Journal of Experimental Child Psychology 166 (2018) 160–177

Contents lists available at ScienceDirect

Journal of Experimental Child Psychology

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

Working memory predicts children's analogical reasoning

Nina K. Simms [a], Rebecca R. Frausel [b,*], Lindsey E. Richland [b]

[a] Spatial Intelligence and Learning Center, Northwestern University, Evanston, IL 60208, USA
[b] Department of Comparative Human Development, University of Chicago, Chicago, IL 60637, USA

A R T I C L E I N F O

Article history:
Received 6 February 2017
Revised 11 August 2017
Available online 15 September 2017

Keywords:
Analogical reasoning
Analogical development
Executive functions
Working memory
Inhibitory control
Relational complexity

A B S T R A C T

Analogical reasoning is the cognitive skill of drawing relationships between representations, often between prior knowledge and new representations, that allows for bootstrapping cognitive and language development. Analogical reasoning proficiency develops substantially during childhood, although the mechanisms underlying this development have been debated, with developing cognitive resources as one proposed mechanism. We explored the role of executive function (EF) in supporting children's analogical reasoning development, with the goal of determining whether predicted aspects of EF were related to analogical development at the level of individual differences. We assessed 5- to 11-year-old children's working memory, inhibitory control, and cognitive flexibility using measures from the National Institutes of Health Toolbox Cognition battery. Individual differences in children's working memory best predicted performance on an analogical mapping task, even when controlling for age, suggesting a fundamental interrelationship between analogical reasoning and working memory development. These findings underscore the need to consider cognitive capacities in comprehensive theories of children's reasoning development.

* Corresponding author.
E-mail addresses: ninasimms@northwestern.edu (N.K. Simms), frausel@uchicago.edu (R.R. Frausel), lrichland@uchicago.edu (L.E. Richland).

http://dx.doi.org/10.1016/j.jecp.2017.08.005
0022-0965/

Introduction

Analogical reasoning—the cognitive process of drawing relationships between representations, often between prior knowledge and new representations—is a fundamental skill that develops dramatically in proficiency and resistance to distraction during childhood. Analogy plays a central role in higher-level cognition, and its ubiquitous and wide-ranging influence makes its developmental underpinnings essential to understanding human cognition more generally (Gentner, 2003; Hofstadter & Sander, 2013). Practically, analogical thinking is an important tool for learning. It enables children to, for example, extend existing knowledge to new contexts even if the representational systems look different. For instance, children might extend what they know about human energy needs (e.g., people need to eat for energy) to plants (e.g., plants likewise need energy input), although humans and plants differ in many respects. Building analogical reasoning skills is also a key objective for educational contexts, where children must build the skills to scaffold their own knowledge, to transfer it to new contexts, to explain new information, and to solve new problems, (Goldwater & Schalk, 2016; Richland & Simms, 2015). Understanding the mechanisms underlying analogical reasoning and development, therefore, is vital to identifying and intervening on points of dysfunction. In this study, we explored how one particular factor, executive function (EF), underlies the development of children's analogical reasoning.

Analogical reasoning and development

Formally, analogies are driven by alignment between systems of relations. Two situations are analogous if they share relational similarities regardless of other superficial properties or similarities. For example, plant stems are like drinking straws because they share functional and mechanistic relationships; both deliver liquid nourishment to a living organism, and both use differential pressure to move the liquid along the shaft. Children who understand how drinking straws work may be able to apply this knowledge to help them understand the less familiar domain of plant stems.

Performing analogical reasoning is not trivial. Assuming that a reasoner has recognized an opportunity for aligning the relationships in two or more analogues (e.g., the relationships between food and humans and between sunlight and plants)—no small feat in itself (Gick & Holyoak, 1980; Loewenstein, Thompson, & Gentner, 1999)—the reasoner must first encode the relational information from both analogues. These relational structures must be mentally maintained and manipulated to find correspondences between them (e.g., between the energy produced when a person metabolizes food and the energy produced when a plant photosynthesizes sunlight). If worthwhile correspondences are not initially found, the analogy must be discarded in favor of another or the representations must be flexibly modified to enable a better alignment (Kurtz, 2005; Yan, Forbus, & Gentner, 2003). For example, children might not initially see how food corresponds to sunlight because food is eaten, whereas sunlight is absorbed. However, this analogy becomes clear when children understand that both eating food and absorbing sunlight are intake processes. And all of this must take place while suppressing attention to irrelevant or extraneous information (Krawczyk et al., 2008).

This research explored whether EF resources can explain patterns of analogical reasoning for children between 5 and 11 years of age given analogy's high cognitive demands as described here. In particular, children's analogical reasoning improves along at least two key dimensions: the ability to resist perceptual distraction to prioritize relational information and the ability to handle and manipulate increasingly complex representations during alignment and mapping. We first describe this developmental trajectory and then explain the rationale for EF as an explanatory mechanism.

Object focus and relational shift

When children are given opportunities to engage in analogical reasoning, they often prioritize salient but nonrelational information over relational information (e.g., Daehler & Chen, 1993; Rattermann & Gentner, 1998; Thibaut, French, & Vezneva, 2010a). In particular, young children tend to rely on perceptually similar objects, which can detract from analogical reasoning performance if those object matches compete with relational matches (Christie & Gentner, 2010; Richland, Morrison, & Holyoak, 2006).

However, during development, children's analogical reasoning skills grow considerably. Children become increasingly oriented toward relations, a pattern described as the *relational shift* (Gentner, 1988; Gentner & Rattermann, 1991). Their early focus on holistic and object similarity gives way to greater appreciation for simple relational similarity and eventually more complex, interconnected relational structures (e.g., Chen, 1996; Gentner & Toupin, 1986; Loewenstein & Gentner, 2005; Richland et al., 2006; Thibaut, French, & Vezneva, 2010b).

Integrating multiple relations

Children also become increasingly adept at engaging with and integrating multiple relational structures as they develop. In Richland et al. (2006) Scene Analogy task, children were asked to align and map relations across two scenes. Some pairs of scenes contained only a single event relation (e.g., woman feeding boy, man feeding bird). Others depicted a more complex event scene consisting of two linked relations (e.g., woman feeding boy feeding dog, man feeding bird feeding hatchlings). Across two studies, younger children were less accurate on complex two-relation trials than on the simpler one-relation trials. For older children, the effect of scene complexity was absent or diminished, suggesting that older children were less impaired by multiple relations than younger children.

Thus, during childhood, children's analogical reasoning improves as children shift their attention from objects to relations and become better at engaging with larger and more complex structures.

Executive function

Development of executive function

Similar to analogical reasoning, EF also develops significantly during childhood. Executive function refers to the coordination of attention and action to carry out intentional goal-directed behavior (for reviews, see Carlson, Zelazo, & Faja, 2013; Diamond, 2013). Over development, children become better able to regulate their behavior under varying circumstances. The processes supporting coordinated goal-directed behavior become increasingly differentiated with age, and three separable but interrelated component functions emerge: *working memory, inhibitory control,* and *cognitive flexibility* (or *shifting*) (Miyake et al., 2000; Wiebe et al., 2011).

Working memory (WM) is the store of information that is active and consciously available at a given time. The amount of information that can be actively held and—importantly—manipulated by individuals is their WM capacity, which is limited (Baddeley, 2012; Cowan, 2010). WM capacity increases as children mature, allowing them to hold and manipulate more information, tackle tasks with greater representational demands, and appear more adult-like (Crone, Wendelken, Donohue, van Leijenhorst, & Bunge, 2006; Gathercole, Pickering, Ambridge, & Wearing, 2004).

Inhibitory control (IC) refers to the ability to suppress attention and action to irrelevant or conflicting information, especially when such responses are prepotent. With age, children's ability to inhibit attention and action improves, and children are better able to focus on relevant information and resolve conflict in service of task goals (Davidson, Amso, Anderson, & Diamond, 2006; Gerstadt, Hong, & Diamond, 1994; Rueda et al., 2004).

