Neural and Cognitive Underpinnings of Counterintuitive Science and Math Reasoning in Adolescence

Iroise Dumontheil[1,2], Annie Brookman-Byrne[1,2], Andrew K. Tolmie[2,3], and Denis Mareschal[1,2]

Abstract

■ Reasoning about counterintuitive concepts in science and math is thought to require suppressing naive theories, prior knowledge, or misleading perceptual cues through inhibitory control. Neuroimaging research has shown recruitment of pFC regions during counterintuitive reasoning, which has been interpreted as evidence of inhibitory control processes. However, the results are inconsistent across studies and have not been directly compared with behavior or brain activity during inhibitory control tasks. In this fMRI study, 34 adolescents (aged 11–15 years) answered science and math problems and completed response inhibition tasks (simple and complex go/no-go) and an interference control task (numerical Stroop). Increased BOLD signal was observed in parietal (Brodmann's area 40) and prefrontal (Brodmann's area 8, 45/47) cortex regions in counterintuitive problems compared with control problems, where no counterintuitive reasoning was required, and in two parietal clusters when comparing correct counterintuitive reasoning to giving the incorrect intuitive response. There was partial overlap between increases in BOLD signal in the complex response inhibition and interference control tasks and the science and math contrasts. However, multivariate analyses suggested overlapping neural substrates in the parietal cortex only, in regions typically associated with working memory and visuospatial attentional demands rather than specific to inhibitory control. These results highlight the importance of using localizer tasks and a range of analytic approach to investigate to what extent common neural networks underlie performance of different cognitive tasks and suggests visuospatial attentional skills may support counterintuitive reasoning in science and math. ■

INTRODUCTION

Health and technological development depend on understanding science and mathematics concepts. These concepts can be abstract, such as thermodynamics, the cellular basis of life, or probabilities, and a poor conceptual understanding can have real-world ramifications for both experts and the rest of the population (e.g., Hotez, 2021; Menge et al., 2018). For example, an understanding of the transmission of illnesses via viruses and bacteria can influence vaccination and hygiene-related behaviors (Strohl et al., 2015). Health-related behaviors more generally, which can impact life expectancy and health outcomes, associate with health literacy (Berkman, Sheridan, Donahue, Halpern, & Crotty, 2011) and education level (Cutler & Lleras-Muney, 2010). Despite the importance of a good conceptual understanding of science and math, research has demonstrated that adults' conceptual understanding of science (Miller, 1998; Furnham, 1992) and math (Spiegelhalter, 2019; National Numeracy, 2017) is limited.

[1]Centre for Brain and Cognitive Development, Birkbeck, University of London, United Kingdom, [2]Centre for Educational Neuroscience, Birkbeck/University College London, United Kingdom, [3]Department of Psychology and Human Development, University College London Institute of Education, United Kingdom

Although lack of instruction can evidently lead to poor science and math conceptual understanding, a limited conceptual understanding of science and math is still present in countries with prolonged compulsory education. Some specific interventions have been shown to improve conceptual understanding, but overall, the evidence is that nonscientific concepts are difficult to replace with scientific ones (Guzzetti, 2000). A fundamental reason for these deficits in conceptual understanding is an intrinsic difficulty to learn science and math concepts (Willingham, 2010; Johnstone, 1991). One source of difficulties is that many of these concepts are counterintuitive: A large proportion of pupils strongly hold a certain understanding—referred to as intuitive understanding, as it is experienced as self-evident and self-consistent (Osman & Stavy, 2006)—acquired through perception, popular belief, or simple heuristics, that disagrees with consensus expert opinion and that needs to be overcome (Mareschal, 2016; Dunbar, Fugelsang, & Stein, 2007; Houdé, 2000). The aim of this study was to investigate the behavioral and neural correlates of counterintuitive science and math reasoning during adolescence, a period during which pupils are taught increasingly complex and abstract science and math content. For many individuals, these will be their final years of formal learning of these subjects.

Journal of Cognitive Neuroscience 34:7, pp. 1205–1229
https://doi.org/10.1162/jocn_a_01854

Role of Inhibitory Control in Counterintuitive Reasoning

Reasoning effectively about counterintuitive concepts or solving counterintuitive problems is thought to require the inhibition of interfering information, misleading perceptual cues, naive theories, prior knowledge, or intuitive rules (Mareschal, 2016; Borst, Poirel, Pineau, Cassotti, & Houdé, 2013; Lubin, Vidal, Lanoë, Houdé, & Borst, 2013; Stavy & Babai, 2010; Dunbar et al., 2007; Osman & Stavy, 2006; Houdé, 2000). Indeed, evidence suggests that rather than being altered or replaced during a process of conceptual change, incorrect intuitive understanding, or misconceptions, coexist with correct scientific representations (see the work of Potvin, Malenfant-Robichaud, Cormier, & Masson, 2020, for a review). Behavioral studies have shown that individuals are slower and less accurate in responding to incongruent trials, where the intuitive response contradicts the scientific response (e.g., two shapes that differ in surface area but not in perimeter, a nonrigid solid, a moving nonliving thing, a heavier ball falling at the same speed as a lighter ball) than to congruent trials, where the responses associated with the intuitive and scientific understandings align (e.g., the shape with the larger surface area also has a larger perimeter, a rigid solid, a moving living thing, two balls of the same weight falling at the same speed; Potvin et al., 2020; Allaire-Duquette, Bélanger, Grabner, Koschutnig, & Masson, 2019; Brault Foisy, Potvin, Riopel, & Masson, 2015; Shtulman & Valcarcel, 2012; Babai, Sekal, & Stavy, 2010; Stavy & Babai, 2010; Babai & Amsterdamer, 2008). Slower and less accurate responses are interpreted as reflecting the conflict arising between intuitive and scientific conceptual representations and associated responses, even when the correct response is eventually provided, and even when participants are experts in the domain of interest (Allaire-Duquette et al., 2021; Potvin et al., 2020; Lewis & Linn, 1994; although see Masson, Potvin, Riopel, & Brault Foisy, 2014).

Is this slowing down and reduced accuracy in counterintuitive or incongruent trials indeed reflecting the involvement of inhibitory control? Inhibitory control is multifaceted, and there is currently no clear agreement regarding which functions fit within the umbrella of inhibitory control and how inhibition-related functions may relate to each other (Diamond, 2013; Friedman & Miyake, 2004; Nigg, 2000). Here, we are considering a distinction between response inhibition, where a dominant behavioral response has to be inhibited, as in the Simon's task, go/no-go task, or Stop signal task, and resistance to distractor interference, where the conflict is observed between representations, such as in the Stroop task or Eriksen Flanker task. Behavioral studies have implicated interference control skills in general science (Latzman, Elkovitch, Young, & Clark, 2010; St Clair-Thompson & Gathercole, 2006) and math performance (Cragg, Keeble, Richardson, Roome, & Gilmore, 2017; Gilmore, Keeble, Richardson, & Cragg, 2015; Latzman et al., 2010; Khng & Lee, 2009; St Clair-Thompson & Gathercole, 2006). Associations with response inhibition have been less consistently observed (no association: Donati, Meaburn & Dumontheil, 2019; Khng & Lee, 2009; association: St Clair-Thompson & Gathercole, 2006). Other aspects of cognitive control such as working memory (Donati et al., 2019; Cragg et al., 2017; Cragg & Gilmore, 2014; Khng & Lee, 2009; St Clair-Thompson & Gathercole, 2006) and more general measures of intelligence (e.g., Donati et al., 2019; Latzman et al., 2010) have also been found to associate with general science and math achievement.

A few studies have focused on reasoning or knowledge of counterintuitive science and math concepts specifically (see the work of Mason & Zaccoletti, 2021, for a review of science studies). A study with toddlers found that response inhibition (gift delay task), but not interference control (reverse categorization task), was associated with reasoning about solidity (Baker, Gjersoe, Sibielska-Woch, Leslie, & Hood, 2011). A study with 5- to 7-year-olds found that a measure of accuracy on task blocks requiring response inhibition and shifting correlated with conceptual understanding of life, death, and bodily function (Zaitchik, Iqbal, & Carey, 2014). In adolescence, one study of counterintuitive science and math reasoning showed that better motor response inhibition was associated with longer response times, suggesting less impulsive responding, whereas better interference control was associated with higher accuracy, indicating effective suppression of intuitive, incorrect responses, and these associations were observed while controlling for general vocabulary and reasoning measures (Brookman-Byrne, Mareschal, Tolmie, & Dumontheil, 2018). Other research found that adolescents with lower inhibitory control (measured by perseverative errors on the Wisconsin Card Sorting Test), planning or working memory, showed poorer scientific reasoning (Kwon & Lawson, 2000) and less benefit of individual tutoring in proportional reasoning (Kwon, Lawson, Chung, & Kim, 2000). Finally, spatial working memory and planning, but not response inhibition (Stop Signal task) associated with conceptual learning in biology (Rhodes et al., 2014) and chemistry (Rhodes et al., 2016) in 12- to 13-year-olds. A single study performed in young adults found that individuals with higher interference control (measured with the color word Stroop task) read texts relating to science misconceptions more slowly than individuals with lower inhibitory control, which was interpreted as reflecting management of interference from reactivated misconceptions from prior knowledge (Butterfuss & Kendeou, 2020).

Overall, these results indicate that individual differences in interference control, as well as other aspects of cognitive control and general intelligence, associate with both counterintuitive science and math reasoning but also more general achievement in these subjects, whereas there is little evidence of associations with simple response inhibition, except at younger ages.

Neural Correlates of Counterintuitive Science and Math Reasoning

A number of studies have tried to complement behavioral research by using fMRI to investigate the neural correlates of counterintuitive science and math reasoning. These studies focused on a single counterintuitive concept: electric circuit wiring (Masson et al., 2014; Potvin, Turmel, & Masson, 2014), falling objects (Brault Foisy et al., 2015), or perimeter-surface area (Stavy & Babai, 2010; Stavy, Goel, Critchley, & Dolan, 2006), or used a range of counterintuitive chemistry (Potvin et al., 2020) or science concepts (Allaire-Duquette et al., 2019, 2021), and investigated various comparisons between trial types and between groups varying in scientific expertise. A common hypothesis was that nonscientific trials would lead to a conflict between intuitive and scientific understandings and recruitment of inhibitory control to support conflict detection and resolution, and that this may be observed to a greater extent for correct than incorrect trials (Stavy & Babai, 2010) and, relatedly, in individuals with greater scientific expertise (and higher accuracy) compared with novices (Allaire-Duquette et al., 2019; although see the work of Allaire-Duquette et al., 2021, for a discussion of alternative hypotheses). Based on their review of the past inhibitory control neuroimaging literature, researchers predicted activation in the ACC, ventrolateral prefrontal cortex (VLPFC), and dorsolateral prefrontal cortex (DLPFC; Allaire-Duquette et al., 2019, 2021; Potvin et al., 2020; Brault Foisy et al., 2015; Masson et al., 2014), and concluded that their results supported their hypotheses.

While the results of these studies seem broadly consistent, the conclusions are weakened by the fact that the studies did not use inhibitory control task localizers in the same participants, nor demonstrate associations between counterintuitive reasoning activation and inhibitory control skills, and the specific neural correlates of inhibitory control are still debated. Although right VLPFC/inferior frontal gyrus (IFG) was initially thought to specifically implement response inhibition, more recent work suggests that it may play a broader role in maintaining goals and modulating activity of other brain regions (see the work of Banich & Depue, 2015; Swick & Chatham, 2014, for a discussion). The right middle frontal gyrus (MFG) has been implicated in inhibiting memory retrieval, suggesting the right pFC may be fractionated into regions supporting motoric versus cognitive inhibitory function (Banich & Depue, 2015). Importantly, inhibitory control tasks are likely to recruit broader attentional and cognitive control processes; indeed, Criaud and Boulinguez (2013) reviewed go/no-go studies and argue that most of the no-go activity typically observed during complex go/no-go tasks in the right DLPFC, right IFG, and pre-SMA, is actually driven by the engagement of high attentional/working memory resources, not by inhibitory processes per se.

An alternative perspective to the view that some prefrontal brain regions specifically implement inhibitory control is that inhibitory control emerges from the main function of the pFC, active goal maintenance (Munakata et al., 2011). In this framework, pFC neurons excite goal-relevant processing areas, allowing them to compete with, and indirectly inhibit, other possible processing pathways. Support for this view comes from behavioral evidence suggesting that inhibitory control cannot always consistently be separated from a general executive function factor (Karr et al., 2018; Friedman & Miyake, 2017; Friedman et al., 2008). Performance on inhibitory control tasks also tends to be less correlated than performance on, for example, working memory tasks (e.g., the works of Hartung, Engelhardt, Thibodeaux, Harden, & Tucker-Drob, 2020; Huizinga, Dolan, & van der Molen, 2006, in developmental samples). This view closely relates to the suggestion that a broad network of brain regions, which includes areas in and around the posterior part of the inferior frontal sulcus, and the anterior insula/frontal operculum, ACC, and pre-SMA, supports the elaboration and maintenance of structured mental programs/subtasks, and underlies the general intelligence factor, *g* (multiple-demand network; Crittenden, Mitchell, & Duncan, 2016; Duncan, 2010).

Another limitation of the neuroimaging literature on counterintuitive reasoning is that the ACC, DLPFC, and VLPFC are broad regions and the similarities in patterns of activation across studies have not been assessed. In an effort to synthesize previous research, Figure 1 illustrates the location of peaks of activation observed when comparing incongruent, or nonscientific, trials—that is, trials where different conceptual understandings may be in conflict—to congruent trials (Figure 1A), and the same contrast comparing patterns of activation in experts versus novices (Figure 1B; see also Table 1). Although activation is observed in pFC in most studies, there is little consistency in the loci of these activations, apart from activation in left VLPFC when comparing counterintuitive reasoning in experts to novices (Figure 1B).

Additional studies focused on slightly different contrasts. For example, Stavy and Babai (2010) observed activation in the orbitofrontal cortex bilaterally in the contrast incongruent correct—incongruent intuitive error in the surface/perimeter task. Potvin et al. (2014) combined congruent and incongruent electric circuit trials and found that humanities and arts college students showed greater activation in the bilateral parietal lobules (Brodmann's area [BA] 19/7), right premotor and motor cortices, and inferior temporal gyrus for correct than incorrect trials when they were certain of their responses. A final study using the electrical circuits and gravity tasks found that after being shown the correct answer, participants who initially showed the misconception had increased activation in the posterior cingulate cortex in both nonscientific and scientific trials, as well as additional activation in bilateral rostrolateral frontal cortex in nonscientific trials, compared with before being shown the answers (Nenciovici et al., 2018).

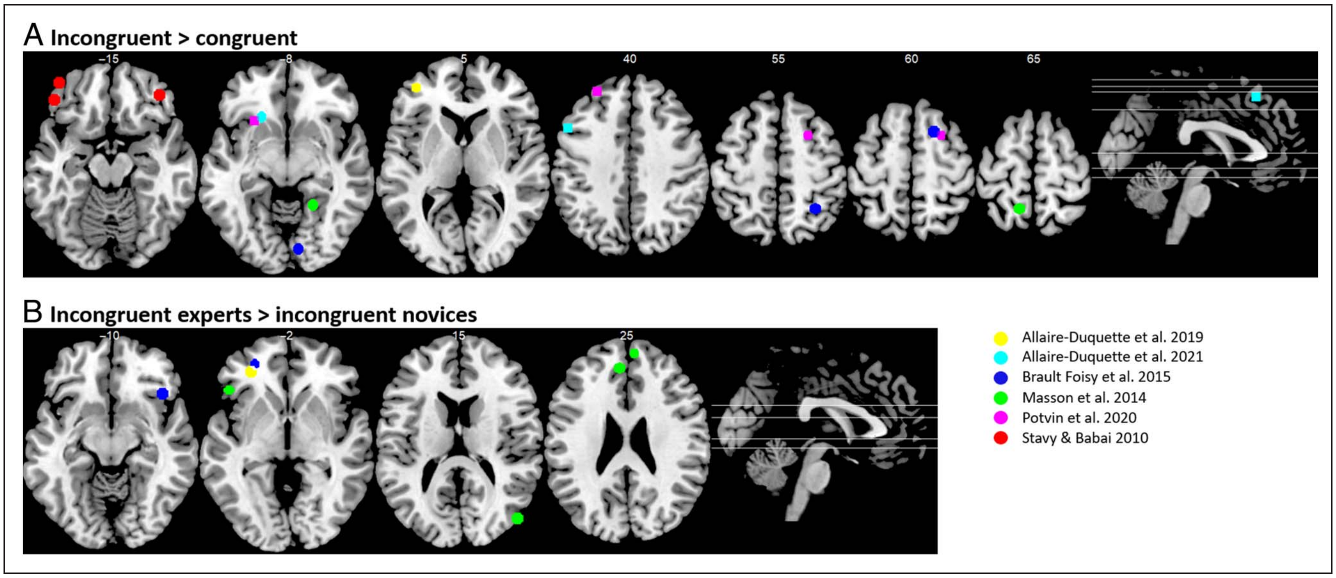

Figure 1. Peak activations reported in previous studies of counterintuitive science and math reasoning. (A) Results of contrasts between incongruent trials, where the intuitive and scientific concepts and associated responses are in conflict, and congruent trials, where they are in agreement. (B) Results of contrasts identifying greater activation in incongruent trials in expert than novice participants with (Allaire-Duquette et al., 2019) or without (Brault Foisy et al., 2015; Masson et al., 2014) using congruent trials as a baseline. See Table 1 for more details. Spheres of 5-mm diameter were drawn around the peak voxels.

