**PAPER**

# Neuroscientific insights into the development of analogical reasoning

Kirstie J. Whitaker[1,*]  |  Michael S. Vendetti[2,3,*]  |  Carter Wendelken[2,4]  |  Silvia A. Bunge[2,5]

[1]Brain Mapping Unit, Department of Psychiatry, University of Cambridge, UK

[2]Helen Wills Neuroscience Institute, University of California at Berkeley, USA

[3]Oracle Corporation, Redwood City, California, USA

[4]Vicarious FPC, Inc, Union City, CA, USA

[5]Department of Psychology, University of California at Berkeley, USA

**Correspondence**
Professor Silvia A. Bunge, Building Blocks of Cognition Laboratory, 134 Barker Hall, Helen Wills Neuroscience Institute, University of California at Berkeley, Berkeley, CA, 94720, USA.
Email: sbunge@berkeley.edu

## Abstract

Analogical reasoning, or the ability to find correspondences between entities based on shared relationships, supports knowledge acquisition. As such, the development of this ability during childhood is thought to promote learning. Here, we sought to better understand the mechanisms by which analogical reasoning about semantic relations improves over childhood and adolescence (e.g. chalk is to chalkboard as pen is to...?). We hypothesized that age-related differences would manifest as differences in the brain regions associated with one or more of the following cognitive functions: (1) controlled semantic retrieval, or the ability to retrieve task-relevant semantic associations; (2) response control, or the ability to override the tendency to respond to a salient distractor; and/or (3) relational integration, or the ability to consider jointly two mental relations. In order to test these hypotheses, we analyzed patterns of fMRI activation during performance of a pictorial propositional analogy task across 95 typically developing children between the ages of 6 and 18 years old. Despite large age-related differences in task performance, particularly over ages 6–10 but through to around age 14, participants across the whole age range recruited a common network of frontal, parietal and temporal regions. However, activation in a brain region that has been implicated in controlled semantic retrieval – left anterior prefrontal cortex (BA 47/45) – was positively correlated with age, and also with performance after controlling for age. This finding indicates that improved performance over middle childhood and early adolescence on this analogical reasoning task is driven largely by improvements in the ability to selectively retrieve task-relevant semantic relationships.

## RESEARCH HIGHLIGHTS

- In a large sample of children spanning ages 6–18 years, this study demonstrates pronounced age-related improvements in analogical reasoning between ages 6 and 10, and continued improvement until mid-adolescence.
- We used neuroimaging to distinguish among several plausible cognitive accounts of the development of analogical reasoning.
- This work demonstrates that the development of analogical reasoning is associated with increased engagement of the left anterior inferior prefrontal cortex (BA 47/45), previously shown to be associated with the ability to select among competing semantic associations.
- Improvements over this age range were not observed in brain regions linked to domain-general processes underlying response control or relational thinking.

## 1 | INTRODUCTION

Analogical reasoning, or the ability to find correspondences between individual objects as well as their relationships (Gentner, 1983;

*Shared first authorship.

Hummel & Holyoak, 1997, 2003), is central to learning and thought (Hofstadter & Sander, 2012; Holyoak & Thagard, 1995; Namy & Gentner, 2002). Across diverse real-world environments, analogies are employed to explain new information (Dunbar & Blanchette, 2001), to solve problems (Gick & Holyoak, 1980, 1983), and to facilitate the learning of new information in educational settings (Richland, Zur, & Holyoak, 2007; Richland & McDonough, 2010; Vendetti, Matlen, Richland, & Bunge, 2015).

Given the ubiquitous use of analogies in understanding new domains, reasoning by analogy is widely considered to be an important tool underlying children's acquisition of knowledge (Gentner, 1988; Gentner & Rattermann, 1991; Goswami, 1996; Halford, 1992; Halford, Wilson, & Phillips, 1998; Kotovsky & Gentner, 1996). Research related to the development of analogical reasoning ability has yielded two major behavioral findings. First, contrary to findings from earlier research on the development of analogical reasoning (e.g. Sternberg & Nigro, 1980), young children can in some cases solve analogy problems (e.g. Holyoak, Junn, & Billman, 1984). Second, content knowledge is an important factor in predicting one's ability to use analogies effectively (Goswami & Brown, 1990; Gentner & Rattermann, 1991).

It is important to note, however, that children are still prone to making certain types of errors during analogical reasoning. For example, one study showed that 3- to 5-year olds will make object-matching errors, choosing an incorrect response based on perceptual similarity (Gentner & Medina, 1998). Another showed that children up to 11 years of age are likely to select items in a propositional analogy task solely on the basis of semantic associations (Sternberg & Nigro, 1980). These patterns of results indicate that there is still much room for improvement after children are first able to make relational matches, and that in order to solve analogies they must learn to look beyond the fact that two items are related and examine *how* they are related. The ability to select the appropriate semantic relation, termed controlled semantic retrieval, is an important mechanism for solving semantic analogy problems. This cognitive process allows the reasoner to select specifically analogous semantic associations in the face of other, more general, semantic associations (Bunge, Wendelken, Badre, & Wagner, 2005; Krawczyk et al., 2010).

Beyond improvements in domain-specific knowledge and controlled semantic retrieval, several other mechanisms have been proposed to underlie the development of analogical reasoning ability, including the ability to maintain multiple mental representations in working memory and integrate distinct mental relations (e.g. Halford, Wilson, & Phillips, 1998; Holyoak & Kroger, 1995) and to suppress interference from irrelevant information (e.g. Richland, Morrison, & Holyoak, 2006; Morrison, Doumas, & Richland, 2011). Thus, although there is wide consensus on the pervasive role of analogical reasoning in children's acquisition and use of knowledge, several domain-general cognitive processes – working memory, interference suppression, and relational integration – have also been shown to contribute to reasoning ability (Cho, Holyoak, & Cannon, 2007), one or more of which could be the key driver(s) of the development of this high-level cognitive ability.

There has been much previous work demonstrating a shift from matching items based on perceptual to more abstract features occurring between the ages of 4 and 6 years (e.g. Rattermann & Gentner, 1998), but much less work examining developmental changes beyond age 6 years. Therefore, we wanted to investigate how the number of semantic versus perceptual errors changes throughout childhood. Beyond choosing a response based on semantic versus purely perceptual features, we were also interested in identifying when children would be able to reliably choose items that matched based on a specific semantic association analogous to that shared between the A and B items.

Preliminary fMRI data from our laboratory (Wright, Matlen, Baym, Ferrer, & Bunge, 2008) pointed to age-related differences in analogical reasoning between 6–13-year-olds (N = 16) and 19–26-year-olds (N = 17), but this earlier work was limited by a small sample size, a child group spanning a broad age range, and a task design wherein analogy trials either contained a semantic or a perceptual lure, but not both. In the current study, we sought to collect data from a much larger pediatric sample, use an up-to-date analytic approach, including appropriate correction for multiple comparisons, and directly pit the influence of semantic and perceptual lures against each other on each analogy trial.

The current study investigated the neural correlates of analogical reasoning ability in over 130 children aged 6 to 18 years. Participants solved two types of trials: (1) semantic match, in which they had to choose a response that was semantically related to the target item; and (2) propositional analogy (i.e. A: B:: C:?), in which they had to choose a response that was semantically similar to an item (C term) in an analogous fashion to a semantic relation displayed between two other items (A and B terms). On each analogy trial, participants had to choose the correct response out of four possible choices. Responses for analogy trials consisted of the correct response (i.e. the answer choice that completes the analogy), a perceptual lure (i.e. an item that was perceptually similar to the C term), a semantic lure (i.e. an item that was semantically related to the C term, but did not complete the analogy), and a lure that was unrelated both perceptually and semantically.

