Cognitive Processing (2019) 20:139–161
https://doi.org/10.1007/s10339-018-0896-z

REVIEW

Brain activations associated with scientific reasoning: a literature review

Lucian Nenciovici[1,2] · Geneviève Allaire-Duquette[1,2] · Steve Masson[1,2]

Received: 30 October 2017 / Accepted: 4 December 2018 / Published online: 8 December 2018
© Marta Olivetti Belardinelli and Springer-Verlag GmbH Germany, part of Springer Nature 2018

Abstract
Scientifically literate individuals are defined as individuals who are able to apply scientific knowledge and use scientific reasoning skills to solve problems. In recent years, cognitive neuroscience has turned its attention to understanding the brain activation patterns associated with scientific reasoning skills, but this work has not been systematically reviewed for more than a decade. The present study reviews neuroimaging studies related to three types of scientific reasoning tasks: overcoming misconceptions, causal reasoning, and hypothesis generation. These studies indicate converging evidence for the involvement of (1) lateral prefrontal areas, reinforcing the idea of an association between scientific reasoning and executive functions, and (2) middle temporal areas, suggesting an association between scientific reasoning and declarative memory. Potential educational implications and leads for future research are discussed.

Keywords Scientific reasoning · Neuroimaging · Overcoming misconceptions · Causal reasoning · Hypothesis generation

Introduction

Science plays a crucial role in most industrialized societies, meaning that being an informed citizen in today's world requires the ability to engage with science-related issues such as climate change, epidemics, genetically modified foods, and nuclear energy (Dragos and Mih 2015; Organisation for Economic Co-operation and Development [OECD] 2017; United Nations Educational Scientific and Cultural Organization [UNESCO] 2010). It is therefore no surprise that leaders in both the European Union and the USA agree that scientific literacy is important, and that improving rates of scientific literacy could be beneficial to society (Ogunkola 2013). Scientifically literate individuals are defined as individuals who are able to correctly apply scientific knowledge and scientific reasoning skills to solve problems and make practical and informed decisions in their personal, civic, and professional lives (Brickhouse et al. 1989; Dragos and Mih 2015; Holbrook and Rannikmae 2009; Laetsch 1987; Laugksch 1998). Dismal scores on international science examinations have led many countries to deplore low levels of scientific literacy and to assert that scientific reasoning skills "are often missing from the scientific learning effort" (Dragos and Mih 2015, p. 170). Better understanding the nature of scientific reasoning could bring an interesting perspective to this global educational concern and eventually help political leaders and educators make more informed decisions on how education can contribute to developing scientific reasoning.

There is a growing corpus of studies that focus on exploring the brain activation patterns associated with scientific reasoning tasks. Pettito and Dunbar conducted a review in 2004 that synthesized the findings of this research. Considering that new studies have since been published, but no other reviews have been conducted, an updated review would appear to be in order. Thus, the aim of the present review is to synthesize findings from studies published both before and after 2004 that examine brain activation patterns associated with scientific reasoning tasks. The potential for a neuroscientific approach to provide insight into learning and reasoning has been acknowledged by several major organizations over the last decade (e.g., OECD 2007; The Royal Society 2011; UNESCO 2013). It is generally accepted that

---

✉ Steve Masson
masson.steve@uqam.ca

[1] Laboratory for Research in Neuroeducation, Département de didactique, Université du Québec à Montréal, P.O. Box 8888, succursale Centre-Ville, Montreal, QC H3C 3P8, Canada

[2] Team for Research in Science and Technology Education, Département de didactique, Université du Québec à Montréal, Montreal, QC, Canada

the neuroscientific approach complements other existing educational approaches (e.g., cognitivism, constructivism, socio-constructivism) and that brain-informed theories of learning could be useful for developing new teaching and educational practices. Thus, the present review has the potential to shed new light on the nature of scientific reasoning and to help educational leaders make better-informed decisions regarding science curricula and pedagogical practices in the science classroom.