This Study

Counterintuitive concepts present a particular learning challenge for adolescents, who are taught science and math content that is increasingly complex and distant from the everyday sensory experiences and lay-beliefs (e.g., things are made of atoms), whereas inhibitory control, and cognitive control more broadly, continue to mature, as shown by behavioral and neuroimaging research (Humphrey & Dumontheil, 2016; Luna, Marek, Larsen, Tervo-Clemmens, & Rajpreet, 2015; Jaeger, 2013; Ordaz, Foran, Velanova, & Luna, 2013; Crone & Dahl, 2012; Luna, Padmanabhan, & O'Hearn, 2010; Leon-Carrion, García-Orza, & Pérez-Santamaría, 2004). U.K. data indicate that there is little progress, but large individual differences, in mathematic abilities in the first phase of secondary school, between 11 and 14 years of age (Ryan & Williams, 2007). For example, 30% of 11-year-olds and 27% of 14-year-olds incorrectly add numerators and denominators when asked to solve 3/8 + 2/8 = ___ (Ryan & Williams, 2007). Other research suggests some improvement in reasoning about counterintuitive concepts, such as buoyancy, during adolescence, but interference from conflict between different conceptual understanding can remain (Potvin & Cyr, 2017). Brookman-Byrne et al. (2018) found accuracy on counterintuitive and control congruent trials combined increased from 66% to 73% in science and 65% to 72% in math between 12 and 15 years of age.

In the current study, we followed up previous behavioral work in a larger sample (Brookman-Byrne et al., 2018) to investigate the potential role of inhibitory control in counterintuitive science and math reasoning in adolescence using fMRI. We sought to go beyond previous neuroimaging research by obtaining, in the same sample of participants, behavioral and neural measures of both inhibitory control and science and math reasoning. As our aim was to further our understanding of counterintuitive reasoning within the broader secondary school education context, we chose to use problems across a range of science and math topics, an approach taken by some other studies (Allaire-Duquette et al., 2019, 2021; Potvin et al., 2020; Brookman-Byrne et al., 2018). We included a confidence rating in the science and math task and predicted that in counterintuitive trials, but not control trials, adolescents may confidently give the incorrect but intuitive answer (see the work of Potvin et al., 2014, for a similar approach). With the caveat that this sample was smaller than previous behavioral research (Brookman-Byrne et al., 2018; Rhodes et al., 2014, 2016; Kwon & Lawson, 2000; Kwon et al., 2000), we first assessed the specificity of behavioral associations between science and math reasoning and inhibitory control performance versus more general vocabulary and *g* factor measures. Second, we identified the neural correlates of reasoning about counterintuitive science and math problems, presented as true or false statements (counterintuitive trials), using science and math problems that were in line with intuition or perceptual cues as a comparison condition (control trials). We further compared neural activation on counterintuitive trials, which were correctly answered, suggesting the intuitive response had been appropriately inhibited, to trials that had been incorrectly answered, a similar contrast to that carried out by Stavy et al. (2006). Third, we compared

Table 1. Summary of the Key Results of fMRI Studies of Counterintuitive Reasoning

Study	*Contrast*	*Group*	*x*	*y*	*z*	*Accuracy* *C*	*Accuracy* *IC*	*Figure*
Allaire-Duquette et al. (2019)	IC > C	High science competency adolescents	−39	48	3	91%	66%	Figure 1A
	IC > C	Low science competency adolescents	–	–	–	80%	43%	
	IC > C high > low competence		−30	42	−3			Figure 1B
Allaire-Duquette et al. (2021)	IC > C	Physics PhD	−21	24	−9	93%	79%	Figure 1A
			0	27	51			Figure 1A
			−51	15	42			Figure 1A
			33	−66	−33			Figure 1A
			6	−84	−24			Figure 1A
Brault Foisy et al. (2015)	IC > C	Physics undergraduates	30	−51	54	99%	94%	Figure 1A
			9	−84	−6			Figure 1A
	IC > C	Humanities undergraduates	−63	−18	36	2%	4%	
			51	24	27			
			−6	3	33			
	IC experts > novices		−27	48	0			Figure 1B
			42	24	−9			Figure 1B
Masson et al. (2014)	IC > C	Physics undergraduates	18	12	60	93%	99%	Figure 1A
			21	−48	−6			Figure 1A
			−12	−51	66			Figure 1A
	IC > C	Humanities undergraduates	–	–	–	4%	2%	
	IC experts > novices		48	−78	15			Figure 1B
			−6	45	24			Figure 1B
			−48	27	0			Figure 1B
			6	57	27			Figure 1B
Potvin et al. (2020)	IC correct > C correct	Chemistry university professors	24	9	57	96%	71%	Figure 1A
			−27	21	−6			Figure 1A
			−27	45	42			Figure 1A

Table 1. (*continued*)

Study	Contrast	Group	x	y	z	Accuracy C	Accuracy IC	Figure
Stavy and Babai (2010)	IC correct > C correct	University educated adults	40	42	−16	92%	62%	Figure 1A
			−46	38	−16			Figure 1A
			−42	52	−14			Figure 1A
	IC correct > IC incorrect	University educated adults	40	32	−16			
			−26	36	−12			

Incongruent (IC; or counterintuitive) trials are those where the intuitive concept and response are in conflict with the scientific concept or response. Congruent (C; or control) trials are those where the intuitive and scientific responses are in agreement.

these contrasts to the neural correlates of a go/no-go task with or without a one-back working memory load, and an interference control numerical Stroop task using both univariate and multivariate approaches. We hypothesized that greater behavioral and neural associations would be observed between the response inhibition task with a working memory load and the interference control task than the simple response inhibition task. As previous neuroimaging results were mixed regarding precise locations of counterintuitive reasoning activations in the pFC (Figure 1, Table 1), we did not have specific predictions regarding regions of activation and used whole-brain analyses.

METHODS

Participants

Thirty-eight pupils (20 girls, 18 boys) aged 11–15 years, with no neurological or developmental disorders, from a range of schools in different demographic areas, participated in the study. Flyers were sent from schools to parents of pupils in Years 7–10, inviting their children to take part. Written informed parental and participant consent was obtained in accordance with the guidelines of the local ethics committee, which approved the study. Participants were given pictures of their brain and £20 for participation, and travel expenses were reimbursed.

One participant was excluded because of low accuracy in the science and math task (15-year-old girl). Three participants were excluded because of movement: one in both the science and math task and the go/no-no (12-year-old girl), one in the go/no-go (15-year-old girl), and one in the Stroop (12-year-old boy). Two participants had just one run excluded because of movement in the science and math task (12-year-old girl, 15-year-old boy), so they were kept in the analysis, discarding the concerning run. Exclusionary criteria relating to movement are described in detail below. The final sample consisted of 34 participants, of which 17 were girls and 17 were boys, with a mean age of 13.4 years (SD = 1.32 years). There was no gender difference in mean age, $t(32) = 0.30, p = .768$.

Tasks

Science and Math

The science and math task was adapted from Brookman-Byrne et al. (2018) and had an event-related design. On each trial, a statement relating to science or math appeared on the screen, and participants pressed a button to indicate whether they thought the statement was definitely true, probably true, probably false, or definitely false. It was explained that the definitely/probably distinction referred to the participant's confidence in their response. Responses were made through two button boxes, and the index and middle fingers of both hands rested on four response buttons.

Forty-eight problem-sets of four problem types were created (Figure 2). Within a set, two problems addressed the same counterintuitive concept, with the counterintuitive-false problem showing a false statement and the counterintuitive-true problem showing a true statement. In both cases, it was anticipated that participants might give the wrong, intuitive answer, as opposed to the correct, counterintuitive answer. Control-false and control-true problems were based on broadly the same topic area, but were not counterintuitive. It was anticipated that participants would be more likely to get these right. Participants were shown one counterintuitive trial and one control trial from each problem-set, giving 96 trials per participant.

A broad spectrum of topics relevant for the Key Stage 3 curriculum for England (Department for Education, 2013a, 2013b) were included, based on literature that suggested these concepts would be counterintuitive to some extent in 11- to 15-year-olds (e.g., Driver, Squires, Rushworth, & Wood-Robinson, 2015; Ryan & Williams, 2007; Stavy & Tirosh, 2000), discussions with science and math teachers, and consultation with study guides (Parsons, 2014; Parsons & Gannon, 2014) and curricula. Half of the problem-sets were scientific: eight in biology, eight in chemistry, and eight in physics. The other half were mathematical and covered number, algebra, ratio, geometry, probability, and statistics (Figure 2). Science and math subjects were combined as there was not enough trials to explore each (e.g., biology, geometry) separately, and the assumption here is that they would require similar types of

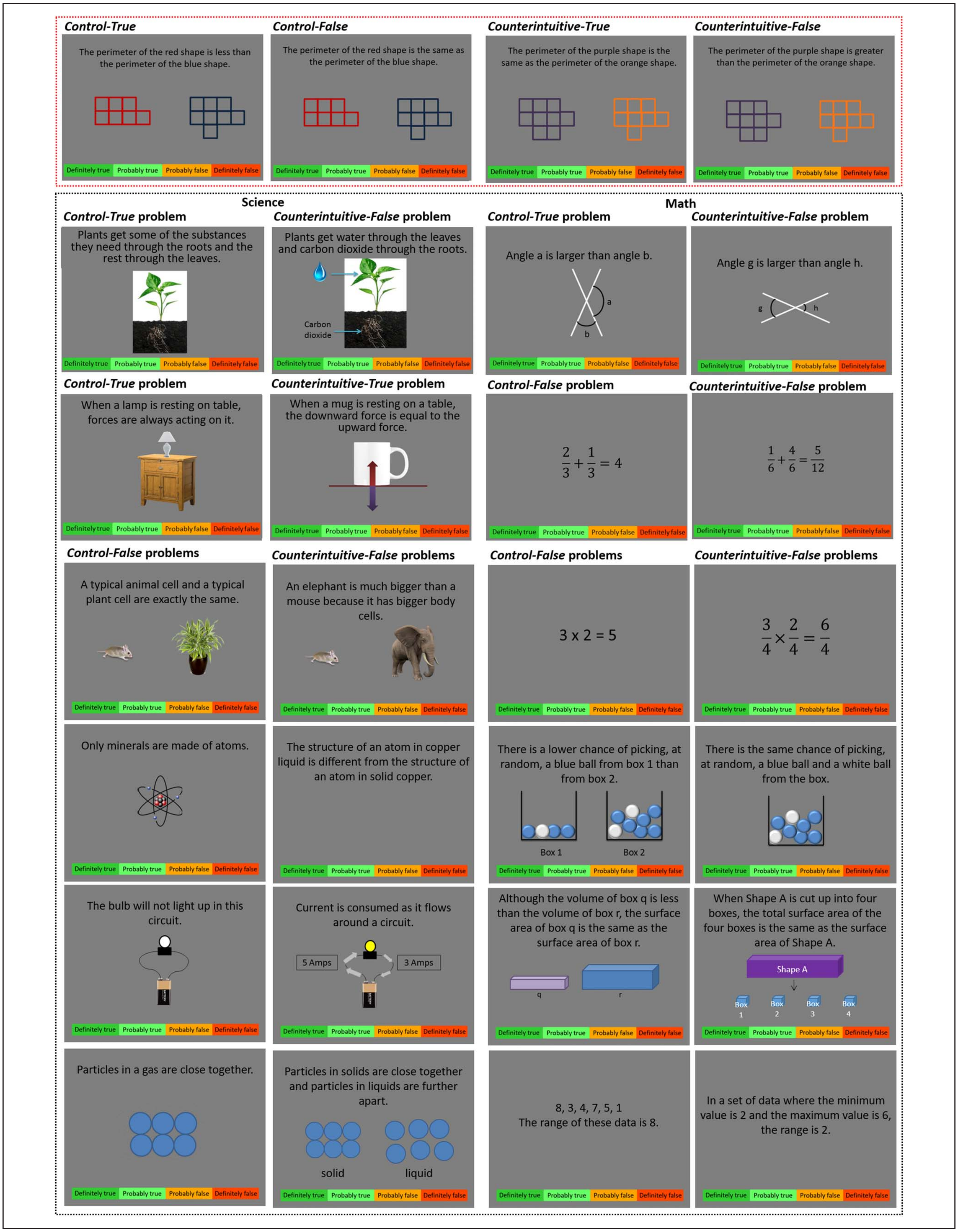

Figure 2. Example problems from the science and math task. The first row shows the full problem set for one math topic, with true and false control and counterintuitive problems. Other rows show an example of a control and a counterintuitive problem for a topic. Each participant only saw one control and one counterintuitive problem of each set. All stimuli are available on-line from osf.io/ytcwk/.

counterintuitive reasoning. Analyses considered science and math problems separately in a first step.

A maximum of 12 sec was allowed for a response to be made on each trial, and if a response had not been made within 9 sec, a red border appeared around the response options on the screen to prompt a response. Each trial lasted 16 sec, and the remaining time following a response was filled with a central fixation cross on a third of trials or a simple arrows task on two thirds of trials. In the arrows task, participants pressed the left or right key with their index fingers according to the direction of arrows on the screen. This constituted an active baseline and was used to ensure participants stopped thinking about the science and math tasks, allowing the BOLD signal to decrease, but remained engaged.

The task consisted of four runs: two runs of science and two runs of math, alternating, and the starting run was counterbalanced across participants. Each run started with an instruction screen that told the participant whether they would be given science or math questions. A central fixation cross appeared for 10 sec at the start and end of each run, and for 15 sec in the middle of each run. Eight fixed trial orders were created, and participants were assigned to one of the orders according to their school year group and gender, ensuring representation of each order across year groups and genders. Each run included 24 trials, with six problems of each type (counterintuitive-true, counterintuitive-false, control-true, control-false). Accuracy and RT were recorded, as well as confidence. All stimuli and a full task description are available on-line (osf.io/ytcwk/).

Inhibitory Control

The go/no-go task, adapted from the works of Watanabe et al. (2002) and Humphrey and Dumontheil (2016), measured response inhibition in a block design with three block types (Figure 3A). Go blocks contained only go trials, where participants used their left or right index finger on the button boxes to indicate the location of a beige square. Simple go/no-no blocks contained 50% go trials (identical to those in go blocks) and 50% no-go trials, where a blue square was presented, and participants should withhold their response. Complex go/no-go blocks contained again 50% go trials and 50% no-go trials. A response indicating the location of pink and yellow squares was required when the color matched the previous trial (go), whereas participants withheld their response when the color did not match the previous trial (no-go). The one-back component taxed working memory. Stimuli remained on the screen for 400 msec, followed by a central fixation cross that remained on screen between 600 and 800 msec. The task was performed in a single run, with four repeats of each block. Each block contained 20 trials, with location, fixation duration, and trial type pseudorandomized. Each block lasted 22 sec and the task lasted approximately 6 min.

