Cognitive Development 38 (2016) 10–26

Contents lists available at ScienceDirect

Cognitive Development

Analogical reasoning, control and executive functions: A developmental investigation with eye-tracking

Jean-Pierre Thibaut*, Robert M. French

LEAD UMR 5022, Univ. Bourgogne Franche-Comté, Dijon, France

ARTICLE INFO

Article history:
Received 1 December 2014
Received in revised form 6 December 2015
Accepted 18 December 2015
Available online 22 January 2016

Keywords:
Analogy
Analogy-making
Development
Processing constraints
Eye-tracking
Executive functions

ABSTRACT

We use eye-tracking to study the development of analogical reasoning in 5-year-olds, 8-year-olds, adolescents and adults in the A:B:C:D paradigm. We observed significant differences between groups in the way they explored the space of possible answers to analogy problems. Looking times showed that adults first studied the possible relations between A and B and, thereafter, they moved to C and the solution set. Children, by contrast, tended to start with the C item and organized their search around this item. Children's and adults' saccade patterns differed at the beginning and the end of the trial. Children monitored their search less efficiently than adults (fewer saccades from the solution set to the A–B pair at the end of the trial). Looking patterns associated with errors and correct trials also differed from the start of the trial, suggesting that different search strategies lead to different outcomes. Results are contrasted with current models of analogical reasoning and are discussed in terms of the interaction between the development of executive functions and the control and integration of the information pertaining to the analogy problem.

1. Introduction

Analogy making is typically conceived of as a process in which a *base* domain and a *target* domain are compared in order to find relational correspondences between them (e.g., Gentner, 1983; Holyoak, 2012). For example, in the classical analogies of the A:B:C:D type, one has to identify a relation which might unify the A:B and the C:D pairs, i.e. to find in what terms the stimuli in both pairs play the same semantic role. It has often been argued that analogies play a central role in development because they increase children's knowledge in various conceptual domains (e.g., Brown & Kane, 1988; Goswami, 1992; Gentner, 2010).

Understanding analogies requires systematic comparisons between the items that are activated by the analogy problem. In that sense, analogical reasoning involves a search through a space of solutions. The present manuscript reports the first developmental account of the temporal organization of the search for a solution, obtained through eye tracking data, focusing on the way children and adults integrate the components of the analogy problem. We describe how four age groups (5-, 8-, 13-year olds, and adults) integrate the relations between A–B and between C and the Target or distractors that are semantically related to C.

For most analogy models, the core of analogical reasoning is the mapping process that takes place between the base and the target domains. Mapping involves a set of one-to-one correspondences that link a particular item in the base with an

* Corresponding author at: UB, Univ. Bourgogne Franche-Comté, LEAD–CNRS UMR 5022, Pôle AAFE–Esplanade Erasme, 21065 Dijon Cedex, France.
E-mail address: jean-pierre.thibaut@u-bourgogne.fr (J.-P. Thibaut).

http://dx.doi.org/10.1016/j.cogdev.2015.12.002
0885-2014/

item in the target. It also involves candidate inferences, i.e., a search for which relation(s) that hold in the base domain (e.g., A and B) can be applied in the target (e.g., C and D) (e.g., Holyoak, 2012, and the Structure mapping theory, SMT, Gentner 1983).

1.1. Analogies and development

Two general hypotheses have been put forward in order to explain the development of analogical reasoning or, to put it in the above terms, mapping failures or successes. These are the development of the knowledge base and the development of cognitive processes, specifically, executive functions. The development of a *knowledge base* is crucial because knowledge is necessary to find or construct the potential relations for any particular semantic analogy. In order to relate "train and railroad tracks" with "car and road", one must know that trains travel on railroad tracks and that cars travel on roads (Gentner, 1988; Goswami & Brown, 1990). Children are more likely to draw correct analogies in conceptual domains that are the most conceptually accessible to them. For example, Rattermann and Gentner (1998), interpret what Gentner (1988) calls the "relational shift in analogies", i.e., a shift from early attention to featural similarities to later attention to common relational structures, in terms of children's knowledge.

A second line of interpretation, the one adopted in the present contribution, explains the development of analogical reasoning by the progressive improvement of *executive functions* (e.g., Halford, Wilson & Phillips, 1998; Richland, Morrison, & Holyoak, 2006; Thibaut, French, & Vezneva, 2010; Thibaut, French, & Vezneva, 2010; see also Morrison et al., 2004; Viskontas, Morrison, Holyoak, Hummel, & Knowlton, 2004, in aging, for similar views). Executive functions include a set of components such as working memory, flexibility, and inhibition (see Zelazo, Carter, Reznick, & Frye, 1997; Brocki & Bolin, 2004; Senn, Espy & Kaufmann, 2004; for a recent review, Carlson, Zelazo, & Faja, 2013). In this context, inhibition plays a central role in analogy comprehension, especially when salient associations come immediately to mind but are irrelevant to the current analogy. These associations must be inhibited (e.g., in the bird: nest: dog:? (doghouse) case, a strongly associated term to be inhibited would be "bone", see Richland et al., 2006; Thibaut, French et al., 2010; for children and Morrison et al., 2004, in aging people). Cognitive flexibility is required in order to find new relations that make sense in the context of the target analogy, especially when the relations that first come to mind are irrelevant (see Glady, Thibaut, French, & Blaye, 2012).

To illustrate, Richland et al. (2006 Fig. 1, p. 255) used scene analogy problems consisting of pairs of scenes illustrating relations among objects. The authors manipulated a featural distractor in the second scene of each problem. For example, if the base scene included a running cat as part of the relation (i.e., dog *chases* cat), they added a distractor object to the target scene (i.e., an object that was not part of the *chase* relation) that was either perceptually similar (a sitting cat) or dissimilar (a sandbox) to the object in the *chase* relation in the base scene. Not surprisingly, results revealed that stimuli with similar distractors elicited more errors than the stimuli with dissimilar distractors. The authors also showed that the number of objects or participants involved in a relation (2 or 3) had a significant effect on performance. The executive function view posits that analogy comprehension involves numerous successive comparisons of the available information, especially for difficult analogies, i.e., when the solution is not obvious or/and when many related distractors are present (e.g., Holyoak, 2012; Morrison et al., 2004; Thibaut, French et al., 2010; Bugaiska & Thibaut, 2015).

1.2. The temporal dynamics of analogical reasoning and its development

Analogical reasoning requires integration of multiple sources of information and various comparisons within and between the item pairs making up the problem. The question of the temporal dynamics of the task is an open issue in the literature. In a developmental perspective, we will use eye-tracking to identify significant differences between adults' and children' search patterns, which cannot be done with the standard static performance measures, such as percentage of correct performance. In the eye tracking literature it is well-known that looking times are highly correlated with the independently assessed informativeness of regions within a scene (e.g., Rayner, Shen, Bai & Yan, 2009).

As far as we are aware, apart from a short article by our group, Thibaut, French, Missault, Gérard, and Glady (2011), the present article reports the first developmental study of the dynamics of analogy making with semantic analogies, comparing various age groups. The eye-tracking literature on analogy is restricted to adults (see Salvucci & Anderson, 2001, for verbal-written analogy problems or Gordon & Moser, 2007; who studied looking times and saccades for scenes from Richland et al., 2006).

If analogy making is characterized by comparisons between and within pairs of items, how are these comparisons organized and how do they develop over time? Thibaut, French et al. (2010) characterized analogy making as a search in a semantic space. The space is itself dynamically constructed during the comparisons. In what follows, we contrast predictions of the two general models of development described above.

The knowledge view of the development of analogical reasoning makes no specific claims regarding the temporal organization of the search. In the present experiment, we ensured that the children had the knowledge necessary to solve the analogy. The executive function view makes the general prediction that younger children will have more difficulties solving analogy problems than adults because of their poorer inhibition capacities. However, this says nothing about the time course of the comparison processes that led to the error (or a correct answer). One crucial question is whether children focus more on irrelevant information than adults when solving a problem (Ratterman & Gentner, 1998; Richland et al., 2006; Thibaut, French et al., 2010). The status of irrelevant information will also be the focus of a comparison between correct trials and errors.

Numerous computational models of analogical reasoning have been proposed and the question of the predictions they make regarding the temporal organization of the search, if any is raised. In this respect, we will compare these predictions to our developmental data. We consider three types of proposals derived from computational models (see French, 2002; Gentner & Forbus, 2011 for reviews), namely the "alignment-first", "projection-first", and "relational-priming" models.

The "Alignment-first" conception (e.g. SME or ACME) is derived from the structural alignment hypothesis (Markman & Gentner, 1993) according to which the items that compose the base and the target are aligned first. Later, inferences are projected from the base pair to the target pair. In the A:B:C:? D paradigm, this means that one would first align A with C and would then look for a solution, D (or Ds) that is aligned with B.

By contrast, "Projection-first" models (e.g., LISA, Hummel & Holyoak, 1997; DORA, Doumas, Hummel, & Sandhofer, 2008) begin by projecting information from the base (i.e., the A–B pair in our case) to the target (i.e., the C-Target pair in our case). This predicts more attention to the A and B pair and more early A–B saccades, followed by more attention to C and Target and C-Target saccades. This contrasts with the alignment-first case which predicts more A–C and B–D saccades.