Cognitive flexibility (CF) is the ability to adaptively switch tasks, broadly construed. This includes being able to think about something in multiple ways, efficiently switch goals or activities, or take multiple perspectives. Like WM and IC, CF improves as children develop, enabling children to successfully negotiate more complex tasks with shifting attentional demands or rule sets (Cepeda, Kramer, & Gonzalez de Sather, 2001; Zelazo, Frye, & Rapus, 1996).

Although separable, WM, IC, and CF work together to support nonautomated goal-directed behavior.

Executive function and analogy

Reflecting the two ways in which analogical reasoning improves over development—integrating multiple relations and shifting attention from objects to relations—gains in WM and IC resources, specifically, have been theorized to support analogical development.

Working memory. Analogies are highly demanding of WM because (a) relational information is inherently more complex, and therefore more costly to represent, than object or featural information (Andrews & Halford, 2002; Halford, Wilson, & Phillips, 1998); (b) multiple relational representations must be held in WM simultaneously; and (c) these representations must be manipulated to, for example, integrate multiple relations into a systematic structure or adjust the representations to match across analogues.

Evidence from work with both children and adults implicates WM in analogical reasoning. When the representational demands of analogy tasks are increased, such as by increasing the size of the relational structures, children show decrements in reasoning that diminish with age (e.g., Richland et al., 2006). In healthy adults, occupying WM resources impairs integration of multiple relations (Waltz, Lau, Grewal, & Holyoak, 2000). Adults with damage to brain areas associated with WM similarly show deficits in processing complex relations (Waltz et al., 1999). And like with children, increasing representational complexity during analogical reasoning decreases accuracy (Cho, Holyoak, & Cannon, 2007; Viskontas, Morrison, Holyoak, Hummel, & Knowlton, 2004).

Inhibitory control. Analogical reasoning also imposes demands on IC processes to override attention to salient object properties or irrelevant perceptual similarities in favor of relational correspondences. This may be especially true for young children, who have a strong tendency to orient toward object and perceptual similarity (Gentner & Rattermann, 1991).

Indeed, introducing conflicting object similarity to an analogical reasoning task hurts children's reasoning, but children become less susceptible to distraction with age (e.g., Loewenstein & Gentner, 2005; Richland et al., 2006). Computational models have also shown that increases in model inhibition, along with relational knowledge, successfully recreate patterns of analogical development (Morrison, Doumas, & Richland, 2011). In adults, damage to brain regions associated with IC is associated with failure to prioritize relational similarity over featural similarity (Krawczyk et al., 2008; Morrison et al., 2004), much as young children do.

The current study

The investigation of executive resources in analogical development has been primarily investigated through manipulation of task demands (e.g., Richland et al., 2006; Thibaut et al., 2010a) or through longitudinal prediction designs (Richland & Burchinal, 2013), but no work has yet clarified that individual differences in children's EF capacities indeed correspond to their analogical reasoning skills. Thus, it remains possible that variations in performance previously attributed to EF could instead be explained by task difficulty or other age-linked factors.

Here, we assessed the EF capacities—including WM, IC, and CF—of children across a wide age range (5–11 years) and used performance on these measures to predict performance on Richland et al. (2006) scene analogy mapping task. This task was selected because (a) it has been successfully used with, and shows variation for, children as young as 3 and as old as 14 years; (b) it strategically manipulates two factors that should impose distinct demands on children's EF resources, namely the presence of an object match distractor and the number of relations that must be integrated to find an alignment; and (c) it uses relations that are familiar to children in our age range, ensuring that variations in relational knowledge could not explain changes in performance across development on these tasks (e.g., Gentner & Rattermann, 1991).

We hypothesized that individual differences in children's EF capacities will predict children's analogy performance. In our account, EF—along with other factors that change with age such as domain knowledge, relational language, attentional biases, and strategy use—interact to support changes in analogical reasoning over development. Although not our central hypothesis, our design allows us to further explore whether WM and IC are differentially related to distinct aspects of analogical reasoning behavior, with WM predicting performance involving relational integration and IC predicting performance involving featural distraction.

This hypothesis stands in contrast to an alternative in which age—but not EF—is related to improvements in children's analogical reasoning. In this alternative account, improvements in analogical reasoning over development are a consequence of these other age-related changes alone, and after controlling for age EF should not be related to analogical performance. Although work with adults has demonstrated a relationship between executive resources and analogical performance (e.g., Morrison et al., 2004; Waltz et al., 2000), these are extreme cases of clinical dysfunction or artificial manipulation that might not reflect the typical relationship between maturational improvements in EF and analogical reasoning over development.

Method

Participants

A total of 67 5- to 11-year-old children were recruited to participate in this study from the child subject pool at the university. This age range was selected because prior work shows age-related changes in patterns of analogical reasoning performance on our outcome measure of interest (the Scene Analogy task, described below) within this range and because this age group could complete multiple tasks in a single session. Two children did not complete the outcome measure of interest. One child scored excessively low on this task (<3 standard deviations from the mean) and was excluded from all further analyses. The remaining 64 children ranged in age from 5.0 to 11.4 years (M = 7.25 years, SD = 1.56). This sample consisted of 36 boys and 28 girls from a variety of racial/ethnic backgrounds (39% White, 31% Black, and 30% other). Mothers' education levels also varied, although the majority had completed college or beyond (41% graduate/professional degree, 37% bachelor's degree, 19% associate's degree, 3% high school/GED or missing information).

Materials and procedure

Data were collected as part of a larger study exploring child and caregiver interactions. The outcome of interest was children's analogical reasoning skill as measured by the Scene Analogy task (Richland et al., 2006). Children also completed three measures of EF and WM from the cognition battery of the National Institutes of Health (NIH) Toolbox (Gershon et al., 2013): the Dimensional Change Card Sort (a measure of CF), the Flanker Inhibitory Control and Attention task, and the List Sorting Working Memory task. These measures were used to predict performance on the Scene Analogy task.

The EF measures were administered in a single session that also included tasks engaging children and their caregivers in independent and joint description of select picture pairs from the Scene Analogy task. Performance on the parent and child scene description tasks is beyond the scope of this study and is not discussed further. All children completed the tasks in a fixed order that interspersed the EF measures and description tasks and ended with the Scene Analogy task.

Scene Analogy task

Children's analogical reasoning skill was measured using the Scene Analogy task, an analogical mapping task (Richland et al., 2006; available from the third author [L.E.R.] on request). In this task, children are presented with 20 pairs of pictures depicting analogous event scenes (Fig. 1). Children's task is to identify the object in the target picture that plays the same role in the events as an object identified in the source picture (the relational match). To do so requires that children identify, align, and prioritize the relations in the scenes.

Picture pairs varied along two dimensions: the presence of an irrelevant and potentially distracting object match in the target (distractor vs. no-distractor problems) and the number of relations to be mapped across pictures (one-relation vs. two-relation problems). Distractor problems were designed to assess children's ability to suppress attention and responses to object similarity in favor of relational similarity. Failure to do so should result in *featural errors,* where children select the object match instead of the relational match. Two-relation problems were designed to assess children's ability to