The numerical Stroop task, adapted from the work of Khng and Lee (2014), measured interference control in a block design with two block types (Figure 3B). Congruent blocks required participants to indicate the number of digits on screen, with number and digits always congruent (e.g., “4 4 4 4” or “1”). Mixed blocks contained 50% congruent trials and 50% incongruent trials where the number and digits did not match (e.g., “1 1 1 1” or “3 3”) and participants were still required to indicate the number of digits. The digits and number of digits ranged between 1 and 4, and the index and middle fingers of both hands were used to respond using the button boxes (left middle finger for 1, right middle finger for 4). Stimuli remained on the screen until a response was made or 1.1 sec had passed. Each trial lasted 1.5 sec, and the remainder of the trial was filled with a central fixation cross. Blocks alternated, and there were five blocks of each type, with a fixed trial order across participants. The task lasted approximately 5 min.

In both inhibitory control tasks, before the start of a block, participants were shown an instruction screen for 2 sec indicating which block type they would be completing . Both tasks had 10 sec of fixation at the beginning and end of a run, with a 15-sec fixation roughly in the middle of a run. Accuracy and RT were recorded.

Additional Behavioral Tasks

The Vocabulary ($M_{\text{standardized score}}$ = 113.8, SD = 8.4) and Matrix Reasoning ($M_{\text{standardized score}}$ = 110.9, SD = 11.8) subtests of the Wechsler Abbreviated Scale of Intelligence (WASI-II; Wechsler, 2011), as well as two working memory tasks and a task of verbal analogical reasoning, were administered to assess general executive and cognitive ability and test for specificity of associations with inhibitory control. Detailed analyses of the association between science and math problem-solving and verbal and visuospatial reasoning measures in this sample were reported in a separate publication (Brookman-Byrne, Mareschal, Tolmie, & Dumontheil, 2019). Visuospatial working memory (VSWM) was measured with an adapted version of the Dot Matrix test of the Automatic Working Memory Assessment (Alloway, 2007). Dots appeared one by one on a grid, and participants clicked in the grid to indicate the order the dots had appeared. One participant had missing data on this task because of a technical problem. Verbal working memory (VWM) was measured with a backward digit span task, where the experimenter read a series of numbers aloud, and the participant verbally repeated the series in reverse order. For both working memory tasks, the load started at three items and increased in sets of four trials until two incorrect responses were given within a load. The total number of correct trials was recorded for each task. The verbal analogical reasoning task was adapted from the work of Leech, Mareschal, and Cooper (2007) and was administered using an on-line Google Form on a laptop. There were 24 questions in the format A is to B as C is to…, with four response options (e.g., Nose is to Smelling as Eye is to…). The total number of correct responses was recorded.

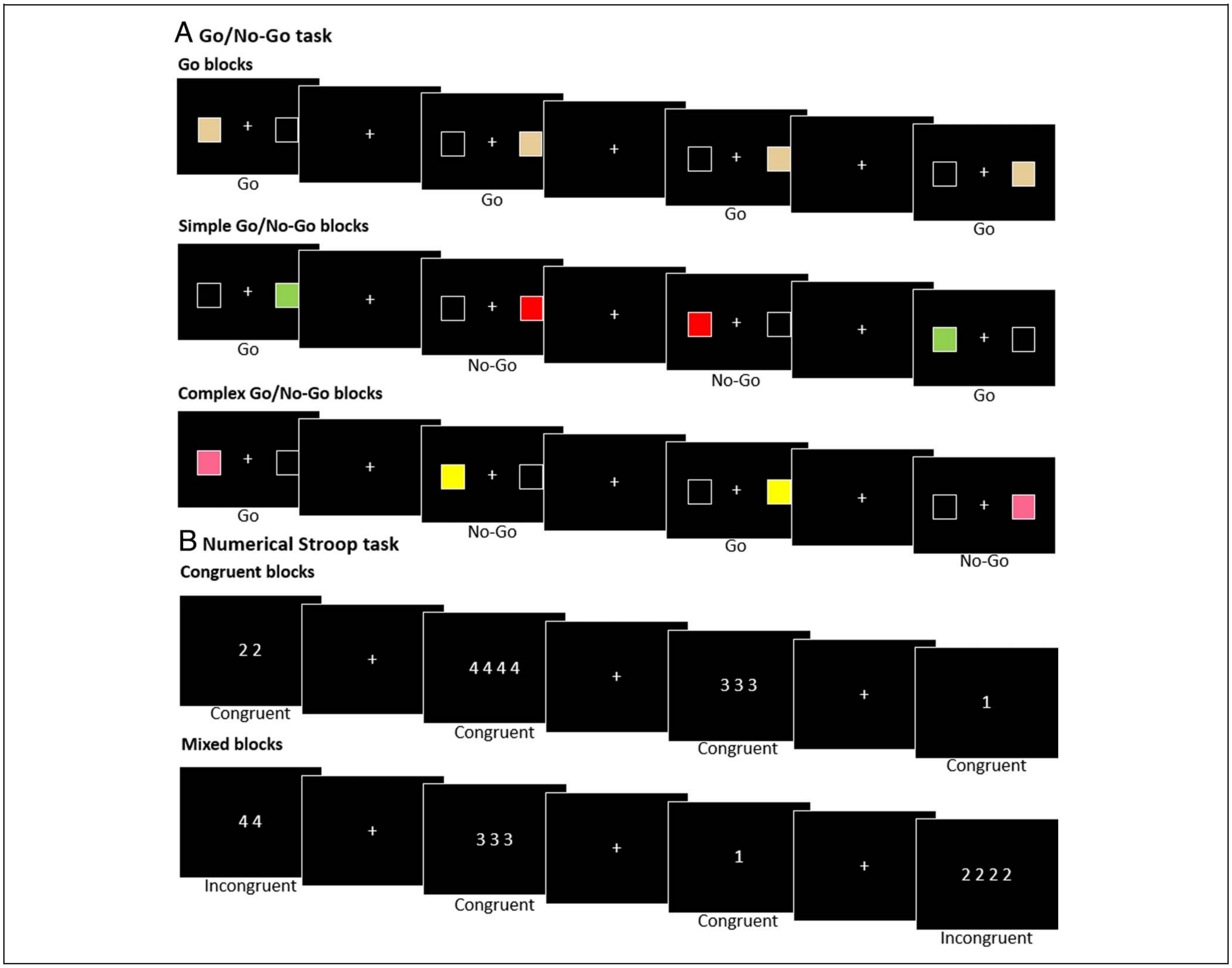

Figure 3. Example sequence of events in (A) the go/no-go task and (B) the numerical Stroop task.

Confirmatory factor analysis was used to create a latent *g* factor using the Lavaan Version 0.6–7 (Rosseel, 2012) structural equation modeling package in R (R Core Team, 2021) with maximum likelihood estimator. One datapoint was missing for the VSWM score, and full information (case-wise rather than list-wise deletion) maximum likelihood estimation was used. Age in months was first regressed out of the VWM, VSWM, and verbal analogical reasoning scores. In a first model, the Vocabulary and Matrix Reasoning WASI standardized scores and the residual VWM, VSWM, and verbal analogical reasoning scores were entered as indicators. Pearson ρ correlation between measures ranged between .08 and .53. The fit of the model was poor, root mean square error of approximation = 0.157, comparative fit index = 0.872, Tucker-Lewis index = 0.744. As the WASI Vocabulary measure had lower correlations with the other variables, this indicator was removed. Correlations between the remaining measures ranged between .21 and .53. The fit of this second model was good, root mean square error of approximation = 0.022, comparative fit index = 0.999, and Tucker-Lewis index = 0.996. Standardized loadings (and R^2) were as follows: WASI Matrix Reasoning, 0.718 (51.6%); verbal analogical reasoning, 0.673 (45.3%); VWM, 0.494 (24.4%); VSWM, 0.663 (44.0%). Subsequent analyses, therefore, used this latent *g* factor and the WASI vocabulary score.

Procedure

The testing session lasted 2 hr. Participants first practiced the tasks outside the scanner to ensure the instructions were understood. The science and math task practice included three science and three math control problems that were not repeated in the scanner. No feedback was provided. Go/no-go practice blocks of 10 trials were repeated if more than one no-go error was made in the simple and complex blocks, and if more than one go error was made in go blocks. A Stroop familiarization phase of eight trials where neutral asterisks were shown instead of digits was repeated if more than one error was made. A practice of eight congruent trials was repeated if more than one error was made. Finally, a practice of eight mixed congruent and incongruent trials was repeated if more than two errors were made.

Once inside the scanner, participants first completed the four runs of the science and math task, followed by a structural scan for 5.5 min, then a single run of the go/no-go task, and finally a single run of the numerical Stroop task. Overall, participants spent approximately 50 min inside the scanner. The additional behavioral tasks were administered outside the scanner either before or after the scanning session. These included science and math anxiety tests, but the results of these are not reported here.

Behavioral Data Analysis

Individual Task Analyses

Repeated-measures ANCOVAs were run on the individual task data with Age in months (*z* score) entered as a covariate to test whether task performance varied as a function of age. Discipline (science, math) and Trial Type (control, counterintuitive) were the within-subject factors for the analyses of accuracy and mean RT (across correct and incorrect trials) in the science and math task. For the inhibitory control tasks, Trial Type was the single within-subject factor, with five levels for the analysis of Accuracy (go trials in go blocks, simple blocks, and complex blocks; no-go trials in simple blocks and complex blocks) and three levels for the analysis of RT (go trials in go blocks, simple blocks, and complex blocks) in the go/no-go task and three levels in the numerical Stroop task (congruent trials in congruent blocks and mixed blocks; incongruent trials in mixed blocks). Mean RT was calculated for correct trials only in the inhibitory control tasks. Main effects were followed up with Bonferroni-corrected pairwise comparisons. In additional analyses of the science and math task, mean accuracy and mean confidence across participants were calculated for each stimulus.

Predictors of Individual Differences in Counterintuitive Reasoning Performance

Hierarchical multiple regressions investigated the extent to which individual differences in inhibitory control specifically or general cognitive abilities could account for individual differences in counterintuitive science and math accuracy and RT. Age in months and science and math control performance (accuracy or RT) were entered as control variables. Then, WASI Vocabulary standardized score, latent *g* factor, and go/no-go and Stroop variables were entered stepwise. Inhibitory control task variables were simple no-go accuracy, complex no-go accuracy, and the residuals of simple go RT and complex go RT cost covarying go RT in go blocks, and of Stroop incongruent accuracy and Stroop incongruent RT covarying, respectively, congruent accuracy and congruent RT, where congruent trials included those in both congruent and mixed blocks (as difference scores are thought to increase measurement errors; see the work of Friedman & Miyake, 2004).

MRI Data Acquisition and Preprocessing

Brain imaging data were acquired on a 1.5 Telsa Siemens Avanto MRI scanner with a 30-channel head coil. Structural data were acquired with a T1-weighted magnetization prepared rapid gradient echo with 2× generalized autocalibrating partially parallel acquisition acceleration, lasting 5.5 min. Functional data were acquired in six sessions using the Center for Magnetic Resonance Research multiband EPI sequence (Xu et al., 2013) 4× acceleration, leak block on (Cauley, Polimeni, Bhat, Wald, & Setsompop, 2014), repetition time = 1 sec, echo time = 45 msec, comprising 44 slices covering most of the cerebrum, with a resolution of $3 \times 3 \times 3$ mm^3.

MRI data were preprocessed and analyzed using SPM12 (www.fil.ion.ucl.ac.uk/spm/software/spm12/). Functional images were realigned to the mean images after the first realignment in a two-pass procedure using a second-degree B-spline interpolation to correct for movement during the session. The bias-field-corrected structural image was coregistered to the mean realigned functional image and segmented on the basis of Montreal Neurological Institute (MNI)-registered International Consortium for Brain Mapping tissue probability maps. Resulting spatial normalization parameters were applied to the realigned images to obtain normalized functional images with a voxel size of $3 \times 3 \times 3$ mm, which were smoothed with an 8-mm FWHM Gaussian kernel. Framewise displacement was calculated for each volume as a scalar measure of head motion across the six realignment estimates (Siegel et al., 2014). Volumes with a framewise displacement greater than 0.9 mm were censored and excluded from the general linear model (GLM) estimation by including a regressor of no interest for each censored volume. Scanning runs with more than 15% of volumes censored or a root mean square movement greater than 1.5 mm were excluded from the analysis.

fMRI Data Analysis

First-level GLMs

Scanning runs were treated as separate time series, each of which was modeled by a set of regressors in the GLM. All regressors were convolved with a canonical hemodynamic response function and, together with the separate regressors representing each censored volume and the session mean, comprised the full model for each session. All coordinates are given in MNI space, region labeling was completed with Automatic Anatomical Labelling (Tzourio-Mazoyer et al., 2002), and BA labeling was completed with MRIcron (Rorden & Brett, 2000).

Science and math task data in each of the four runs were modeled by box-car regressors separately representing counterintuitive and control trials using each trial's RT as the duration. The arrows task and fixation phases were not modeled explicitly and served as a baseline. A first-level contrast of the difference between counterintuitive and control trials was calculated, as well as the difference between counterintuitive correct and incorrect trials.

In the go/no-go task, box-car regressors modeled go, simple go/no-go, and complex go/no-go blocks (22-sec duration), as well as fixation blocks (10- or 15-sec duration). First-level contrasts between simple go/no-go and go blocks, and complex go/no-go and go blocks were calculated. In the numerical Stroop task, box-car regressors modeled congruent and mixed blocks (21-sec duration), as well as fixation blocks (10- or 15-sec duration). First-level contrasts between mixed and congruent blocks were calculated. For both tasks, there was an additional single event-related regressor of duration zero for all errors and a box-car regressor of duration 2 sec modeling the instructions presented at the start of each block.

Intersection of Reasoning and Inhibitory Control Contrasts

First-level contrasts were entered into one-sample t tests to create SPM maps, which were thresholded at $p < .001$ uncorrected at the voxel level and FWE-corrected $p < .05$ at the cluster level. Voxels surviving a voxel-level FWE-corrected $p < .05$ are also reported. There was no difference between science and math in the counterintuitive > control contrast; therefore, science and math conditions were collapsed in subsequent fMRI data analyses.

In order to identify overlapping activations, inclusive masking was used to identify brain areas of increased BOLD signal in the counterintuitive > control contrast or the counterintuitive correct > incorrect contrast and in either of the simple go/no-go > go, complex go/no-go > go, and mixed > congruent numerical Stroop contrasts, using the same statistical threshold of $p < .001$ at the voxel-level and FWE-corrected $p < .05$ at the cluster level. MNI coordinates and cluster size of overlapping regions were obtained using MarsBaR (Brett, Anton, Valabregue, & Poline, 2002). Follow-up Pearson correlations were run in SPSS 26, averaging the data over each overlapping cluster using MarsBar, to test whether individual differences in the inhibitory control tasks contrasts (complex go/no-go > go, Stroop mixed > congruent) correlated across individuals with individual differences in the science and math task contrasts (counterintuitive > control and counterintuitive correct > counterintuitive incorrect).

Similarity Analysis

Overlapping univariate activations do not necessarily reflect the same underlying neural circuits. A multivariate approach was used to further investigate potential neural overlaps between counterintuitive reasoning and inhibitory control. In a first set of analyses, we used the clusters from the science and math counterintuitive > control contrast. For each participant, we (1) extracted parameter estimates for the contrasts complex go/no-go > go, numerical Stroop mixed > congruent, and science and math counterintuitive > control, in each voxel of each cluster; and (2) calculated the correlation of these estimates, across voxels, pairwise for each cluster. This allowed us to assess the similarity of the patterns of activation across voxels between tasks. As the data deviated from a normal distribution, Kendall's tau correlations were used, which were then transformed into Pearson's r using the equation $r = \sin(0.5 \times \pi \times \tau)$ and then using Fisher's transformation into $z_r = 0.5 \times \ln((1+ r)/(1 - r))$ (Walker, 2003). One-tailed one-sample t tests were carried out to test which correlations were significantly above zero at the group level. For reference, a similar comparison was performed between the science counterintuitive > control and math counterintuitive > control contrasts. In a second step of analyses, the same approach was taken using the clusters from the science and math counterintuitive correct > incorrect contrast and comparing patterns of activation across voxels in this contrast and the two inhibitory control task contrasts.

