

# Children's Generalization of Novel Object Names in Comparison Contexts: An eye tracking analysis

Ella Stansbury, Arnaud Witt and Jean-Pierre Thibaut  
([ellastansbury@gmail.com](mailto:ellastansbury@gmail.com); [arnaud.witt@u-bourgogne.fr](mailto:arnaud.witt@u-bourgogne.fr); [jean-pierre.thibaut@u-bourgogne.fr](mailto:jean-pierre.thibaut@u-bourgogne.fr))

LEAD-CNRS, UMR 5022, Université de Bourgogne Franche-Comté,  
Pôle AAFE – Esplanade Erasme, 21065 Dijon, France

## Abstract

A common result is that comparison settings (i.e., several stimuli introduced simultaneously) favor conceptualization and generalization. In a comparison setting, we manipulated the semantic distance between the two training items (e.g., two bracelets versus a bracelet and a watch), and the semantic distance between the training items and the test items (e.g., a pendant versus a bow tie). We tested 5- and 8-year-old children's generalization of novel names for objects. This study is the first one to study the temporal dynamics of comparison in a generalization task with eye-tracking data. The eye movement data revealed clear patterns of exploration in which participants first focused on the training items and compared them with each of the choice options. We also compared the search profiles for correct answers and errors. The results show that participants first found commonalities in the learning items, which they compared with each item in the solution set. This pattern is consistent with an alignment view of generalization.

**Key words:** comparisons; conceptual distance; generalization; strategies; eye tracking measures

## Introduction

Children usually learn the reference of novel words with a limited number of stimuli which are associated with these words. Which learning stimuli lead to accurate generalizations and which mode of presentation would be optimal to achieve this goal are crucial issues for concept learning. A large set of recent studies have shown that comparison settings lead to better generalization results than no-comparison learning conditions. In the latter case, young children tend to generalize novel words to objects that are perceptually similar to the learning items rather than to conceptually related ones (Imai, Gentner, & Uchida, 1994). By contrast, comparison settings favor conceptually based generalizations because they enable children to neglect irrelevant perceptual dimensions and highlight non-obvious properties that need to be identified to choose a taxonomic match (e.g., (Augier & Thibaut, 2013; Gentner & Namy, 1999; Namy & Gentner, 2002).

However, still little is known of the solving strategies used to process comparison settings and generalize novel words, or of the steps that lead to generalization. In the present study we use eye tracking data to identify these strategies and get a better understanding of the cognitive processes that undergo comparison and generalization during learning.

## Comparison and generalization

A large body of research demonstrates the benefits of comparison settings for learning novel object names (e.g., (Graham, Namy, Gentner, & Meagher, 2010), adjectives (e.g., Waxman & Klibanoff, 2000), action verbs (e.g., (Childers & Paik, 2009), objects (Thibaut, 1991; 1995) relational nouns (Gentner, Anggoro, & Klibanoff, 2011; Thibaut & Witt, 2015; see (Alfieri, Nokes-Malach, & Schunn, 2013). For example, Gentner and Namy (1999) presented 4-year olds with familiar objects with an imaginary name and asked them to extend the name. Children had to choose between two pictures, a taxonomic match and a perceptual match. Results showed that children preferred the perceptual match when they had only seen one object during the learning phase but preferred the taxonomic match when they had the opportunity to compare two objects with the same name, introduced simultaneously during the learning phase. The conditions under which comparisons lead to better learning and generalization have received much attention in recent years.

One crucial point is that comparisons generate cognitive costs (e.g., Richland, Morrison, & Holyoak, 2006; Thibaut, French, & Vezneva, 2010b) in the field of analogical reasoning). The hypothesis is that comparing multiple items, and choosing a match, while neglecting irrelevant dimensions including salient dimensions such as perceptual similarities may generate cognitive costs because of the inhibition, decision making and flexibility involved in the task. For example, Augier and Thibaut (2013) studied conceptualization of unfamiliar objects in a comparison paradigm and manipulated the number of exemplars shown during the comparison phase. They tested 4- and 6-year olds and compared a no comparison condition, a 2-item comparison condition and a 4-item comparison condition. Interestingly, all children benefited from the comparison conditions compared to the no-comparison conditions. However only older children benefited from the four-item comparisons compared to the two-item comparisons. This suggests that cognitive control is necessary to succeed the task as suggested by contributions in numerous domains involving comparisons and integration of multiple information (see (Wiebe & Karbach, 2018).