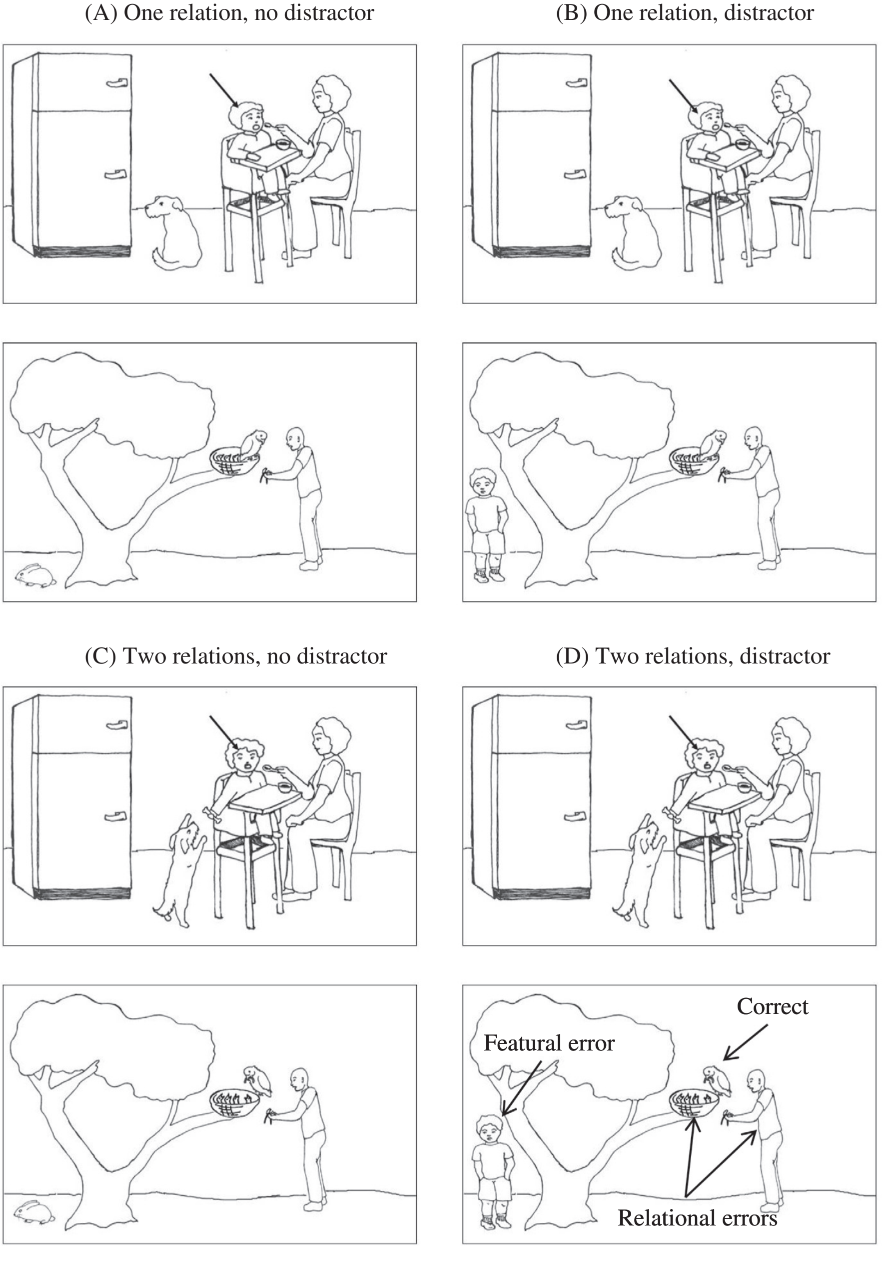

Fig. 1. Examples of the four types of Scene Analogy task problems, which varied scene complexity (A,B: one relation; C,D: two relations) and the presence of a competing object match distractor (A,C: no distractor; B,D: distractor). Children were asked to find the object in the target/bottom picture that corresponded with the object of interest in the base/top picture (identified here with an arrow). The coding of possible responses in the target—correct relational match, featural error, and relational errors—is labeled in Panel D.

integrate multiple relations and align larger relational structures. On these (and all) trials, noticing the relation but failing to properly align the corresponding roles should result in *relational errors,* where children select a participant in the event in a different role (Fig. 1D).

The task was administered to participants in paper form individually by an experimenter. Participants received instructions and two sample problems with feedback to ensure that they understood the goal of the task. While first presenting the one-relation sample problem, the experimenter explained, "There is a certain pattern in the top picture, and the same pattern happens in the bottom picture. [...] First I will tell you what pattern is happening in the top picture. Then I am going to put a sticker on one thing in the top picture, and your job is to tell me what is in the same part of the pattern in the bottom picture, so I can put a sticker on that too." The events in both pictures were described for children and used to illustrate the task. Only one child failed to select the correct match on the first sample picture, and that child was corrected and heard the correspondences between pictures described again.

The second sample problem introduced the two-relation problem. While presenting the second sample, the experimenter explained, "Now sometimes the pattern will have two parts, like the one you just saw [...] and sometimes the pattern will have three parts [referring to the number of participants in the event]. Let me show you what I mean." Again, the experimenter described the events in both pictures to illustrate. All participants correctly selected the relational match on the second sample problem.

Next, children solved 20 test problems. For each problem, the experimenter first described the event and identified the key object in the source picture and then asked children to find the corresponding object from the target picture (e.g., "In the top picture, there is a woman feeding a boy. If I put a sticker on the boy in the top picture, where should I put a sticker in the bottom picture?").

The 20 problems consisted of 5 of each of the four types of problems generated from crossing distractor (versus no distractor) and complexity (one versus two relations), and each participant saw only one version of each event. Trials were administered in five semi-randomized orders, and the version of each event pair was counterbalanced across orders. Children had seen some of these pictures before in a description task with their parents; some pictures they saw were identical to the test problem, and some were different versions of the event to be mapped. Regardless, the pattern of responding when including all 20 problems did not differ qualitatively overall from the pattern found when previously seen events were excluded, so all analyses of the Scene Analogy task include all 20 problems.

NIH toolbox tasks

Three aspects of children's EFs were measured using three tasks from the NIH Toolbox (Gershon et al., 2013): the Dimensional Change Card Sort (DCCS) task (a measure of CF and task switching), the Flanker Inhibitory Control and Attention task (Flanker), and the List Sorting Working Memory task (List Sort). The NIH Toolbox is a validated and normed computerized battery to assess cognitive, emotional, motor, and sensory functions. Tasks can be administered from 3 to 85 years of age and take only a few minutes each.

NIH toolbox DCCS (CF). This task is a measure of children's ability to flexibly switch between tasks. In this test, children must match a test picture (e.g., a yellow ball) to one of two target pictures, one of which matches the test picture in color (e.g., a yellow truck) and the other of which matches it in shape (e.g., a blue ball). Participants first sort the test pictures along one dimension, and then they must switch sorting rules and sort along the other dimension. For example, they may be asked to match the test and target pictures by shape for four trials, then to match by color for one trial, and then to resume matching by shape. The program tells participants which dimension to sort by at the start of each trial with written and verbal dimension labels (e.g., "COLOR"). Task administration took approximately 4 min.

The DCCS task is scored by combining two scoring vectors: accuracy and reaction time. The score for each vector ranges from 0 to 5 points; thus, the full computed score ranges from 0 to 10. Accuracy is always considered first; if accuracy does not exceed 80%, only the accuracy vector is reported (i.e., the score will be less than 5, with chance performance at 2.5). If accuracy is greater than 80%, the reaction time and accuracy vectors are combined into the computed score.

Test–retest reliability for the DCCS task is high for 3- to 15-year-olds, intraclass correlation coefficient (*ICC*) = .92 (95% confidence interval [CI] .86, .95), and the measure correlates highly with measures of the same construct for both younger children (3- to 6-year-olds: $r = .69$, $p < .0001$) and older children (8- to 15-year-olds: $r = .64$, $p < .0001$), demonstrating good convergent construct validity (Zelazo et al., 2013).

NIH toolbox flanker (IC and attention). This test is a version of the Eriksen flanker task derived from the Attention Network Task (Rueda et al., 2004). This task tests the ability to inhibit attention and behavioral responses to irrelevant features. On each trial, participants see a horizontal line made up of five fish (for children younger than 8 years) or five arrows (for participants 8 years or older). Participants are instructed to push a button corresponding to the direction of the central object (e.g., push the left arrow key if the center fish is pointing left) while ignoring the direction of the flanking objects. On congruent trials, all the fish/arrows point the same way; on incongruent trials, the middle fish/arrow points in the opposite direction of the fish/arrows on either side. Task administration took approximately 2–4 min.

For children younger than 8 years, if accuracy on the 20 fish trials meets or exceeds 90%, they are also given 20 arrow trials. Older children and adults complete only the 20 arrow trials, and their scores are computed as though they were 100% accurate on the fish trials. These age groups typically perform at ceiling on the fish trials.

As in the DCCS task, the Flanker task uses a two-vector scoring method that combines accuracy and reaction time. The score for each vector ranges from 0 to 5 points, and computed scores range from 0 to 10. Reaction time is considered only if accuracy exceeds 80%; otherwise, only the accuracy vector is included and the computed score will be less than 5 (with chance performance at 2.5).