Our aims in this study were twofold. First, given that prior developmental research on analogical reasoning focused largely on improvements in early childhood (~ages 4–6), we sought to determine whether improvements would also be observed later in childhood, and potentially even in adolescence. In addition to testing for age differences in overall accuracy, we sought to test for age differences in the types of errors participants made, namely in the selection of semantic, perceptual, or unrelated distractors. Second, we sought to pinpoint the key neurocognitive mechanism(s) that underlie improvements in analogical reasoning over childhood.

In prior adult fMRI research using a difficult verbal propositional analogy task (e.g. 'agile: acrobat:: eloquent: orator?' yes/no), we had dissociated the roles of three subregions within lateral prefrontal cortex (PFC) by manipulating task difficulty in several ways (Bunge et al., 2005). We found that left anterior inferior prefrontal cortex (aLIPC; ~BA 45/47) was sensitive to a manipulation of semantic relatedness between the A and B words; it was engaged more strongly when it was necessary to retrieve a relation among more

weakly associated terms than among strongly associated ones, consistent with prior work showing that this region is associated with effortful, or controlled, semantic memory retrieval (e.g. Wagner, Paré-Blagoev, Clark, & Poldrack, 2001). Both left aLIPC and right dorsolateral PFC (DLPFC; ~BA 9) – a region implicated in cognitive control (e.g. Miller & Cohen, 2001; Stuss & Knight, 2012) – were engaged more strongly when it was necessary to reject an invalid analogy involving mismatched pairs of semantic associates (e.g. 'variable: equation: clay: sculpture?') than to accept a valid one. Finally, left aLIPC and left rostrolateral prefrontal cortex (RLPFC; ~BA 10) – a region that had been hypothesized to play a role in the ability to jointly consider two mental relations (e.g. Christoff et al., 2001; Kroger et al., 2002) – were engaged more strongly when participants were asked to consider whether two word pairs formed a valid analogy than when they were asked to conduct a semantic relatedness judgment on one of the two word pairs. In summary, then, we found that several PFC regions contributed in different ways to analogical reasoning in adults. Left aLIPC was sensitive to the semantic relatedness manipulation, implicating it in the process of retrieving semantic relations among words in a pair. Neither RLPFC nor DLPFC was sensitive to this manipulation, consistent with their purported involvement in post-retrieval, domain-independent processes such as relational integration (Krawczyk, 2012) and response selection (Aron, Robbins, & Poldrack, 2014).

Another fMRI study of analogical reasoning in young adults, using a different paradigm, independently manipulated demands for interference resolution and relational integration (Cho et al., 2010). In this study, bilateral DLPFC and posterior left inferior prefrontal cortex (pLIPC), as well as right aLIPC, were sensitive to the need to ignore distracting information; by contrast, RLPFC was specifically sensitive to the need to integrate multiple relations. Thus, consistent with our prior work, multiple PFC regions are recruited during analogical reasoning, and their roles are dissociable.

Here, we sought to test whether improvements in analogical reasoning over ages 6–18 years were related to changes in one or more of these PFC regions. If age-related changes over this age range are driven in large part by an improved ability to access and select task-relevant semantic knowledge (e.g. Gentner & Ratterman, 1991), then we would expect age-related improvements in analogical reasoning to be most closely linked to changes in left aLIPC, a region that has been implicated in controlled semantic retrieval, or the process of retrieving task-relevant semantic relations while ignoring task-irrelevant ones (e.g. Badre, Poldrack, Paré-Blagoev, Insler, & Wagner, 2005; Bunge et al., 2005; Poldrack et al., 1999; Souza, Donohue, & Bunge, 2009; Wagner et al., 2001). Left aLIPC has, beyond its role in selecting among competing semantic relations, been implicated more generally in the ability to resolve conflict among conflicting working memory representations (e.g. Novick, Kan, Trueswell, & Thompson-Schill, 2009; Thompson-Schill et al., 2002). In the context of this task, if we were to find that age-related improvements in analogical reasoning were linked primarily to left aLIPC, we would infer that they are driven largely by an improved ability to retrieve and/or select task-relevant semantic relations.

By contrast, if age-related changes over the age range investigated in this study are driven largely by an improved ability to inhibit impulsive responses to distractors (e.g. Richland et al., 2006; Morrison et al., 2011), then we would expect to find age-related improvements associated with activation in PFC regions implicated in voluntary control over action (e.g. Passingham, 1995). One such region is right VLPFC (both posterior and anterior inferior PFC; BA 44; 45, 47), which has been strongly linked to response inhibition (aLIPC as well as pLIPC; for a review, see Aron et al., 2014). Other candidate regions include bilateral DLPFC and pLIPC, which have been implicated more broadly in goal-directed behavior – including but not limited to selection among competing responses on the basis of task rules (e.g. see Bunge, 2004). Thus, if we were to find that age-related improvements in analogical reasoning were linked primarily to right VLPFC, left pLIPC, and/or DLPFC, we would infer that they are driven largely by improvements in the ability to select between competing responses, overriding the tendency to respond to a salient distractor (Cho et al., 2010).

Finally, if age-related changes in analogical reasoning over ages 6–18 years are driven by improvements in a domain-general capacity for relational thinking (e.g. Halford et al., 1998; Dumontheil, 2014), we would predict age-related improvements associated with activation in RLPFC. This region and the inferior parietal lobule (IPL) have been implicated in relational integration across a range of tasks (for a review, see Vendetti & Bunge, 2014). RLPFC and IPL have been hypothesized to play a domain-general role in relational thinking, as they are engaged when encoding both semantic and visuospatial relations (e.g. Wendelken, Chun, & Bunge, 2012). In addition, the collective results of several studies have shown that developmental improvements in non-semantic relational reasoning can be linked strongly to changes in the activation and functional connectivity of RLPFC and IPL (Dumontheil et al., 2010; Wendelken, O'Hare, Whitaker, Ferrer, & Bunge, 2011; Wendelken, Ferrer, Whitaker, & Bunge, 2016). Thus, if we were to find that age-related improvements in analogical reasoning were linked primarily to these regions, we would infer that they are driven largely by improvements in relational thinking.

Notably, although these hypotheses are not mutually exclusive – that is, developmental improvements in analogical reasoning could be the result of changes in multiple underlying neurocognitive processes, given that analogical reasoning draws on multiple lower-level cognitive functions (e.g. Krawczyk, 2012) – each prediction is associated with a spatially dissociable pattern of brain activation. Our preliminary developmental fMRI study of analogical reasoning (Wright et al., 2008) pointed to age-related changes in the activation of a number of the brain regions outlined above, but – given its small sample size and the strong correlation between age and task performance – could not distinguish between brain regions whose activation is associated with task performance and those that merely become more active with age while performing the task. In the current study, we sought to identify the key drivers of age-related and individual differences in analogical reasoning over ages 6 to 18 years: improvements in controlled semantic memory retrieval, response inhibition, and/or relational integration.
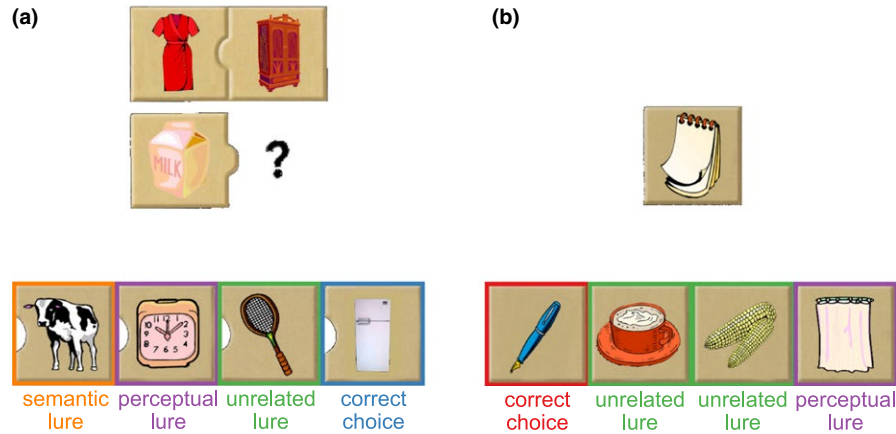
**FIGURE 1** (a) Analogy Task. Participants indicated which of the four choices was associated with an item in analogous fashion to the relation shared between the top two items. In this example, the refrigerator is associated with the milk carton (i.e., refrigerator stores the milk carton) in an analogous way to the dress and the closet items. On each trial, a semantic lure (e.g., cow), a perceptual lure (e.g., clock), and an unrelated lure (e.g., tennis racket) were included. Thus, participants' correct choice was based on understanding the correct semantic association and disregarding irrelevant semantic or perceptual information. (b) Semantic matching task. Participants decide which of the four choices share a semantic relationship with a target object. In this example, the pen is used to write on the notepad, and thus is the item with the strongest semantic association. On each semantic trial, a perceptual lure (e.g., a shower curtain) as well as two unrelated lures were included. Thus, participants had to understand that the correct choice was based on semantic, rather than just perceptual, associations