Consistent with the projection-first view, Gordon and Moser (2007) found that *adults* initially focused on what they called "actor-patient" pairs (e.g., a dog – the actor – chasing a cat – the patient –, which is analogous to our A and B terms) in the source image and then looked for the solution in the target image (a second actor-patient pair, e.g., a girl chasing a boy) (analogous to C and D terms in the A:B:C:D format). The authors also studied saccades involving the distractors (i.e., saccades between the actor and the perceptually similar distractor in the target pair) and showed that these saccades also occurred after saccades towards relational matches, suggesting that distractors are not systematically processed before relational matches. However, this study was limited to *adults* and provided information only for item-to-item saccades between the source scene and the target scene for a limited number of pairs. In our developmental context, we collected more extensive data about the temporal organization of the comparisons between A–C, the Target and the semantic distractors. It might be that participants apply a mixture of alignment and projection (see Copycat, Mitchell & Hofstadter, 1990; or Tabletop, French, 1995). Participants might first analyze the A–B pair then apply the relation found for A–B to C–D. Later they might align A with C and B with D as a means of verifying their answers.

By contrast with the above models, a third view, the "relational priming" model proposed by Leech, Mareschal, and Cooper (2008), gives no role to mapping. According to this view, children first study the A–B pair. Then, the relation found between A and B (authors appeal to the retrieval of a relevant transformation between A and B, e.g., "cuts" for knife and bread) *automatically primes* the retrieval of a relationship between C and the target item into which it is transformed. Interestingly, this model is a developmental account of children's data obtained in the A:B:C:D paradigm (e.g., Goswami & Brown's data, 1990) which will be used here. With respect to the time course of analogy comprehension, this model predicts that participants should study the A–B pair first, turning, later, to C and the solution set, (like the projection-first model). However, in contrast with the projection-first model, children should not actively compare C with the semantic distractors or the semantic distractor with the Target because the C-Target pair would be *automatically primed* by the A–B pair relation. The solution is found through *priming* without systematic comparisons within and between the pairs, thus no mapping. Gordon and Moser (2007) data do not support the relational priming view. They found actor-patient and actor-distractor saccades in the target picture during the second half of the trial. They also found actor-distractor and actor-patient saccades in both the base and target domains. This suggests that there are comparisons *outside* the target actor-patient pair. However, their data were obtained with *adults*, whereas Leech et al. (2008) have a developmental perspective that we will confront to our data. In short, we will compare our data to the predictions of the "align-first", "project-first", and "relational priming" views.

1.3. Mapping and the role of distractors in analogical reasoning

From a developmental perspective, another important issue for most models of analogical reasoning is how children and adults deal with distractors that are semantically related to C or that share perceptual features with C. In most developmental models, it is assumed that younger children are, first, attracted by perceptually or semantically similar distractors. The knowledge view (see Gentner, 1988) predicts that increasing knowledge in the domain of the analogy will give rise to more relational choices.

The executive function view of development predicts that highly associated semantic distractors will either prevent or interfere with the discovery of the analogical solution (see Morrison et al., 2004; Richland et al., 2006; Thibaut et al., 2010). Studies show that most children' errors involve selecting one of these conceptually related matches. The executive function view predicts a large number of C-Distractor saccades in the younger groups, because of their difficulty in inhibiting irrelevant but salient information (Brocki & Bolin, 2004; Senn et al., 2004).

Both the knowledge view and the executive-function accounts predict that adults will focus less on distractors, especially, once they have discovered a reasonable A–B relation. However, this broad description does not answer the "when" question or speak to the difference between correct answers and errors.

In the present executive control context, the outcome of a trial (i.e., correct or incorrect), depends on how distractors are processed. In this context, we will compare the gaze-profiles associated with error trials and correct trials. They should differ in terms of the allocation of attention towards the distractor and the target. One crucial question is whether these differences occur early or, rather, do they appear late, at the decisional stage, with similar fixation and saccade patterns prior to that? More early attention paid to the distractors and more C-distractor saccades in the case of errors would mean that errors might result from a search strategy that differs from the one followed for the correct answers.

Differences between correct trials and error trials might also have to do with the amount of "monitoring" of the chosen solution. Indeed, participants need to check whether their chosen solution-that-goes-with-C is consistent with the relation holding between A and B. This monitoring should come in the form of saccades from the solution set back to the A–B pair. If such saccades exist, they might appear at the end of the trial. If children monitor their choice less efficiently than adults (see Zelazo et al., 2014) they will have fewer saccades from C, Target and the Semantic Distractor (SemDis) to A and B than adults.

1.4. Exploring the temporal dynamics of analogy making

To summarize, our main purpose is to study how children and adults integrate information in the classical A:B:C:?(D) paradigm This paradigm puts emphasis on the processes leading to the construction of a relation that makes sense of the entire set of stimuli. We will study four related main questions. First, we will study the development of the temporal organization of the search, comparing the fixation times for the stimuli defining a trial. This will tell us whether children and adults organize their search pattern in the same way. It might be that adults follow a more systematic search strategy than children. Second, we will look at the gaze-switches (transitions) between these stimuli across age groups. This will tell us which comparisons were the most important for reaching a solution. Third, we will examine the development of information monitoring over the course of the trial, specifically looking at how the solution set and the A–B pair are compared during the trial by children and adults. And, fourth, we will compare the search profiles for correct answers and errors to look at how relevant and irrelevant stimuli are processed in these two types of trials.

2. Methods

The present study used a 4 (Age: 5, 8, 13, adults) × 2 (Number of distractors: 1 semantic distractor, no semantic distractor) × 2 (Association strength: Weak or strong) mixed factorial design with Age as a between-participants factor, and the Number of distractors and Association strength as within-participants factors.

2.1. Participants

A total of 117 children and adults took part in this experiment. Thirty-eight 5-year-old children (*M* = 5;7, age range, 4;9–6;2), thirty-seven 8-year-old children (*M* = 8;8, range 8;0–8;10), 22 adolescents (*M* = 13;9, range 13;3–14;3) and 20 adults (*M* = 21;7, range 19;4–25;4) participated in the experiment. We tested a larger number of children participants because we expected a greater loss of eye-tracking data for the younger groups (see analyses below). Informed consent was obtained from their parents.

2.2. Materials

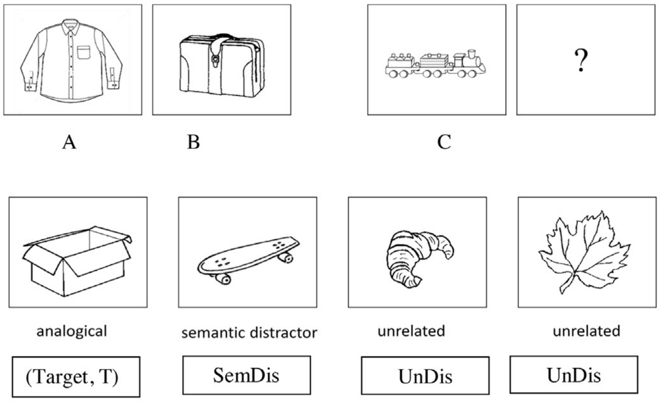
The experiment consisted of a total of 14 trials divided into 2 practice trials and 12 experimental trials (see Fig. 1B for the list of analogy problems). Each of the two distractor conditions (No or 1 semantic distractor) consisted of 6 trials (3 weak and 3 strong, see hereafter). Each trial contained 7 drawings (see Fig 1A), namely the items corresponding to the A–C items and the 4 drawings that were shown as the solution set, including the analogical match (hereafter, "Target"). In the one-semantic-distractor case, there was one semantically related distractor (hereafter "Semantic distractor", SemDis), and two distractors that were semantically unrelated to C (hereafter, "Unrelated distractors", UnDis). In the no-semantic-distractor condition the three distractors were semantically unrelated to C.

We also controlled for the association strength between A and B, between C and D (and between C and the semantic distractor in the one-semantic-distractor case). Thibaut, French et al. (2010) showed that analogies built around weakly semantically associated pairs were more difficult than analogies built around strongly associated pairs. In the present experiment, we used an equivalent number of "weak" and "strong" analogies, that is 3 weak and 3 strong analogies in both the No semantic distractor case and 1 semantic distractor case (see Thibaut, French et al., 2010; for a description of how the two types of analogies were constructed). In Fig. 1, the weak analogies are in the first half of the table, and the strong analogies in the second half. The experiment was run with E-prime® software. The experiment was run with a Tobii T120.

2.3. Procedure

The experimenter saw the children individually at their school in a quiet room or, for the adults, in our laboratory. Participants were seated in front of the Tobii screen with their eyes at a distance of approximately 40 cm. For each participant, the experiment started with a calibration phase which followed the protocol specified for the apparatus.

The participants were then shown picture cards of each of the items used in the experiment and were asked to give their names. When they did not know an item's name, they were asked to describe it functionally. The children knew, on average, A.



B.

1 Distractor				No distractor		
Analogy problem (A:B::C)	Target (T)	Distractors: Semantic (SemDis)	Distractors: Unrelated (Undis)	Analogy problem (A:B::C)	Target (T)	Distractors: Unrelated(UnDis)
shirt : suitcase :: toy car	box	fuel pump	tree, monkey	man : nose :: moose	muzzle	pen, ball, bell
child : bed :: cat	pillow	whiskers	guitar, strawberry	glass : dresser :: ring	case	rope, bowling pin, scooter
pig : trough :: man	dish	watch	plane, tractor	pineapple:bottle :: orange	pitcher	feather, bench, envelope
train : rail :: boat	sea	crab	leave, banana	bird : nest :: dog	doghouse	chair, glasses, saucepan
mitten : hand :: shoe	foot	footprint	trumpet, croissant	spider : web :: bee	beehive	hat, ball, kayak
lamp : outlet :: remote control	battery	stereo system	bucket, boot	door lock : key :: bottle	bottle opener	toy train, mouse, cap

Fig. 1. (A) One trial from the weak association-1-distractor condition. The first line of drawings displays stimuli A–C, and the second line displays the Target (analogical solution), the semantic distractor (SemDis) and two semantically unrelated distractors (UnDis). (B) The set of analogy problems used in the experiment. The left part of Table gives the 1-distractor trials, the right part lists the No-distractor trials. The first three lines are weak analogies and the last three are strong analogies.