The semantic distance between the compared items might contribute to increase the cognitive costs of comparison. For example, Green, Kraemer, Fugelsang, Gray, and Dunbar (2010) have shown that analogies based on distant domains were more difficult than equivalent analogies connecting closer domains because distant analogies involved more creativity, which was related to the central role of the prefrontal cortex in cognitive control. In children, Thibaut, French, and Vezneva, (2010a) have shown that semantic analogies based on weakly associated

relations are more difficult than those based on strongly associated relations. The authors interpreted this result in terms of the necessity to inhibit strongly associated but irrelevant items in the context at hand or in terms of the necessity to generate new candidate relations, which requires cognitive flexibility in the case of distant semantic domains.

In this cognitive control framework, it is argued that aligning semantically distant training items might involve deeper conceptual encoding. Indeed, for semantically close items, perceptual similarities are aligned with conceptual similarities (e.g. two apples) whereas for semantically more distant items alignable perceptual similarities are less synonymous of conceptually alignable similarities: aligning surface similarities does not entail an alignment of conceptual similarities or surface similarities are less correlated with conceptual similarities (e.g. a bracelet might look like a watch, but the nature of a watch is strongly connected with a device giving the time, which can have a low saliency). On the other hand, if alignable perceptual similarities are well correlated with deep similarities for close learning items, the fact that these deeper similarities are embedded in perceptual similarities might prevent them from being easily aligned with conceptually important features when the generalization items are perceptually dissimilar. In that case, generating conceptual similarities might be difficult because the conceptual space cannot be grounded on perception and thus requires more extensive conceptual analysis.

### **Exploring children's strategies with eye tracking movements in a learning-generalization task**

The present study's aim is to analyze the temporal dynamics of a comparison task, from the study of learning items to the selection of a candidate generalization stimulus, which, to the best of our knowledge has never been done. We will use materials by Thibaut and Witt (2017). They manipulated the semantic distance between the learning items (e.g., two bracelets versus a bracelet and a watch), and the semantic distance between the learning items and the generalization items (e.g., a jewel, near distance, versus a bow tie, far distance), and analyzed which combination of conditions would lead to more taxonomic choices. Four-year-old children made less taxonomic choices in the far generalization condition than the close generalization condition whatever the learning distance, whereas only six-year-old children got better results in the far learning distance, a condition in which participants had to coordinate information coming from very different domains. In the above cognitive control context, the authors argued that, as executive functions develop, children are able to compare stimuli from remote conceptual spaces more systematically. Indeed, the common features between two items may be found more easily with semantically close items than with semantically distant items. In the latter case, these features might be less salient and require more comparisons to be noticed. Also, in a broader conceptual space, the set of irrelevant properties to inhibit is likely to be larger than in a close domain.

Recent eye-tracking research on analogical reasoning tasks (another generalization task) have shown that during development younger children's solving strategies differ from older children's and adults' strategies (J.-P. Thibaut & French, 2016). They confronted two main hypotheses to the data, the projection first and the alignment first strategies. Projection-first refers to an initial analysis of the learning domain, in search of a relation connecting A and B. Once a relation is found it is projected on the generalization domain (which generalization item goes with C with the same relation). The alignment-first strategy refers to the alignment of equivalent stimuli (i.e., that play the same role) in the learning and the generalization domains (A with C, and B with D in a A:B::C:D proportional analogy). The authors showed that adults and children followed different search strategies.

In the type of comparison task we use, successful learning requires the learning items (L1 and L2) to be compared and conceptually aligned. Generalization requires switches (witnessing comparisons) between L1-L2 and the Taxonomic target. How one reaches the taxonomic solution (or fails to) will be reflected in the transitions between L1-L2 and the set of the available options (taxonomic, thematic, and perceptual). The set of transitions and the time spent on each item will illustrate the search-construction of a solution.

Among the potential strategies, the projection-first strategy predicts early L1-L2 transitions (finding commonalities between L1 and L2), followed by comparisons between the three generalization options in terms of the features they actually share (either thematic, or perceptual, or taxonomic) with the common feature they have discovered for L1 and L2. The alignment-first hypothesis predicts early comparisons between learning items but also between the learning items and the generalization items, in order to find conceptually analogous items in the transfer set. One additional prediction is that participants compare each learning stimulus with each of the options

Another strategy contrast exists between constructive matching or response elimination (Bethell-Fox, Lohman, & Snow, 1984). Constructive matching predicts early L1-L2 comparisons followed by comparisons between generalization items that may reveal a careful construction of a solution and the application of the solution to the generalization set. This hypothesis makes similar predictions to those from the projection-first hypothesis. Response elimination predicts L1-L2 comparisons followed by back and forth switches between L1-L2 items and generalization options that may reveal a systematic response elimination strategy until a final choice is made. A difference between alignment first and response elimination, is that alignment-first predicts a progressive convergence towards the solution whereas the response elimination predicts no systematic search pattern.