Test–retest reliability for the Flanker task is high for 3- to 15-year-olds, $ICC = .92$ (95% CI .86, .95), and the measure correlates highly (for 3- to 6-year-olds: $r = .60$, $p < .0001$) and moderately (for 8- to 15-year-olds: $r = .34$, $p = .002$) with measures of the same construct, demonstrating good convergent construct validity (Zelazo et al., 2013).

NIH toolbox list sort (WM). This task is adapted from Mungas and colleagues' List Sorting task from the Spanish and English Neuropsychological Assessment Scales (Mungas, Reed, Marshall, & González, 2000). After a brief training session for children aged 7 years and older (a more elaborate training session is used for children aged 3–6 years to ensure task comprehension), children hear a list of either animals or fruit in a consistent random order. Each word is accompanied by a picture of the named object, which varies in size relative to the other objects in the list.

Participants are directed to report the items in size order, from smallest to largest. Once participants reach the threshold with single-category sorting (by getting two wrong of a given list size), children then see and hear lists of both fruit and animals intermixed and are directed to report first the fruit items in size order, from smallest to largest, followed by the animals, from smallest to largest. When children reach the threshold on the two-category sorting, the task ends. The approximate time for administration was 7 min.

The List Sort task is scored by summing the total number of items correctly recalled, which can range from 0 to 26.

Test–retest reliability for the List Sort task is high in this age group (3–15 years), $ICC = .86$ (95% CI .78, .91). For both younger children (3–6 years: $r = .57$, $p < .0001$) and older children (8–15 years: $r = .57$, $p < .0001$), the measure correlates with other measures of the construct, suggesting that the measure has good convergent construct validity in this age range (Tulsky et al., 2013).

Results

Before analyzing the relationship between EF capacities and analogical reasoning performance, we first summarize performance on each set of tasks.

Analogical reasoning (scene analogy)

Analyses were conducted to confirm that our participants replicated the patterns found in Richland et al. (2006). Statistical tests were conducted using all participants with usable scene analogy data, who were grouped by age into three categories: 5- and 6-year-olds ($n = 25$), 7- and 8-year-olds ($n = 29$), and 9- to 11-year-olds ($n = 10$). We elected to treat age categorically in our analysis in order to facilitate direct comparison with the previous work. However, it is important to note that these groupings result in low numbers for the 9- to 11-year-olds, potentially limiting the power to detect patterns of performance in this group.

As shown next, accuracy and error data replicated previous findings (Richland et al., 2006), which demonstrated that analogical skill increased with age. In particular, our results mirror those of Richland et al. (2006) in that (a) over time children became less susceptible to featural distraction, with accuracy improving with age on problems with distracting object matches and the greatest improvements being between the youngest and middle age groups; and (b) over development children improved in their ability to create internally consistent mappings, with accuracy on problems with greater complexity improving with age and the greatest improvements being between the middle and oldest age groups.

Accuracy

To explore the effects of object distractors and relational complexity on children's relational matching performance, the proportion of children's correct relational responses for each type of trial was entered into a mixed-measures analysis of variance (ANOVA) with age as a between-participants variable and distractor and complexity as within-participants variables (Fig. 2 and Table 1). In addition to main effects for age, $F(2, 61) = 12.78$, $p < .001$, $\eta^2_p = .295$, distractor, $F(1, 61) = 5.83$, $p < .05$, $\eta^2_p = .087$, and complexity, $F(1, 61) = 26.96$, $p < .001$, $\eta^2_p = .307$, the interaction between age and distractor was significant, $F(2, 61) = 4.55$, $p < .05$, $\eta^2_p = .130$. No other interactions were significant.

To further explore the Age $\times$ Distractor interaction from the main analysis, as well as more precisely characterize age-related differences in patterns of accuracy on the task, a 2×2 within-participants ANOVA with distractor and complexity was calculated for each age group separately. The 5- and 6-year-olds showed main effects of distractor, $F(1, 24) = 19.03$, $p < .001$, $\eta^2_p = .442$, and complexity, $F(1, 24) = 21.05$, $p < .001$, $\eta^2_p = .467$. They were more accurate on no-distractor problems than on distractor problems and were more accurate on one-relation problems than on two-relation problems. For 7- and 8-year-olds, the analysis yielded a main effect of complexity, $F(1, 28) = 16.92$, $p < .001$, $\eta^2_p = .377$. They were more accurate on one-relation problems than on two-relation problems but showed no main effect of distractor. Finally, the 9- to 11-year-old group did not have any significant main effects or interactions.

The oldest group might not have had sufficient numbers to detect differences across problem types. However, these children's consistently high performance also suggested the possibility of ceiling performance. Comparisons against perfect accuracy for each problem type revealed that this group's accuracy did differ significantly from ceiling on two-relation problems (no distractor: $t(9) = -2.69$, $p < .05$; distractor: $t(9) = -3.00$, $p < .05$) but not on one-relation problems (no distractor: $t(9) = -1.49$, $p = .17$; distractor: $t(9) = -1.41$, $p = .19$).

Featural errors

To examine children's susceptibility to competing object matches, the proportion of featural errors on each problem type was entered into a 3 (Age) $\times$ 2 (Distractor) $\times$ 2 (Complexity) mixed-measures ANOVA (Table 1). Note that a featural error—selecting the object distractor—was not possible on the no-distractor trials (meaning that we would expect more featural errors on distractor problems).

In addition to main effects of distractor, $F(1, 61) = 30.48$, $p < .001$, $\eta^2_p = .333$, and age, $F(2, 61) = 7.27$, $p < .01$, $\eta^2_p = .193$, the Age $\times$ Distractor interaction was significant, $F(2, 61) = 7.27$, $p < .01$, $\eta^2_p = .193$.

To further explore the Age $\times$ Distractor interaction, a 2×2 within-participants ANOVA with distractor and complexity was calculated for each group separately. A main effect of distractor was found for the 5- and 6-year-olds, $F(1, 24) = 38.78$, $p < .001$, $\eta^2_p = .618$, and the 7- and 8-year-olds, $F(1, 28) = 13.39$, $p < .01$, $\eta^2_p = .324$, but not for the 9- to 11-year-olds, $F(1, 9) = 2.25$, $p = .17$, $\eta^2_p = .200$ The two younger groups made more featural errors on the distractor problems than on the no-distractor problems, but this difference was not significant for the oldest group. No other main effects or interactions were significant.

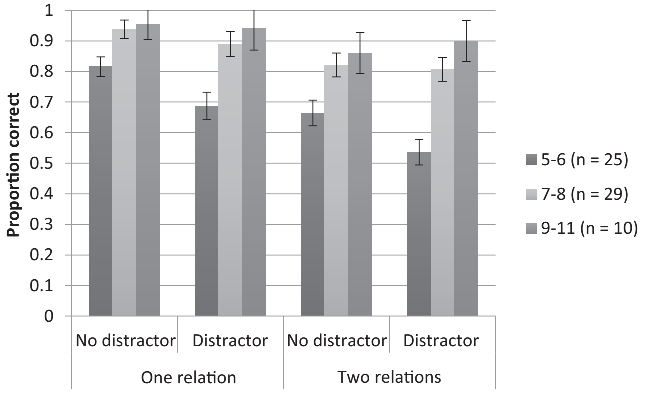

Fig. 2. Proportions of correct relational matches on the Scene Analogy task.

Table 1
Proportions of response types on the scene analogy task.