Table 1
Percentage correct (standard deviation) per age group, and per condition, Association strength (weak or strong) and number of distractors (0 or 1).

Association strength	Weak	Weak	Strong	Strong
Number of distractors	0	1	0	1
5-year-old	87 (22)	49 (40)	98 (4)	65 (31)
8-year-old	94 (13)	46 (39)	100 (0)	64 (24)
13-year-old	100 (0)	86 (24)	100 (0)	82 (20)
Adults	100 (0)	97 (10)	100 (0)	95 (12)

96% of the names and when they did not, in most of the cases, they were able to give a description showing that they knew the stimulus. Overall, the percentage of cases in which children could not name the item or could not provide a correct description (functional or contextual) was 1%. In these cases, the experimenter gave the children the missing information and the corresponding data were kept in the data set.

Each trial began when the experimenter pressed the space-bar. The 7 stimuli for each trial were displayed simultaneously on the screen. The A:B pair and the C item were shown in an array with the first two items grouped together to the left of the screen. The C item was alone on the right of the screen and next to C there was a box with a question mark. The four solution items were displayed on a separate row, beneath the A–C? row (see Fig. 1A). Participants were asked to point to the item in the lower row that went with C in the same way as A went with B. The first two trials were training trials and participants received feedback. In these two training trials, for children, the experimenter explained why the Target was the correct solution and incorporated the relation holding in the A–B pair in his/her demonstration and carefully checked that the child understood the solution. In the experimental trials, the experimenter gave no information to children. Adults were assumed to fully understand the task. For each experimental trial, the reaction times were recorded by the experimenter who started timing at the beginning of each trial. Participants were instructed that they were to point to the stimulus on the screen corresponding to their choice "as soon as they had found the solution". They were told that they were to point to only one stimulus per trial. The experimenter stopped timing the participant when he/she pointed to a solution.

Afterwards, for all the experimental trials, children's understanding of the semantic relation between A and B and between C and D was assessed. They were shown the A:B pairs and were asked *why* the two items of each pair went together. The same was true for the C–D pairs (see Thibaut, French et al., 2010 for more details). Adolescents and adults did not perform this phase because their performance was at ceiling in the analogy task.

3. Results

First we conducted analyses on the performance data (correct-response percentages). Second, we analyzed item looking times and item-to-item saccades.

3.1. Performance data

Performance was measured as the percentage of valid relational (i.e., most obvious relational) matches. We eliminated all trials in which either the children could not name at least one item or did not understand the semantic relation between the A and B items or between the C and D items, in order to avoid cases in which failure was due to a lack of the relevant knowledge (see Thibaut et al., 2010). These misidentifications of stimuli or of relations were rare, and appeared in the younger group only. We discarded 4% of the strong trials and 6% of the weak trials in this latter group.

We performed a three-way mixed ANOVA on the data with Age (5, 8, 13, adults) as a between-participants factor and Number-of-Distractors (0 or 1 semantic distractor) and Association Strength (Weak or Strong) as within-participants factors. There were main effects of Age $F(3, 112) = 27.32$, $p < .0001$, $\eta_p{}^2 = .42$, and of Number-of-distractors, $F(3, 112) = 132{,}68$, $p < .001$, $\eta_p{}^2 = .54$. The two most interesting results were interactions involving age, first, a significant interaction between Age and Association Strength, $F(3, 112) = 4{,}36$, $p = 0{,}01$, $\eta_p{}^2 = .10$ with near perfect performance in both conditions for the adult and adolescent groups (91% correct or more). Both groups of children had around 69% correct trials in the weak condition and 81% in the strong condition (see Thibaut, French et al., 2010 for similar results in 5-year-olds).

There was also an interaction between Number-of-Distractors and Age, $F(3, 112) = 16.32$, $p < .0001$, $\eta_p{}^2 = .30$. There was virtually perfect performance in both conditions (0 and 1 distractor) in adolescents and adults, whereas both groups of children were at 92% correct in the no-distractor condition and 58% correct in the 1-distractor condition (see Table 1). These results confirm previous studies showing that the presence of distractors decreased performance for children (e.g., Richland et al., 2006; Thibaut, French et al., 2010).

3.2. Eyetracking data: defining a criterion for inclusion

We first had to define a criterion for data inclusion in the analysis. Indeed, there were cases in which participants looked away from the screen during a trial or in which participants, especially 5-year-old, seemed to be looking correctly at the screen but their gazes were not recorded for various reasons (reflections on glasses, body movements, suboptimal orientation of the Tobii screen, see Duchowski, 2007; for a discussion of these typical difficulties). A priori, our criterion was to discard a trial from the data set if more than 50% of looking-time data for that trial were absent. Moreover, since the analyses were performed on correct trials, we removed around 30% of the data in the two younger groups, corresponding to their errors (see above). Overall, in the main analysis performed below, we lost no adult, no adolescent, 8 eight-year-olds and 14 five-year-olds (the decrease in children was expected, which is why we had increased the size of the children groups).

3.3. Eye-tracking data: looking times and number of saccades

We performed separate analyses on both the percentage of looking time for each item and the number of item-to-item saccades in the case of *correct* trials. Looking times for areas of interests (AOIs, the items themselves) tell which items are attended to while solving the problem. However, looking times say nothing about comparisons between items. This information is provided by saccades between items, e.g., when a participant goes from item A to item C (see Duchowski, 2007). The two measures allow an item-based and a relation-based analysis of the trial.

3.3.1. Looking times

The analyses considered the looking times at the AOIs corresponding to the locations of the six types of items: A–C, Target (T), the Semantic Distractor (SemDis), and the two Unrelated Distractors (UnDis) items. The two unrelated distractors were equivalent in experimental terms and were averaged into a single score. An item location was defined by a square 240 pixels on a side that corresponded to the area occupied by an item in the trial.

3.3.2. Time course of looking times during the trial

We compared the mean proportion of time spent on each stimulus (i.e., A–C, T, SemDis and UnDis) for each age group (5-, 8-, 13-year olds and Adults). Another central goal was to devise a means of capturing the time dynamics of the trials. We, therefore, divided each trial into 3 equal time slices corresponding to the beginning, middle and end of a trial.

The main purpose of this analysis was to compare the four age groups in terms of the proportion of search time they allocated to each type of stimulus as a function of time slice. Longer gaze times at the beginning of the trial in favor of A and B at the expense of C and the solution set would support the Projection-first model rather than the Alignment-first model, as the latter predicts early alignment of A with C, and B with a potential Target candidate (see above). The relational priming view (Leech et al., 2008) predicts a majority of gazes towards A and B at the beginning of trial and a majority of gazes towards C and target later, with no or very few gazes at the related distractor (since analogy operates through priming), which distinguishes it from the Projection-first view.

First, we restricted this first analysis to the 1-distractor trials because we wanted to study the influence of the semantically related distractor compared to the Target across age groups and because the respective influence of these two types of stimuli is a major issue for models of analogy. Also, we wanted to characterize the search patterns obtained for errors and correct answers. However, errors occurred almost exclusively when a semantic distractor was present, as shown by Table 1 (except for the five-year-old group in the weak condition, 87% correct), which did not allow us to analyze the error pattern in the strong condition. Indeed, in the 0-distractor case, there is no ambiguity regarding the relational solution.

Second, we will present two analyses of the data. The first analysis included the Association Strength factor, the second one, on which our discussion of the results will be based, did not include this factor. In the materials, we controlled for the association strength within the pairs composing the analogies (weakly and strongly associated pairs) for the sake of representativeness, because Thibaut, French et al. (2010) showed that children's performance was influenced by this factor. However, our main purpose in the present study was to compare the time course of children' and adults' search across the stimuli, which explains our focus on the stimulus types over time. We did this in order to compare these gaze profiles to the predictions of existing models of analogy (see above). In this theoretical context, we had no specific hypotheses regarding association strength or interactions involving this factor. As a result, we will first report the omnibus ANOVA including Association strength, followed by the ANOVA with this factor collapsed. Collapsing this factor also contributed to decrease the problem of missing data cells in repeated-measures ANOVAs. This problem can be severe in eye-tracking analyses, especially with children, since if a child looks away, even for a moment, during the trial, the result is usually that all of the data from that child must be eliminated entirely from the analysis. (We chose not to use data imputation techniques to artificially fill in missing values in our analyses.) Thus, in our case, there were three within-participant variables, one with 8 levels (Items), another with 2 levels (Association Strength), and another with 3 levels (Time Slice). If there was a single missing data cell for any of these 48 levels in all the trials defining one level of the Association Strength factor, the participant was not able to be included in the repeated-measures analysis, which resulted in a serious loss of power in our analysis. Therefore, we collapsed the data over Association Strength in order to increase statistical power.