The present study's main goal is to describe the strategies used by children to generalize correctly in a comparison setting, by analyzing eye movement data from two groups of children (5- and 8-year olds). We selected these two age

groups because previous research has shown that participants eye-tracking methods can be used with complex tasks with 5-year olds. Also, (J.-P. Thibaut et al., 2010b) showed that five-year olds might adapt their search strategy to the difficulty of the task in a less systematic way than 8-year olds. Thus, a priori, these two age groups were good candidates for studying strategy differences (if any exist). Also Thibaut and French (2016) showed that reliable results could be obtained with 5-year olds in an eye-tracking task.

We will confront our data with the strategies mentioned above. One hypothesis is that age matters: younger participants should use the response elimination strategy more often than the older group because it is cognitively less demanding: participants compare each transfer with the learning items, one by one, rather than store the found dimensions in working memory and compare all the transfer items in a row. They should also produce less systematic search patterns. For example, correct trials should start with L1L2 transitions less often for young children. There should also be differences depending on the difficulty of the task: easier conditions should elicit less transitions than difficult ones. Far generalization should be more difficult and should result in a larger proportion of comparisons between the options compared to the learning items.

Of particular interest are the differences, if any, between correct and error trials. Do strategy differences between errors and correct trials appear at the onset of the trial (thus, with significant differences in the first slices) or do they result from a wrong decision at the end of the trial (i.e. differences in the last slice), once all the options have been considered.

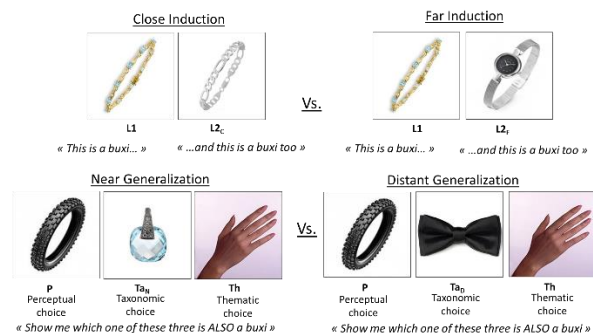
## Methods

**Participants** 109 French speaking children were tested individually in a quiet room at their school. Two age groups were tested, five year olds, and eight year olds. Forty-nine younger children were recruited (mean age = 5;3; range: 4;11 to 5;9), and 60 children for the older age group (mean age = 8;4; range: 7;11 to 9;4). Informed consent was obtained from their school and their parents.

**Materials** Fourteen experimental sets of pictures were built, plus three warm-up trials. Each set was associated with a category (e.g., clothing, food, tools, accessories, animals), and was composed of 7 pictures. Two learning objects, either from the same basic level category (close learning, L1 and L2<sub>C</sub>) or from the same superordinate category (far learning, L1 and L2<sub>F</sub>) (see Figure 1). The test pictures subsets were composed of three pictures: a taxonomically related generalization object (Ta), either near Ta<sub>N</sub>, or distant, Ta<sub>D</sub>, see Figure 1), an object perceptually similar to the initial learning object (P) and an object thematically related to the category (Th) (see Figure 1). This design worked as follows. For each object category (e.g., clothing accessories), the close learning objects (L1, L2<sub>C</sub>) were composed of perceptually and semantically close items (e.g., a bracelet - a curb chain), while the far pairs (L1, L2<sub>F</sub>) were composed of perceptually similar but conceptually more distant items (e.g., a bracelet – a watch).

The three test pictures consisted of three objects in both the near and the distant generalization conditions. The perceptual match (P) was perceptually similar but semantically unrelated to the two training items (e.g., a tire in our bracelet case), the taxonomic choice (Ta) was perceptually dissimilar but taxonomically related to the learning objects and a thematically related object that was not perceptually related but thematically related (Th, e.g., a hand). Depending on the generalization condition, near or distant, the taxonomic choice was semantically near (Ta<sub>N</sub>) or more distant (Ta<sub>D</sub>) to the learning items (e.g., a jewel pendant in the near generalization case, or a bow tie in the distant generalization case). See Figure 1 for the "clothing accessories" category. Thus, a trial was composed of 5 pictures, L1, L2 (L2<sub>C</sub> or L2<sub>F</sub>), Th, Ta (Ta<sub>N</sub> or Ta<sub>D</sub>) and P, resulting in four possibilities (Close learning - Near or Distant generalization; Far learning - Near or Distant generalization).