## 2 | METHODS

### 2.1 | Participants

This study included 138 typically developing individuals (81 males; 1234 right-handed) from the Neurodevelopment of Reasoning Ability (NORA) study, a project designed to examine the behavioral and neural factors that underlie changes in reasoning ability through childhood and adolescence. Analyses of the development of functional specialization and white matter microstructure using data from the NORA cohort have previously been published (Ferrer et al., 2013; Wendelken et al., 2011). The age range was 6.2 to 18.9 years ($M$ = 11.0, $SD$ = 3.6). Participants were excluded if they had a neurological impairment, psychiatric illness, or a history of learning disability and/or developmental delay. Participants' fMRI data were included if at least two of four runs of usable fMRI data. Each run was considered usable if it covered the whole brain and had a root mean square translational movement of less than 1 mm. All participants and their parents gave their informed assent (under 12 years of age) or consent (aged 12 or older) to participate in the study, which was approved by the Committee for Protection of Human Subjects at the University of California at Berkeley.

### 2.2 | Study design

#### 2.2.1 | Experimental task

The task included two conditions: analogy and semantic match problems. The task was based on a test from the Kaufman Brief Intelligence Test, 2nd edition (KBIT-2), designed for use with children. Our laboratory designed the analogy and semantic trials using Adobe Photoshop, and made use of line drawings from 'The Big Box of Art: 1 Million'. All

stimuli were pictures of common objects known to young children, as judged by age-of-acquisition psycholinguistic norms for the words that they depicted (Gilhooly & Logie, 1980).

On analogy trials (Figure 1a), participants were shown an image containing an incomplete propositional analogy (i.e. A : B :: C :?) above a row of four items. Participants were asked to indicate with a button press which of the four possible answers best completed the array (e.g. dress is to closet as milk carton is to refrigerator). Specifically, participants were asked to choose which of the bottom pictures best fills in the question mark. They were told that it should go with the middle picture (e.g. milk carton) in the same way that the top two pictures go together (i.e. in the same way that the dress goes with the closet). They were then told that, 'there may be other pictures that you think go with the C term (e.g. the milk carton), but you should choose the picture that goes with the milk carton in the same way that the dress goes with the closet'. Each trial contained a perceptual lure; for example, in Figure 1a, the clock has a similar shape and color to the milk carton. In addition, analogy trials had a semantic lure that did not complete the analogy; in Figure 1a, the cow represented the semantic lure. Finally, the response choices contained an unrelated lure that was neither perceptually nor semantically related to the terms in the analogy problem.

On semantic match trials (Figure 1b), participants saw one item (e.g. notepad) presented above a row of four items. Participants were asked to indicate with a button press which of four possible answers 'went best with it' (e.g. pen). Response choices on semantic trials included the correct semantic associate, a perceptual lure (e.g. a shower curtain with a similar shape and color to the notepad), and two perceptually and semantically unrelated lures. The experimenter reinforced the rule by stating the correct choice (i.e. the semantic association) for each example after the participant had responded.

## 2.3 | Data acquisition

### 2.3.1 | Training procedure

On an initial visit to the laboratory, prior to the day of scanning, children between ages 6 and 12 years undertook a mock scan designed to minimize movement and participant anxiety during the MRI scan. The mock scan was conducted in a decommissioned Varian MRI scanner, which did not have an active magnetic field, but accurately recreated the experience of lying in an MRI scanner. Participants wore headphones and listened to recorded sounds of MRI acquisition sequences, and watched a cartoon movie on a mirror system. The researcher reminded children to stay still when they moved their arms or legs, and a head-tracker was used to provide feedback when the participant moved their head more than 2 mm. This mock scan lasted about 15 minutes. On the day of the MRI scan, a research assistant explained the task outside the scanner. Participants practiced five semantic and five analogy questions. If they answered a problem incorrectly, the researcher corrected them and the task was explained again. Once in the scanner, participants practiced using the button box to respond to the same practice questions before the scan acquisition.

During the fMRI scan, stimulus presentation and button-press responses were collected using the Presentation psychological experimentation software (Neurobehavioral Systems). Semantic and analogy trials were presented in a random order in a fast event-related design. Each of the four task runs consisted of 10 semantic and 10 analogy trials and lasted 4 minutes. Participants were given up to 10 s to answer each question; once they answered, a fixation cross appeared and stayed on the screen until the end of the 10-s interval. Children were encouraged to respond as soon as they knew the answer, but were reminded that picking the right answer was more important than responding quickly.

### 2.3.2 | MRI data acquisition

Brain imaging data were collected at UC Berkeley on a 3-T Siemens Trio TIM MR scanner using a 12-channel head coil with a maximum gradient strength of 40mT/m. Functional MRI data of the whole brain was acquired using echo-planar imaging (TR = 2000 ms; TE = 25 ms; 2.0 × 1.8 × 3.0 mm voxels; 33 slices oriented along the anterior commissure–posterior commissure axis; no interslice gap; flip angle = 90°; field of view = 230 mm; 120 volumes per run). Parallel acquisition (GRAPPA) was used with an acceleration factor of 2, and a gradient-echo echo-planar pulse Prospective Acquisition Correction (3D-PACE) sequence was used to minimize motion artifacts by prospectively adjusting scan parameters throughout a run on the basis of real-time assessment of head motion (Siemens Medical Solutions; Thesen, Heid, Mueller, & Schad, 2000). A T1-weighted image was also acquired in each participant for image registration (MPRAGE; TR = 2300 ms; TE = 2.98 ms; 1 mm$^3$ isotropic voxels).

### 2.3.3 | FMRI pre-processing

FMRI data processing was carried out using FEAT (FMRI Expert Analysis Tool) Version 6.00, part of FSL (FMRIB's Software Library, www.fmrib.ox.ac.uk/fsl). The following pre-statistics processing was applied: motion correction using MCFLIRT (Jenkinson, Bannister, Brady, & Smith, 2002); slice-timing correction using Fourier-space time-series phase-shifting; non-brain removal using BET (Smith et al., 2004); spatial smoothing using a Gaussian kernel of FWHM 5 mm; grand-mean intensity normalization of the entire 4D dataset by a single multiplicative factor; high-pass temporal filtering (Gaussian-weighted least-squares straight line fitting, with sigma = 50.0 s). Runs were excluded at this stage if the root mean square translational movement between volume acquisitions was greater than 1 mm (83 runs excluded out of a possible 524; 16%). We chose this liberal inclusion criterion a priori in order to minimize data loss, while also ensuring that data quality was not excessively degraded (Galván, Leijenhorst, & McGlennan, 2012). If a participant did not have useable data for at least two of the four runs, they were excluded from further analyses (35 participants excluded; 25%), including 20 runs that were not collected because the participant asked to end the acquisition early (4%). A further 8 participants (6%) had miscellaneous errors in their processing pipeline which meant they could not be included in group analyses. The final neuroimaging sample included 95 participants (age range: 6.3 to 18.9 years, M = 11.7, SD = 3.5).