Thus, we first conducted a four-way mixed ANOVA on the proportions of looking times with Age (5-, 8-, 13-year olds, and adults) as a between factor, Items (A–C, Target, Semantic Distractor–SemDis, Unrelated distractor–UnDis), Association strength (Weak vs Strong) and Slice (1st, 2nd, 3rd) as within factors. It revealed a significant effect of Items (5, 275) = 93.52, $p < .0001$, $\eta_p^2 = .63$, three significant two-way interactions, Items × Age, $F(15, 275) = 3.04$, $p < 0.001$, $\eta_p^2 = .14$, Slice × Items, $F(10, 550) = 49.54$, $p < .0001$, $\eta_p^2 = .47$; Slice × Age, $F(6, 110) = 2.26$, $p < .05$, $\eta_p^2 = .11$. There was also a three way interaction Slice × Association Strength × Age, $F(6, 110) = 2.26$, $p < .05$, $\eta_p^2 = .11$. However, as mentioned above, the most important result, was the three way interaction, Slice × Items × Age, $F(30, 550) = 3.44$, $p < .0001$, $\eta_p^2 = .16$. No other main effect or interaction was significant, Age, $F(3, 55) = 2.26$, $p = .09$, $\eta_p^2 = .04$, Association Strength, $F(1, 55) = 2.96$, $p = .09$, $\eta_p^2 = .05$, Slice $F(2, 110) = 2.96$, $p = .09$, $\eta_p^2 = .05$. The same was true for the following interactions, Age × Association strength, $F(3, 55) = 2.26$, $p = .09$, $\eta_p^2 = .11$, Slice × Association strength, $F(2, 110) = 2.96$, $p = .06$, $\eta_p^2 = .05$, Items × Association strength, $F(5, 275) = 1.64$, $p = .15$, $\eta_p^2 = .03$, Items × Age × Association strength, $F(15, 275) = 0.56$, $p = .9$, $\eta_p^2 = .03$. Slice × Association strength × Items, $F(10, 550) = 1.16$, $p = .31$, $\eta_p^2 = .02$, and Slice × Association strength × Items × Age, $F(30, 550) = 1.11$, $p = .31$, $\eta_p^2 = .06$.

Second, we conducted a three-way mixed ANOVA with Age as a between factor, Items and Slice as within factors in which, as mentioned above, the "Association Strength" factor was collapsed. The ANOVA revealed a main effect of Items, $F(5, 285) = 129.1$, $p < .0001$, $\eta_p^2 = .59$ and two significant two-way interactions, Items × Age, $F(15, 460) = 6.91$, $p < .0001$, $\eta_p^2 = .18$, Slice × Items, $F(10, 920) = 71.11$, $p < .00001$, $\eta_p^2 = .44$. There were subsumed by the three-way interaction between Age, Slice, and Items, $F(30, 920) = 3.91$, $p < .0001$, $\eta_p^2 = .11$ (Fig. 2), the most interesting result. All the other effects were non-significant, Age, $F(3, 92) = .95$, $p = .42$, $\eta_p^2 = .029$. Slice, $F(2, 184) = .90$, $p = .41$, $\eta_p^2 = .009$, and the Slice × Age interaction, $F(6, 184) = .95$, $p = 0.46$, $\eta_p^2 = .029$.

The significant triple interaction between Age, Slice, and Items, which is the main focus of the present analysis, shows a clear pattern of development with the most important difference between the two age groups located in the *first* time slice. A contrast analysis revealed that the most interesting differences between the four age groups were located in the first slice. In this *first* slice, -5- and 8-year olds paid significantly less attention to A and B and significantly more attention to C, Target and SemDis than Adults (all contrast analyses, $p < .05$), more attention to C and Target than adolescents. As shown by Fig. 2, for 5- and 8-year olds, the highest proportion peak was on C; the attention to A–C and SemDis was lower. However, in the first slice, Adults and Adolescents focused on A and B, and paid significantly less attention to the other stimuli. The second and the third time slices had the same peak centered on Target, becoming sharper during the third slice (no significant difference between the four age groups, contrast, $p > 0.05$). In the third slice, 5- and 8-year olds focused significantly more on C and significantly less on B than adults (see Fig. 2) (see Tables 1 and 2 for the list of significant contrasts in Supplementary materials). We will interpret this result at the light of the saccade pattern presented hereafter.

These results demonstrate that the temporal search organization of young children's search differs from adolescents' and adults' search (see Thibaut et al., 2011; Glady, Thibaut & French, 2013). The two major findings are, first, that adults and adolescents, in the first time slice, examine the A–B pair at the expense of the other stimuli.

This is consistent with the projection-first hypothesis and with Gordon and Moser's data (2007, see above), and this is not consistent with the alignment-first view. Second, compared to the older groups, 5- and 8-year olds start a trial focusing on C and distribute their attention evenly to the other stimuli. The two younger groups paid attention to C across all three time slices and more than adults in the first and third slices, suggesting that C plays a pivotal role for young children. This preeminence of C fits well with an executive function view, particularly the idea that children have difficulties inhibiting the main, explicit, purpose of the task, i.e.,"to look for the item that goes with C" (see General Discussion). Adolescents' patterns were close to adults' patterns, often between adults' patterns and the two younger groups' patterns (see first slice, Fig. 2).

3.3.3. Saccades: time course of comparisons during the entire trial

Like Gordon and Moser (2007), we also analyzed the *saccade frequencies* between items which occurred during the entire trial. Together with previous authors, Gordon and Moser argued that a first-order Markov model was the best model. This means that the probability of fixating an object (e.g., N) was mostly influenced by the object fixated just before (i.e., N-1)(see Josephson & Holmes, 2002) and much less by objects that were fixated before (i.e., N-2, N-3, etc.). We analyzed the entire trial as in the analysis performed on gazes (see 1). Thus, we divided each trial into three time slices and compared the four age groups. The entire set of first-order saccades can be represented as a 7 × 7 matrix of zones of interest corresponding to the 7 individual items, i.e. 49 possibilities. To reduce this number, we grouped symmetrical saccades together (e.g., A–B and B–A are considered to be equivalent, see Gordon & Moser, 2007). We also grouped the two unrelated distractors and averaged saccades from each stimulus to either of these two unrelated distractors.

The analysis was done on the 8 most relevant saccade types, A–B, A–C, B–C, B–T, C–T, C–SemDis, T-SemDis and C-UnDis. Together with the Slice factor, they depict the time course of comparisons involving A–C and Target on the one hand, and saccades involving C and the solution set items on the other hand. As in the previous analysis, we were interested by the interaction between Age, Slice and Saccade type, which revealed which stimuli were compared by participants during the trial. Again, even though the Association strength factor was controlled for, it generated no specific prediction. Thus, we started with the omnibus ANOVA, including Association strength. Then, as in the analysis above, we analyzed our data with the two levels of this factor collapsed.

We started the analysis with a four-way mixed ANOVA with Age (5-, 8-, 13-year olds and Adults) as a between factor, Slice (First, Second, Third), Association strength (Strong, weak) and Saccade type (A–B, A–C, B–C, B–T, C–T, C-SemDis, C-UnDis, and T-SemDis) as within factors. We are interested in interactions involving Age and Saccade type.

This Anova revealed a main effect of Age, $F(3, 55) = 11.41$, $p < .0001$, $\eta_p^2 = .38$, of Slice, $F(2, 110) = 11.15$, $p < .0001$, $\eta_p^2 = .17$, and Saccade Type, $F(7, 385) = 80.77$, $p < .0001$, $\eta_p^2 = .59$, a significant interaction between Saccade Type and Age, $F(21, 385) = 3.45$, $p < .0001$, $\eta_p^2 = .16$, between Slice and Saccade Type, $F(14, 770) = 24.20$, $p < .0001$, $\eta_p^2 = .31$, Slice, Association Strength and Age, $F(6, 110) = 2.91$, $p < .05$, $\eta_p^2 = .15$, Saccade Type × Association Strength, $F(7, 385) = 3.30$, $p < .005$, $\eta_p^2 = .13$, and the most interesting result, an interaction between Age, Slice and Saccade Type. $F(42, 770) = 3.14$, $p < .0001$, $\eta_p^2 = .15$,