RESULTS

Behavioral Results

Science and Math Task

As expected, participants were both more accurate, $F(1, 32) = 323.44$, $p < .001$, $\eta_p^2 = .910$, and faster, $F(1, 32) = 228.80$, $p < .001$, $\eta_p^2 = .877$, in control trials, $M_{accuracy} = 85.6\%$, $M_{RT} = 5448$ msec, than in counterintuitive trials, $M_{accuracy} = 59.9\%$, $M_{RT} = 6515$ msec. There was a main effect of Age on accuracy, $F(1, 32) = 8.17$, $p = .007$, $\eta_p^2 = .203$, which was modulated by a significant interaction with Trial Type, $F(1, 32) = 4.39$, $p = .044$, $\eta_p^2 = .121$. Follow-up analyses indicated that there was a significant positive correlation between Age and Accuracy in counterintuitive trials, $r = .47$, $p = .005$, which did not reach significance in control trials, $r = .30$, $p = .091$ (Figure 4A). There was no main effect of Age on RT, $\eta_p^2 = .006$, $p = .659$.

There were also differences between disciplines. Participants were both more accurate, $F(1, 32) = 11.43$, $p = .002$, $\eta_p^2 = .263$, and faster, $F(1, 32) = 24.60$, $p < .001$, $\eta_p^2 = .435$, in science, $M_{accuracy} = 75.0\%$, $M_{RT} = 5{,}726$ msec, than math trials, $M_{accuracy} = 70.5\%$, $M_{RT} = 6{,}237$ msec. An additional Trial Type × Discipline interaction, $F(1, 32) = 4.35$, $p = .045$, $\eta_p^2 = .120$, indicated that accuracy was in fact higher in science than math for counterintuitive trials, $F(1, 32) = 13.48$, $p = .001$, $\eta_p^2 = .296$, but not for control trials, $p = .497$ (Figure 4B). Finally, there was a significant interaction between Trial Type, Discipline, and Age, $F(1, 32) = 4.92$, $p = .034$, $\eta_p^2 = .133$. Follow-up analyses showed the interaction between Discipline and Age was significant in control trials, $F(1, 32) = 6.481$, $p = .016$, $\eta_p^2 = .168$, and not in counterintuitive trials, $p = .718$. However, age did not significantly correlate with RT in any of the four trial types, r range $[-.25\text{–}.07]$, ps $> .15$.

The range of mean accuracy for each of the 192 stimuli was wider for counterintuitive trials than control trials

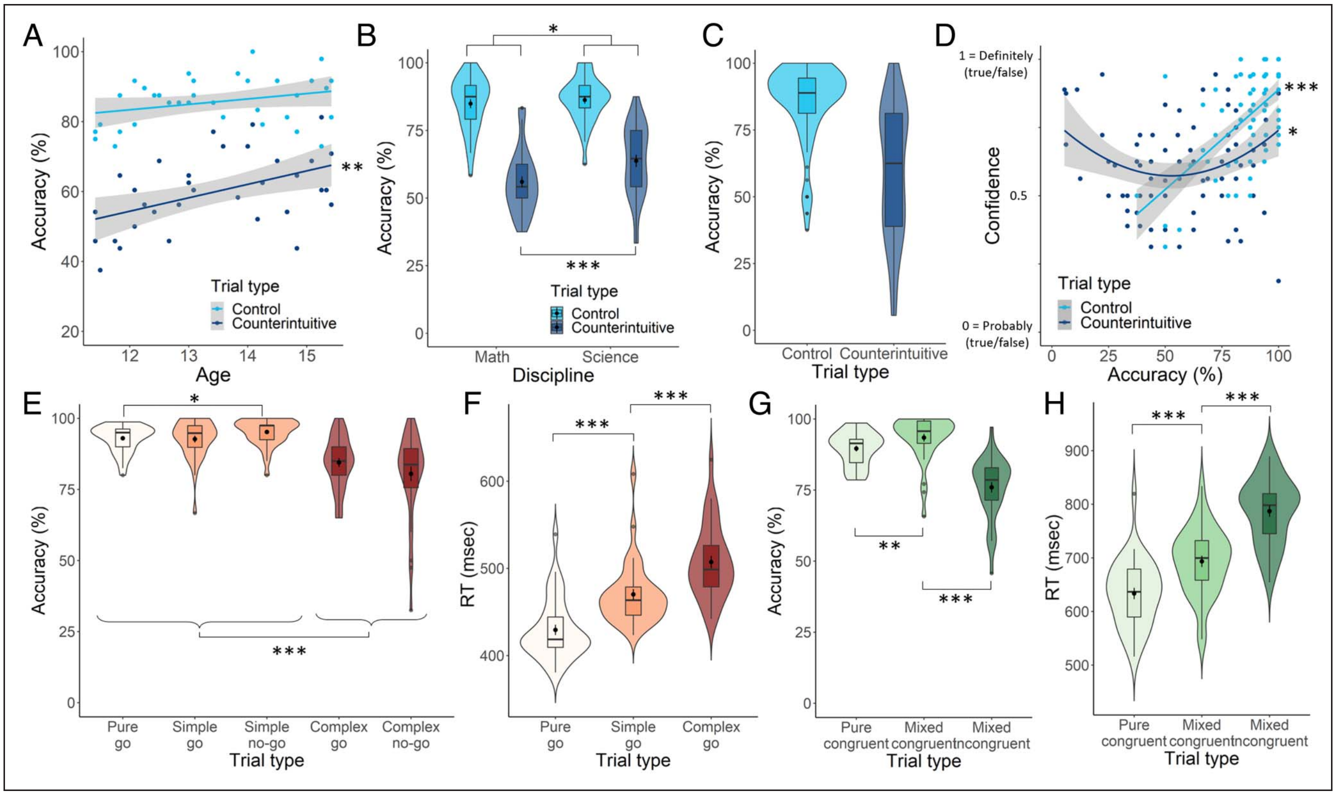

Figure 4. Behavioral results. (A) Scatterplot illustrating the Trial Type × Age interaction observed in the science and math accuracy data. Graphs (B–C, E–H) are a combination of violin plots, boxplots, and estimated marginal means (± *SE*) used to show the distribution of the data. (B) Science and math task accuracy (averaged across stimuli) as a function of trial type and discipline. (C) Science and math task accuracy (averaged across participants) as a function of trial type across the 192 stimuli. (D) Scatterplot illustrating the association between confidence and accuracy across the 192 stimuli, split by trial type. (E) Go/no-go task accuracy as a function of trial type. (F) Go RT as a function of trial type. (G) Numerical Stroop accuracy as a function of trial type. (H) Numerical Stroop RT as a function of trial type. Asterisks denote significance of interaction or pairwise comparisons: **p* < .05, ***p* < .01, ****p* < .001.

(Figure 4C); the pattern was very similar for science and math. It was predicted that for counterintuitive problems, participants may confidently choose the incorrect (intuitive) answer. Mean confidence for each stimulus was plotted as a function of accuracy. Whereas confidence increased linearly with accuracy in control trials, $F(2, 94) = 54.24$, $p < .001$, $\beta_{\text{Accuracy}} = .605$, $R^2 = 36.6\%$, in counterintuitive trials, confidence was highest for stimuli associated with low or high accuracy than for stimuli with intermediary accuracy and a quadratic function best fitted the data, $F(2, 93) = 3.62$, $p = .030$, $\beta_{\text{Accuracy}} = -1.12$, $p = .016$, $\beta^2_{\text{Accuracy}} = 1.20$, $p = .010$, $R^2 = 7.2\%$ (Figure 4D).

Inhibitory Control Tasks

In the go/no-go task, trial types differed in accuracy, $F(4, 128) = 21.47$, $p < .001$, $\eta_p^2 = .402$, and RT, $F(2, 64) = 130.62$, $p < .001$, $\eta_p^2 = .803$. Bonferroni-corrected pairwise comparisons indicated that accuracy did not differ between pure go and simple go trials, $p = 1.00$, but participants were slower in the latter, $p < .001$. Participants had a higher accuracy in simple no-go trials than pure go, $p = .012$, and simple go trials, trend, $p = .052$ (Figure 4E). These results suggest that participants slowed down in the simple go/no-go blocks compared with the go blocks to ensure good no-go trials accuracy. As expected, accuracy was lower in complex go/no-go blocks than pure go and simple go/no-go blocks, *p*s < .001 (Figure 4E), and RTs were slower in complex go than simple go and pure go trials, *p*s < .001 (Figure 4F). Complex go and no-go trials did not differ in accuracy, $p = 1.00$. There was no main effect of Age or interaction between Age and Trial Type for either accuracy, *p*s > .06, or RT, *p*s > .69.

In the numerical Stroop, trial types also differed in accuracy, $F(2, 64) = 74.84$, $p < .001$, $\eta_p^2 = .700$, and RT, $F(2, 64) = 301.11$, $p < .001$, $\eta_p^2 = .904$. Bonferroni-corrected pairwise comparisons indicated that participants, as expected, were slower and less accurate on incongruent than congruent trials, *p*s < .001. Participants were also slower, $p < .001$, but more accurate, $p = .008$, in congruent trials in mixed blocks than in pure blocks (Figure 4G–4H). There was no significant effect of Age and no interaction with Age for either Accuracy or RT, *p*s > .06.

Regression Analyses

Regression analyses were carried out to assess whether general cognitive (WASI Vocabulary and latent *g* factor) or inhibitory control measures (go/no-go, numerical Stroop) explained variance (i) in science and math reasoning performance after controlling for age in months, and (ii) in science and math counterintuitive reasoning specifically after controlling for age in months and performance on control trials. The latent factor *g* predicted both science and math accuracy and RT, and complex no-go accuracy and vocabulary explained additional variance in accuracy (Table 2). The only variable that explained additional specific variance in counterintuitive reasoning was vocabulary (Table 2). The same pattern of results, with vocabulary predicting counterintuitive accuracy, was observed when analyzing science, $\beta = .398$, $\Delta R^2 = 11.4\%$, and math, $\beta = .320$, $\Delta R^2 = 9.8\%$, performance separately. (Note that, although in these analyses stepwise entry selects the most significant predictors at each step, we reran the multiple regressions with the two interference control measures added in a first step as predictors of science and math accuracy or RT, and the results did not change.)

Overall, these behavioral results showed the predicted patterns of poorer performance in counterintuitive than control problems, and cases where participants are confidently incorrect. Counterintuitive problem accuracy improved with age, but there were no other age effects. Vocabulary, *g*, and complex no-go accuracy predicted overall science and math performance; however, only vocabulary specifically predicted variance in counterintuitive reasoning performance.

Neuroimaging Results

Science and Math Counterintuitive Reasoning

The contrast counterintuitive > control trials (Table 3A, Figure 5A) showed increased BOLD signal in bilateral supramarginal gyrus (BA 40), extending into the inferior parietal lobule and angular gyrus, as well as in superior and middle frontal gyri (predominantly BA 8, extending into BA 9) and middle and inferior frontal gyri (BA 45, 47, 11). The right hemisphere activation extended in both DLPFC and VLPFC along BA 45. There was an additional small cluster in the left lingual gyrus. Mean parameter estimates for control and counterintuitive trials in the six larger clusters are plotted in Figure 5B. One-sample *t* tests indicated that there were increases in BOLD signal

Table 2. Results of Multiple Regression Assessing Predictors of Individual Differences in Science and Math Reasoning Overall, and in Science and Math Counterintuitive Reasoning Specifically, Controlling for Performance in Control Trials

	Predictor	β	*t*	*p*
DV: Science & Math Acc.				
$F(4, 29) = 15.20$, $p < .001$, $R^2 = 63.2\%$, $\Delta R^2 = 45.4\%$ (vs. model with age only)	Age	.497	4.67	< .001
	g	.422	3.67	< .001
	Complex no-go acc.	.257	2.33	.027
	WASI Vocabulary[a]	.260	2.21	.035
DV: Science & Math RT				
$F(2, 31) = 4.95$, $p = .014$, $R^2 = 19.3\%$, $\Delta R^2 = 21.8\%$ (vs. model with age only)	Age	−.117	0.75	.460
	g	−.487	3.11	.004
DV: Counterintuitive Trials Acc.				
$F(3,30) = 10.76$, $p < .001$, $R^2 = 47.0\%$, $\Delta R^2 = 6.6\%$ (vs. model with age and control acc.)	Age	.401	2.94	.006
	Control trials acc.	.344	2.34	.026
	WASI Vocabulary[a]	.312	2.21	.035
DV: Counterintuitive Trials RT				
$F(2,31) = 52.08$, $p < .001$, $R^2 = 75.6\%$	Age	.024	0.28	.780
	Control trials RT	.880	10.18	< .001

[a] Standardized score.

Acc. = accuracy. We report adjusted R^2 and change in adjusted R^2 (ΔR^2) between models.

Table 3. Regions Showing Differences in BOLD Signal in the Science and Math Task when Comparing (A) Counterintuitive Trials to Control Trials, (B) Counterintuitive Correct Trials to Incorrect Trials

			MNI				
Brain Region	*L/R*	*BA*	*x*	*y*	*z*	*Z*	*k*
(A) Science & Math Counterintuitive > Control							
Supramarginal gyrus	R	40	60	−31	50	4.94[a]	527[b]
Lingual gyrus	L	37	−27	−49	−7	4.67[a]	26
Superior frontal gyrus	R	8	18	20	62	4.61[a]	220[b]
MFG	R	9	36	17	59	4.13	
Supramarginal gyrus	L	40	−63	−28	44	4.58[a]	163[b]
Inferior parietal lobule	L	40	−57	−37	47	3.99	
MFG	L	8	−36	14	62	4.56[a]	176[b]
Superior frontal gyrus	L	8	−21	26	62	3.73	
MFG	R	45	45	44	11	4.49	337[b]
IFG (orbital)	R	47	42	47	−10	4.32	
MFG (orbital)	R	11	27	41	−19	4.18	
IFG (orbital)	L	47	−42	47	−10	3.68	134[b]
IFG	L	45	−45	41	14	3.48	
IFG	L	46	−45	50	5	3.39	
(B) Science & Math Counterintuitive Correct > Incorrect							
Precuneus	L	7	−9	−67	56	4.08	108[b]
Inferior parietal lobule	L	40	−27	−40	38	3.76	100[b]

[a] $p_{FWE} < .05$ at the voxel level.

[b] $p_{FWE} < .05$ at the cluster level, cluster defining threshold: $p_{uncorr} < .001$.

k = cluster size; L/R = left/right hemisphere.

in control trials in all regions compared with the implicit baseline (which includes fixation and the arrows task), all $ts(33) > 2.47$, $ps < .019$, except in the right parietal BA 40 cluster, $t(33) = 0.55$, $p = 0.58$, which was therefore the only region showing specificity of increase in BOLD signal for counterintuitive trials.

The contrast counterintuitive correct > incorrect aimed to identify brain regions that may support overcoming intuitive responses, beyond a more general greater involvement in counterintuitive trials than control trials. Two brain regions showed greater BOLD signal in correct than incorrect counterintuitive trials: the precuneus (BA 7) and inferior parietal lobule (BA 40) in the left hemisphere (Table 3B).

There was no association with age for either contrast.

Inhibitory Control Tasks

The simple go/no-go > go contrast showed no region of increased BOLD signal. The complex go/no-go > go contrast (Figure 5C, Table 4A) revealed increased BOLD signal in a large bilateral parietal cluster covering the inferior parietal lobules, superior parietal gyri, and precuneus, in a large bilateral frontal cluster covering the middle frontal gyri, precentral gyri, SMA, and extending into the right anterior insula, as well as in smaller clusters in the left insula and caudate nucleus, right middle orbital frontal gyrus, and in the cerebellum.

The mixed > congruent numerical Stroop contrast (Figure 5D, Table 4B) similarly showed increased BOLD signal in parietal clusters covering the inferior parietal lobules, superior parietal gyri, and precuneus, in frontal clusters located in the middle/inferior frontal gyri, extending into the precentral gyri, and in a cluster in the right inferior temporal gyrus.

