090 p = .220		r = .184 p = .056	**r = .198** **p = .043**
Vocabulary score (EVIP)	r = -.070 p = .725	r = -.044 p = .649	**r = .279** **p = .007**	r = .059 p = .306	r = .184 p = .056		r = -.070 p = .727
First trial	**r = .656** **p < .001**	**r = .264** **p = .010**	r = -.010 p = .534	r = .014 p = .452	**r = .198** **p = .043**	r = -.070 p = .727	

Table 3: Goodness of fit of the regression linear model

Model	Degree of freedom	AIC	F value	Adjusted R^2	p value
M1: Categorization score ~ Age + Working memory + Flexibility score + Inhibition Score + Vocabulary score + First trial	69	734.87	10.66	.436	< .001
M2: Categorization score ~ Age + Working memory + Flexibility score + Inhibition Score + First trial	70	732.89	12.97	.444	< .001
M3: Categorization score ~ Age + Flexibility score + Inhibition Score + First trial	71	730.93	16.43	.451	< .001
M4: Categorization score ~ Age + Flexibility score + First trial	72	729.86	21.64	.452	< .001
M5: Categorization score ~ Flexibility score + First trial	**73**	**729.51**	**31.44**	**.448**	**< .001**
M6: Categorization score ~ First trial	74	731.97	55.89	.423	< .001

Discussion

This study investigated the link between executive functions and concept learning and generalization in a novel name learning task comparison design. The purpose was to assess whether executive function would correlate (and in the positive case, which one) with children's generalization performance in our generalization task. We also assessed the role of another explanatory contender, world knowledge which was assessed with a vocabulary test.

Results revealed a correlation between flexibility and concept generalization performance, and no correlation with inhibition and working memory. We hypothesize that despite the fact that executive functions are moderately correlated (Miyake et al., 2000), or less differentiated in young children (Wiebe & Karbach, 2017), inhibition might be involved in the inhibition of the salient but irrelevant common shape. Working memory might be involved in keeping different hypotheses active and in keeping the relevant dimension in working memory from one trial to the next. Last, flexibility was expected to be important when salient dimensions were inhibited or found irrelevant, because participants had to generate novel representations of the stimuli.

We found a positive correlation between categorization performance and flexibility and no significant correlation with the other tested executive components. In our framework, this result is compatible with the idea that performance differences were mainly associated with the necessity to rerepresent the stimuli when this was necessary. However, we also found a correlation between the score in the first trial and inhibition. This result is interesting because it is compatible with the idea that the early inhibition of a salient information (shape) is an important step towards the discovery of the relevant less salient dimension. Indeed the first trial score is also strongly correlated with the categorization score. Our results show that flexibility is important to find the correct solution together with the early inhibition of the irrelevant information.

One interesting feature of the present comparison design was its simplicity. The available information (constitutive dimensions) is relatively small (mainly shape and texture) and the number of putative processes involved at the conceptual level of comparisons is most likely relatively low: looking at a salient shape, comparing the first standard with the second standard, considering shape as a potential hypothesis, discarding (inhibiting) it and look for other dimensions (flexibility) and keeping the information in mind across trials. This low complexity makes it relatively easier to relate an executive component with a particular task component than in other cases in which complex skills (e.g., reading comprehension) are correlated with cognitive processes.

Vocabulary-world knowledge did not explain the results. The absence of correlation involving vocabulary and, thus, world knowledge in our task is an interesting result because world knowledge provides a rich interpretation of the stimuli when participants relate new stimuli with known stimuli, is a common hypothesis. Also, a richer world database might also provide a richer set of relevant encoding dimensions: the more a child knows, the more dimensions this child can use in order to encode the stimuli. This absence of a relationship does not confirm the hypothesis that world knowledge is the major determinant of conceptual learning and abstraction (Gentner & Hoyos, 2017).

Here we sought to correlate executive functions with a learning task, which is by nature a dynamic task. This is analogous to former studies searching for correlations between conceptual tasks such as analogical tasks and executive functions (see Simms et al. 2018). However, analogical tasks are tasks that require the relevant background knowledge in order to find the solution. The challenge is to find the relation which is common to both compared domains. In our case, we managed to correlate a novel conceptual content learning task with executive functions. In many other studies aiming at finding correlations between academic competences such as mathematics and executive functions, the relation between the competence and executive functions is often more obscure, since the competence is multidimensional, with both declarative knowledge and processing. Our task was much simpler and showed that inhibition and flexibility were at play, meaning that children were trying to inhibit part of the activated dimensions and redescribe the stimuli.