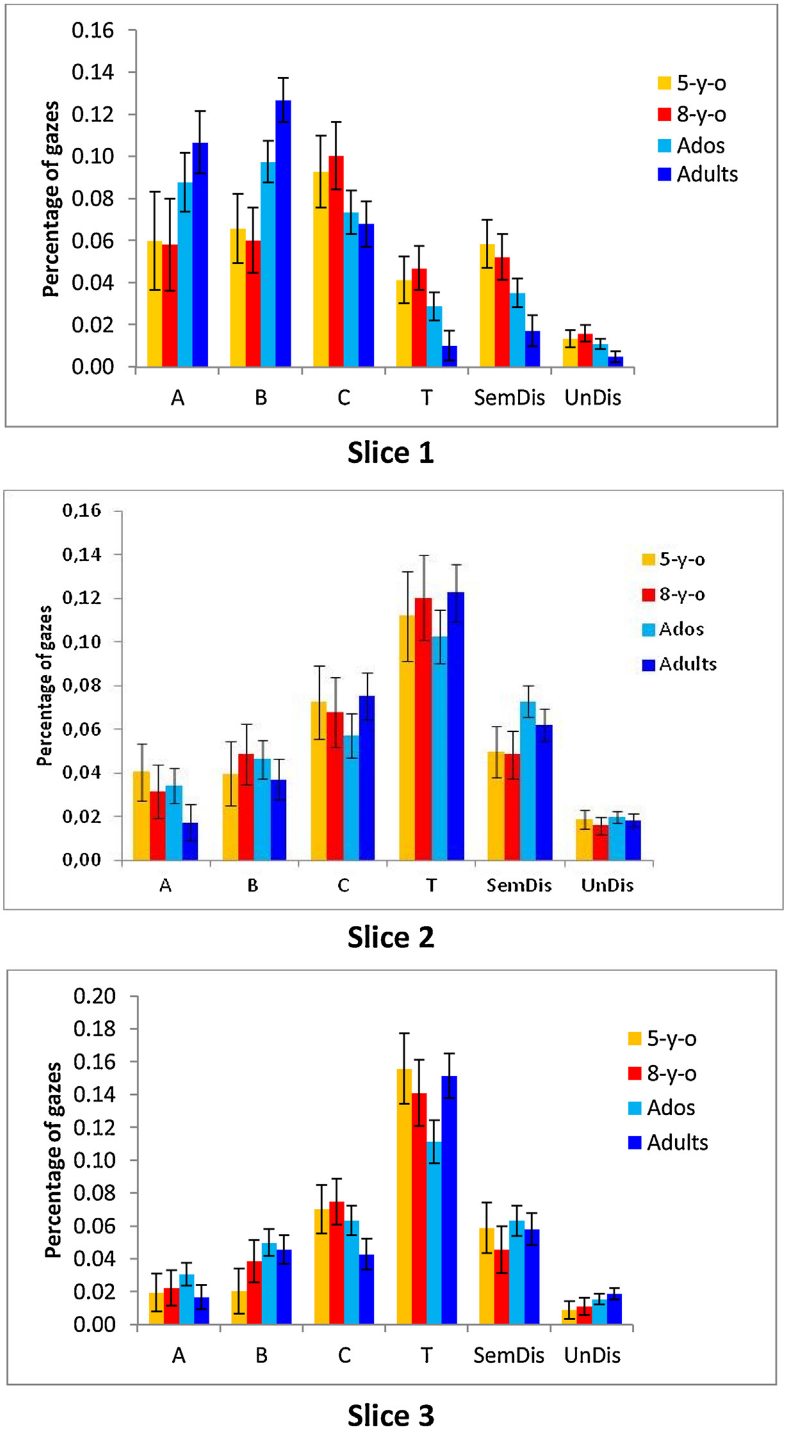


Fig. 2. Percentage of gazes as a function of Age, Slice, and Items. The analysis was performed on the entire trial time (One standard deviation error bars were used.).

All the other main effects and interactions were not significant. The main effect of Association Strength, $F(1, 55) = .04$, $p = .85$, $\eta_p^2 < .001$, the interaction between Age and Slice, $F(6, 110) = .72$, $p = .64$, $\eta_p^2 = 04$, Association strength and Age, $F(3, 55) = 2.46$, $p = .07$, $\eta_p^2 = .12$, or Association strength and Slice, $F(2, 110) = 2.22$, $p = .11$, $\eta_p^2 = .04$ or between Slice × Age and Association Strength, $F(21, 385) = 1.32$, $p = .16$, $\eta_p^2 = .07$, Slice, Association Strength, and Saccade Type, $F(14, 770) = 1.05$, $p = .40$, $\eta_p^2 = .02$ and between the four factors, $F(42, 770) = 1.37$, $p = .07$, $\eta_p^2 = .07$.

Then, we conducted a three-way mixed ANOVA in which Association strength was removed. This ANOVA gave significant main effects of Age, $F(3, 91) = 4.34$, $p < .01$, $\eta_p^2 = .13$, of Slice, $F(2, 182) = 14.28$, $p < .0001$, $\eta_p^2 = .14$, and saccade type, $F(7, 399) = 70.1$, $p < .00001$, $\eta_p^2 = .44$. There were also two two-way interactions, Saccade type × Age, $F(21, 637) = 5.23$, $p < .0001$, $\eta_p^2 = .15$, and Slice × Saccade type, $F(14, 1274) = 31.17$, $p < .00001$, $\eta_p^2 = .26$, which were subsumed by the triple interaction between Slice, Saccade type and Age, $F(14, 1274) = 3.15$, $p < .0001$, $\eta_p^2 = .094$, the most interesting result. The interaction Slice × Age was not significant. $F(6, 182) = .41$, $p < .87$, $\eta_p^2 = .01$. We will analyze the triple interaction between Slice, Saccade Type and Age at the light of the "projection first", "alignment first", "relational priming" models, which generated different predictions.

As for the items A–C and T, the data show that, in the four groups, significant differences were always in favor of A–B and C–T saccades over A–C and B–T saccades. This is consistent with *the projection first* hypothesis. A–B saccades progressively decreased across slices and were replaced by C–T in the second slice which was dominant in the last slice (see Fig. 3). By contrast, A–C and B–T saccades were lower than A–B and C–T saccades in the four age groups and were scarce in the three slices, at the same level as irrelevant saccades (i.e., C-UnDis) (see Table 3 in Supplementary Materials for the list of significant contrasts, $p < 0.05$).

The target and the distractor items allow us to study how participants control the solution options and to test the relational priming view. We consider the saccades involving the solutions set items, i.e. C–T, C-SemDis, Target-SemDis and C-Undis, plus B–C (see Fig. 3).

The relational priming model (Leech et al., 2008) predicts that once one has identified the A–B relation, priming would lead to the correct C–T solution, which means very few or no saccades from C towards SemDis or UnDis, or between SemDis and Target. As shown by Fig. 3, this is clearly not the case: B–C, C-SemDis, T-SemDis were frequent in the three slices, especially in young children.

By contrast, if we do not endorse the *relational priming* model, the crucial developmental question is how and when participants compare C, Target and SemDis. Both the relational shift view (Rattermann & Gentner, 1998) and the executive function view predict that the distractors will be more focused on by young children than by adults. However, the question of the time course of C–T and C-SemDis saccades remains.

Results showed, in the *first* slice (1) that even though A–B were the most numerous saccades in all age groups, children had more "other saccades" than adults, especially B–C, T-SemDis and C-SemDis; (2) when there were differences between age groups, most often, they followed a smooth developmental progression. For A–B, there was a smooth increase with increasing age whereas for T-SemDis, C-SemDis, C-UnDis, and C–T there was a smooth increase with increasing age. We will argue that children are less able to prioritize the comparisons: they compare *all* the items related to C whereas adults concentrate on the A–B subgoal. Later in the trial, the number of C–T and T-SemDis saccades progressively increases in children whereas it reaches its maximum in the second slice for adults. This suggests that the main saccades are T-SemDis and C–T but that they are done earlier in the trial by adults. The larger difference between C–T and T-SemDis and other saccades in younger children compared to older participants suggests that 5- and 8-year olds need more comparisons than adults (see Tables 3 and 4 in Supplementary Materials).

To summarize, there are two important results. First, in both children and adults, there was a move from an initial predominance of A–B saccades to a predominance of C–T saccades, which is consistent with a focus on A–B then C–T pattern. However, 5- and 8-year-old children also had saccades involving C and the solution set in the first slice, consistent with the looking times showing that gazes towards A and B were less dominant in children than in adults. There was no evidence of systematic A–C or B–T alignments at any point in the trial. The second important result is that the four groups had many T-SemDis saccades throughout the trial together with C–T saccades. This means that both children and adults look at C and, then, compare T and SemDis in order to check which one goes with C.

3.3.4. Reviewing the chosen solution: Backward saccades from C, T and SemDis to A and B.

It is also important to monitor the ongoing comparisons between items, specifically to assess the chosen solution. Indeed, most models assume that, at some point, an evaluation of the coherence of a potential solution is taking place. One hypothesis is that the evaluation takes place at the end of the trial and that it should involve saccades between A and B on the one hand, and C, T and SemDis on the other hand. More specifically, this check has to do with saccades from C, T and SemDis back to A and B to be sure that the chosen solution is consistent with the relation found for the A–B pair, that is "backward" saccades such as C–A, C–B, etc. We focused on the interaction between Age and slices, to describe the temporal organization of this check. A three-way ANOVA, with Age as a between factor, Slice, and Saccade type (C-, C–B, T–A, T–B, SemDis-A and SemDis-B) as between factors, revealed a significant main effect of Saccade Type, $F(5, 455) = 32.1$, $p < .0001$, $\eta_p^2 = .26$, a Saccade type × Slice interaction, $F(10, 910) = 1.96$, $p < .05$, $\eta_p^2 = .02$ and a significant Saccade type × Slice × Age interaction, $F(30, 910) = 1.99$, $p < .005$, $\eta_p^2 = .06$, which are less informative in the present context. We were interested by the significant Age × Slice interaction, $F(6, 182) = 4{,}49$, $p < .001$, $\eta_p^2 = .13$. As shown by Fig. 4, adults had few saccades towards A and B in the first and more in the last slice whereas the reverse is true for the five-year-olds. This shows that children's choice evaluation is less developed than adults', most likely for less developed executive functions reasons. A Tukey HSD test confirmed these observations with adults producing significantly less backward switches than 5- and 8-year olds in the first slice whereas the other age groups were close one to the other. Adults and adolescents produced significantly more backward switches in the third slice than. This is a very interesting pattern showing that only the adults had very few backward switches at