Overlapping Activation

Of the six clusters observed in the counterintuitive > control contrast, all except the left IFG showed partial overlap

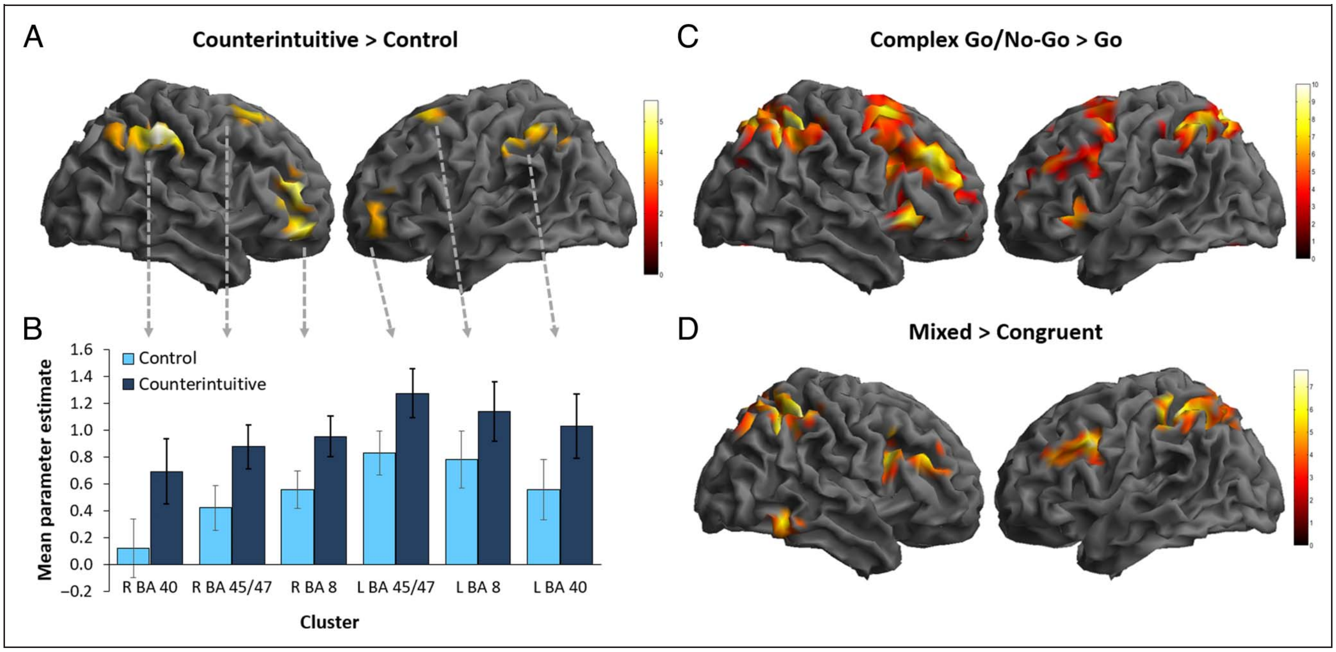

Figure 5. Main contrasts of interest in the science and math, go/no-go, and numerical Stroop tasks. (A) Regions of increased BOLD signal in the right and left hemispheres in the counterintuitive > control contrast of the science and math task. (B) Mean parameter estimates (± *SE*) in counterintuitive and control trials for the six main clusters (see Table 2). Zero represents the implicit baseline of the model, which includes the arrows task and fixation phases. (C) Regions of increased BOLD signal in the right and left hemispheres in the complex go/no-go blocks > go blocks contrast of the go/no-go task. (D) Regions of increased BOLD signal in the right and left hemispheres in the mixed blocks > congruent blocks contrast of the numerical Stroop task. R = right; L = left. For all contrasts $p_{\text{uncorr}} < .001$ at the voxel level and $p_{\text{FWE}} < .05$ at the cluster level.

with the complex go/no-go > go contrast, and all except the left middle and right superior frontal gyri also showed partial overlap with the numerical Stroop mixed > congruent contrast (Table 5, Figure 6A–6B). There was overlap between all three contrasts in the right intraparietal sulcus (BA 40) and MFG (BA 45) and left inferior parietal lobule (BA 40; Figure 7). However, this overlap was not complete, the network of brain regions showing increased BOLD signal in the inhibitory control tasks was broader, and part of the increased BOLD signal in the six science and math clusters was unique to the counterintuitive > control contrast.

In contrast, both clusters observed in the counterintuitive correct > incorrect contrast, which did not overlap with regions revealed by the counterintuitive > control contrast (Figure 6C), mostly fell within the brain regions showing increased BOLD signal in the complex go/no-go > go contrast (Figure 6D) and the numerical Stroop mixed > congruent contrast (Figure 6E). Overlap between all three contrasts was found in the left precuneus (BA 7) and inferior parietal lobule (BA 40; Figure 7).

Correlations of individual differences in mean contrast parameter estimates for each of these overlapping clusters showed no significant association between activation in the science and math reasoning task contrast and activation in the complex go/no-go > go or numerical Stroop mixed > congruent contrast, *r* range (−.17, .27).

Univariate analyses of the fMRI data therefore showed that although some spatial overlap in activation was observed, those participants who showed greater activation during counterintuitive reasoning did not necessarily show greater activation during inhibitory control tasks.

Voxel-level Correlation Analyses

In a second step, multivariate analyses were used to explore similarities in the patterns of activation across voxels between the contrasts of interest. First, voxel-level data were extracted for each of the six clusters from the science and math counterintuitive > control contrast. Kendall's τ correlations were performed to assess similarities in the pattern of activation between the science and math counterintuitive > control, complex go/no-go > go and numerical Stroop mixed > congruent contrasts. One-tailed one-sample *t* tests performed on the Z_r values at the group level tested whether at the group level, there were significant positive correlations between the patterns of activation across voxels. Results indicated that the patterns of activation across voxels in the right supramarginal gyrus/inferior parietal lobule (BA 40) cluster were similar between the counterintuitive > control and both the Stroop mixed > congruent, $z_r = .19$, $t(33) = 2.34$, $p = .013$, and the complex go/no-go > go contrasts, $z_r = .16$, $t(33) = 1.77$, $p = .043$ (Figure 8). Correlations between the two inhibitory control tasks were much larger and all significant, $ts(33) > 5.12$, $ps < .001$, whereas correlations between science and math counterintuitive > control contrasts were significantly greater than zero in the left MFG (BA 47), right MFG (BA 45/47), and right

Table 4. Regions Showing Changes in BOLD Signal in (A) the Complex Go/No-Go > Go Blocks Contrast, and (B) the Numerical Stroop Mixed Congruent and Incongruent > Pure Congruent Blocks Contrast of the Inhibitory Control Tasks

			MNI				
	L/R	*BA*	*x*	*y*	*z*	*Z*	*k*
(A) Complex go/no-go > go							
Inferior parietal lobule	R	40	39	−49	50	7.20[a]	2555[b]
Inferior parietal lobule	L	40	−42	−46	50	6.67[a]	
Precuneus	R	7	9	−67	56	6.58[a]	
Precuneus	L	7	−6	−70	53	6.43[a]	
MFG	R	45	45	35	35	6.66[a]	3898[b]
SMA	L	32	−3	14	50	6.51[a]	
Superior frontal gyrus	R	6/8	33	8	65	6.42[a]	
Anterior insula	R		33	23	−4	6.14[a]	
MFG	R	46	36	50	20	6.14[a]	
MFG	L	6	−30	2	56	5.98[a]	
IFG	R	44	45	8	26	5.39[a]	
Precentral gyrus	L	6	−42	2	35	5.35[a]	
MFG	L	46	−33	56	17	5.06[a]	
Insula	L	48	−30	20	5	6.17[a]	191[b]
Caudate nucleus	L		−18	2	20	3.65	
Crus I of cerebellum	L		−33	−58	−34	6.10[a]	590[b]
Crus II of cerebellum	L		−6	−79	−28	5.43[a]	
Crus I of cerebellum	R		36	−61	−31	6.00[a]	186[b]
Lobule III of vermis			3	−43	−19	5.72[a]	228[b]
Middle orbital frontal gyrus	R	11	27	44	−19	5.47[a]	70
(B) Numerical Stroop Mixed > congruent							
Precentral gyrus	L	44	−42	8	32	5.89[a]	517[b]
IFG	L	45	−51	29	32	5.10[a]	
MFG	L	46	−48	50	14	4.62[a]	
Inferior parietal lobule	L	40	−45	−43	56	5.88[a]	1007[b]
Inferior parietal lobule	L	40	−48	−37	44	5.59[a]	
Precuneus	L	7	−6	−64	50	4.81[a]	
Superior parietal gyrus	L	7	−27	−61	53	4.74[a]	
Inferior parietal lobule	R	2/40	48	−37	53	5.82[a]	888[b]
Angular gyrus	R	40/7	33	−52	44	5.73[a]	
Inferior parietal lobule	R	40	39	−43	44	5.68[a]	
Superior occipital gyrus	R	7	30	−64	41	5.30[a]	

Table 4. (*continued*)

	L/R	*BA*	*MNI*			*Z*	*k*
			x	*y*	*z*		
IFG	R	44	45	8	23	5.46[a]	564[b]
MFG	R	45	51	35	23	4.97[a]	
Inferior temporal gyrus	R	20	51	−52	−13	4.96[a]	143[b]

[a] $p_{FWE} < .05$ at the voxel-level.

[b] $p_{FWE} < .05$ at the cluster-level, cluster defining threshold: $p_{uncorr} < .001$.

k = cluster size; L/R = left/right hemisphere.

supramarginal gyrus/inferior parietal lobule (BA 40), $t(33) > 3.06$, *p*s $< .002$ (Figure 8).

Similar analyses were carried out for the two clusters from the science and math counterintuitive correct > incorrect contrast. While the patterns of activation across voxels were again most similar when comparing the complex go/no-go > go and numerical Stroop mixed > congruent contrasts to each other , BA 40: $z_r = .92$, $t(33) > 12.28$, $p < .001$, BA 40: $z_r = .73$, $t(33) = 8.82$, $p < .001$, both right inferior parietal lobule (BA 40) and precuneus (BA 7) clusters also showed significant correlations in patterns of activation for correct versus incorrect counterintuitive science and math trials and the Stroop mixed > congruent contrast , BA 40: $z_r = .28$, $t(33) = 2.45$, $p = .010$; BA 7: $z_r = .20$, $t(33) = 2.39$, $p = .011$, and the complex go/no-go > go contrast , BA 40: $z_r = .28$, $t(33) = 2.53$, $p = .008$; BA 7: $z_r = .32$, $t(33) = 3.84$, $p < .001$ (Figure 8).

All correlations remained significantly greater than zero after applying the false discovery rate Benjamini–Hochberg procedure to correct for multiple comparisons except for the correlation between counterintuitive > control and complex go/no-go > go contrasts in right BA 40.

Voxel-level correlational analyses therefore indicated that there were strong similarities in pattern of activation between the two inhibitory control tasks contrasts, and similarities, of a smaller magnitude, with the counterintuitive reasoning contrasts in parietal clusters.

Table 5. Overlapping Activation between the Inhibitory Control Tasks Contrasts and the Science and Math Task Counterintuitive > Control Contrast

Brain Region	*L/R*	*BA*	*MNI x*	*MNI y*	*MNI z*	*k*
Science & Math Counterintuitive > Control ∩ Complex go/no-go > go						
Superior frontal gyrus	R	8	25	15	61	127
IFG	R	11	28	44	−18	8
MFG	R	45	45	43	21	69
Inferior parietal lobule	R	40	49	−43	47	333
MFG	L	8	−23	13	63	20
Inferior parietal lobule	L	40	−52	−44	49	66
Science & Math Counterintuitive > Control ∩ Mixed > Congruent Numerical Stroop						
MFG	R	45	46	41	21	67
Inferior parietal lobule	R	40	47	−41	48	188
IFG	L	45	−47	47	9	9
Inferior parietal lobule	L	40	−53	−40	49	58

k = cluster size; L/R = left/right hemisphere. Coordinates are the center of mass of each cluster as calculated by MarsBaR.

DISCUSSION

Previous research has proposed that inhibitory control plays an important role in counterintuitive reasoning by allowing the selection of scientific theories and suppression of misleading perceptual cues, naive theories, or prior knowledge (Potvin et al., 2020; Mareschal, 2016; Houdé, 2000). Neuroimaging studies have demonstrated increased pFC activation, although in inconsistent locations, during the resolution of counterintuitive science and math problems compared with intuitive problems, which was interpreted as reflecting inhibitory control. The current study aimed to compare, within the same group of adolescent participants, the behavioral and neural correlates of science and math counterintuitive reasoning and inhibitory control. Behavioral data in this relatively small sample of 34 participants did not replicate previous findings of association between complex response inhibition and interference control and counterintuitive reasoning performance. Instead, complex response inhibition accuracy, along with more general measures of *g* and vocabulary, predicted overall science and math reasoning accuracy. Univariate neuroimaging data analyses showed partial overlap between regions activated in tasks requiring response inhibition combined with a working load or

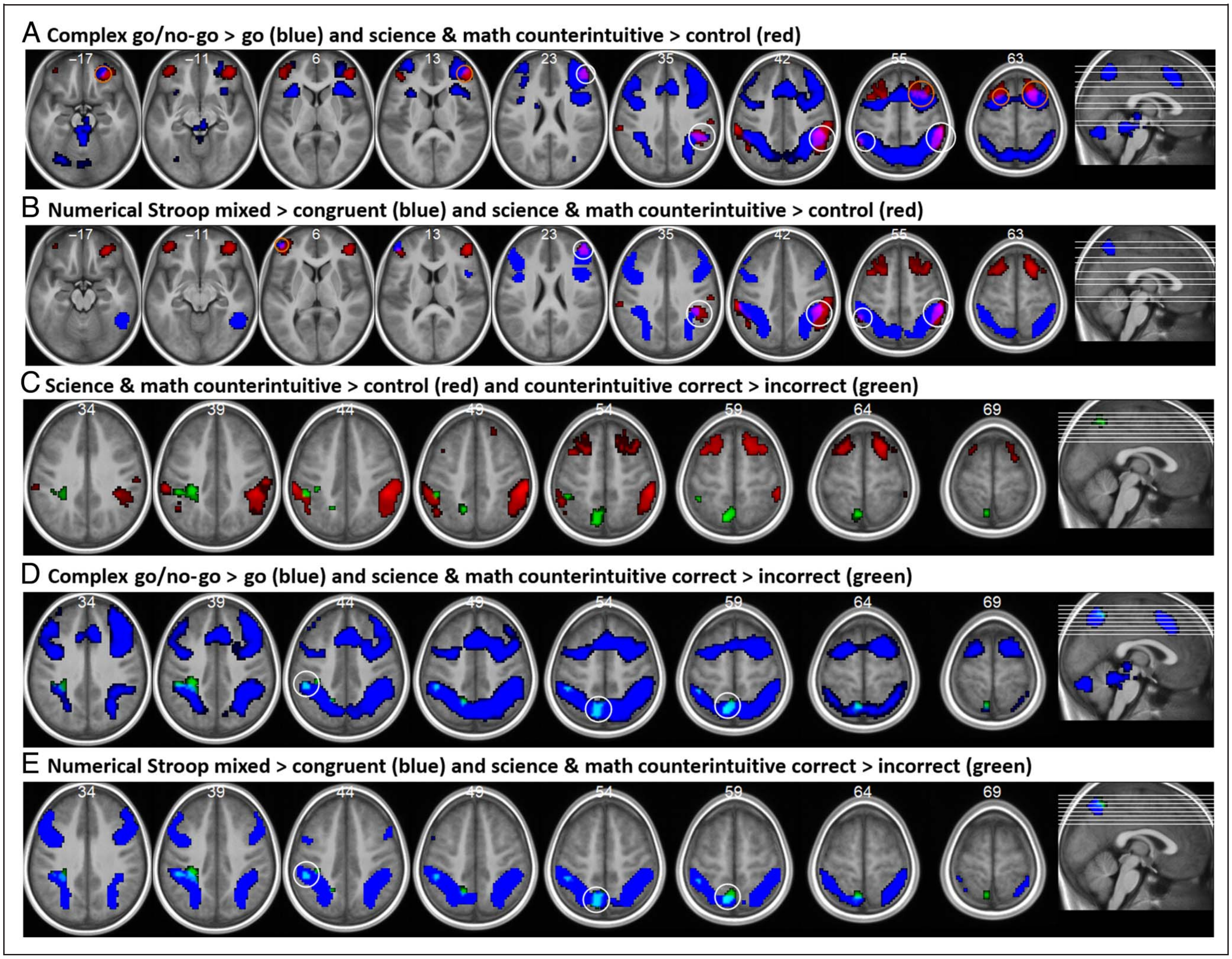

Figure 6. Overlapping activation between the science and math and inhibitory control tasks contrasts. Overlap between increases in BOLD signal in science and math counterintuitive > control contrast and (A) complex go/no-go > go and (B) numerical mixed > congruent contrasts. Inhibitory control tasks contrasts are shown in blue, whereas the science and math counterintuitive > control contrast is shown in red, and regions of overlap are shown in purple. Note that the slices shown are the same in (A) and (B) to enable comparison between images, and the *z* coordinate is indicated at the top. Orange circles highlight regions of overlap between the two contrasts, whereas white circles highlight regions common to all three contrasts. Overlap between increases in BOLD signal in the science and math counterintuitive correct > incorrect and the (C) science and math counterintuitive > control contrast, (D) complex go/no-go > go, and (E) numerical mixed > congruent contrasts. The counterintuitive correct > incorrect contrast is shown in green, and regions of overlap are shown in cyan. Contrasts are overlaid using MRIcron onto an image of the mean normalized structural brain image of participants created using *ImCalc* in SPM.

interference control, but not simple response inhibition, and during counterintuitive reasoning or when comparing correct counterintuitive reasoning to intuitive errors. Multivariate analyses suggested that the overlapping activation may reflect overlapping neural populations in the parietal cortex only. Overall, these results provide only limited evidence for a role of domain-general inhibitory control mechanisms in counterintuitive science and math reasoning.