## 2.4 | Statistical analyses

### 2.4.1 | Behavioral analyses

We analyzed participants' accuracy and response times (RTs) on analogy and semantic trials. We tested the following hypotheses using paired t-tests: participants would be more accurate and faster on semantic than on analogy trials, and they would make more semantic errors than perceptual errors and more perceptual errors than unrelated errors.

We fitted linear and quadratic models to predict, based on their age, participants' accuracy, RTs on correct trials, and number of semantic, perceptual and unrelated lure errors. The best-fitting model was selected according to the adjusted $R^2$ values of the models. We also investigated whether analogy accuracy explained the relationship between age and semantic accuracy, or whether semantic accuracy explained the relationship between age and analogy accuracy.

All behavioral analyses were two-tailed and assessed at $\alpha$ = .05. Although we do not correct for multiple comparisons, as the hypotheses were non-independent, we note that the Bonferroni correction for all tests described above is .003 (.05/16).

### 2.4.2 | Individual subject analyses

Time-series statistical analyses for individual runs were carried out using FILM with local autocorrelation correction (Woolrich et al., 2009). Although runs with large amounts of motion had been excluded from the analysis, there remained the possibility that motion could corrupt the fMRI signal. In order to alleviate this concern, six motion regressors (indexing translation and rotation in the x-, y- and z-dimensions for each 2-s TR) were included in the analysis as covariates
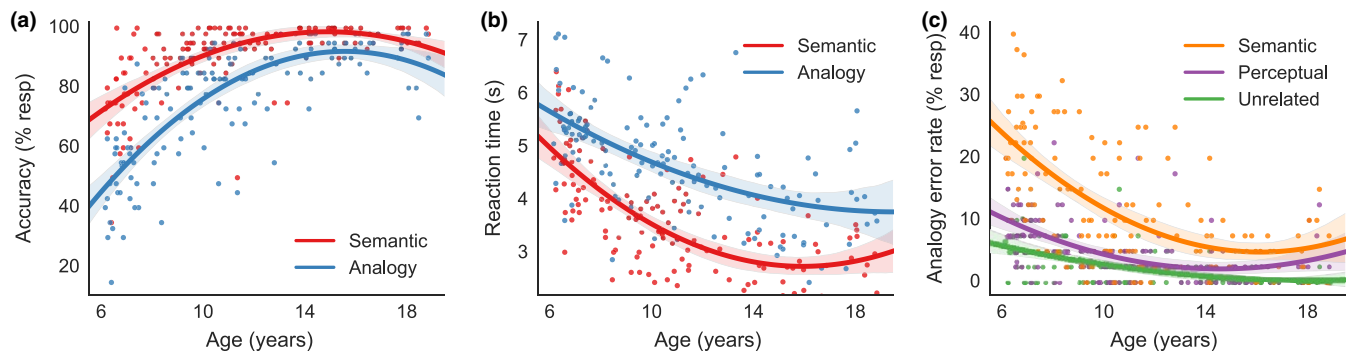
**FIGURE 2** Accuracy improved with age (a), while response times decreased (b) for both semantic (red) and analogy (blue) trials. The number of errors on analogy trials decreased with age (c), and semantic lures (yellow) were more common than perceptual lures (purple), which were themselves more common than unrelated lures (green) across all ages (c). Lines for all plots represent 95% confidence intervals around predicted values using a quadratic model

of no interest, along with time-series data representing signal in white matter, cerebrospinal fluid, and areas outside the brain. Regressors representing correct trials and the four types of error trials (semantic, perceptual, unrelated distractor, and omission) were modeled separately as events starting when the question appeared on the screen and ending at the time the subject answered the question (or when the question disappears in the case of omitted responses). Analogy and semantic trials and their associated errors were modeled separately, yielding up to nine task regressors. The temporal derivative of these modeled events was also included.

### 2.4.3 | Group fMRI statistics

Scanning runs that passed our motion criterion were combined for each participant using a fixed-effects model, after which group-level analyses were performed using a random-effects model (Beckmann, Jenkinson, & Smith, 2003; Woolrich, Behrens, Beckmann, Jenkinson & Smith, 2004; Woolrich, 2008). For the fixed-effects analyses, the contrasts of correct semantic events > fixation, correct analogy events > fixation and correct analogy events > correct semantic events were averaged over each participant's completed runs.

First, we quantified mean activation across all participants on each of three whole-brain contrasts: analogy > fixation, semantic > fixation, and analogy > semantic. We then used these results to generate a liberal mask of regions that were not deactivated by either the semantic or the analogy tasks relative to fixation, by excluding regions that showed effects significantly less than zero for either semantic or analogy tasks. Next, we computed two whole-brain regressions across all participants to examine effects of age, as well as effects of accuracy while correcting for any effects of age. We limited our analyses to regions that were not significantly deactivated by the task using the mask described above. (For readability, we refer to this mask as showing regions activated by the task.) Z (Gaussianised T/F) statistical images were assessed for significance using permutation tests at a cluster defining threshold of $Z > 2.3$ and family-wise error correction at $p < .05$ (Eklund, Nichols, & Knutsson, 2016; Worsley, 2001). All thresholded and un-thresholded maps were uploaded to

NeuroVault (Gorgolewski et al., 2016). Code to reproduce the figures in this manuscript can be found at https://github.com/KirstieJane/NORA_WhitakerVendetti_DevSci2017.

## 3 | RESULTS

### 3.1 | Behavioral performance

Participants across the entire age range were faster and more accurate on semantic than on analogy problems (mean ± standard deviation; semantic: accuracy = 89.38 ± 12.52%, RT = 3.60 ± 0.99s; analogy: accuracy = 74.95 ± 18.96 %, RT = 4.67 ± 1.01 s; paired $t$-tests: accuracy: $t(137) = 15.01$, $p < .001$; RT: $t(137) = 23.72$, $p < .001$). For both semantic and analogy problems, we observed significant increases in performance across the age range of our participants (Figure 2a). The effect of age on accuracy on both trial types was best fitted with a quadratic model, such that the greatest improvements occurred through the younger ages, with a plateau by the end of the teenage years (semantic: $\beta_{age}^2 = -0.34$, $p_{age}^2 < .001$; $\beta_{age} = 9.99$, $p_{age} < .001$; $R^2_{adj} = .43$, $F(2, 135) = 52.90$, $p < .001$; analogy: $\beta_{age}^2 = -0.51$, $p_{age}^2 < .001$; $\beta_{age} = -15.96$, $p_{age} < .001$; $R^2_{adj} = .61$, $F(2, 135) = 110.10$, $p < .001$). Similarly, we saw a decrease in RTs for both semantic and analogy trials (Figure 2b) that were best fitted with quadratic models (semantic: $\beta_{age}^2 = 0.02$, $p_{age}^2 < .001$; $\beta_{age} = -0.73$, $p_{age}^2 < .001$; $R^2_{adj} = .52$, $F(2, 135) = 73.69$, $p < .001$; analogy: $\beta_{age}^2 = 0.01$, $p_{age}^2 = .07$; $\beta_{age} = -0.40$, $p_{age} = .004$; $R^2_{adj} = .33$ ($R^2_{adj}$ for linear model = .32), $F(2, 135) = 34.79$, $p < .001$). Thus, we observed age-related improvements in accuracy and RTs on both semantic and analogy problems, particularly during childhood. Most strikingly, accuracy on analogy trials increased from about 50% in the youngest participants to close to 100% in the oldest.