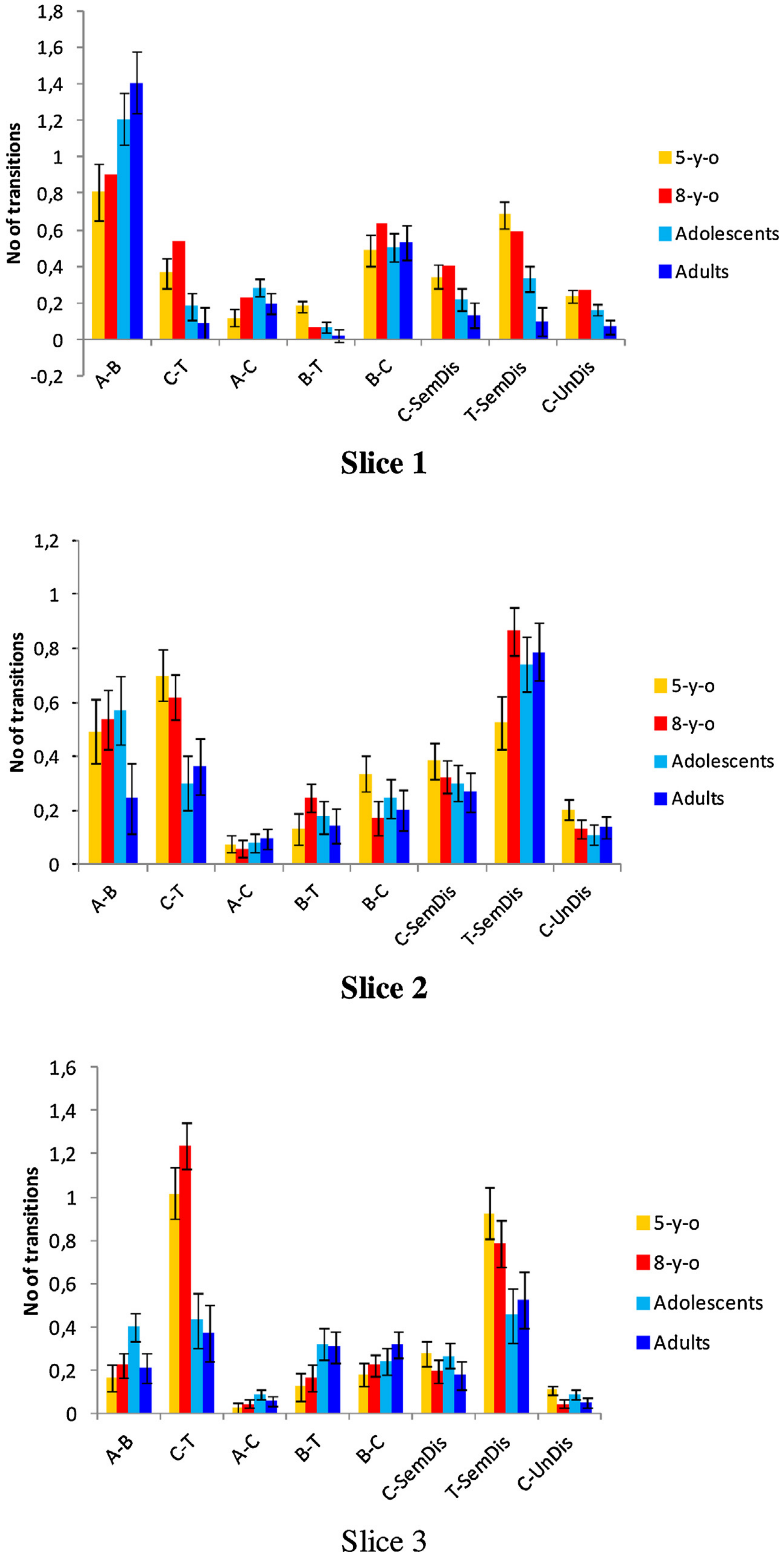


Fig. 3. Number of saccades, as a function of Saccade type, Time slice and Age group. T stands for Target. (One standard deviation error bars were used.).

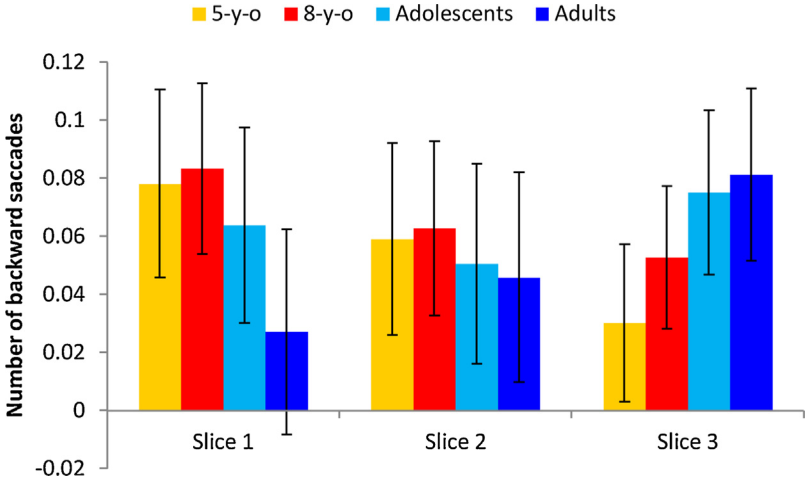


Fig. 4. Number of backward saccades from C, Target, and SemDis towards A and B as a function of time slice. The graph shows that younger children had more backwards saccades in the first slice and less in the last slice, whereas the opposite pattern was found in adults and adolescents.

the beginning whereas both the adults and the adolescents had significantly more of these switches at the end of the trial, which is a very nice picture of a progressive setting up of control processes.

3.3.5. Profiles for correct answers and errors: analyses of the saccades

The previous analyses have revealed striking differences between adults and children and allowed us to test models of mapping. The analysis on performance data also showed that children had more errors than adults. Eyetracking data suggest that children integrated the A–B pair less than adults and/or performed some comparisons between stimuli at a less appropriate time during the trial. Our previous analyses dealt with correct answers in all age groups. It might be that children's errors also result from less efficient comparisons during the trial. For example, the difference between errors and correct answers might be analogous to the ones we observed when we compared adults and children for correct answers.

Most current conceptions and models account for errors either in terms of difficulties inhibiting an irrelevant solution, or in terms of an irrelevant mapping with a semantically related item (e.g., Hummel & Holyoak, 1997; Holyoak, 2012; Morrison, Doumas, & Richland, 2011, Thibaut, French et al., 2010). However, these broad descriptions tell us nothing about the search dynamics that lead to erroneous solutions versus those that lead to correct answers. We considered two hypotheses. First, the *decisional hypothesis*, claims that both types of trials rely on similar processes and comparisons until the end of the trial. Participants decide correctly (correct answer) or not (error) between two related-to-C related (Target and SemDis), when all the comparisons have been performed. Differences between errors and correct trials, if any, should appear at the end, at the decisional stage. Second, the *comparison hypothesis*, by contrast, claims that errors and correct answers differ, from the outset, on crucial comparisons, especially those involving the A–B pair or the analogical match and the semantic distractor. It predicts more A–B, C–T, T-SemDis and less C-SemDis, for correct trials than for errors. This analysis showed that there were significant different errors and correct answers in terms of saccades, including during the first time slice.

The following analysis was restricted to children because performance was virtually perfect for adults in the One-semantic-distractor condition. For children, the error rate was around 40% allowing comparison with correct answers. We compared item-to-item saccade profiles leading to errors or correct answers. As in previous analyses, we conducted a four-way mixed ANOVA on the same set of saccades as above (A-B, A–C, B–C, B–T, C–T, C-SemDis, C-UnDis, T-SemDis) with Age (5- vs. 8-year old) Type of Answer (Correct vs. error) as a between factor, Saccade Type (A–B, A–C, B–C, B-Target, C-Target, C-SemDis, C-UnDis, T-SemDis), and Slice (first, second, third) as within factors. There was no main effect of Age and no interaction involving this factor. Results revealed main effects of Slice, $F(2, 142) = 9.83$, $p < .001$, $\eta_p^2 = .12$, and Transition, $F(7, 497) = 44.42$, $p < .0001$, $\eta_p^2 = .38$. There were two two-way interactions, (a) between Slice and Transition, $F(14, 994) = 6.37$, $p < .00001$, $\eta_p^2 = .082$, (b) between Transitions and Type of answer, $F(7, 497) = 13.44$, $p < .0001$, $\eta_p^2 = .16$. These interactions were subsumed by the following three-way interaction between Slice, Saccade type and Type of answer which was the key result, $F(14,994) = 3.38$, $p < .0001$, $\eta_p^2 = .045$ (see Fig. 5). Tukey HSD a posteriori analyses ($p < .05$) revealed that, compared to errors, correct trials had (1), in the 1st slice, significantly more CT, BT less CN and C-SemDis for correct trials, (2) in the 2nd slice, more A–B, C-Target, and less C-UnDis and C-SemDis in correct trials, and,(3) in the 3rd slice, more C–T and T-SemDis, and, again, less C-UnDis and C-SemDis for correct trials. Overall, correct trials were characterized by more relevant comparisons (C-Target, A–B and T-SemDis) and fewer irrelevant comparisons (i.e., C–N and C-SemDis). Importantly, these differences were present from the start of the trial, suggesting that errors are "set up" well before the end of the trial. In general terms errors are characterized by an imbalance between the main goal "finding the one that goes with C" and the secondary goal "taking into consideration the A–B relation". For example, French and Thibaut (2014) have recently shown