Science and Math Counterintuitive Reasoning

As expected, better average performance was observed in control trials than counterintuitive trials, with mean accuracy above 75% for 79 of 96 control problems and 33 of 96 counterintuitive problems. Whereas confidence increased with accuracy for control trials, a quadratic association was found for counterintuitive trials. This pattern fits with the proposal that individuals are confident in their intuitive (and incorrect) response. A previous study on counterintuitive electrical circuits in adults similarly found that participants claimed to be certain of half of their incorrect answers (Potvin et al., 2014). Accuracy showed small improvements in counterintuitive but not control trials with age, from 54.5% at age 12 years to 66.0% at age 15 years, a finding in line with previous reports of continuing difficulties with counterintuitive concepts during adolescence (Brookman-Byrne et al., 2018; Potvin & Cyr, 2017; Ryan & Williams, 2007). The behavioral age effect

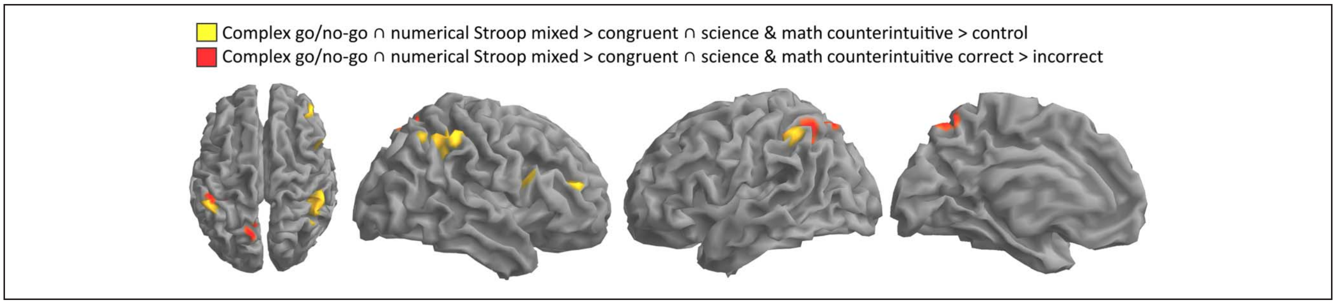

Figure 7. Render of the overlapping activation between the science and math and the two inhibitory control tasks contrasts. For all contrasts $p_{\mathrm{uncorr}} < .001$ at the voxel level and $p_{\mathrm{FWE}} < .05$ at the cluster level.

was not reflected in age differences in BOLD signal in either of the science and math task contrasts.

Reasoning about counterintuitive science and math concepts, compared with solving control problems that did not draw on counterintuitive concepts, was associated with increased BOLD signal in bilateral supramarginal gyrus (BA 40), extending into the inferior parietal lobule and angular gyrus, the superior and middle frontal gyri (BA 8/9) and middle and inferior frontal gyri (BA 45/47/11). Although these activations were mostly bilateral, they tended to be stronger in the right hemisphere than the left hemisphere. There was no difference between science

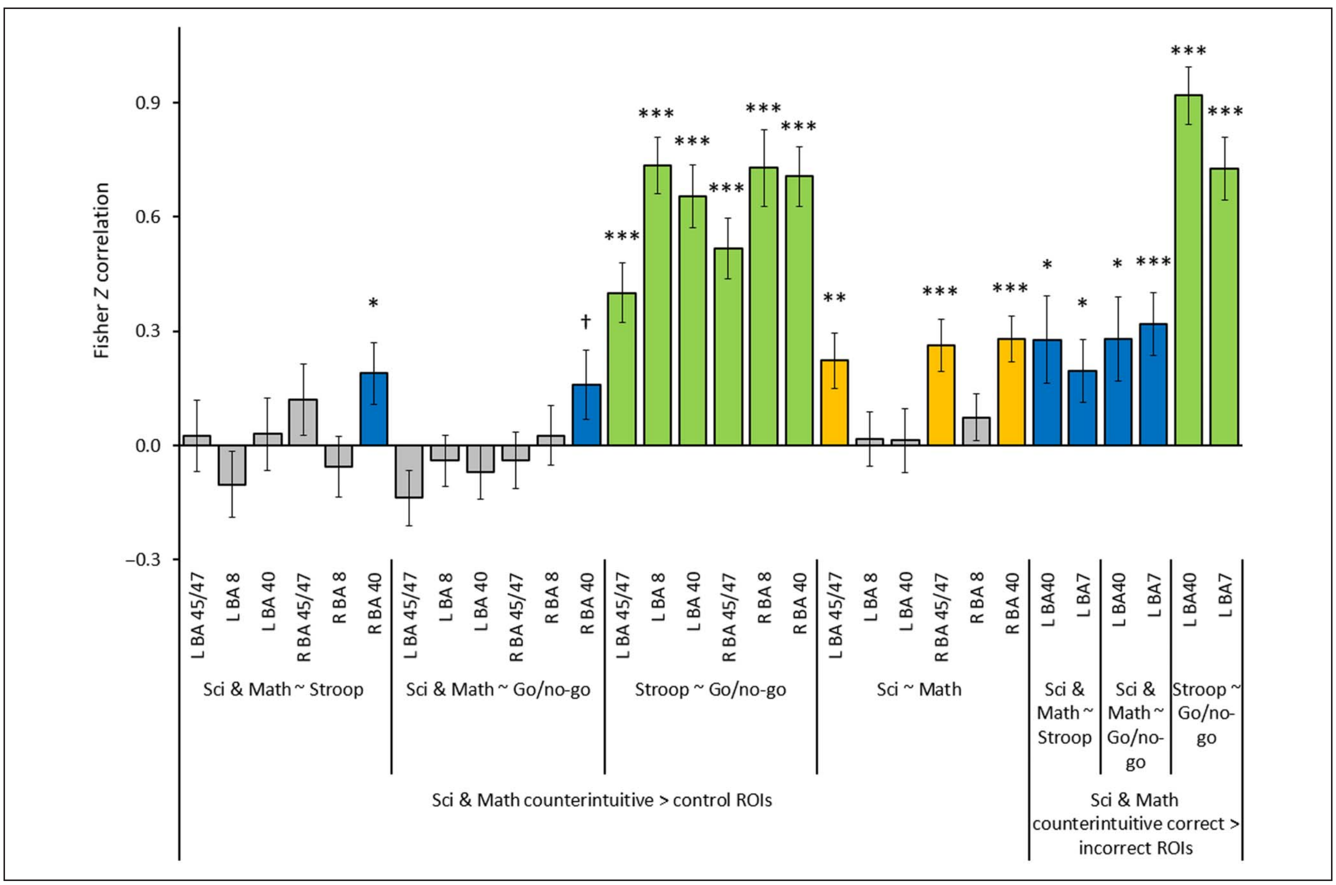

Figure 8. Mean (± *SE*) Fisher *z*-transformed voxel-wise correlations between the key contrasts of interest within each cluster activated in the science and math counterintuitive > control contrast (left) or the two clusters activated in the science and math counterintuitive correct > incorrect contrast (right). Note that although the same broad "L BA 40" label was used, the clusters differed in the two sets of analyses (see Table 3A and B, and Figure 6C). Go/no-go: Complex go/no-go > go contrast; Math: math counterintuitive > control contrast; Sci & Math: science and math counterintuitive > control contrast (left) or science and math counterintuitive correct > incorrect (right); Science: science counterintuitive > control contrast; Stroop: numerical Stroop mixed > congruent contrast. L = left hemisphere; R = right hemisphere. † $p < .10$, *$p < .05$, **$p < .01$, ***$p < .001$; p values are for one-tailed one-sample t tests testing whether the mean correlations were greater than zero, with false discovery rate correction for multiple comparisons. Blue bars indicate significant correlations between the science and math and either inhibitory control task, green bars between the two inhibitory control tasks, and yellow bars between science and math, whereas gray bars indicate nonsignificant correlations.

and math for this contrast, suggesting solving counterintuitive problems in these two subjects draw on similar brain regions. However, multivariate analyses indicated that only the bilateral anterior frontal clusters (peaking in BA45/47) and right supramarginal cluster (BA 40) showed positive correlations between science and math changes in BOLD signal across voxels, suggesting shared neural substrates in these regions only. Although the counterintuitive reasoning problems were more difficult than the control problems, the activation clusters did not overlap with the peaks of the multidemand network (Duncan, 2010). Note that additional analyses showed that when combining control and counterintuitive trials and comparing them against the implicit baseline, there was widespread activation that did overlap with the multidemand network in the DLPFC, pre-SMA and anterior insula/frontal operculum.

There was some overlap between the activations observed in the science and math counterintuitive versus control trials contrast and previous studies of counterintuitive reasoning synthesized in Figure 1 and Table 1. There was overlap with activation peaks observed by Stavy and Babai (2010) bilaterally in the inferior frontal cortex in BA 47. The prefrontal cluster in the left hemisphere also overlapped with the peak activation reported by Allaire-Duquette et al. (2019) in BA 47/46. Finally, there was overlap with peak activations reported in the right superior frontal gyrus in the posterior part of BA 8 by Brault Foisy et al. (2015) and Potvin et al. (2020).

The comparison between counterintuitive correct and incorrect trials assessed which regions may be implicated when participants successfully inhibited the intuitive response and retrieved the correct scientific response. This contrast revealed two clusters, which did not overlap with the counterintuitive versus control contrast, located in the precuneus (BA 7) and inferior parietal lobule (BA 40) in the left hemisphere. One other study ran this contrast and reported greater activation in bilateral BA 11/47, with the cluster in the right hemisphere in a similar region to the contrast counterintuitive correct versus intuitive correct (Table 1; Stavy & Babai, 2010).

Science and Math Counterintuitive Reasoning and Inhibitory Control

The regression analyses indicated that complex no-go accuracy predicted overall science and math accuracy, along with vocabulary and a general *g* factor combining visuospatial and VWM and visuospatial matrix reasoning and verbal analogical reasoning measures. Only vocabulary predicted variance in counterintuitive trials specifically, which means we did not replicate results from a previous behavioral study in a larger sample of adolescents that showed response and interference control to predict counterintuitive performance (Brookman-Byrne et al., 2018). We attribute this lack of replication to the smaller sample size in this study, the small effect size observed in the original study, as well as to the differences in task design for the go/no-go task in particular—the original study had 25% no-go trials compared with 50% no-go trials in mixed go/no-go blocks in this study. Our results highlight that counterintuitive science and math reasoning builds on cognitive capacities recruited for science and math reasoning more broadly, which is consistent with the neuroimaging results described above, indicating that most of the regions showing increased BOLD signal when reasoning about counterintuitive concepts also showed increased BOLD signal when adolescents were answering control questions that did not draw on counterintuitive science and math concepts.

Univariate analyses showed increases in BOLD signal in the counterintuitive versus control contrast overlapped to some extent with increased BOLD signal in both response inhibition in the context of a small working memory load (complex go/no-go) and interference control (numerical Stroop); however, the overlap was far from complete. There were small clusters of overlap across all three contrasts in the inferior parietal lobule bilaterally (BA 40) and the right MFG (BA 45). There was further overlap between the science and math counterintuitive > control contrast and the complex go/no-go task in the superior/MFG bilaterally (BA 8), close to the location of overlap with the studies by Brault Foisy et al. (2015) and Potvin et al. (2020) in the right hemisphere, and in the left IFG, close to the location of overlap with the study by Stavy and Babai (2010). Similarly, additional overlap with the numerical Stroop task was found in the left IFG (BA 45), close to the location of overlap with the study by Allaire-Duquette et al. (2019). While these results are encouraging and suggest that previous interpretation of increased pFC activation during counterintuitive reasoning may indeed reflect the recruitment of domain-general inhibitory control mechanisms, the additional multivariate analyses performed do not support this interpretation.

Indeed, the multivariate analyses showed the right supramarginal gyrus/intraparietal sulcus BA 40 cluster was the only cluster showing a positive association between increases in BOLD signal across voxels during counterintuitive reasoning and during the complex go/no-go and numerical Stroop task. Interestingly, the right parietal cluster was also the only cluster showing specific activation for counterintuitive trials and no increase in BOLD signal in control trials in the science and math task (Figure 5B). The center of mass of the overlapping cluster, averaging across the two IC tasks, was (47 −41 48), and Neurosynth reports association with the terms "working memory," "calculation," "symbolic," "attention," "visually," and "spatial," suggesting a role that goes beyond inhibitory control. Indeed, Criaud and Boulinguez (2013) identified a large cluster including the right inferior parietal lobule and supramarginal gyrus as specifically activated in a complex go/no-go task with a working memory requirement compared with other go/no-go tasks. Therefore, the neural activity common to all three tasks may reflect the engagement of high attentional/working memory resources or mental imagery, rather than inhibitory processes per se. This interpretation would fit with the previously discussed suggestion that activation in complex go/no-go tasks in the frontal lobes is driven by the engagement of high attentional and working memory resources (Criaud & Boulinguez, 2013).

Although the multivariate analysis does not provide evidence of the recruitment of broadly similar neural networks for science and math counterintuitive reasoning and domain-general inhibitory control, the specific association observed in the right BA 40 is interesting. The intraparietal sulcus is thought to support the representation or comparison of magnitude, including numerical magnitude (Peters & De Smedt, 2018; Cohen Kadosh, Lammertyn, & Izard, 2008; Dehaene, Molko, Cohen, & Wilson, 2004). Magnitude information was relevant in many trials of the science and math task used in this study (e.g., comparing the size of animals, concentrations, perimeters, surface areas, volumes, pressure, equations), and this may have driven the overlap observed. However, the IPS is also involved in top–down control of visual attention (Corbetta & Shulman, 2002) and working memory (Killebrew, Mruczek, & Berryhill, 2015), with links observed between these different processes. For example, activation in the IPS in a VSWM task during childhood and adolescence predicts arithmetical performance 2 years later (Dumontheil & Klingberg, 2012). The precise location of brain activation within the IPS varies between studies, and more research is needed to directly compare parietal cortex activation during counterintuitive math and science reasoning and in tasks of numerical magnitude comparison, visual attention, and VSWM.