Importantly, although we observed age-related improvements on the semantic task, improvements on the analogy task were significant even when including semantic accuracy as a covariate in the quadratic model for analogy accuracy ($\beta_{sem} = 0.81$, $p_{sem} < .001$, $\beta_{age}^2 = -0.24$, $p_{age}^2 < .001$; $\beta_{age} = 7.82$, $p_{age} < .001$; $R^2_{adj} = .78$, $F(3, 134) = 159.97$, $p < .001$). Thus, improved analogical reasoning over this age range cannot be fully accounted for by improved knowledge of

**FIGURE 3** Whole-brain activation patterns on average across all participants for the contrasts of semantic greater than fixation (yellow), analogy greater than fixation (red), and analogy greater than semantic (blue). Overlapping regions are shown in orange, green, purple and pink according to the Venn diagram, and all results are constrained to be within regions liberally activated by the semantic or analogy tasks. Statistical tests are permutation tests, and thresholds are set using a cluster defining threshold of $Z > 2.3$ and are family-wise error-corrected at $p < .05$. Thresholded and unthresholded maps are available in NeuroVault at http://neurovault.org/collections/1658
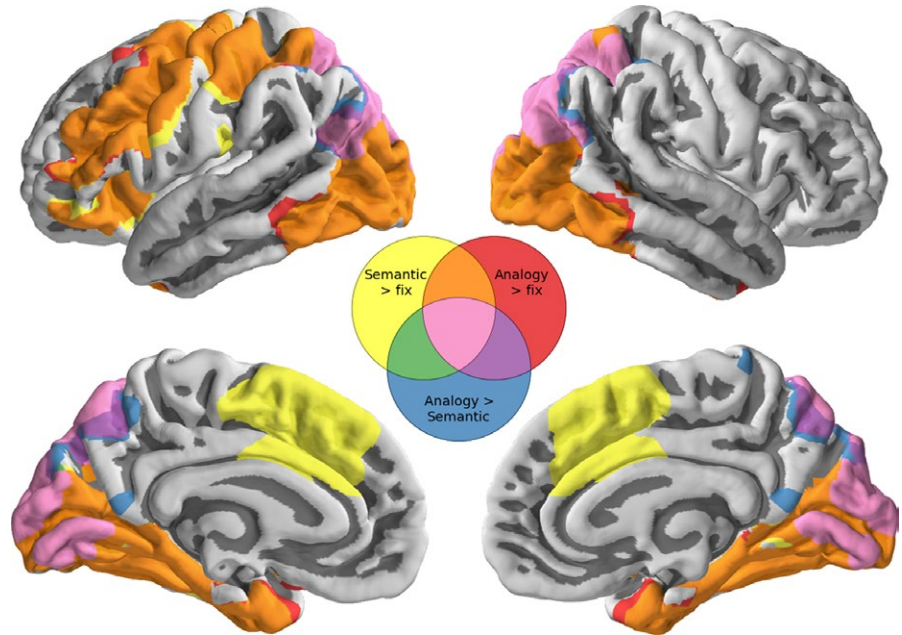
semantic associations between familiar visual stimuli. Rather, the ability to retrieve a semantic associate is necessary but not sufficient for analogical reasoning.

The number of semantic, perceptual, and unrelated lures that participants erroneously selected on analogy trials decreased with age (Figure 2c; semantic: $\beta_{age}^2 = 0.19$, $p_{age}^2 < .001$; $\beta_{age} = -6.04$, $p_{age} < .001$; $R_{adj}^2 = .40$, $F(2, 135) = 47.15$, $p < .001$; perceptual: $\beta_{age}^2 = 0.11$, $p_{age}^2 < .001$; $\beta_{age} = -3.29$, $p_{age} < .001$; $R_{adj}^2 = .27$, $F(2, 135) = 26.41$, $p < .001$; unrelated: $\beta_{age}^2 = -0.04$, $p_{age}^2 = .06$; $\beta_{age} = -1.41$, $p_{age} = .005$; $R_{adj}^2 = .26$, $Rsq_{adj}$ for linear model $= .25$, $F(2, 135) = 25.07$, $p < .001$). However, when participants of all ages made errors, they were predominantly semantic lures ($12.24\% \pm 9.62$), followed by perceptual lures ($5.07\% \pm 4.83$), and then unrelated lures ($2.81\% \pm 3.49$; paired $t$-tests: semantic > perceptual: $t(137) = 10.64$, $p < .001$; perceptual > unrelated: $t(137) = 5.42$, $p < .001$). Thus, participants across the age range most often selected an answer choice that was semantically related to the C item, be it the target stimulus or a semantic lure – and errors of all types decreased with age.

## 3.2 | Functional MRI analyses

We first investigated which brain regions were active during semantic and analogy trials, averaged across all participants (Figure 3). Overall, the same regions are active for both trial types, specifically inferior occipital, temporal, lateral prefrontal, and parietal regions (see Tables S1-2). While prefrontal, inferior occipital, and temporal regions did not show a difference in activation between these semantic and analogy trials, parietal cortex and medial occipital regions were more active for analogy than for semantic trials, perhaps related to the difference in stimulus complexity between the conditions (blue and pink in Figure 3).

In order to characterize age-related changes in activation during analogical reasoning, we correlated activation for these three contrasts

of interest (semantic > fixation, analogy > fixation and semantic > analogy) with age across all participants (Figure 4; Tables S3–4). A broad swath of visual cortex showed increased activation with age for both semantic and analogy trials, and left aLIPC (~BA 47; 45) and left DLPFC (~BA 46; 9), among other regions, additionally showed a strong age-related increase in activation for analogy trials.

We extracted parameter estimates for each comparison from within this region of a priori interest in left aLIPC/DLPFC. As defined, estimates of analogy > fixation were correlated with age ($r = .59$); however, this analysis revealed that semantic > fixation also showed this pattern ($r = .51$, $p < .001$). The correlation between analogy > semantic activation and age was significant but very weak ($r = .258$, $p = .012$). In sum, we observed age-related increases in activation on analogy and semantic trials, and the within-person increase in analogy activation was greater than the within-person increase in semantic activation, in left lateral PFC as well as visual cortex.

Our final analyses were aimed at identifying regions for which activation tracked individual differences in analogy accuracy even after correcting for the effects of age. The power of a whole-brain partial correlation analysis in this dataset is limited as a result of the strong correlations between accuracy and age, but given individual differences in performance across our large sample, we nonetheless sought to disambiguate age-related and performance-related differences in activation. We tested whether analogy accuracy related to activation on analogy trials. We also tested whether analogy accuracy would be related to activation on semantic trials, given that the ability to retrieve a semantic associate is necessary for analogical reasoning. The first of these whole-brain analyses revealed no relationship between analogy accuracy and analogy trial activation controlling for age, after correcting for multiple comparisons. However, there was a positive relationship between semantic trial activation and analogy accuracy that survived after controlling for age ($r_{acc|age} = .47$; Figure 5bii; Table S5). When we extracted individual contrast of parameter estimates
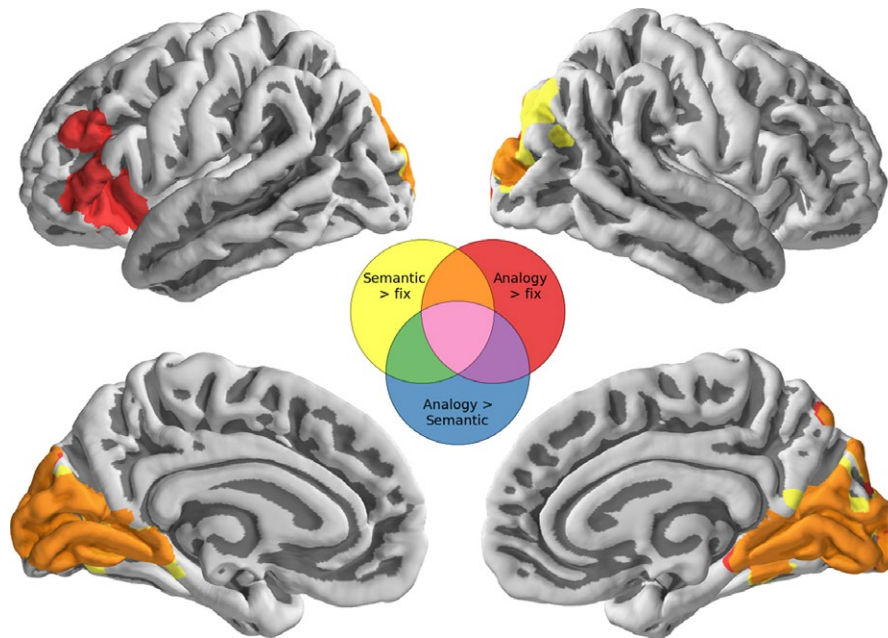
**FIGURE 4** Whole-brain activation patterns showing regions that show an increase in activation with age across all participants. Results for the semantic greater than fixation contrast are shown in yellow, analogy greater than fixation in red, and regions for which both are increasing are shown in orange. There were no regions that showed a within-person differential increase in activation during analogy trials compared to semantic trials. All results are constrained to be within regions liberally activated on average by the semantic or analogy tasks. Statistical tests are permutation tests, thresholds are set using a cluster defining threshold of $Z > 2.3$ and are family-wise error-corrected at $p < .05$. Thresholded and unthresholded maps are available in NeuroVault at http://neurovault.org/collections/1658