				5- and 6-year-olds $n = 25$	7- and 8-year-olds $n = 29$	9- to 11-year-olds $n = 10$
One relation	No distractor	Correct	*M*	.82	.94	.96
			(*SD*)	(.22)	(.11)	(.10)
		Relational error	*M*	.09	.05	.03
			(*SD*)	(.13)	(.09)	(.08)
		Other error	*M*	.10	.01	.02
			(*SD*)	(.18)	(.05)	(.06)
	Distractor	Correct	*M*	.69	.89	.94
			(*SD*)	(.30)	(.16)	(.13)
		Featural error	*M*	.14	.06	.02
			(*SD*)	(.16)	(.11)	(.06)
		Relational error	*M*	.15	.04	.02
			(*SD*)	(.19)	(.10)	(.06)
		Other error	*M*	.02	.01	.02
			(*SD*)	(.09)	(.04)	(.06)
Two relations	No distractor	Correct	*M*	.66	.82	.86
			(*SD*)	(.27)	(.16)	(.16)
		Relational error	*M*	.23	.16	.08
			(*SD*)	(.21)	(.15)	(.10)
		Other error	*M*	.10	.02	.06
			(*SD*)	(.16)	(.06)	(.14)
	Distractor	Correct	*M*	.54	.81	.90
			(*SD*)	(.25)	(.20)	(.11)
		Featural error	*M*	.18	.08	.02
			(*SD*)	(.17)	(.15)	(.06)
		Relational error	*M*	.26	.10	.06
			(*SD*)	(.20)	(.13)	(.10)
		Other error	*M*	.03	.01	.02
			(*SD*)	(.09)	(.04)	(.06)
Overall		Correct	*M*	.68	.86	.00
			(*SD*)	(.22)	(.11)	(.00)
		Featural error	*M*	.08	.04	.01
			(*SD*)	(.06)	(.05)	(.02)
		Relational error	*M*	.18	.09	.05
			(*SD*)	(.12)	(.07)	(.04)
		Other error	*M*	.06	.01	.03
			(*SD*)	(.11)	(.02)	(.04)

Relational errors

The proportion of children's relational errors on each problem type was entered into a 3 (Age) × 2 (Distractor) × 2 (Complexity) mixed-measures ANOVA (Table 1). Note that there was one possible relational error on the one-relation problems and two possible relational errors on the two-relation problems (meaning that we might expect more relational errors on two-relation problems).

The analysis yielded significant main effects of complexity, $F(1, 61) = 23.46$, $p < .001$, $\eta_p^2 = .278$, and age, $F(2, 61) = 11.53$, $p < .001$, $\eta_p^2 = .274$, and a marginally significant Age × Distractor interaction, $F(2, 61) = 2.78$, $p < .10$, $\eta_p^2 = .084$. (Post hoc comparisons suggested that this unexpected interaction was driven by the 5- and 6-year-olds, who were marginally less likely to make relational errors on distractor problems than on no-distractor problems, Bonferroni, $p < .10$, presumably because they were selecting the object match distractor more frequently on the distractor problems.)

To understand these patterns more clearly, a 2 × 2 within-participants ANOVA with distractor and complexity was calculated separately for each age group. A significant main effect of complexity was found for both the 5- and 6-year-olds, $F(1, 24) = 14.31$, $p < .01$, $\eta_p^2 = .374$, and the 7- and 8-year-olds, $F(1, 28) = 21.99$, $p < .001$, $\eta_p^2 = .440$, but not for the 9- to 11-year-olds, $F(1, 9) = 2.98$, $p = .12$, $\eta_p^2 = .249$. The younger groups made more relational errors on two-relation problems than on one-relation problems, but this difference was not significant for the oldest group. No other main effects or interactions were significant.

Executive functions (Flanker, List Sort, and DCCS)

Due to fatigue, experimenter error, and technological failure, not all participants had data for all three EF tasks. Among the 64 children with the outcome measure, 5 did not complete any of the EF tasks and so were excluded from further analyses. Among the 59 remaining children, none was missing the DCCS task, 13 (22%) were missing the Flanker task, and 9 (15%) were missing the List Sort task. Six of the participants with missing data were missing both measures. Age in months was related to missing the Flanker task ($r = -.297$, $p = .023$) and the List Sort task ($r = -.317$, $p = .016$).

One participant who was missing only one EF measure (the List Sort task) was excluded after visual inspection of scatterplots indicated, and correlations confirmed, that the child's consistently low performance across measures could skew the model used to impute missing data (see below) as well as the relationship between the Scene Analogy task and the EF measures.[1]

Missing scores for the 15 participants who were missing one or two EF measures were imputed using multiple imputation using the following predictors: age in months, available scores on the remaining EF tasks, race (dummy coded), and maternal education. We include race and maternal education because previous research has shown effects of race and maternal education on EF (Hackman, Gallop, Evans, & Farah, 2015; Little, 2017).

Multiple imputation relies on the assumption that data are missing at random, which is a strong assumption to be making given that in our data missingness is related to age. To compensate for this, and to satisfy the recommendation that the number of imputations exceed the percentage of missing data (White, Royston, & Wood, 2011), we generated 40 multiply imputed datasets using the built-in functionality in SPSS Version 22.0. The final statistics reported are the pooled estimates of the coefficients (β, standard error, and corresponding p value) and the average for the F statistic and R-square (and corresponding p value). After imputation, 58 participants had complete data with which to analyze the relationships between performance on the EF measures and performance on the Scene Analogy task (Table 2).[2]

[1] Prior to imputation, the correlation between the Flanker and DCCS tasks was significant when this participant's data were included, $r = .35$, $p < .05$, but was only marginally significant when they were excluded, $r = .26$, $p < .10$. Thus, we excluded this participant from the imputation model to avoid skewing the imputed data. Furthermore, proportion correct on the Scene Analogy task was highly correlated with pre-imputation Flanker task performance when this participant was included, $r = .53$, $p < .001$, but after exclusion the effect size was much smaller, $r = .36$, $p < .05$. Therefore, we decided to also remove this participant from all further predictive analyses.

[2] Despite the change in sampling due to missing EF data, patterns of performance on the Scene Analogy task for the 58 participants analyzed in this part of the Results section were indistinguishable from those for the 64 participants reported in the first part of this Results section.

To confirm that performance on the EF measures improved with age in our sample, as well as explore the relationships among performance on the three tasks, bivariate correlations between age in months, Flanker task, List Sort task, and DCCS task scores were conducted (Table 3). As expected, age was significantly correlated with performance on all three EF measures. The List Sort task and DCCS task were significantly correlated. However, neither task significantly correlated with the Flanker task. Partial correlations controlling for age in months suggested that the collinearity between List Sort task and DCCS task performance could be accounted for by changes in performance with age; correlations among the three tasks were not significant when controlling for age.

It is somewhat surprising that we did not find relationships between performance on the Flanker task and the other two EF measures before controlling for age, or between any of the three measures after controlling for age, because measures of different EF constructs often correlate (e.g., Blackwell & Munakata, 2014; Carlson, Moses, & Breton, 2002; Koch, Gade, Schuch, & Philipp, 2010). However, the three tasks were designed to measure distinct aspects of the overall cognitive resource system, which may differ within an individual (Blackwell, Chatham, Wiseheart, & Munakata, 2014). Furthermore, given the strong correlations between age and each of the EF measures (e.g., Davidson et al., 2006), it may be difficult to detect variability beyond those age-associated changes.

Relations between analogical reasoning and executive functions

To explore whether and how individual differences in EFs related to analogical reasoning performance, two linear regression models using performance on the three EF tasks and age were built to predict performance on key measures within the Scene Analogy task. Model 1 included age only, and Model 2 included both EF and age[3] (Table 4). The goal was to determine whether including the three EF constructs explained significantly more variability on the Scene Analogy task than age alone. In both models, variables were entered simultaneously.

We summarize two main patterns in advance. First, the model including both EF and age always provided the best fit for Scene Analogy task performance, although it was not statistically better than the age-alone model for all measures. Second, only the List Sort task emerged as a significant independent predictor of analogy performance.

Overall response types

The proportions of children's correct responses, featural errors, and relational errors on the Scene Analogy task were predicted using multiple linear regression models with age in months and scores from the three EF measures—Flanker (IC and attention) task, List Sort (WM) task, and DCCS (CF) task—as predictors.

Both models significantly predicted children's overall accuracy on the Scene Analogy task [age only: $F(1, 56) = 19.538$; EF + age: $F(4, 53) = 7.685$]. The EF + age model (Model 2) was the best predictor of children's overall accuracy on the Scene Analogy task, accounting for 36.4% of the variance. Among the three EF measures, only the List Sort task was a significant individual predictor of accuracy. In addition, the EF + age model was a marginally significant better predictor of accuracy than a model including only age (Model 1).

Both the EF + age model, $F(4, 53) = 2.948$, and the age-only model, $F(1, 56) = 11.124$, were significant predictors of children's featural errors. In the EF + age model, which provided the best fit and accounted for 18.2% of the variance, age was a marginally significant individual predictor. The model with both EF and age was not significantly better than the model with age alone.