Independent similarity ratings were obtained from fifty-four university undergraduate students. They are described in Thibaut and Witt (2017). They revealed that the close learning objects in a pair were conceptually closer one to the other than the objects composing the far learning pairs ( $p < .01$ , see Thibaut & Witt, 2017, for details) and that close generalization stimuli were semantically more similar to the two learning stimuli than far generalization stimuli were (see Thibaut & Witt for details  $p < .01$ ). The same is true for perceptual similarity ratings which also revealed that the perceptual choices were more perceptually similar to the learning material than the objects used to instantiate taxonomic choices, ( $p < .01$ ). For example, for the accessories category, a jewel pendant (near generalization object) or a bow tie (distant generalization object)



**Figure 1:** Example of a stimulus set and instructions adapted for the fourteen experimental conditions resulting from crossing Learning distance (Close vs. Far comparison) and Generalization distance (Near vs. Distant generalization) factors.

We forged 14 different bisyllabic labels (pseudo-words) which are, as shown by Gathercole and Baddeley (1993), easier to remember than monosyllabic pseudo-words (e.g., buxi, dajo, zatu, xanto, vira). Syllables were of the CV type which is the dominant word structure in French (from Lexique.org, New, Pallier, Brysbaert, & Ferrand, 2004).

The pictures were displayed on a Tobii T120 eye-tracker device with a 1024x768 screen resolution. The five pictures

of a trial were displayed simultaneously until the answer was chosen. Between each trial a standard fixation cross was shown for 3 seconds. Each experimental session started with a standard calibration phase, after the three warm-up trials. The experiment was run with E-prime®.

The five areas of interest (AOI, L1, L2, Ta, Th or P) of a trial had a size of 500 by 500 pixels regardless of the object size inside the frame. The frame was chosen as the AOI's outline instead of the picture's outline, to standardize the AOI size.

**Procedure** The learning pair was displayed at the top of the screen and the test objects at the bottom. First the experimenter introduced the experiment as a game, using the following instructions. "Hello, we are going to play together, and we are going to play with a bear called Sammy. Look, this is Sammy, he lives far away from here and speaks a different language, we are going to learn his language" Then the children saw all three warm-up sets, with the trial instruction, which were followed by eye-tracking calibration. The experimenter then showed the fourteen trials, with the following instructions: "See Sammy's mummy says this is a buxi. And this is a buxi too. Sammy must find another buxi. Can you show which one is also a buxi, to help Sammy? Can you point to the other buxi?". Children chose one of the three test objects by pointing to it on the screen and the experimenter selected it with the mouse.

The presentation order of the fourteen experimental trials, the learning pair objects' position and the generalization objects' position were randomly assigned by the program (e.g., L1 L2, left right or right left, on the top of the screen; Th Ta P, Ta Th P for generalization objects). The names were assigned randomly to each trial. Participants were supposed to know the items. Indeed, these items were calibrated for knowledge by Thibaut and Witt (2017) and, in their experiment, were used with younger children. In their case, the percentage of unknown items was very low.

**Design:** Five and eight-year-old children were compared. Children were randomly assigned to one of the two experimental conditions (*close* comparison, 55 children or *far* comparison, 54 children). Age was crossed with Learning distance (*close* vs. *far* comparison, between-subject factor) and Generalization distance (*near* vs. *distant*, within-subject factor).

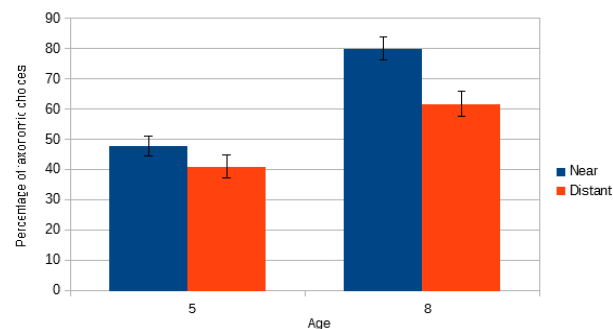
## Results

Our first point of interest was the strategies used by the participants to compare and generalize the novel word to the taxonomic item.