from this cluster in left aLIPC (~BA 47/45) – largely corresponding to the inferior portion of the prefrontal cluster identified in Figure 4 – we found that activation on analogy trials was in fact correlated with increased analogy accuracy controlling for age, as we had observed at the whole-brain level for semantic trials ($r_{acc|age}$ = .36, $p < .001$; Figure 5cii; Table S6). This region did not show a pattern of increasing specificity of activation for analogy over semantic trials ($r_{acc|age}$ = −.14, $p = .17$). Thus, left aLIPC activation on both semantic and analogy trials was correlated with analogy performance over and above effects of age.

## 4 | DISCUSSION

The current study was designed to test which mechanisms support age-related improvements in analogical reasoning from middle childhood to adolescence. Successful performance on analogy problems relied on participants' ability to choose an item whose semantic relation was analogous to that expressed in the A:B pair while avoiding choosing a non-analogous semantic associate or a perceptually similar lure. In the current study, we extended the design in Wright et al. (2008) to include both a semantic lure and a perceptual lure among the four answer choices for each analogy problem. If there is not a semantic lure present in the analogy trial, then it would not be possible to know whether children were basing their answers on semantic relatedness or actually engaging in higher-order relational reasoning. In other words, given the presence of semantic and perceptual distractors, correct performance on analogy trials should incorporate mechanisms that

specifically support retrieval of analogical relationships, rather than mechanisms that support more general retrieval processes.

Behaviorally, performance on both semantic and analogy trials improved with age, with the greatest improvements occurring between the ages of 6 and 9 years. The percentage of errors for semantic, perceptual and unrelated lures in the analogy task significantly decreased with age, but the most pronounced age-related improvement was observed for semantic lures. Even the youngest children chose semantic lures more often relative to perceptual and unrelated lures, but the number of semantic lure errors decreased dramatically over middle childhood. While the perceptual errors were subject to restricted range effects, it remains the case that even the youngest children studied, who made many errors, were more likely to choose semantically relevant items (either targets or semantic lures) than semantically irrelevant ones (perceptual or unrelated lures). This finding fits well with previous research demonstrating an ability to avoid matching based on purely perceptual attributes occurring by the age of 6 (Kotovsky & Gentner, 1996).

Although participants of all ages performed well on the semantic match task, the youngest participants in our study (~ages 6–7) correctly answered only about 50% of the analogy problems, which is lower than expected based on prior studies in younger children, such as a study reporting analogical reasoning performance of around 60% in 4-year-olds with a similar paradigm (Goswami & Brown, 1990), which may indicate that some of our reasoning problems required more advanced semantic knowledge. Our analogy task revealed the biggest age-related differences between the ages 6 and 10, but was sensitive to changes until around age 14.
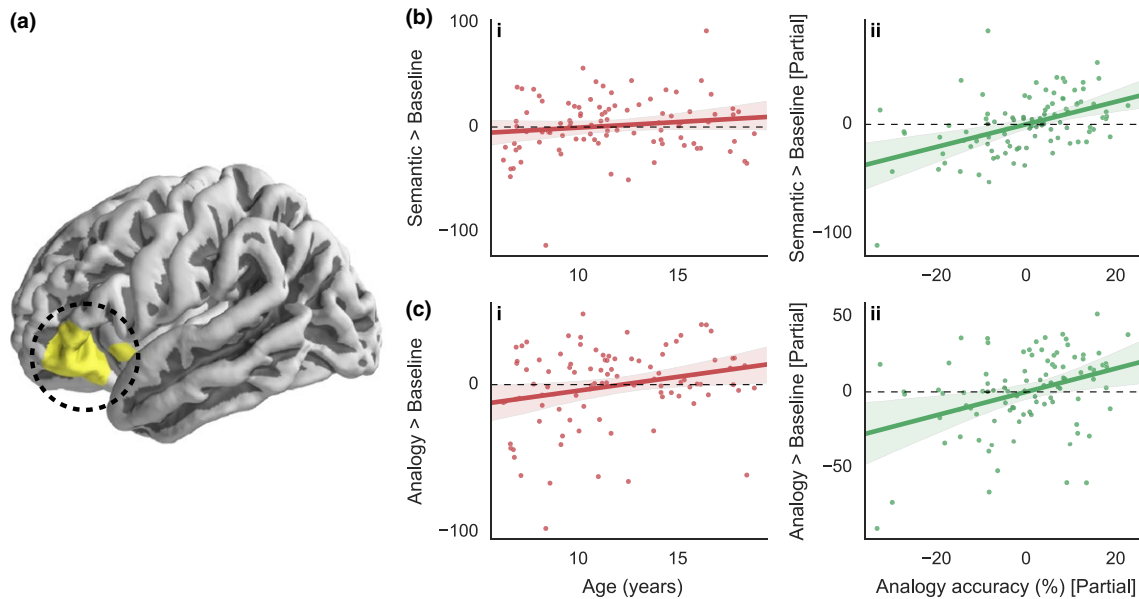
**FIGURE 5** Whole-brain analysis showing regions that demonstrate an increase in activation with accuracy on analogy trials across all participants after correcting for the effects of age. Only the contrast of semantic greater than fixation (shown in yellow) showed a significant correlation in left aLIPC (a). Panel (b) illustrates how activation in this region correlates with (i) age and (ii) accuracy on analogy trails after correcting for the effects of age. Very similar patterns are shown when activation on analogy trials are extracted from this region (c). All results are constrained to be within regions liberally activated on average by the semantic or analogy tasks. No regions in the right hemisphere or the medial aspect of the left hemisphere passed the threshold for significance. Statistical tests are permutation tests, thresholds are set using a cluster defining threshold of Z > 2.3 and are family-wise error-corrected at p < .05. Thresholded and unthresholded maps are available in NeuroVault at http://neurovault.org/collections/1658

What drives the observed improvements in analogical reasoning across development? We addressed this question by looking at our neuroimaging data collected while participants solved analogies. Specifically, we looked at patterns of brain activation to identify which mechanisms were associated with age-related differences in reasoning performance. Using fMRI, we were able to demonstrate that children as young as 6 years of age engage a network consisting of frontal, parietal and occipital regions while solving both analogy and semantic problems, echoing patterns found in previous work both in adults (e.g. Watson & Chatterjee, 2012) and in children (e.g. Crone et al., 2009; Wright et al., 2008). This pattern of results suggests that, as early as we have studied here, children rely on regions qualitatively similar to those found in adults.