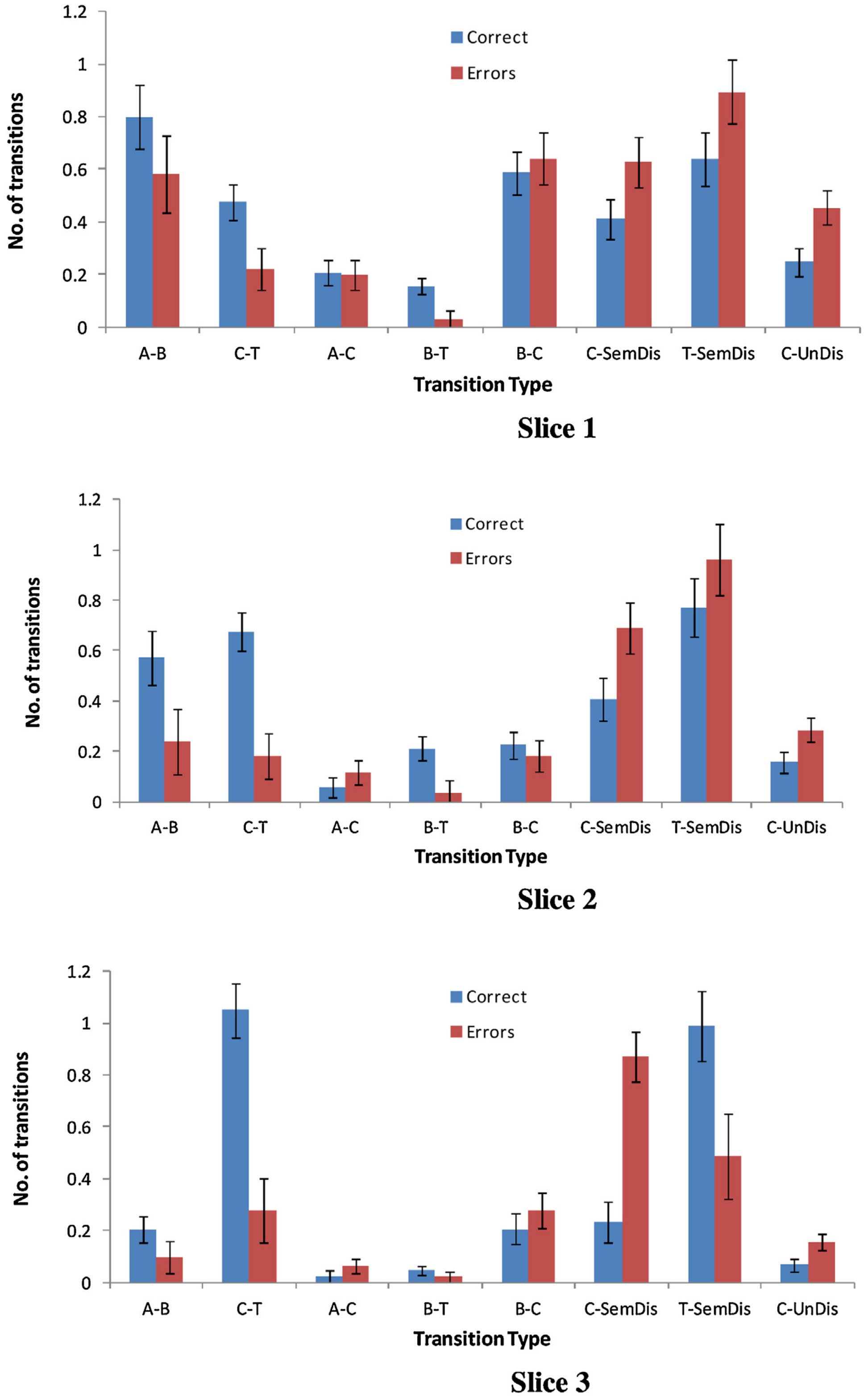


Fig. 5. Transition times for Correct answers and Errors as a function of Slice and Transition type. Differences between errors and correct trials are located on A–B, B–T, C–T, and T-SemDis. Specifically errors were biased towards the solution set with more saccades involving this set and less saccades involving A and B.

that the pattern of saccades during the first slice allowed them to predict the outcome of a trial, i.e., correct or incorrect, with an accuracy well above chance (67%).

4. General discussion

We provide the first developmental fine-grained analysis of the temporal dynamics of the search for an analogical answer in the classic A:B:C:D paradigm looking at three components of the comparisons involved in analogy making: (1) how people compare the base and the target pairs, (2) how they deal with irrelevant but salient options/distractors and eventually come up with a correct solution or an error, and (3) assess the quality of their choice/answer (see Gentner & Forbus, 2011; Holyoak, 2012). Our results show that children's search patterns differ from those of adults, a difference we explain in terms of executive function development.

4.1. Mapping A–C and target; finding common relations between the base and the target

We compared children's and adults' search strategies in the context of "alignment first", "projection-first" and "relational-priming" models. Adults' data are generally consistent with the "projection first" view which predicts early fixations on items A and B and saccades between them, and later fixations on C and the solution set and saccades between these items. This is what we observed and found no or little evidence of A–C and B–T saccades, which would have been predicted by the alignment-first view. This is consistent with Gordon and Moser (2007) data showing a peak of fixations for "actor-patient" saccades in the source picture (similar to our A–B saccades) at the beginning of the trial, followed by an "actor–patient" peak (similar to our C–T saccades) in the target picture later in the trial. By contrast, very few A–C saccades, B–T, or B-SemDis saccades were observed.

By contrast, for 5- and 8-year-old children, while there was little or no evidence of A–C and B-Target alignments, there was also little evidence for the "projection first" model which predicted A–B-followed-by-C-Target sequences. Children, especially 5-year-olds, tended to organize their search from the start around C, with their highest peak of fixation on C in the first slice, and saccades involving C (with B, T and -SemDis). In the last slice children moved predominantly to the Target, like adults. They had many C–T and T-SemDis switches, showing that the way they check their choice is via a comparison of T and the SemDis item.

In other words, the primary difference between age groups is that younger children focus on C more than adults and this plays a pivotal role from the beginning. With development, the predominant pattern becomes "focus first on the A–B pair at the expense of the other stimuli, then turn your attention to C and the solution set." We will argue that children's initial focus on C to the relative exclusion of A and B results from their difficulty in temporarily inhibiting the main explicit task goal—namely, "find something that goes with C". This means that they immediately organize their search around C. Adults, by contrast, temporarily set aside the main goal and focus on the A–B pair. These data clearly show that there are distinct ways of organizing the search for a solution and that analogical search is not organized in the same way for children and adults.

The above data also do not support the relational priming view (Leech et al., 2008) which postulates that the relation between A and B, once discovered, would automatically prime the C-Target relation. Indeed, we found evidence of systematic comparisons taking place between C, the Target and SemDis and between the Target and SemDis. Since Leech et al. (2008) model is a developmental model of analogies devised to account for data such as Goswami and Brown (1990), with the A:B:C:D paradigm, it is not consistent with our children's search pattern organized around C (C–T, C-SemDis, T-SemDis), from the beginning of the trial.

4.2. Task control: semantic distractors, choice evaluation and errors

Our second main issue was to explore the processing of associated distractors, the nature of errors, and choice monitoring. We believe that the differences between adults and children on correct trials also explain young children's errors. We suggest that children monitor comparisons less efficiently than adults. Five-year-olds exploration started around C and had more fixations of the Target and SemDis and fewer backward saccades at the end of the trial. These search features suggest that children have trouble parsing the problem in a systematic way. Indeed, studying related-to-C stimuli is an inefficient strategy if exploration is not guided by candidate relations. This open-ended search centered on C leads to a broader set of semantically unconstrained possibilities which, if the analogy is to be solved correctly, require the constraints provided by the A–B pair. Thus, children's smaller number of backward saccades at the end of the trial means that the information provided by the A–B pair is less likely to be integrated in the search. Finally, children's parallel explorations of the available stimuli might interfere with the construction/discovery of the A–B relation.

We believe that the above patterns provide a natural explanation of the search patterns associated with errors (choice of semantic distractors). If participants pay too much attention to the instruction "Find something that goes with C" and less attention to the A–B pair, the probability of an error increases. In sum, even when answering a problem correctly, 5-year-old children had fewer fixations on the A–B pair, less backward control saccades, and their search was less well focused on successive subgoals than adults. This pattern (less focus on A–B, fewer backward saccades) was amplified in the case of errors. With age, the patterns became closer to the adults' pattern. This description of both correct answers and errors, and predictions based on comparisons early in the trial is the first one in the literature on children.

According to the executive function approach, young children should have more difficulty in temporarily detaching their attention from the main goal of the task because of their less developed inhibition and flexibility capacities (Brocki & Bolin, 2004; Senn et al., 2004; Richland et al., 2006; Augier & Thibaut, 2013; Thibaut & Witt, 2015). Or children might forget the secondary and less salient goal "in the same way as A goes with B". This is consistent with the literature on goal forgetting (see Blaye & Chevalier, 2011). Our results are also consistent with the problem solving literature showing children's difficulties in achieving intermediate subgoals that temporarily require moving away from the main goal (inhibition of the "find the stimulus that goes with C" goal) (see, e.g., Siegler & Alibali, 2005).

It might also be argued that younger children's looking-time patterns reflect a failure to engage in analogical reasoning. Indeed, the behavioral results (58% performance on distractor trials) is very much in line with what might expected if children were mainly looking for Semantic associates and chose among them, in this case, either the correct analogical answer or the semantic distractor. We believe that this is not the case, for several reasons. First, it is unlikely that 8-year-old children would be unable to understand the paradigm. This paradigm has been used for children as young as 3 or 4 in similar situations with semantic materials (see other studies by Gentner and colleagues, see introduction above; Goswami & Brown, 1990; Thibaut et al., 2010). Second, in order to more directly test this hypothesis, we performed an analysis in which we divided children into low- and high-performing groups (L–P and H–P groups). Results revealed that high-performing participants still displayed the same C-oriented peak, distributing their attention evenly over A, B, Target and Semantically-related Distractor in the first time slice (see Fig. 1S in the Supplementary materials), even though this profile was less pronounced than for the low-performing participants (i.e., L–P participants had a higher C-peak and slightly fewer gazes towards A and B than H–P children). This stands in clear contrast with the adults' pattern which revealed an A–B peak with virtually no gazes towards the solution set during the first time slice (see Fig. 2). Since they were beyond 66% correct, it cannot be argued that high-performing children did not understand the task. Rather, their results are consistent with difficulties in inhibiting the main goal of the task (i.e., finding what goes with C). Finally, the transition pattern revealed clear differences between errors and correct trials and confirmed that correct answers were based on a larger number of relevant comparisons than errors (C-Target, A–B and T-SemDis), whereas errors were driven by a less accurate search organization (C-SemDis, C–N). This means that children's overall performance cannot be explained by a failure to understand the task. The difference between the search profile for errors and correct answers was confirmed by French and Thibaut (2014). Using a Linear Discriminant Analysis (LDA) and a neural network classifier, they showed that switches in the first third of the trial predicted with a 64% accuracy whether or not the problem will be answered correctly. If performance was driven only by semantic relatedness, the profile for correct answers and errors would be similar.