**Performance data** We ran a three-way ANOVA on the percentage of correct taxonomic answers with Age (5, 8 years), Learning distance (*close*, *far*) as a between factor, Generalization distance (*Near*, *Distant*) as a within factor. This ANOVA revealed a significant main effect of Age  $F(1,101) = 29.41, p < .01, \eta_p^2 = .23$  (5-y-o.,  $M = 70.91\%$ ;  $SD = 3.31$ ; 8-y-o.,  $M = 44.37\%$ ;  $SD = 3.61$ ). The main effect of the Generalization Distance was significant,  $F(1,101) = 31.04, p < .01, \eta_p^2 = .24$ . Age and Generalization Distance interacted,  $F(1,102) = 6.61, p < .05, \eta_p^2 = .06$  (Figure 2). A

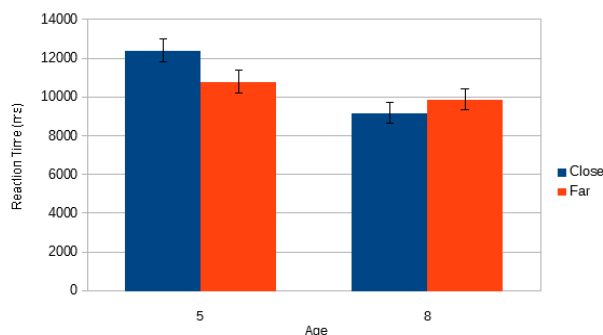
posteriori Tukey analyses showed that both generalization levels did not differ significantly in the younger group ( $p = .18$   $M_{Near} = 45.9\%$   $M_{Distant} = 38.9\%$ ) whereas 8-year olds had better results for near generalization stimuli ( $p < .001$   $M_{Near} = 76.1\%$ ;  $M_{Distant} = 58.6\%$ ). Both age groups answered significantly above chance (5-year olds,  $p < .001$ ; 8-year olds,  $p < .001$ ).

A one sample t-test revealed that the majority of errors were perceptual matches, (5-year olds:  $t = 5.18, p < .001, M_P = 5.7, M_{Th} = 2.42$ ; 8-year olds:  $t = 2.81, p < .01, M_P = 3.12, M_{Th} = 1.49$ )



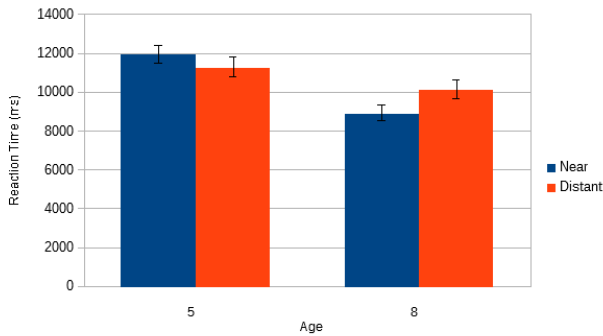
**Figure 2:** Percentage of taxonomic choices as a function of Age (5 and 8 years) and Generalization Distance (Near, Distant). Error bars are SEM.

We ran the same 3-way ANOVA on reaction times for the items that were correctly answered (see Thibaut & French, 2016). This ANOVA revealed the effect of Age,  $F(1,101) = 13.35, p < .01, \eta_p^2 = .12$ , the older children made faster choices ( $M = 9539.83$  ms) than the younger children ( $M = 11618$  ms). Age interacted with Learning distance  $F(1,101) = 4.01, p < .05, \eta_p^2 = .04$  (Figure 3), and with Generalization distance  $F(1,101) = 9.62, p < .01, \eta_p^2 = .09$  (Figure 4). As shown by Figures 3 and 4, the interactions resulted from an opposite pattern in the two age groups, longer RTs in the close and near conditions for the



**Figure 3:** Reaction times (in ms) as a function of Age (5 or 8 years) and Learning Distance (Close or Far). Error bars are SEM.

younger group, and the opposite in the older group. One interpretation of this pattern of results is that 8-year olds had a high level of performance in all conditions, but that the distant generalization condition was more difficult than the near generalization condition. The higher RTs reflect this higher level of difficulty. For younger children,



**Figure 4:** Reaction times (in ms) as a function of Age (5 or 8 years) and Generalization Distance (Near or Distant). Error bars are SEM.