Children's relational errors were significantly predicted by both models [age only: $F(1, 56) = 23.015$; EF + age: $F(4, 53) = 9.408$]. The best-fitting model, EF + age, accounted for 41.3% of the variance. Only the List Sort (WM) task was a significant individual predictor of relational errors in this model. Including both EF and age significantly improved the fit of the model compared with age alone.

[3] We also ran a model including EF only. The results were largely the same as the EF and age model, with WM emerging as the only significant predictor of analogy performance. Only one scene analogy measure—2N problem accuracy—was significantly predicted by WM in the EF-only model but not in the EF and age model, so we report only the EF and age model here.

Table 2
Means and standard deviations for the three EF measures before and after imputation.

		Before imputation			After imputation		
		DCCS	Flanker	List Sort	DCCS	Flanker	List Sort
		Possible range			Possible range		
		0–10	0–10	0–26	0–10	0–10	0–26
5- and 6-year-olds	*M* (*SD*)	3.14 (2.02)	5.22 (1.07)	9.73 (2.84)	3.14	5.15	9.77
	n	21	12	15	21	21	21
7- and 8-year-olds	*M* (*SD*)	5.52 (1.50)	5.58 (1.30)	14.08 (2.38)	5.52	5.61	14.17
	n	27	24	25	27	27	27
9- to 11-year-olds	*M* (*SD*)	6.42 (0.57)	6.46 (0.72)	16.50 (2.99)	6.42	6.49	16.50
	n	10	9	10	10	10	10
Total	*M* (*SD*)	4.81 (2.06)	5.66 (1.20)	13.26 (3.61)	4.81	5.60	12.98
	N	58	45	50	58	58	58

Table 3
Bivariate and partial correlations between age in months and the EF measures.

			Flanker	List Sort	DCCS
Age (months)	*r*		.42*	.64**	.56**
	n		58	58	58
Flanker	*r*	Controlling for age (months)		.25	.26
	n			58	58
List Sort	*r*		−.02		.46***
	n		58		58
DCCS	*r*		.03	.16	
	n		58	58	

Note. Cells in top half of diagonal show bivariate correlations; cells in bottom half of diagonal show correlations controlling for age in months.
* $p < .05$.
** $p < .01$.
*** $p < .001$.

Correct responses on different problem types

To better understand how EF relates to the distinct demands of different types of problems, the same multiple linear regression models with age in months and scores from the three EF measures were used to predict accuracy on each type of Scene Analogy task problem. These models offer a more detailed breakdown of how EF and age relate to overall accuracy on the Scene Analogy task.

For one-relation, no-distractor problems (1N), the EF + age model, $F(4, 53) = 3.169$, significantly predicted accuracy. The age-only model was only marginally predictive, $F(1, 56) = 4.014$. The EF + age model provided the best fit and accounted for 19.1% of the variance on this type of problem. Only the List Sort (WM) task was a significant individual predictor of accuracy in this model. The model including both EF and age was a marginally better predictor of 1N performance than a model including only age.

For one-relation, distractor problems (1D), both models were significant [age only: $F(1, 56) = 14.992$; EF + age: $F(4, 53) = 4.442$]. The EF + age model accounted for 25.0% of the variance on this type of problem and provided the best fit. Age was a marginally significant individual predictor in the model. Including both EF and age in the model did not significantly improve the fit compared with a model with only age.

For two-relation, no-distractor problems (2N), both models were again significant [age only: $F(1, 56) = 7.753$; EF + age: $F(4, 53) = 2.962$]. Accounting for 18.2% of the variance, the EF + age model was

Table 4
Summary of linear regression models predicting scene analogy task performance.

	Model 1	Model 2	ΔR^2	Model 1	Model 2			
	Age	Age + 3 EFs	(Model 1 to Model 2)	Age	Age	DCCS	Flanker	List Sort
	R^2	R^2		β (SE)	β (SE)	β (SE)	β (SE)	β (SE)
Overall accuracy	.259***	.364***	.105^	.005*** (.001)	.002 (.002)	.000 (.012)	.012 (.022)	.018* (.008)
Overall featural errors	.166**	.182*	.016	−.001** (.000)	−.001^ (.001)	.001 (.005)	−.003 (.008)	−.002 (.003)
Overall relational errors	.291***	.413*	.122*	−.003*** (.001)	−.001 (.001)	.000 (.007)	−.011 (.012)	−.011* (.004)
1N accuracy	.067^	.191*	.124^	.002* (.001)	.000 (.002)	−.006 (.012)	.013 (.024)	.018* (.008)
1D accuracy	.211***	.250**	.039	.006*** (.002)	.004^ (.002)	−.004 (.016)	.028 (.029)	.008 (.011)
2N accuracy	.122**	.182*	.060	.004** (.001)	.001 (.002)	.004 (.016)	.010 (.028)	.015 (.011)
2D accuracy	.265***	.389***	.124^	.007*** (.002)	.003 (.002)	.006 (.017)	−.001 (.030)	.029* (.011)

^ $p < .10$.
* $p < .05$.
** $p < .01$.
*** $p < .001$.

the best-fitting model of 2N accuracy. However, neither age nor any of the EF measures was a significant (or marginal) individual predictor in this model. Including both EF and age in the model did not explain more of the variance compared with an age-only model.

Finally, for two-relation, distractor problems (2D), the age-only model, $F(1, 56) = 20.228$, and the EF + age model, $F(4, 53) = 8.537$, both significantly predicted accuracy. The model with both EF and age provided the best fit, accounting for 38.9% of the variance on this problem type. Performance on the List Sort (WM) task was a significant individual predictor of accuracy in the EF + age model. Including age and EF measures marginally improved the model's fit over a model with just age.

Replication with a separate dataset

As part of a different study on long-term memory and analogical reasoning, an additional 24 children (15 girls) aged 5 or 6 years ($M = 6.03$, $SD = 0.61$, range = 5.05–6.95 years) completed the Flanker Inhibitory Control and Attention task, the List Sort Working Memory task, and the Scene Analogy task, using the same methods outlined in this article. To test the replicability of our findings, we ran two linear regression models to predict performance on the Scene Analogy task: one with age in months (age alone) and a second with age and the scores on the two EF tasks entered simultaneously (EF + age). Overall, we replicated our main patterns of results with these participants. Specifically, WM predicted performance on the Scene Analogy task even when controlling for age and IC. In addition, including the EF measures significantly improved the fit of the models over the models using age alone as a predictor.

Discussion

In this study, we examined the relationships among age, individual differences in EF, and 5- to 11-year-olds' analogical mapping performance. As prior work has found, age was highly predictive of children's success on the analogical reasoning task. For the first time, we also showed that children's WM capacity—an important component of EF—also predicts analogy performance. Together, age and performance on the EF measures provided the most accurate predictions of children's analogical reasoning. Even after controlling for age, which is highly related to EF development, WM remained a significant predictor of performance, suggesting that this relationship was not due to other age-related changes that may be involved in analogical development. In addition, we were able to replicate these general patterns in a separate dataset of 5- and 6-year-olds, providing further assurance that the relationship between WM and analogy performance was genuine. These findings illuminate the mechanisms underlying analogical development by (a) demonstrating a within-individual relationship between WM capacity and analogical reasoning skill and (b) providing specificity about what circumstances and behaviors are most related to individual capacity.

We further considered the speculation that WM and IC might predict distinct aspects of analogical reasoning behavior. In particular, we conjectured that WM capacity would be especially related to problems with multiple relations and to errors involving relational integration (see Halford, 1993), whereas IC would be more highly related to problems with competing object matches and to errors involving featural distraction (Richland et al., 2006). We found that WM was related to children's reasoning more broadly, at both high and low levels of relational complexity, suggesting that individual differences in WM may be relevant even when representational demands are not being strained. We did not find evidence in this study that IC (or CF) was a significant individual predictor of children's analogical performance, although this could have been explained by the nature of the specific tasks that were administered.