Developmental differences in brain activation patterns do exist, however. Left lateral PFC became more active while solving analogy trials over childhood and adolescence, and – within this region – left aLIPC showed increasing activity during semantic trials, with increasing accuracy on analogy trials after correcting for the effects of age. The same pattern of results was replicated in activation during analogy trials, although this result did not survive whole-brain correction for multiple comparisons – perhaps because analogy task performance was related to age over a broader age range than semantic task performance.

Following from numerous studies showing the importance of this region for controlled semantic retrieval (e.g. Badre et al., 2005; Bunge et al., 2005; Poldrack et al., 1999; Thompson-Schill et al., 2002; Wagner et al., 2001), this pattern of results indicates that participants are able to engage left aLIPC more as they get older. The positive correlation between age and brain activation in this region, as well as the correlation with task performance over and above the effects of age, suggests that a key driver of improved analogical reasoning over middle childhood and early adolescence in an increase in the engagement of controlled semantic retrieval. However, this region was active across the entire age range. Therefore, we think that this neurocognitive process is already present by middle childhood, but is fine-tuned thereafter.

We had predicted that age-related changes in RLPFC and IPL activation would contribute to improved performance, given other research involving a largely overlapping sample of participants that implicated these regions in the development of non-semantic relational reasoning (Wendelken et al., 2011, 2016), as well as our prior adult fMRI research showing that these regions were active when adults made both semantic and non-semantic relational comparisons (Wendelken et al., 2012). However, the current results are compatible with another prior adult fMRI study (Wendelken et al., 2008). We had found that RLPFC was more active when participants evaluated the validity of an analogy (e.g. painter: brush:: writer: pen?) than when they attempted to complete it (e.g. painter: brush:: writer: ___?). We argued that being asked to evaluate complete analogies promoted relational comparison between pairs of stimuli, whereas being asked to complete analogies promoted sequential retrieval of semantic associations (painter: brush -> type of implement relied on in a profession -> writer -> pen). Given that the task used in the present study requires relational completion rather than comparison, this prior study could explain why the present study implicates controlled semantic retrieval

but not domain-general relational thinking as a key driver of developmental improvements. Alternatively, it could be that RLPFC and IPL contribute indirectly to the ability to complete analogies, via a change in temporal coupling with aLIPC; this hypothesis should be tested in the future with task-related functional connectivity analyses.

To conclude, probing the underlying mechanisms supporting the development of analogical reasoning can inform researchers regarding how children approach analogical reasoning at different ages, and provide insight as to which particular mechanisms underlie age-related improvements in reasoning performance. Our work demonstrates that there is much improvement in analogical reasoning occurring over middle childhood, and to a lesser extent through early adolescence. Across our broad age range, participants engaged a common network of brain regions, thus suggesting that the neural circuitry supporting analogical reasoning is already in place by the time children enter elementary school. However, this circuitry is still being refined as children mature and garner experience. Specifically, our results indicate that much of the refinement is occurring in regions that support the retrieval of task-relevant semantic associations. This finding supports the idea that, although younger children can solve analogies, they have difficulty selecting the appropriate semantic information in the presence of competing semantic associates. This work highlights the promise of developing pedagogical approaches for focusing students' attention on the relevant aspects of analogies that are used to scaffold learning in the classroom (Vendetti et al., 2015).

## REFERENCES

Aron, A., Robbins, T.W., & Poldrack, R.A. (2014). Inhibition and the right inferior frontal cortex: One decade on. *Trends in Cognitive Sciences*, 18, 177–185.

Badre, D., Poldrack, R.A., Paré-Blagoev, E.J., Insler, R.Z., & Wagner, A.D. (2005). Dissociable controlled retrieval and generalized selection mechanisms in ventrolateral prefrontal cortex. *Neuron*, 47, 907–918.

Beckmann, C.F., Jenkinson, M., & Smith, S.M. (2003). General multilevel linear modeling for group analysis in FMRI. *Neuroimage*, 20, 1052–1063.

Bunge, S.A. (2004). How we use rules to select actions: a review of evidence from cognitive neuroscience. *Cognitive, Affective & Behavioral Neuroscience*, 4, 564–579.

Bunge, S.A., Wendelken, C., Badre, D., & Wagner, A.D. (2005). Analogical reasoning and prefrontal cortex: Evidence for separable retrieval and integration mechanisms. *Cerebral Cortex*, 15, 239–249.

Cho, S., Holyoak, K.J., & Cannon, T.D. (2007). Analogical reasoning in working memory: Resources shared among relational integration, interference resolution, and maintenance. *Memory and Cognition*, 35, 1445–1455.

Cho, S., Moody, T.D., Fernandino, L., Mumford, J.A., Poldrack, R.A., Cannon, T.D., ... & Holyoak, K.J. (2010). Common and dissociable prefrontal loci associated with component mechanisms of analogical reasoning. *Cerebral Cortex*, 20, 524–533.

Christoff, K., Prabhakaran, V., Dorfman, J., Zhao, Z., Kroger, J.K., Holyoak, K.J., & Gabrieli, J.D.E. (2001). Rostrolateral prefrontal cortex involvement in relational integration during reasoning. *Neuroimage*, 14, 1136–1149.

Crone, E.A., Wendelken, C., Van Leijenhorst, L., Honomichl, R.D., Christoff, K., & Bunge, S.A. (2009). Neurocognitive development of relational reasoning. *Developmental Science*, 12, 55–66.

Dumontheil, I. (2014). Development of abstract thinking during childhood and adolescence: the role of rostrolateral prefrontal cortex. *Developmental Cognitive Neuroscience*, 10, 57–76. http://doi.org/10.1016/j.dcn.2014.07.009

Dumontheil, I., Houlton, R., Christoff, K., & Blakemore, S.J. (2010). Development of relational reasoning during adolescence. *Developmental Science*, 13, 15–24.

Dunbar, K., & Blanchette, I. (2001). The in vivo/in vitro approach to cognition: The case of analogy. *Trends in Cognitive Sciences*, 5, 334–339.

Eklund, A., Nichols, T.E., & Knutsson, H. (2016). Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates. *Proceedings of the National Academy of Sciences of United States of America*, 113, 7900–5. doi:10.1073/pnas.1602413113

Ferrer, E., Whitaker, K.J., Steele, J.S., Green, C.T., Wendelken, C., & Bunge, S.A. (2013). White matter maturation supports the development of reasoning ability through its influence on processing speed. *Developmental Science*, 16, 941–951.

Galván, A., Van Leijenhorst, L., & McGlennen, K.M. (2012). Considerations for imaging the adolescent brain. *Developmental Cognitive Neuroscience*, 2, 293–302. http://doi.org/10.1016/j.dcn.2012.02.002

Gentner, D. (1983). Structure-Mapping: A theoretical framework for analogy. *Cognitive Science*, 7, 155–170.

Gentner, D. (1988). Metaphor as structure mapping: The relational shift. *Child Development*, 59, 47–59.

Gentner, D., & Medina, J. (1998). Similarity and the development of rules. *Cognition*, 65, 263–297.

Gentner, D., & Rattermann, M.J. (1991). Language and the career of similarity. In S.A. Gelman & J.P. Byrnes (Eds.), *Perspectives on thought and language: Interrelations in development* (pp. 225–277). London: Cambridge University Press.

Gick, M., & Holyoak, K.J. (1980). Analogical problem solving. *Cognitive Psychology*, 12, 306–355.

Gick, M., & Holyoak, K.J. (1983). Schema induction and analogical transfer. *Cognitive Psychology*, 15, 1–38.

Gilhooly, K.J., & Logie, R.H. (1980). Age-of-acquisition, imagery, concreteness, familiarity, and ambiguity measures for 1,944 words. *Behavior, Research Methods, and Instrumentation*, 12, 395–427.