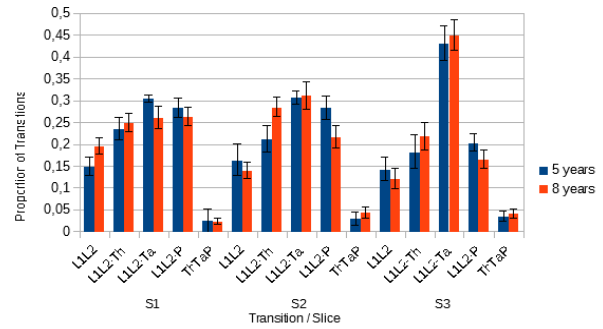
the level of performance was close to chance, and lower RTs might reflect a tendency to answer quickly when the answer was difficult to find, or was less obvious, resulting in shorter RTs

Eight-year-olds had a higher level of performance in both conditions compared to 5-year-olds. Younger children's RTs are lower than the 8 year-old's RTs. However, the younger group does not significantly differ from chance. Chance performance might reflect a tendency to answer too quickly whereas the 8-year-olds RTs are likely to reflect the time necessary for the children to perform a more systematic analysis of a trial before giving an answer.

#### Eye tracking analyses on transitions (saccades)

The design of the analysis is complex. Since we focus on the temporal dynamic of the search for a solution, interactions involving transitions and time slices are central. A transition (or switch) was defined as a saccade between two stimuli. Each trial was decomposed into 3 time slices (S1-beginning, S2-middle, S3-end) of equal size. We ran a five-way analysis of variance (ANOVA) on the proportions of transitions for *correct* answers, with Age (5 and 8 years), Learning Distance (*Close*, *Far*) as between factors, Generalization distance (*Near*, *Distant*), slice (S1, S2, S3), and Transition type (L1L2, L1L2-Ta, L1L2-P, L1L2-Th, Ta-P-Th) as within factors. There was a main effect of the Transition type factor,  $F(8,640) = 111, p < .01, \eta_p^2 = .58$ . Transition type and Slice interacted,  $F(8,640) = 22, p < .01, \eta_p^2 = .21$ . The ANOVA revealed two three-way interactions. The most interesting was the interaction between Slice, Transition type, and Age:  $F(8,640) = 2.19, p < .05, \eta_p^2 = .03$  (see Figure 5). Slice, Transition type, Learning distance also interacted:  $F(8,640) = 2.41, p < .05, \eta_p^2 = .03$ , an interaction that we will not analyze here.

Figure 5 shows that all the transition types appeared in the first slice, at the same level, except transitions Th-Ta-P (i.e., between Thematic, Taxonomic and Perceptual generalization items) which are virtually absent in the threeslices. This absence of between-solution transitions is important because it shows that the alignment hypothesis is confirmed (ie. back and forth transitions between learning and generalization items). Second, overall, the general search profile was similar in both age groups. They compared L1 with L2 and each option with L1 and L2 in



**Figure 5:** Proportion of transitions as a function of the Slice (S1, S2 or S3) and the Transition type (L1L2, L1L2-Th, L1L2-Ta, L1L2-P, ThTaP) for correct trials.

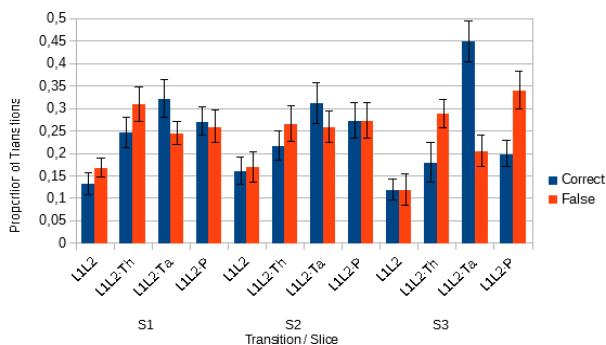
Note: L1L2 are transitions between Learning1 and Learning2; L1L2-Th, between L1 or L2 and Thematic; L1L2-Ta, between L1 or L2 and Taxonomic; L1L2-P, between L1 or L2 and Perceptual; ThTaP, between Th, Ta and P) (Error bars are SEM)

the first slice and then progressively converged on the correct solution. The large proportion of saccades between L1-L2 and each option is a bit unexpected, since we expected more L1-L2 transitions than any other type. However, it might mean either that from the onset of the trial participants actually looked at all the options at the same rate or that participants looked at L1 and L2 first, and very slightly later transitioned between generalization options and L1-L2 during the first time slice.