The observed relationship between WM and analogical performance is consistent with findings from the adult literature. In adults, when WM resources are compromised, both the ability to integrate multiple relations *and* the ability to resist featural distraction on analogy tasks suffer (Morrison et al., 2004; Waltz et al., 1999, 2000). Interactions between increasing representational and inhibitory demands on performance indicate that in the context of analogical reasoning, WM capacity and inhibitory processes in WM share a common resource (Cho et al., 2007); thus, it may be difficult to disentangle the distinct contributions from each.

In adults, it is the inhibitory processes in WM that have been shown to be predictive of analogical reasoning. One explanation for why we did not find that IC was related to Scene Analogy task performance in this study is that our measure of IC, the Flanker task, does not need to be done in WM and, therefore, might not measure the type of inhibition recruited in analogical reasoning (or the important interaction between WM and inhibition; e.g., Cho et al., 2007). It is possible that a task requiring IC in WM (e.g., the Dots task used in Davidson et al., 2006, which requires children to remember a set of response rules as well as to inhibit prepotent responses) would be more related to children's analogical performance.

The Flanker task also showed only moderate convergent validity with another measure of IC, the Delis–Kaplan Executive Function System (D-KEFS) Color Word Interference test (Delis, Kaplan, & Kramer, 2001), in 8- to 15-year-olds. The D-KEFS requires children to ignore a written color word in order to name the (incongruent) color of the text instead. It is conceptually quite similar to the Flanker task in that children must selectively attend to certain aspects of the stimulus and override a conflicting prepotent response associated with the unattended information. However, the conflicting information on the Flanker task is spatially separated; narrowing the focus of attention to the center of the screen could bypass attention to the conflicting information, reducing the need for IC. In contrast, children performing the D-KEFS must attend to the same spatial location and, thus, cannot avert the conflict resolution requirements. Spatially separating conflicting information, so that attention to the conflicting information can be avoided, has been shown to benefit children's executive control performance (Chevalier, Blaye, Dufau, & Lucenet, 2010); likewise, featural distraction creates decrements in adults' performance only when that information has been attended to (Cho et al., 2007). Any single task introduces unique task-specific variance that can mask the underlying construct and skew its relationship to other tasks and capacities or obscure important interactions between the construct and its context. Thus, it may be that the Flanker task was not an ideal measure of IC for this study.

Another possibility is that IC would be associated with avoiding featural distraction for younger children but not for children in the age range tested here. In fact, only the youngest age group in our study (5- and 6-year-olds) showed a decrement in performance on distractor problems, and many of the demonstrations of the object bias in analogical development have been conducted with younger preschool-aged children (e.g., Christie & Gentner, 2010; Richland et al., 2006; Thibaut et al., 2010b). Substantial gains in IC also occur during the preschool years, just younger than most of the children in our study (Davidson et al., 2006), and IC has been shown to be predictive of problem solving earlier, rather than later, in children's development (Blakey, Visser, & Carroll, 2016; Senn, Espy, & Kaufmann, 2004). Potentially, gains in other capacities, such as WM, alleviate demands on IC (Engle, 2002), so that for children of the ages in this study individual variability in IC capacity is no longer strongly predictive of analogical performance.

Regardless, our findings do not support the alternative hypothesis that age, but not components of EF, is related to children's analogical reasoning performance. This result would have been expected if other age-related changes accounted wholly for developmental improvements in analogical reasoning. In that case, after controlling for age (and, correspondingly, other age-linked factors related to analogical performance such as knowledge accretion), there should not have been any additional variance predicted by the EF measures. Instead, we found that differences in WM were related to individual differences in analogical reasoning beyond effects of age. Of course, our results are not sufficient to conclude that increases in WM themselves cause improvements in analogical reasoning over development, but they are consistent with a view supported by prior developmental and adult research that WM—and possibly other executive control resources, although we did not find evidence for that in our study—provides part of the foundation on which analogical reasoning ability is built.

We consider this research to be an important initial exploratory step toward understanding how these skills codevelop within individual children. Future work will need to include a greater variety of EF and analogy tasks to provide more robust measurement of these skills. Future work should also strive to look more broadly at a full range of development. As we have already alluded to, different aspects of EF may play more or less prominent roles in analogical reasoning as children's capacities develop.

Nonetheless, this work highlights the necessity of seriously considering the constraints imposed by the structure and limits of EF resources like WM and when and how changes in these resources influence analogical development. A comprehensive account of analogical development will need to specify how WM and other contributors to EF interact with other factors involved in development, including relational knowledge, relational language, strategy use, and patterns of attention. For example, relational language may allow children to construct efficient, robust representations in WM, reducing load and helping children to direct attention (Morrison & Cho, 2008; Vales & Smith, 2015).

From an applied perspective, understanding how EF resources limit or support analogical reasoning has implications for educational practice by helping curriculum designers and teachers to create and implement more effective instruction that takes cognitive demands and student capacity into account (Begolli, Richland, & Jaeggi, 2015; Richland & McDonough, 2010; Richland & Simms, 2015). For example, making the representations of to-be-compared analogues visible simultaneously during analogical instruction can alleviate WM demands, freeing those resources for forging the connections that make analogical learning so powerful (Begolli et al., 2015).

Acknowledgments

The research reported here was partially supported by grants from the National Science Foundation (SMA-1548292) and the Institute of Education Sciences, U.S. Department of Education (R305A170488), to the University of Chicago. We are grateful to the children and families who participated in this research as well as to Carey DeMichelis who aided in data collection and project management. The opinions expressed are those of the authors and do not necessarily represent views of the Institute of Education Sciences or the U.S. Department of Education.

References

Andrews, G., & Halford, G. S. (2002). A cognitive complexity metric applied to cognitive development. *Cognitive Psychology, 45*, 153–219.

Baddeley, A. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology, 63*, 1–29.

Begolli, K. N., Richland, L. E., & Jaeggi, S. (2015). The role of executive functions for structure-mapping in mathematics. In D. C. Noelle, R. Dale, A. S. Warlaumont, J. Yoshimi, T. Matlock, C. D. Jennings, & P. P. Maglio (Eds.), *Proceedings of the 37th annual meeting of the cognitive science society* (pp. 208–213). Austin, TX: Cognitive Science Society.

Blackwell, K. A., Chatham, C. H., Wiseheart, M., & Munakata, Y. (2014). A developmental window into trade-offs in executive function: The case of task switching versus response inhibition in 6-year-olds. *Neuropsychologia, 62*, 356–364.