Gorgolewski, K.J., Varoquaux, G., Rivera, G., Schwartz, Y., Sochat, V. V., Ghosh, S. S., ...Poldrack, R.A. (2016). NeuroVault.org: A repository for sharing unthresholded statistical maps, parcellations, and atlases of the human brain. *Neuroimage*, 124, 1242–1244.

Goswami, U. (1996). Analogical reasoning and cognitive development. *Advances in Child Development and Behavior*, 26, 91–138.

Goswami, U., & Brown, A.L. (1990). Melting chocolate and melting snowmen: Analogical reasoning and causal relations. *Cognition*, 35, 69–95.

Halford, G.S. (1992). Analogical reasoning and conceptual complexity in cognitive development. *Human Development*, 35, 193–217.

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. (2012). *Surfaces and essences: Analogy as the fuel and fire of thinking*. New York: Basic Books.

Holyoak, K.J., Junn, E.N., & Billman, D.O. (1984). Development of analogical problem-solving skill. *Child Development*, 2042–2055.

Holyoak, K.J., & Kroger, J.K. (1995). Forms of reasoning: Insight into prefrontal functions? *Annals of the New York Academy of Sciences*, 769, 253–264.

Holyoak, K.J., & Thagard, P. (1995). *Mental Leaps*. Boston: MIT Press.

Hummel, J.E., & Holyoak, K.J. (1997). Distributed representations of structure: A theory of analogical access and mapping. *Psychological Review*, 104, 427–466.

Hummel, J., & Holyoak, K.J. (2003). A symbolic-connectionist theory of relational inference and generalization. *Psychological Review*, 110, 220–264.

Jenkinson, M., Bannister, P.R., Brady, J.M., & Smith, S.M. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage*, 17, 825–841.

Kotovsky, L., & Gentner, D. (1996). Comparison and categorization in the development of relational similarity. *Child Development*, 67, 2797–2822.

Krawczyk, D.C. (2012). The cognition and neuroscience of relational reasoning. *Brain Research*, 1428, 13–23.

Krawczyk, D.C., McClelland, M.M., Donovan, C.M., Tillman, G.D., & Maguire, M.J. (2010). An fMRI investigation of cognitive stages in reasoning by analogy. *Brain Research*, 1342, 63–73.

Kroger, J.K., Sabb, F.W., Fales, C.L., Bookheimer, S.Y., Cohen, M.S., & Holyoak, K.J. (2002). Recruitment of anterior dorsolateral prefrontal cortex in human reasoning: A parametric study of relational complexity. *Cerebral Cortex*, 12, 477–485.

Miller, E.K., & Cohen, J.D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, 24, 167–202.

Morrison, R.G., Doumas, L.A.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.

Namy, L.L., & Gentner, D. (2002). Making a silk purse out of two sow's ears: Young children's use of comparison in category learning. *Journal of Experimental Psychology: General*, 131, 5–15.

Novick, J.M., Kan, I.P., Trueswell, J.C., & Thompson-Schill, S.L. (2009). A case for conflict across multiple domains: Memory and language impairments following damage to ventrolateral prefrontal cortex. *Cognitive Neuropsychology*, 26, 527–567.

Passingham, D. (1995). *The frontal lobes and voluntary action*. Oxford, UK: Oxford University Press.

Poldrack, R.A., Wagner, A.D., Prull, M.W., Desmond, J.E., Glover, G.H., & Gabrieli, J.D. (1999). Functional specialization for semantic and phonological processing in the left inferior prefrontal cortex. *Neuroimage*, 10, 15–35.

*Presentation (Version 0.80) [Computer software]*. (2004). Albany, CA: NeuroBehavioral Systems. Available from: www.neurobs.com.

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., & 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., Zur, O., & Holyoak, K.J. (2007). Cognitive supports for analogies in the mathematics classroom. *Science*, 316, 1128–1129.

Smith, S.M., Jenkinson, M., Woolrich, M.W., Beckmann, C.F., Behrens, T.E.J., Johansen-Berg, H., ... & Flitney, D.E. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage*, 23, 208–219.

Souza, M.J., Donohue, S.E., & Bunge, S.A. (2009). Controlled retrieval and selection of action-relevant knowledge mediated by partially overlapping regions in left ventrolateral prefrontal cortex. *Neuroimage*, 46, 299–307.

Sternberg, R.J., & Nigro, G. (1980). Developmental patterns in the solution of verbal analogies. *Child Development*, 51, 27–38.

Stuss, D.T., & Knight, R.T. (2012). (Eds.) *Principles of frontal lobe functions* (2nd edn). Oxford, UK: Oxford University Press.

Thesen, S., Heid, O., Mueller, E., & Schad, L.R. (2000). Prospective acquisition correction for head motion with image-based tracking for real-time fMRI. *Magnetic Resonance in Medicine*, 44, 457–465.

Thompson-Schill, S.L., Jonides, J., Marshuetz, C., Smith, E.E., D'Esposito, M., Kan, I.P., Knight, R.T., & Swick, D. (2002). Effects of frontal lobe damage on interference effects in working memory. *Cognitive, Affective, and Behavioral Neuroscience*, 2, 109–120.

Vendetti, M.S., & Bunge, S.A. (2014). Evolutionary and developmental changes in the lateral frontoparietal network: A little goes a long way for higher-level cognition. *Neuron*, 84, 906–917.

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.

Wagner, A.D., Paré-Blagoev, E.J., Clark, J., & Poldrack, R.A. (2001). Recovering meaning: Left prefrontal cortex guides controlled semantic retrieval. *Neuron*, 31, 329–338.

Watson, C.E., & Chatterjee, A. (2012). A bilateral frontoparietal network underlies visuospatial analogical reasoning. *Neuroimage*, 59, 2831–2838.

Wendelken, C., Bunge, S.A., & Carter, C.S. (2008). Maintaining structured information: an investigation into functions of parietal and lateral prefrontal cortices. *Neuropsychologia*, 46, 665–678.

Wendelken, C., Chung, D., & Bunge, S.A. (2012). Rostrolateral prefrontal cortex: Domain-general or domain-sensitive? *Human Brain Mapping*, 33, 1952–1963.

Wendelken, C., Ferrer, E., Whitaker, K.J., & Bunge, S.A. (2016). Fronto-Parietal network reconfiguration supports the development of reasoning ability. *Cerebral Cortex*, 26, 2178–2190.

Wendelken, C., O'Hare, E.D., Whitaker, K.J., Ferrer, E., & Bunge, S.A. (2011). Increased functional selectivity over development in rostrolateral prefrontal cortex. *Journal of Neuroscience*, 31, 17260–17268.

Woolrich, M. (2008). Robust group analysis using outlier inference. *NeuroImage*, 41, 286–301. http://doi.org/10.1016/j.neuroimage.2008.02.042

Woolrich, M. W., Behrens, T. E. J., Beckmann, C. F., Jenkinson, M., & Smith, S. M. (2004). Multilevel linear modelling for FMRI group analysis using Bayesian inference. *NeuroImage*, 21, 1732–1747. http://doi.org/10.1016/j.neuroimage.2003.12.023

Woolrich, M.W., Jbabdi, S., Patenaude, B., Chappell, M., Makni, S., Behrens, T., ... & Smith, S. M. (2009). Bayesian analysis of neuroimaging data in FSL. *Neuroimage*, 45, 173–186.

Worsley, K. J. (2001). Statistical Analysis of Activation Images. In P. Jezzard, P. M. Matthews & S. M. Smith (Eds.), *Functional MRI: An introduction to methods* (pp. 251–270). Oxford, UK: OUP Oxford.

Wright, S. B., Matlen, B. J., Baym, C. L., Ferrer, E., & Bunge, S. A. (2008). Neural correlates of fluid reasoning in children and adults. *Frontiers in Human Neuroscience*, 1(March), 8. http://doi.org/10.3389/neuro.09.008.2007

## SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.