In order to disentangle these two possibilities, we ran an ANOVA on the fixation times towards the five AOIs in the first time slice, for correct trials. The analysis revealed a significant effect of AOI,  $F(4, 388) = 28.864, p < .0001, \eta_p^2 = .22$ ,  $M = L1 = 27\%$ ,  $L2 = 27\%$ ,  $Th = 15\%$ ,  $Ta = 15.5\%$ ,  $P = 16\%$ . Tukey HSD revealed that L1 and L2 looking times were significantly larger than the other three AOIs. These results show that children gazed more at L1 and L2 than at the other stimuli at the beginning of the trial, but switched to the options quite early in the trial.

**Correct answers and errors profile** A last analysis compared the search profiles for correct answers and errors in the younger group only (5-year olds), because the number of errors was low for 8-year olds. An error was either a thematic or a perceptual choice. Two options are possible. First, errors and correct answers have similar search profiles: errors would be the result of a correct search, but followed by a wrong decision. Second, errors might result from different search patterns, which would differ from the onset of the trial. We ran a five-way ANOVA with Learning distance (*Close*, *Far*) as a between factor, and Accuracy (*Correct*, *False*), Generalization Distance (*Near*, *Distant*), Slice (S1, S2, S3), Transition type (L1L2, L1L2-Th, L1L2-Ta, L1L2-P) as within factors. Time slices were defined as the 1<sup>st</sup>, 2<sup>nd</sup>, and 3<sup>rd</sup> thirds of a trial. We excluded the transition Th-Ta-P from the analysis, because its frequency was close to 0. The ANOVA revealed a main effect of Transition type  $F(6,162) = 19, p < .01, \eta_p^2 = .41$ ; an interaction between Accuracy and Transition type  $F(6,162) = 6.64, p < .01, \eta_p^2 = .19$ ; an interaction between Generalization distance, Slice and Learning distance,

$F(6,162) = 3.65, p < .05, \eta_p^2 = .12$ . The main result was the interaction between Accuracy, Slice and Transition type:  $F(6,162) = 6.70, p < .0001, \eta_p^2 = .19$  (Figure 6). It shows that the main difference between errors and correct answers takes place during the third slice of the trial, participants focusing on the selected option, error or correct. Note that there were two peaks for errors, on Th and P. This can be related to the predominance of perceptual errors, in the 3<sup>rd</sup> slice. This suggests that participants studied both incorrect options but, eventually, went for the most salient one. Another interesting feature of this interaction is that the first two slices of the error trials had a flatter pattern than the



**Figure 5:** Proportion of transitions as a function of Accuracy (Correct, False), Transition type (L1L2, L1L2-Th, L1L2-Ta, L1L2-P) and Slice (S1, S2, S3). Error bars are SEM

correct answers. This might suggest that errors take hold in the 1<sup>st</sup> and 2<sup>nd</sup> slice, that is earlier than at the decisional stage. A priori contrasts between correct answers and errors revealed significantly more L1L2-Th transitions in error patterns rather than in correct trials in slice 1 and significantly more L1L2-Th and L1L2-P for error than for correct trials in slice 3, and significantly more L1L2-Ta in correct than in errors.

### Discussion

First, we assessed which learning and generalization conditions would give the best generalization results as a function of conceptual distance between learning items and between learning items and generalization items. Second, we characterized the temporal dynamics of a solution, as a function of age and learning and generalization conditions with eye tracking movements. Generalization was better for near items than for distant items, with a larger difference in older children. This was confirmed by the higher RTs in the distant generalization case, in the 8-year-old group (younger participants results are difficult to interpret since they were at chance). These results might seem straightforward, at first glance. However, we predicted that the difference between near and distant trials should decrease for older children, because they should have a deeper conceptual understanding in the far learning case. Older children's performance significantly differed from chance in the two generalization conditions whereas younger children were at chance in the four conditions. This pattern of results suggests that some of the younger children encountered difficulties to integrate the information resulting from the comparisons, but does not mean that

younger children answered randomly, as shown by the difference between error-correct gaze patterns.

The other main contribution, the eye-tracking analysis, revealed a consistent pattern of results across ages. All the comparisons were between learning items (L1-L2) and each of the three types of options, with virtually no comparison of the three options (i.e., no Th-Ta-P transitions). In the 1<sup>st</sup> slice, the remaining four transitions were equally distributed. However, the looking times on each of the five AOIs in the 1<sup>st</sup> slice showed that both age groups spent significantly more time on the training items than on the three generalization items, which is consistent with an initial search of commonalities between the learning items before considering the options.