In sum, our data provide evidence that components of executive functions might contribute to learning and generalizing a novel name.

Supplementary information

Note that our study started in the beginning of 2020 and was stopped because of the Covid-19 pandemic. The absence of correlation between learning-generalization, and vocabulary, for example might be due to the small number of participants. Also, it will be interesting to measure the correlations for our two difficulty levels separately. We hoped to have completed data collection for our paper resubmission. This was not possible so the present data must be considered as preliminary.

Acknowledgments

The authors wish to thank the National Agency of Research (ANR) for their financial support (COMPARE project). They also would like to thank Damien Foinant, Ella Stansbury Juliette Taillandier-Coindard, Florine Crotet and Audrey Monterrat for their help in running the experiment.

References

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

Ashton, M. C., Lee, K., Vernon, P. A., & Jang, K. L. (2000). Fluid intelligence, crystallized intelligence, and the openness/intellect factor. *Journal of Research in Personality*, 34(2), 198–207.

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.

Best, J. R., & Miller, P. H. (2010). A developmental perspective on executive function. *Child development*, 81(6), 1641-1660.

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.

Childers, J. B., & Paik, J. H. (2009). Korean-and English-speaking children use cross-situational information to learn novel predicate terms. *Journal of Child Language*, 36(1), 201–224.

Dunn, L. M., & Dunn, D. M. (2007). *PPVT-4: Peabody picture vocabulary test*. Pearson Assessments.

Gentner, D., Anggoro, F. K., & Klibanoff, R. S. (2011). Structure mapping and relational language support children's learning of relational categories. *Child Development*, 1173–1188.

Gentner, D., & Christie, S. (2010). Mutual bootstrapping between language and analogical processing. *Language and Cognition*, 2(2), 261-283.

Gentner, D., & Colhoun, J. (2010). Analogical Processes in Human Thinking and Learning. In B. Glatzeder, V. Goel, & A. Müller (Éd.), *Towards a Theory of Thinking* (p. 35-48). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-03129-8_3

Gentner, D., & Hoyos, C. (2017). Analogy and abstraction. *Topics in cognitive science*, 9(3), 672–693.

Gentner, D., Loewenstein, J., & Hung, B. (2007). Comparison Facilitates Children's Learning of Names for Parts. *Journal of Cognition and Development*, 8(3), 285-307. https://doi.org/10.1080/15248370701446434

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

Hu, S. (2007). Akaike information criterion. *Center for Research in Scientific Computation*, *93*.

Lagarrigue, Y., Augier, L., Thibaut, J.-P. (2020). Dimensional Distinctiveness Constraints on Comparison Processing Across Development, *submitted for publication*.

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.

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.

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.

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

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.

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.

Son, J. Y., Smith, L. B., & Goldstone, R. L. (2011). Connecting instances to promote children's relational reasoning. *Journal of experimental child psychology*, 108(2), 260-277.

Thibaut, J.-P., French, R., & Vezneva, M. (2010a). The development of analogy making in children: Cognitive load and executive functions. *Journal of Experimental Child Psychology*, 106(1), 1-19.

Thibaut, J. P., French, R., & Vezneva, M. (2010b). Cognitive load and semantic analogies: Searching semantic space. *Psychonomic bulletin & review*, 17(4), 569-574.

Thibaut, J.-P., & Witt, A. (2015). Young children's learning of relational categories: Multiple comparisons and their cognitive constraints. *Frontiers in psychology*, 6, 643.

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

Wiebe, S. A., & Karbach, J. (2017). Executive function: Development across the life span. Routledge.

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.

Zelazo, P. D., & Müller, U. (2007). Executive Function in Typical and Atypical Development. *In Blackwell Handbook of Childhood Cognitive Development* (p. 445-469). John Wiley & Sons, Ltd. https://doi.org/10.1002/9780470996652.ch20