Blackwell, K. A., & Munakata, Y. (2014). Costs and benefits linked to developments in cognitive control. *Developmental Science, 17*, 203–211.
Blakey, E., Visser, I., & Carroll, D. J. (2016). Different executive functions support different kinds of cognitive flexibility: Evidence from 2-, 3-, and 4-year-olds. *Child Development, 87*, 513–526.
Carlson, S. M., Zelazo, P. D., & Faja, S. (2013). Executive function. In P. D. Zelazo (Ed.), *The Oxford handbook of developmental psychology*, Vol. 1: *Body and mind* (pp. 706–743.) New York: Oxford University Press.
Carlson, S. M., Moses, L. J., & Breton, C. (2002). How specific is the relation between executive function and theory of mind? Contributions of inhibitory control and working memory. *Infant and Child Development, 11*, 73–92.
Cepeda, N. J., Kramer, A. F., & Gonzalez de Sather, J. (2001). Changes in executive control across the life span: Examination of task-switching performance. *Developmental Psychology, 37*, 715–730.
Chen, Z. (1996). Children's analogical problem solving: The effects of superficial, structural, and procedural similarity. *Journal of Experimental Child Psychology, 62*, 410–431.
Chevalier, N., Blaye, A., Dufau, S., & Lucenet, J. (2010). What visual information do children and adults consider while switching between tasks? Eye-tracking investigation of cognitive flexibility development. *Developmental Psychology, 46*, 955–972.
Cho, S., Holyoak, K. J., & Cannon, T. D. (2007). Analogical reasoning in working memory: Resources shared among relational integration, interference resolution, and maintenance. *Memory & Cognition, 35*, 1445–1455.
Christie, S., & Gentner, D. (2010). Where hypotheses come from: Learning new relations by structural alignment. *Journal of Cognition and Development, 11*, 356–373.
Cowan, N. (2010). The magical mystery four: How is working memory capacity limited, and why? *Current Directions in Psychological Science, 19*, 51–57.
Crone, E. A., Wendelken, C., Donohue, S., van Leijenhorst, L., & Bunge, S. A. (2006). Neurocognitive development of the ability to manipulate information in working memory. *Proceedings of the National Academy of Sciences of the United States of America, 103*, 9315–9320.
Daehler, M. W., & Chen, Z. (1993). Protagonist, theme, and goal object: Effects of surface features on analogical transfer. *Cognitive Development, 8*, 211–229.
Davidson, M. C., Amso, D., Anderson, L. C., & Diamond, A. (2006). Development of cognitive control and executive functions from 4 to 13 years: Evidence from manipulations of memory, inhibition, and task switching. *Neuropsychologia, 44*, 2037–2078.
Delis, D. C., Kaplan, E., & Kramer, J. H. (2001). *Delis-Kaplan Executive Function System (D-KEFS)*. New York: Psychological Corporation.
Diamond, A. (2013). Executive functions. *Annual Review of Psychology, 64*, 135–168.
Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science, 11*, 19–23.
Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. *Developmental Psychology, 40*, 177–190.
Gentner, D. (1988). Metaphor as structure mapping: The relational shift. *Child Development, 59*, 47–59.
Gentner, D. (2003). Why we're so smart. In D. Gentner & S. Goldin-Meadow (Eds.), *Language in mind: Advances in the study of language and thought* (pp. 195–235). Cambridge, MA: MIT Press.
Gentner, D., & Rattermann, M. J. (1991). Language and the career of similarity. In S. A. Gelman & J. P. Byrnes (Eds.), *Perspectives on language and thought: Interrelations in development* (pp. 225–277). New York: Cambridge University Press.
Gentner, D., & Toupin, C. (1986). Systematicity and surface similarity in the development of analogy. *Cognitive Science, 10*, 277–300.
Gershon, R. C., Wagster, M. V., Hendrie, H. C., Fox, N. A., Cook, K. F., & Nowinski, C. J. (2013). NIH Toolbox for assessment of neurological and behavioral function. *Neurology, 80*(11, Suppl 3), S2–S6.
Gerstadt, C. L., Hong, Y. J., & Diamond, A. (1994). The relationship between cognition and action: Performance of children 3½–7 years old on a Stroop-like day–night test. *Cognition, 53*, 129–153.
Gick, M. L., & Holyoak, K. J. (1980). Analogical problem solving. *Cognitive Psychology, 12*, 306–355.
Goldwater, M. B., & Schalk, L. (2016). Relational categories as a bridge between cognitive and educational research. *Psychological Bulletin, 142*, 729–757.
Hackman, D. A., Gallop, R., Evans, G. W., & Farah, M. J. (2015). Socioeconomic status and executive function: Developmental trajectories and mediation. *Developmental Science, 18*, 686–702.
Halford, G. S. (1993). *Children's understanding: The development of mental models*. Hillsdale, NJ: Lawrence Erlbaum.
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.
Hofstadter, D., & Sander, E. (2013). *Surfaces and essences: Analogy as the fuel and fire of thinking*. New York: Basic Books.
Koch, I., Gade, M., Schuch, S., & Philipp, A. M. (2010). The role of inhibition in task switching: A review. *Psychonomic Bulletin & Review, 17*, 1–14.
Krawczyk, D. C., Morrison, R. G., Viskontas, I., Holyoak, K. J., Chow, T. W., Mendez, M. F., ... Knowlton, B. J. (2008). Distraction during relational reasoning: The role of prefrontal cortex in interference control. *Neuropsychologia, 46*, 2020–2032.
Kurtz, K. J. (2005). Re-representation in comparison: Building an empirical case. *Journal of Experimental & Theoretical Artificial Intelligence, 17*, 447–459.
Little, M. (2017). Racial and socioeconomic gaps in executive function skills in early elementary school: Nationally representative evidence from the ECLS-K: 2011. *Educational Researcher, 46*(2), 103–109.
Loewenstein, J., & Gentner, D. (2005). Relational language and the development of relational mapping. *Cognitive Psychology, 50*, 315–353.
Loewenstein, J., Thompson, L., & Gentner, D. (1999). Analogical encoding facilitates knowledge transfer in negotiation. *Psychonomic Bulletin & Review, 6*, 586–597.
Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology, 41*, 49–100.
Morrison, R. G., & Cho, S. (2008). Neurocognitive process constraints on analogy: What changes to allow children to reason like adults? *Behavioral and Brain Sciences, 31*, 391–392.

Morrison, R. G., Doumas, L. A., & 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.
Morrison, R. G., Krawczyk, D. C., Holyoak, K. J., Hummel, J. E., Chow, T. W., Miller, B. L., & Knowlton, B. J. (2004). A neurocomputational model of analogical reasoning and its breakdown in frontotemporal lobar degeneration. *Journal of Cognitive Neuroscience, 16*, 260–271.
Mungas, D., Reed, B. R., Marshall, S. C., & González, H. M. (2000). Development of psychometrically matched English and Spanish language neuropsychological tests for older persons. *Neuropsychology, 14*, 209–223.
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*, 453–478.
Richland, L. E., & Burchinal, M. R. (2013). Early executive function predicts reasoning development. *Psychological Science, 24*, 87–92.
Richland, L. E., & McDonough, I. M. (2010). Learning by analogy: Discriminating between potential analogs. *Contemporary Educational Psychology, 35*, 28–43.
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.
Richland, L. E., & Simms, N. (2015). Analogy, higher order thinking, and education. *Wiley Interdisciplinary Reviews: Cognitive Science, 6*, 177–192.
Rueda, M. R., Fan, J., McCandliss, B. D., Halparin, J. D., Gruber, D. B., Lercari, L. P., & Posner, M. I. (2004). Development of attentional networks in childhood. *Neuropsychologia, 42*, 1029–1040.
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.
Thibaut, J. P., French, R., & Vezneva, M. (2010a). Cognitive load and semantic analogies: Searching semantic space. *Psychonomic Bulletin & Review, 17*, 569–574.
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*, 1–19.
Tulsky, D. S., Carlozzi, N. E., Chevalier, N., Espy, K. A., Beaumont, J. L., & Mungas, D. (2013). NIH Toolbox Cognition Battery (CB): Measuring working memory. *Monographs of the Society for Research in Child Development, 78*(4), 70–87.
Vales, C., & Smith, L. B. (2015). Words, shape, visual search and visual working memory in 3-year-old children. *Developmental Science, 18*, 65–79.
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.
Waltz, J. A., Knowlton, B. J., Holyoak, K. J., Boone, K. B., Mishkin, F. S., de Menezes Santos, M., ... Miller, B. L. (1999). A system for relational reasoning in human prefrontal cortex. *Psychological Science, 10*, 119–125.
Waltz, J. A., Lau, A., Grewal, S. K., & Holyoak, K. J. (2000). The role of working memory in analogical mapping. *Memory & Cognition, 28*, 1205–1212.
White, I. R., Royston, P., & Wood, A. M. (2011). Multiple imputation using chained equations: Issues and guidance for practice. *Statistics in Medicine, 30*, 377–399.
Wiebe, S. A., Sheffield, T., Nelson, J. M., Clark, C. A., Chevalier, N., & Espy, K. A. (2011). The structure of executive function in 3-year-olds. *Journal of Experimental Child Psychology, 108*, 436–452.
Yan, J., Forbus, K., & Gentner, D. (2003). A theory of rerepresentation in analogical matching. In R. Alterman & D. Kirsh (Eds.), *Proceedings of the 25th annual meeting of the Cognitive Science Society* (pp. 1265–1270). Austin, TX: Cognitive Science Society.
Zelazo, P. D., Anderson, J. E., Richler, J., Wallner-Allen, K., Beaumont, J. L., & Weintraub, S. (2013). NIH Toolbox Cognition Battery (CB): Measuring executive function and attention. *Monographs of the Society for Research in Child Development, 78*(4), 16–33.
Zelazo, P. D., Frye, D., & Rapus, T. (1996). An age-related dissociation between knowing rules and using them. *Cognitive Development, 11*, 37–63.