The contrast identifying neural regions supporting correct counterintuitive reasoning, compared with providing an incorrect intuitive response, showed two clusters of activation that overlapped with both the complex go/no-go and numerical Stroop contrasts. Multivariate analyses further showed significant associations between patterns of activation across voxels in the three contrasts, providing support for shared neural substrates. Neurosynth reports the BA 7 peak associates with the terms "working memory," "calculation," "visuospatial," "spatial," "executive," "encoding retrieval," and "demands," suggesting this brain region has a broad visuospatial executive role. We note, however, that meta-analyses of working memory tasks have identified precuneus activations that tended to be more lateral and inferior in their location (Wang et al., 2019; Daniel, Katz, & Robinson, 2016). The BA 40 peak associates with the terms "passively," "hands," "gestures," and "video clips" in Neurosynth, which is more difficult to interpret in the context of this study.

Both behavioral and neuroimaging results provide evidence that simple response inhibition may not play an important role in science and math reasoning (cf. mixed results observed by Donati et al., 2019; Khng & Lee, 2009; St Clair-Thompson & Gathercole, 2006) but that the capacity to inhibit a dominant response within the context of a working memory load, which would be more akin to science or math problem-solving contexts, is more relevant to science and math reasoning. Indeed, the behavioral association was found with complex no-go accuracy but not simple no-go accuracy, and the univariate neuroimaging analyses showed overlapping activation with the complex go/no-go task. These results are also aligned with previous research suggesting that response inhibition may play a greater role in counterintuitive reasoning in childhood (Zaitchik et al., 2014; Baker et al., 2011) than adolescence (Rhodes et al., 2014, 2016), when interference control may play a greater role (Kwon & Lawson, 2000; Kwon et al., 2000). One caveat is that the simple go/no-go blocks, when contrasted with go blocks, showed no increased BOLD signal. Although 50% no-go trials are common in fMRI studies of inhibitory control (e.g., Tamm, Menon, & Reiss, 2002) to increase the proportion of trials requiring inhibitory control per task block, this high percentage may have meant that the task became too simple to demonstrate strong response inhibition effects.

The limited neurocognitive overlap between counterintuitive reasoning and the two inhibitory control tasks used in this study may be driven by the fact that the conflict detection and resolution processes involved in inhibitory control are domain-specific (Egner, 2008). Indeed, inhibitory control is multifaceted (Banich & Depue, 2015) and the neural implementation of inhibitory control may be domain-specific, depending on, for example, the biasing of stimulus representation in the sensory cortex, response preparation processes in the motor cortex, or the prioritization of emotionally salient stimuli in the limbic system (see the work of Egner, 2008, for review and discussion). This was in fact one of the premises behind the development of the Stop & Think intervention, which encourages children to use inhibitory control to solve counterintuitive problems and is embedded within the domains of science and math specifically (Wilkinson et al., 2020; Palak et al., 2019). The numerical Stroop task used numerical stimuli; there was therefore some overlap with the type of representation manipulated in some of the science and math task problems. However, the problems also required the consideration of more abstract concepts (e.g., concentration) and resistance to interference from misleading visual information (e.g., surface area, when considering parameters). Future research could try to separate counterintuitive problems on the basis of the nature of the information that elicits a conflict; however, conflict will likely arise at a combination of level for a given problem.

Conclusions

This study aimed to go beyond previous research, investigating the role of inhibitory control in science and math counterintuitive reasoning by studying associations between the behavioral and neural correlates of both inhibitory control and counterintuitive reasoning tasks within the same participants. Our review of the literature showed little consistency in patterns of activation observed during counterintuitive reasoning across studies. The results of our own study provide little evidence that inhibitory control, whether simple response inhibition, response inhibition combined with a working memory load, or interference control, specifically supports counterintuitive reasoning during adolescence. However, it is possible that domain-specific inhibitory control processes are at play. Although some univariate overlaps were observed, evidence from multivariate analyses was limited to parietal clusters that likely reflect general visuospatial attentional executive processes. These results highlight the importance of using localizer tasks and a range of analytic approaches to investigate to what extent common neural networks underlie performance of different cognitive tasks. Further research will be needed to investigate the underlying cognitive mechanisms contributing to effective counterintuitive reasoning in science and math. Our results suggest that future research may benefit from considering visuospatial attentional skills rather than focusing on inhibitory control.

Acknowledgments

A. B. B. received funding from the UK Economic and Social Research Council, grant number ES/J500021/1, and the Wellcome Trust / Birkbeck ISSF scheme, grant number 204770/Z/16/Z. D. M. was also supported by a Wolfson Royal Society research merit award. Birkbeck, University of London contributed funding for the scanning. We thank Jack White-Foy and Minh Nguyen for help with data collection, schools for handing out flyers, teachers for discussing stimuli with us, and the adolescents who took part.

Reprint requests should be sent to Iroise Dumontheil, Centre for Brain and Cognitive Development, Department of Psychological Sciences, Birkbeck, University of London, London, UK, or via e-mail: i.dumontheil@bbk.ac.uk.

Author Contributions

Iroise Dumontheil: Conceptualization; Formal analysis; Investigation; Supervision; Visualization; Writing—Original draft; Writing—Review & editing. Annie Brookman-Byrne: Conceptualization; Investigation; Project administration; Writing—Original draft; Writing—Review & editing. Andrew K. Tolmie: Conceptualization; Investigation; Supervision; Writing—Review & editing. Denis Mareschal: Conceptualization; Funding acquisition; Supervision; Writing—Review & editing.

Funding Information

Denis Mareschal, Royal Society (https://dx.doi.org/10.13039/501100000288), grant number: Wolfson Royal Society research merit award. Annie Brookman-Byrne, Wellcome Trust (https://dx.doi.org/10.13039/100010269), grant number: 204770/Z/16/Z. Annie Brookman-Byrne, Economic and Social Research Council (https://dx.doi.org/10.13039/501100000269), grant number: ES/J500021/1.

Diversity in Citation Practices

Retrospective analysis of the citations in every article published in this journal from 2010 to 2021 reveals a persistent pattern of gender imbalance: Although the proportions of authorship teams (categorized by estimated gender identification of first author/last author) publishing in the *Journal of Cognitive Neuroscience* (*JoCN*) during this period were M(an)/M = .407, W(oman)/M = .32, M/W = .115, and W/W = .159, the comparable proportions for the articles that these authorship teams cited were M/M = .549, W/M = .257, M/W = .109, and W/W = .085 (Postle and Fulvio, *JoCN*, 34:1, pp. 1–3). Consequently, *JoCN* encourages all authors to consider gender balance explicitly when selecting which articles to cite and gives them the opportunity to report their article's gender citation balance. The authors of this article report its proportions of citations by gender category to be as follows: M/M = .415; W/M = .215; M/W = .123; W/W = .246.

REFERENCES

Allaire-Duquette, G., Bélanger, M., Grabner, R. H., Koschutnig, K., & Masson, S. (2019). Individual differences in science competence among students are associated with ventrolateral prefrontal cortex activation. *Journal of Neuroscience Research*, *97*, 1163–1178. https://doi.org/10.1002/jnr.24435, PubMed: 31077448

Allaire-Duquette, G., Foisy, L.-M. B., Potvin, P., Riopel, M., Larose, M., & Masson, S. (2021). An fMRI study of scientists with a Ph.D. in physics confronted with naive ideas in science. *NPJ Science of Learning*, *6*, 11. https://doi.org/10.1038/s41539-021-00091-x, PubMed: 33976228

Alloway, T. P. (2007). *Automated working memory assessment manual*. Harcourt.

Babai, R., & Amsterdamer, A. (2008). The persistence of solid and liquid naive conceptions: A reaction time study. *Journal of Science Education and Technology*, *17*, 553–559. https://doi.org/10.1007/s10956-008-9122-6

Babai, R., Sekal, R., & Stavy, R. (2010). Persistence of the intuitive conception of living things in adolescence. *Journal of Science Education and Technology*, *19*, 20–26. https://doi.org/10.1007/s10956-009-9174-2

Baker, S. T., Gjersoe, N. L., Sibielska-Woch, K., Leslie, A. M., & Hood, B. M. (2011). Inhibitory control interacts with core knowledge in toddlers' manual search for an occluded object. *Developmental Science*, *14*, 270–279. https://doi.org/10.1111/j.1467-7687.2010.00972.x, PubMed: 22213900

Banich, M. T., & Depue, B. E. (2015). Recent advances in understanding neural systems that support inhibitory control. *Current Opinion in Behavioral Sciences*, *1*, 17–22. https://doi.org/10.1016/j.cobeha.2014.07.006

Berkman, N. D., Sheridan, S. L., Donahue, K. E., Halpern, D. J., & Crotty, K. (2011). Low health literacy and health outcomes: An updated systematic review. *Annals of Internal Medicine*, *155*, 97–107. https://doi.org/10.7326/0003-4819-155-2-201107190-00005, PubMed: 21768583

Borst, G., Poirel, N., Pineau, A., Cassotti, M., & Houdé, O. (2013). Inhibitory control efficiency in a Piaget-like class-inclusion task in school-age children and adults: A developmental negative priming study. *Developmental Psychology*, *49*, 1366–1374. https://doi.org/10.1037/a0029622, PubMed: 22889392

Brault Foisy, L.-M., Potvin, P., Riopel, M., & Masson, S. (2015). Is inhibition involved in overcoming a common physics misconception in mechanics? *Trends in Neuroscience and Education*, *4*, 26–36. https://doi.org/10.1016/j.tine.2015.03.001

Brett, M., Anton, J. L., Valabregue, R., & Poline, J.-B. (2002). *Region of interest analysis using an SPM toolbox*. Paper Presented at the 8th International Conference on Functional Mapping of the Human Brain. Sendai, Japan.

Brookman-Byrne, A., Mareschal, D., Tolmie, A. K., & Dumontheil, I. (2018). Inhibitory control and counterintuitive science and maths reasoning in adolescence. *PLoS One*, *13*, e0198973. https://doi.org/10.1371/journal.pone.0198973, PubMed: 29927969

Brookman-Byrne, A., Mareschal, D., Tolmie, A. K., & Dumontheil, I. (2019). The unique contributions of verbal analogical reasoning and nonverbal matrix reasoning to science and maths problem-solving in adolescence. *Mind, Brain, and Education*, *13*, 211–223. https://doi.org/10.1111/mbe.12212, PubMed: 32362934

Butterfuss, R., & Kendeou, P. (2020). Reducing interference from misconceptions: The role of inhibition in knowledge revision. *Journal of Educational Psychology*, *112*, 782–794. https://doi.org/10.1037/edu0000385

Cauley, S. F., Polimeni, J. R., Bhat, H., Wald, L. L., & Setsompop, K. (2014). Interslice leakage artefact reduction technique for simultaneous multislice acquisitions. *Magnetic Resonance in Medicine*, *72*, 93–102. https://doi.org/10.1002/mrm.24898, PubMed: 23963964

Cohen Kadosh, R., Lammertyn, J., & Izard, V. (2008). Are numbers special? An overview of chronometric, neuroimaging, developmental and comparative studies of magnitude representation. *Progress in Neurobiology*, *84*, 132–147. https://doi.org/10.1016/j.pneurobio.2007.11.001, PubMed: 18155348

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience*, *3*, 201–215. https://doi.org/10.1038/nrn755, PubMed: 11994752

Cragg, L., & Gilmore, C. (2014). Skills underlying mathematics: The role of executive function in the development of mathematics proficiency. *Trends in Neuroscience and Education*, *3*, 63–68. https://doi.org/10.1016/j.tine.2013.12.001

Cragg, L., Keeble, S., Richardson, S., Roome, H. E., & Gilmore, C. (2017). Direct and indirect influences of executive functions on mathematics achievement. *Cognition*, *162*, 12–26. https://doi.org/10.1016/J.COGNITION.2017.01.014, PubMed: 28189034

Criaud, M., & Boulinguez, P. (2013). Have we been asking the right questions when assessing response inhibition in go/no-go tasks with fMRI? A meta-analysis and critical review. *Neuroscience and Biobehavioral Reviews*, *37*, 11–23. https://doi.org/10.1016/j.neubiorev.2012.11.003, PubMed: 23164813

Crittenden, B. M., Mitchell, D. J., & Duncan, J. (2016). Task encoding across the multiple demand cortex is consistent with a frontoparietal and cingulo-opercular dual networks distinction. *Journal of Neuroscience*, *36*, 6147–6155. https://doi.org/10.1523/JNEUROSCI.4590-15.2016, PubMed: 27277793

Cutler, D. M., & Lleras-Muney, A. (2010). Understanding differences in health behaviors by education. *Journal of Health Economics*, *29*, 1–28. https://doi.org/10.1016/j.jhealeco.2009.10.003, PubMed: 19963292

Crone, E. A., & Dahl, R. E. (2012). Understanding adolescence as a period of social-affective engagement and goal flexibility. *Nature Reviews Neuroscience*, *13*, 636–650. https://doi.org/10.1038/nrn3313, PubMed: 22903221

Daniel, T. A., Katz, J. S., & Robinson, J. L. (2016). Delayed match-to-sample in working memory: A BrainMap meta-analysis. *Biological Psychology*, *120*, 10–20. https://doi.org/10.1016/j.biopsycho.2016.07.015, PubMed: 27481545

Dehaene, S., Molko, N., Cohen, L., & Wilson, A. J. (2004). Arithmetic and the brain. *Current Opinion in Neurobiology*, *14*, 218–224. https://doi.org/10.1016/j.conb.2004.03.008, PubMed: 15082328

Department for Education. (2013a). *Mathematics programmes of study: Key stage 3*.

Department for Education. (2013b). *Science programmes of study: Key stage 3*.

Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, *64*, 135–168. https://doi.org/10.1146/annurev-psych-113011-143750, PubMed: 23020641

Donati, G., Meaburn, E. L., & Dumontheil, I. (2019). The specificity of associations between cognition and attainment in English, maths and science during adolescence. *Learning and Individual Differences*, *69*, 84–93. https://doi.org/10.1016/j.lindif.2018.11.012

Driver, R., Squires, A., Rushworth, P., & Wood-Robinson, V. (2015). *Making sense of secondary science: Research into children's ideas*. New York: Routledge.

Dumontheil, I., & Klingberg, T. (2012). Brain activity during a visuospatial working memory task predicts arithmetical performance 2 years later. *Cerebral Cortex*, *22*, 1078–1085. https://doi.org/10.1093/cercor/bhr175, PubMed: 21768226

Dunbar, K. N., Fugelsang, J. A., & Stein, C. (2007). Do naïve theories ever go away? Using brain and behavior to understand changes in concepts. In M. Lovett & P. Shah (Eds.), *Thinking with data* (pp. 193–205). Mahwah, NJ: Lawrence Erlbaum Associates.