4.3. *Conceptions of analogy*

We believe that our results add constraints to general developmental models of analogy making and how stimuli are processed. Leech et al. (2008) relational priming model does not fit our data and we found no sign of alignments between the base and the target set (A–C, B–T or B-SemDis) (Markman & Gentner, 1993).

As for projection-first (A–B then C–D) models, children's search patterns depart from their controlled-by-the-driver hypothesis. For example, in LISA (Hummel & Holyoak, 1997), mappings are performed as a form of guided pattern matching. Since all analogs share the same semantic units, patterns generated in one analog will activate similar propositions in other analogs. In the bird:nest example, the activation of the proposition "lives in (bird, nest)" should be followed by many C–T saccades (corresponding to "lives in (dog, doghouse)). How would the model account for the peak, at all age groups, of T-SemDis comparisons (e.g., doghouse-bone), two items which are semantically unrelated? Incorporating developmental data would require models to account for monitoring during a trial, in particular, backward comparisons at the end of the trial in adults, or the differing numbers of T-SemDis and A–B saccades in adults and children (see Morrison et al., 2004; in ageing, for a developmental model.)

4.4. *Remaining issues*

The generality of this description of mapping remains an open question. It might well be that different kinds of analogy problems lead to different exploration patterns. For example, it might be that scene analogies, such as those in Richland et al. (2006) and Gordon and Moser (2007) are more compatible with an 'alignment-first' strategy because, in order to look for what is equivalent in both pairs, one needs to align the potential equivalent candidates in both scenes. Gordon and Moser do not, however, address this point. Our data (Glady, French & Thibaut, 2014) show that adults made essentially no systematic A–C and B-Target comparisons.

5. Conclusion

The main purpose of the present study was to better understand the developmental dynamics of search patterns in an analogy task. First, item looking times and item-to-item saccades revealed significant differences in the search strategies adopted by adults and children. Second, we suggest how these data constrain our conceptions of mapping and models of analogy making.

Acknowledgement

This research has been supported by a French ANR Grant for the "ANAFONEX" project ANR-10-BLAN-1908-01. Part of our results was presented at the 2011 Annual Conference of the Cognitive Science Society. We wish to thank Yannick Gérard, Yannick Glady, Angélique Missault for their help in data gathering and analysis.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.cogdev.2015.12.002.

References

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.
Blaye, A., & Chevalier, N. (2011). The role of goal representation in preschoolers' flexibility and inhibition. *Journal of Experimental Child Psychology*, *108*(3), 469–483.
Brown, A. L., & Kane, M. J. (1988). Preschool children can learn to transfer: learning to learn and learning from example. *Cognitive Psychology*, *20*(4), 493–523.
Brocki, K. C., & Bolin, G. (2004). Executive functions in children aged 6 to 13: a dimensional and developmental study. *Developmental Neuropsychology*, *26*, 571–593.
Bugaiska, A., & Thibaut, J. P. (2015). Analogical reasoning and aging: the processing speed and inhibition hypothesis. *Aging, Neuropsychology & Cognition*, *22*(3), 340–356.
Carlson, S. M., Zelazo, P. D., & Faja, S. (2013). Executive function. Oxford handbook of developmental psychology, 1, 706-742.
Doumas, L. A. A., Hummel, J. E., & Sandhofer, C. M. (2008). A theory of the discovery and predication of relational concepts. *Psychological Review*, *115*, 1–43.
Duchowski, A. J. (2007). *Eye-tracking methodology: theory and practice*. London: Springer-Verlag.
French, R. M. (1995). *The subtlety of sameness*. Cambridge, MA: The MIT Press.
French, R. M. (2002). The computational modeling of analogy-making. *Trends in cognitive Sciences*, *6*(5), 200–205.
French, R. M., & Thibaut, J.-P. (2014). Using eye-tracking to predict children's success or failure on analogy tasks. In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), *Proceedings of the thirty-sixth annual meeting of the cognitive science society Austin*. TX: Cognitive Science Society.
Gentner, D. (1983). Structure-mapping: a theoretical framework for analogy-making. *Cognitive Science*, *7*, 155–170.
Gentner, D. (1988). Metaphor as structure mapping: the relational shift. *Child Development*, *59*, 47–59.
Gentner, D., & Forbus, K. D. (2011). Computational models of analogy. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*(3), 266–276.
Glady, Y., Thibaut, J. P., French, R. M., & Blaye, A. (2012). Explaining children's failure in analogy making tasks: a problem of focus of attention? In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the Thirty-fourth Annual Meeting of the Cognitive Science Meeting* (pp. 384–389).
Glady, Y., Thibaut, J. P., & French, R. M. (2013). Visual strategies in analogical reasoning development: A new method for classifying scanpaths. In *In Proceedings of the Thirty-fifth Annual Meeting of the Cognitive Science Society* (pp. 2398–2403).
Gordon, P. C., & Moser, S. (2007). Insight into analogies: evidence from eye movements. *Visual Cognition*, *15*(1), 20–35.
Goswami, U. (1992). *Analogical reasoning in children*. Mahwah, NJ: Erlbaum.
Goswami, U., & Brown, A. L. (1990). Higher-order structure and relational reasoning: contrasting analogical and thematic relations. *Cognition*, *36*, 207–226.
Halford, G. S., Wilson, W. H., & Phillips, S. (1998). Processing capacity defined by relational complexity: implications for comparative, developmental, and cognitive psychology. *Behavioral and Brain Sciences*, *21*, 803–831.
Holyoak, K. J. (2012). Analogy and relational reasoning. In K. J. Holyoak, & R. G. Morrison (Eds.), *The Oxford handbook of thinking and reasoning* (pp. 234–259). New-York: Oxford University Press.
Hummel, J. E., & Holyoak, K. J. (1997). Distributed representations of structure: a theory of analogical access and mapping. *Psychological Review*, *104*, 427–466.
Josephson, S., & Holmes, M. E. (2002). Attention to repeated images on the World-Wide Web: another look at scanpath theory. *Behavior Research Methods, Instruments, & Computers*, *34*(4), 539–548.
Leech, R., Mareschal, D., & Cooper, R. (2008). Analogy as relational priming: a developmental and computational perspective on the origins of a complex cognitive skill. *Behavioral and Brain Sciences*, *31*, 357–414.
Markman, A. B., & Gentner, D. (1993). Structural alignment during similarity comparisons. *Cognitive Psychology*, *25*, 431–467.
Mitchell, M., & Hofstadter, D. R. (1990). The emergence of understanding in a computer model of concepts and analogy-making. *Physica D*, *42*, 322–334.
Morrison, R. G., Krawczyk, D. C., Holyoak, K. J., Hummel, J. E., Chow, T. W., Miller, B. L., et al. (2004). A neurocomputational model of analogical reasoning and its breakdown in frontotemporal lobar degeneration. *Journal of Cognitive Neuroscience*, *16*, 260–271.
Morrison, R. G., Doumas, L. L., & Richland, L. E. (2011). A computational account of children's analogical reasoning: balancing inhibitory control in working memory and relational representation. *Developmental Science*, *14*, 516–529.
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.
Rayner, K., Shen, D., Bai, X., & Yan, G. (Eds.). (2009). Cognitive and cultural influences on eye movements. In. London: Psychology 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*, 249–273.
Salvucci, D. D., & Anderson, J. R. (2001). Integrating analogical mapping and general problem solving: the path-mapping theory. *Cognitive Science*, *25*(1), 67–110.
Senn, T. E., Espy, K. A., & Kaufmann, P. M. (2004). Using path analysis to understand executive function organization in preschool children. *Developmental Neuropsychology*, *26*, 445–464.
Siegler, R. S., & Alibali, M. W. (2005). Information-processing theories of development. *Children's Thinking*, 65–106.
Thibaut, J.-P., French, R. M., & Vezneva, M. (2010). The development of analogy-making in children: cognitive load and executive functions. *Journal of Experimental Child Psychology.*, *106*, 1–19.
Thibaut, J.-P., French, R. M., & Vezneva, M. (2010). Cognitive load and semantic analogies: searching semantic space. *Psychonomic Bulletin and Review*, *17*, 569–574.
Thibaut, J.-P., French, R. M., Missault, A., Gérard, Y., & Glady, Y. (2011). In the Eyes of the Beholder: What Eye-Tracking Reveals About Analogy-Making Strategies in Children and Adults. *Proceedings of the Thirty-Third Annual Cognitive Science Society Conference*, 453–458.

Thibaut, J. P., & Witt, A. (2015). Young children's learning of relational categories: multiple comparisons and their cognitive constraints. *Frontiers in Psychology*, *6*, 643.
Viskontas, I. V., Morrison, R. G., Holyoak, K. J., Hummel, J. E., & Knowlton, B. J. (2004). Relational integration, inhibition and analogical reasoning in older adults. *Psychology and Aging*, *19*, 581–591.
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.