Overall, in terms of the compared solving strategies, the results are consistent with an alignment view rather than with a projection view: participants first compare the learning stimuli, that is align each one with the other. Then comparisons between the learning pair and each option show that participants align commonalities found in L1-L2 with each of the options. A projection interpretation would be compatible with a high number of Th-Ta-P transitions, with participants comparing the three solution options one with the other in terms of the commonalities initially extracted from L1-L2, which occurred very rarely. In a similar way, results are not compatible with a constructive matching strategy. Indeed, as Figure 4 and 5 show, and the discussion above suggests, participants keep on looking at L1-L2 during the entire trial, while testing each option, the latter occurring very early.

There was no interaction between generalization distance and the transitions and slices. The significant interaction between learning distance, transitions and slices, though significant had a small effect size, and seemed to result from small differences, essentially in slice 1. This seems to suggest that generalization distance did not affect the search strategy in a systematic way.

**The anatomy of errors** Do search patterns for correct answers differ from those for errors? Much of the triple interaction between accuracy, slice and transitions seems to be explained by the distribution of taxonomic, thematic and perceptual choices in correct trials and errors in the 3<sup>rd</sup> slice (see Figure 5). This pattern would mostly reflect decisional processes at the end of the trial rather than early differences. However, the flatter profile in slice 1 for errors together with the significant difference between errors and correct trials for thematic answers suggest that errors might be prepared early on. This would be consistent with Thibaut and French (2016) who, in their eye-tracking study of analogical reasoning, showed that children's errors differed from correct answers in significant respects even at the onset of the trials.

In sum, younger children had difficulties across conditions whereas the older group could reliably extract the relation in most conditions, especially in the close generalization cases. The eye-tracking measures revealed similar search patterns in both groups of children, with early transitions between L1 and L2 and L1-L2 towards each solution option. Errors seemed to result from an incorrect decision but seemed to be prepared early on, maybe by a less systematic analysis of the taxonomic choice. A more extensive analysis of AOIs looking times and of the order

of the initial gazes should give us a more refined picture of early search steps. We might also analyze the response evaluation processes, for example with an analysis of the distribution of the backward transitions from the options towards L1-L2 separately. These transitions might reflect evaluation of participants' choices.

### Acknowledgments

The authors wish to thank the Conseil Regional of the Bourgogne Franche-Comté (PARI program) and the FEDER, for their financial support to this project.

### References

- 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>
- 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.
- Bethell-Fox, C. E., Lohman, D. F., & Snow, R. E. (1984). Adaptive reasoning: Componential and eye movement analysis of geometric analogy performance. *Intelligence, 8*(3), 205-238. [https://doi.org/10.1016/0160-2896\(84\)90009-6](https://doi.org/10.1016/0160-2896(84)90009-6)
- 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*(01), 201-224.
- 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., & 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](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.
- Green, A. E., Kraemer, D. J. M., Fugelsang, J. A., Gray, J. R., & Dunbar, K. N. (2010). Connecting Long Distance: Semantic Distance in Analogical Reasoning Modulates Frontopolar Cortex Activity. *Cerebral Cortex, 20*(1), 70-76. <https://doi.org/10.1093/cercor/bhp081>
- 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.
- 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. <https://doi.org/10.1037/0096-3445.131.1.5>
- 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>
- Thibaut, J. P. (1991). Récurrence et variations des attributs dans la formation des concepts. *Unpublished doctoral thesis, University of Liège, Liège.*
- Thibaut, J.-P. (1995). The abstraction of relevant features by children and adults: The case of visual stimuli. *Proceedings of the Seventeenth Annual Conference of the Cognitive Science Society, 17*, 194. Psychology Press.
- 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., French, R., & Vezneva, M. (2010a). Cognitive load and semantic analogies: Searching semantic space. *Psychonomic Bulletin & Review, 17*(4), 569-574. <https://doi.org/10.3758/PBR.17.4.569>
- Thibaut, J.-P., French, R., & Vezneva, M. (2010b). The development of analogy making in children: Cognitive load and executive functions. *Journal of Experimental Child Psychology, 106*(1), 1-19. <https://doi.org/10.1016/j.jecp.2010.01.001>
- 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. *Livre/Conférence Proceedings of the 40th Annual Meeting of the Cognitive Science Society, 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*, 643.
- Waxman, S. R., & Klibanoff, R. S. (2000). The role of comparison in the extension of novel adjectives. *Developmental Psychology, 36*(5), 571-581. <https://doi.org/10.1037/0012-1649.36.5.571>
- Wiebe, S. A., & Karbach, J. (2018). *Executive Function / Development Across the Life Span*. New-York, Routledge.