Duncan, J. (2010). The multiple-demand (MD) system of the primate brain: mental programs for intelligent behaviour. *Trends in Cognitive Sciences*, *14*, 172–179. https://doi.org/10.1016/j.tics.2010.01.004, PubMed: 20171926

Egner, T. (2008). Multiple conflict-driven control mechanisms in the human brain. *Trends in Cognitive Sciences*, *12*, 374–380. https://doi.org/10.1016/J.TICS.2008.07.001, PubMed: 18760657

Friedman, N. P., & Miyake, A. (2004). The relations among inhibition and interference control functions: A latent-variable analysis. *Journal of Experimental Psychology: General*, *133*, 101–135. https://doi.org/10.1037/0096-3445.133.1.101, PubMed: 14979754

Friedman, N. P., & Miyake, A. (2017). Unity and diversity of executive functions: Individual differences as a window on cognitive structure. *Cortex*, *86*, 186–204. https://doi.org/10.1016/j.cortex.2016.04.023, PubMed: 27251123

Friedman, N. P., Miyake, A., Young, S. E., Defries, J. C., Corley, R. P., & Hewitt, J. K. (2008). Individual differences in executive functions are almost entirely genetic in origin. *Journal of Experimental Psychology: General*, *137*, 201–225. https://doi.org/10.1037/0096-3445.137.2.201, PubMed: 18473654

Furnham, A. (1992). Lay understanding of science: Young people and adults' ideas of scientific concepts. *Studies in Science Education*, *20*, 29–64. https://doi.org/10.1080/03057269208560003

Gilmore, C., Keeble, S., Richardson, S., & Cragg, L. (2015). The role of cognitive inhibition in different components of arithmetic. *ZDM Mathematics Education*, *47*, 771–782. https://doi.org/10.1007/s11858-014-0659-y

Guzzetti, B. J. (2000). Learning counter-intuitive science concepts: What have we learned from over a decade of research? *Reading and Writing Quarterly*, *16*, 89–98. https://doi.org/10.1080/105735600277971

Hartung, J., Engelhardt, L. E., Thibodeaux, M. L., Harden, K. P., & Tucker-Drob, E. M. (2020). Developmental transformations in the structure of executive functions. *Journal of Experimental Child Psychology*, *189*, 104681. https://doi.org/10.1016/j.jecp.2019.104681, PubMed: 31648081

Hotez, P. J. (2021). Anti-science kills: From Soviet embrace of pseudoscience to accelerated attacks on US biomedicine. *PLoS Biology*, *19*, e3001068. https://doi.org/10.1371/journal.pbio.3001068, PubMed: 33507935

Houdé, O. (2000). Inhibition and cognitive development: Object, number, categorization, and reasoning. *Cognitive Development*, *15*, 63–73. https://doi.org/10.1016/S0885-2014(00)00015-0

Huizinga, M., Dolan, C. V., & van der Molen, M. W. (2006). Age-related change in executive function: Developmental trends and a latent variable analysis. *Neuropsychologia*, *44*, 2017–2036. https://doi.org/10.1016/j.neuropsychologia.2006.01.010, PubMed: 16527316

Humphrey, G., & Dumontheil, I. (2016). Development of risk-taking, perspective-taking, and inhibitory control during adolescence. *Developmental Neuropsychology*, *41*, 59–76. https://doi.org/10.1080/87565641.2016.1161764, PubMed: 27070826

Jaeger, A. (2013). Inhibitory control and the adolescent brain: A review of fMRI research. *Psychology & Neuroscience*, *6*, 23–30. https://doi.org/10.3922/j.psns.2013.1.05

Johnstone, A. H. (1991). Why is science difficult to learn? Things are seldom what they seem. *Journal of Computer Assisted Learning*, *7*, 75–83. https://doi.org/10.1111/j.1365-2729.1991.tb00230.x

Karr, J. E., Areshenkoff, C. N., Rast, P., Hofer, S. M., Iverson, G. L., & Garcia-Barrera, M. A. (2018). The unity and diversity of executive functions: A systematic review and re-analysis of latent variable studies. *Psychological Bulletin*, *144*, 1147–1185. https://doi.org/10.1037/bul0000160, PubMed: 30080055

Khng, K. H., & Lee, K. (2009). Inhibiting interference from prior knowledge: Arithmetic intrusions in algebra word problem solving. *Learning and Individual Differences*, *19*, 262–268. https://doi.org/10.1016/j.lindif.2009.01.004

Khng, K. H., & Lee, K. (2014). The relationship between Stroop and stop-signal measures of inhibition in adolescents: Influences from variations in context and measure estimation. *PLoS One*, *9*, e101356. https://doi.org/10.1371/journal.pone.0101356, PubMed: 24992683

Killebrew, K., Mruczek, R., & Berryhill, M. E. (2015). Intraparietal regions play a material general role in working memory: Evidence supporting an internal attentional role. *Neuropsychologia*, *73*, 12–24. https://doi.org/10.1016/j.neuropsychologia.2015.04.032, PubMed: 25940098

Kwon, Y.-J., & Lawson, A. E. (2000). Linking brain growth with the development of scientific reasoning ability and conceptual change during adolescence. *Journal of Research in Science Teaching*, *37*, 44–62. https://doi.org/10.1002/(SICI)1098-2736(200001)37:1<44::AID-TEA4>3.0.CO;2-J

Kwon, Y.-J., Lawson, A. E., Chung, W.-H., & Kim, Y.-S. (2000). Effect on development of proportional reasoning skill of physical experience and cognitive abilities associated with prefrontal lobe activity. *Journal of Research in Science Teaching*, *37*, 1171–1181. https://doi.org/10.1002/1098-2736(200012)37:10<1171::AID-TEA8>3.0.CO;2-A

Latzman, R. D., Elkovitch, N., Young, J., & Clark, L. A. (2010). The contribution of executive functioning to academic achievement among male adolescents. *Journal of Clinical and Experimental Neuropsychology*, *32*, 455–462. https://doi.org/10.1080/13803390903164363, PubMed: 19813129

Leech, R., Mareschal, D., & Cooper, R. P. (2007). Relations as transformations: Implications for analogical reasoning. *Quarterly Journal of Experimental Psychology*, *60*, 897–908. https://doi.org/10.1080/17470210701288599, PubMed: 17616908

Leon-Carrion, J., García-Orza, J., & Pérez-Santamaría, F. J. (2004). Development of the inhibitory component of the executive functions in children and adolescents. *International Journal of Neuroscience*, *114*, 1291–1311. https://doi.org/10.1080/00207450490476066, PubMed: 15370187

Lewis, E. L., & Linn, M. C. (1994). Heat energy and temperature concepts of adolescents, adults, and experts: Implications for curricular improvements. *Journal of Research in Science Teaching*, *31*, 657–677. https://doi.org/10.1002/TEA.3660310607

Lubin, A., Vidal, J., Lanoë, C., Houdé, O., & Borst, G. (2013). Inhibitory control is needed for the resolution of arithmetic word problems: A developmental negative priming study. *Journal of Educational Psychology*, *105*, 701–708. https://doi.org/10.1037/a0032625

Luna, B., Marek, S., Larsen, B., Tervo-Clemmens, B., & Rajpreet, C. (2015). An integrative model of the maturation of inhibitory control. *Annual Review of Neuroscience*, *38*, 151–170. https://doi.org/10.1146/annurev-neuro-071714-034054, PubMed: 26154978

Luna, B., Padmanabhan, A., & O'Hearn, K. (2010). What has fMRI told us about the development of cognitive control through adolescence? *Brain and Cognition*, *72*, 101–113. https://doi.org/10.1016/j.bandc.2009.08.005, PubMed: 19765880

Mareschal, D. (2016). The neuroscience of conceptual learning in science and mathematics. *Current Opinion in Behavioral Sciences*, *10*, 114–118. https://doi.org/10.1016/j.cobeha.2016.06.001

Mason, L., & Zaccoletti, S. (2021). Inhibition and conceptual learning in science: A review of studies. *Educational Psychology Review*, *33*, 181–212. https://doi.org/0.1007/s10648-020-09529-x

Masson, S., Potvin, P., Riopel, M., & Brault Foisy, L.-M. (2014). Differences in brain activation between novices and experts in science during a task involving a common misconception in electricity. *Mind, Brain, and Education*, *8*, 44–55. https://doi.org/10.1111/mbe.12043

Menge, D. N. L., MacPherson, A. C., Bytnerowicz, T. A., Quebbeman, A. W., Schwartz, N. B., Taylor, B. N., et al. (2018). Logarithmic scales in ecological data presentation may cause misinterpretation. *Nature Ecology and Evolution*, *2*, 1393–1402. https://doi.org/10.1038/s41559-018-0610-7, PubMed: 30013132

Miller, J. D. (1998). The measurement of civic scientific literacy. *Public Understanding of Science*, *7*, 203–223. https://doi.org/10.1088/0963-6625/7/3/001

Munakata, Y., Herd, S. A., Chatham, C. H., Depue, B. E., Banich, M. T., & O'Reilly, R. C. (2011). A unified framework for inhibitory control. *Trends in Cognitive Sciences*, *15*, 453–459. https://doi.org/10.1016/j.tics.2011.07.011, PubMed: 21889391

National Numeracy. (2017). *The essentials of numeracy: A new approach to making the UK numerate*. https://www.nationalnumeracy.org.uk/essentials-numeracy

Nenciovici, L., Brault Foisy, L.-M., Allaire-Duquette, G., Potvin, P., Riopel, M., & Masson, S. (2018). Neural correlates associated with novices correcting errors in electricity and mechanics. *Mind, Brain, and Education*, *12*, 120–139. https://doi.org/10.1111/MBE.12183

Nigg, J. T. (2000). On inhibition/disinhibition in developmental psychopathology: Views from cognitive and personality psychology and a working inhibition taxonomy. *Psychological Bulletin*, *126*, 220–246. https://doi.org/10.1037/0033-2909.126.2.220, PubMed: 10748641

Ordaz, S. J., Foran, W., Velanova, K., & Luna, B. (2013). Longitudinal growth curves of brain function underlying inhibitory control through adolescence. *Journal of Neuroscience*, *33*, 18109–18124. https://doi.org/10.1523/JNEUROSCI.1741-13.2013, PubMed: 24227721

Osman, M., & Stavy, R. (2006). Development of intuitive rules: Evaluating the application of the dual-system framework to understanding children's intuitive reasoning. *Psychonomic*

Bulletin & Review, *13*, 935–953. https://doi.org/10.3758/BF03213907, PubMed: 17484417
Palak, R., Rutt, S., Easton, C., Sims, D., Bradshaw, S. & McNamara, S. (2019). Stop and think: Learning counterintuitive concepts evaluation report. Retrieved from https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/learning-counterintuitive-concepts/
Parsons, R. (2014). *KS3 math: Complete study and practice*. Newcastle upon Tyne, UK: Coordination Group Publications Ltd.
Parsons, R., & Gannon, P. (2014). *Key stage three science: The study guide*. Newcastle upon Tyne, UK: Coordination Group Publications Ltd.
Peters, L., & De Smedt, B. (2018). Arithmetic in the developing brain: A review of brain imaging studies. *Developmental Cognitive Neuroscience*, *30*, 265–279. https://doi.org/10.1016/j.dcn.2017.05.002, PubMed: 28566139
Potvin, P., & Cyr, G. (2017). Toward a durable prevalence of scientific conceptions: Tracking the effects of two interfering misconceptions about buoyancy from preschoolers to science teachers. *Journal of Research in Science Teaching*, *54*, 1121–1142. https://doi.org/10.1002/TEA.21396
Potvin, P., Malenfant-Robichaud, G., Cormier, C., & Masson, S. (2020). Coexistence of misconceptions and scientific conceptions in chemistry professors: A mental chronometry and fMRI study. *Frontiers in Education*, *5*, 180. https://doi.org/10.3389/feduc.2020.542458
Potvin, P., Turmel, É., & Masson, S. (2014). Linking neuroscientific research on decision making to the educational context of novice students assigned to a multiple-choice scientific task involving common misconceptions about electrical circuits. *Frontiers in Human Neuroscience*, *8*, 1–13. https://doi.org/10.3389/fnhum.2014.00014, PubMed: 24478680
R Core Team. (2021). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/
Rhodes, S. M., Booth, J. N., Campbell, L. E., Blythe, R. A., Wheate, N. J., & Delibegovic, M. (2014). Evidence for a role of executive functions in learning biology. *Infant and Child Development*, *23*, 67–83. https://doi.org/10.1002/icd.1823
Rhodes, S. M., Booth, J. N., Palmer, L. E., Blythe, R. A., Delibegovic, M., & Wheate, N. J. (2016). Executive functions predict conceptual learning of science. *British Journal of Developmental Psychology*, *34*, 261–275. https://doi.org/10.1111/bjdp.12129, PubMed: 26751597
Rorden, C., & Brett, M. (2000). Stereotaxic display of brain lesions. *Behavioural Neurology*, *12*, 191–200. https://doi.org/10.1155/2000/421719, PubMed: 11568431
Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*, 1–36. https://doi.org/10.18637/jss.v048.i02
Ryan, J., & Williams, J. (2007). *Children's mathematics 4–15: Learning from errors and misconceptions*. Maidenhead, UK: Open University Press.
Shtulman, A., & Valcarcel, J. (2012). Scientific knowledge suppresses but does not supplant earlier intuitions. *Cognition*, *124*, 209–215. https://doi.org/10.1016/j.cognition.2012.04.005, PubMed: 22595144
Siegel, J. S., Power, J. D., Dubis, J. W., Vogel, A. C., Church, J. A., Schlaggar, B. L., et al. (2014). Statistical improvements in functional magnetic resonance imaging analyses produced by censoring high-motion data points. *Human Brain Mapping*, *35*, 1981–1996. https://doi.org/10.1002/hbm.22307, PubMed: 23861343
Spiegelhalter, D. (2019). *The art of statistics: Learning from data*. Penguin Random House UK.
St Clair-Thompson, H. L., & Gathercole, S. E. (2006). Executive functions and achievements in school: Shifting, updating, inhibition, and working memory. *Quarterly Journal of Experimental Psychology*, *59*, 745–759. https://doi.org/10.1080/17470210500162854, PubMed: 16707360
Stavy, R., & Babai, R. (2010). Overcoming intuitive interference in mathematics: Insights from behavioural, brain imaging and intervention studies. *ZDM Mathematics Education*, *42*, 621–633. https://doi.org/10.1007/s11858-010-0251-z
Stavy, R., Goel, V., Critchley, H., & Dolan, R. (2006). Intuitive interference in quantitative reasoning. *Brain Research*, *1073–1074*, 383–388. https://doi.org/10.1016/j.brainres.2005.12.011, PubMed: 16443204
Stavy, R., & Tirosh, D. (2000). *How students (mis-)understand science and mathematics: Intuitive Rules*. New York: Teachers College Press.
Strohl, A. E., Mendoza, G., Ghant, M. S., Cameron, K. A., Simon, M. A., Schink, J. C., et al. (2015). Barriers to prevention: Knowledge of HPV, cervical cancer, and HPV vaccinations among African American women. *American Journal of Obstetrics and Gynecology*, *212*, 65.e1–65.e5. https://doi.org/10.1016/j.ajog.2014.06.059, PubMed: 24983684
Swick, D., & Chatham, C. H. (2014). Ten years of inhibition revisited. *Frontiers in Human Neuroscience*, *8*, 329. https://doi.org/10.3389/fnhum.2014.00329, PubMed: 24904369
Tamm, L., Menon, V., & Reiss, A. L. (2002). Maturation of brain function associated with response inhibition. *Journal of the American Academy of Child & Adolescent Psychiatry*, *41*, 1231–1238. https://doi.org/10.1097/00004583-200210000-00013
Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., et al. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage*, *15*, 273–289. https://doi.org/10.1006/nimg.2001.0978, PubMed: 11771995
Walker, D. A. (2003). JMASM9: Converting Kendall's tau for correlational or meta-analytic analyses. *Journal of Modern Applied Statistical Methods*, *2*, 525–530. https://doi.org/10.22237/jmasm/1067646360
Wang, H., He, W., Wu, J., Zhang, J., Jin, Z., & Li, L. (2019). A coordinate-based meta-analysis of the n-back working memory paradigm using activation likelihood estimation. *Brain and Cognition*, *132*, 1–12. https://doi.org/10.1016/J.BANDC.2019.01.002, PubMed: 30708115
Watanabe, J., Sugiura, M., Sato, K., Sato, Y., Maeda, Y., Matsue, Y., et al. (2002). The human prefrontal and parietal association cortices are involved in no-go performances: An event-related fMRI study. *Neuroimage*, *17*, 1207–1216. https://doi.org/10.1006/nimg.2002.1198, PubMed: 12414261
Wechsler, D. (2011). *Wechsler Abbreviated Scale of Intelligence (WASI-II)* (2nd ed.). San Antonio, TX: Pearson.
Wilkinson, H. R., Smid, C., Morris, S., Farran, E. K., Dumontheil, I., Mayer, S., et al. (2020). Domain-specific inhibitory control training to improve children's learning of counterintuitive concepts in mathematics and science. *Journal of Cognitive Enhancement*, *4*, 296–314. https://doi.org/10.1007/s41465-019-00161-4, PubMed: 32832846
Willingham, D. T. (2010). Is it true that some people just can't do math? *American Educator, Winter*, *14–19*, 39.
Xu, J., Moeller, S., Auerbach, E. J., Strupp, J., Smith, S. M., Feinberg, D. A., et al. (2013). Evaluation of slice accelerations using multiband echo planar imaging at 3 T. *Neuroimage*, *83*, 991–1001. https://doi.org/0.1016/j.neuroimage.2013.07.055, PubMed: 23899722
Zaitchik, D., Iqbal, Y., & Carey, S. (2014). The effect of executive function on biological reasoning in young children: An individual differences study. *Child Development*, *85*, 160–175. https://doi.org/10.1111/cdev.12145, PubMed: 23889035