Methods

Neuroimaging articles focusing on scientific reasoning were identified from three databases (i.e., ERIC, Google Scholar, and PsycINFO) using an algorithm combining sets of keywords related to neuroimaging techniques and scientific disciplines. The complete algorithm used to search databases was as follows:

> (“computerized tomography” OR tomography OR “magnetic resonance imaging” OR MRI OR fMRI OR “electrophysiological imaging” OR “metabolic imaging” OR “computerized electroencephalog*” OR electroencephalog* OR EEG OR “evoked potentials” OR “event-related potentials” OR ERP OR “spectral analysis” OR spectroscopy OR “topographic brain mapping” OR “positron emission tomography” OR PET OR “single photon emission computerized tomography” OR magnetoencephalog* OR MEG)
> AND
> (science* OR physic* OR chemist* OR biolog* OR geography OR geology OR “earth science” OR astronomy)

This search led to 581 articles, which were screened for inclusion criteria by reading titles and abstracts (Mateen et al. 2013). Studies included in this review satisfied the following criteria:

1. Involve a cognitive task performed in a scientific discipline (e.g., physics, chemistry, biology).
2. Present brain activation data during task execution.
3. Comprise a comparison consisting of either (1) comparing brain activation of two groups (e.g., a two-group design in which participants in both groups answered the same types of questions), (2) comparing brain activation in two conditions (e.g., a single-group design in which participants answered at least two types of questions), or (3) comparing brain activation at two moments (e.g., a single-group design in which participants answered the same types of questions at a pretest and then at a posttest).
4. Be published in a peer-reviewed journal.

This first step led to a body of 8 apparently relevant studies. Then, each article’s reference list was screened for additional studies (Greenhalgh and Peacock 2005). For the sake of completeness, the reference list of Pettito and Dunbar’s (2004) review was also screened. This second step led to a new total of 22 apparently relevant studies. The full texts of these 22 studies were read, and inclusion criteria were applied, leading to a final count of 10 relevant studies. Although several types of classification of these 10 studies appeared possible, such as categorizing them by scientific discipline or by type of comparison involved, the most coherent type of classification in line with the purpose of this review was considered to be a categorization by type of cognitive task involved. Thus, the 10 studies were classified into three categories of cognitive tasks emerging from keywords provided by the studies’ authors to describe their research: overcoming misconceptions, causal reasoning, and hypothesis generation.

Results and discussion

Overcoming misconceptions

Misconceptions are erroneous representations regarding natural phenomena that seem intuitively true to the learner, but are discordant with scientific representations (Shtulman and Valcarcel 2012). Misconceptions are persistent (Wandersee et al. 1994), resistant to change (Treagust and Duit 2008), and need to be overcome in order for learning of several counterintuitive scientific concepts to occur (Stavy and Babai 2010). Recently, diSessa (2017, p. 5) emphasized the importance of a better understanding of the learning mechanisms involved in overcoming misconceptions and indicated that such understanding remains “a compelling challenge for the field to resolve.”

Mechanisms involved in overcoming misconceptions have been examined by several studies from the standpoint of cognitive psychology (e.g., Babai and Amsterdamer 2008; Shtulman and Harrington 2015; Kelemen and Rosset 2009). These studies all conclude that overcoming misconceptions is related to the implementation of inhibitory control. At the cognitive level, inhibitory control is a process that involves the ability to resist a habit, a spontaneous and tempting response, a strategy, or a conception that is inappropriate or erroneous in some contexts, such as misconceptions in science (Dempster 1995; Dempster and Corkill 1999; Diamond 2013; Houdé and Borst 2014). The process of inhibitory control is related to error detection and suppression. It is part of a family of top-down mental processes known as executive functions: high-level cognitive processes, often associated with the frontal lobes, which control lower-level processes in the service of goal-directed behavior (Diamond 2013; Friedman and Miyake 2017).

One of the first studies to establish a link between overcoming misconceptions and inhibition using neuropsychological tests was conducted by Kwon and Lawson (2000). This study focused on students aged 13–16 and observed a positive correlation between three variables: the students' inhibitory control ability, measured using the Wisconsin Card Sorting Test (Berg 1948); scientific reasoning ability, measured using Lawson's Classroom Test (Lawson 1978); and scientific concept acquisition, measured using a test on air pressure concepts. The results of this study showed that inhibitory control explained 29% of concept acquisition. Later studies that also used a cognitive psychology approach (Babai and Amsterdamer 2008; Babai et al. 2006, 2010; Kelemen and Rosset 2009; Kelemen et al. 2013; Shtulman and Harrington 2015; Shtulman and Valcarcel 2012) reported that participants with several years of science education took more time to answer test items containing a misconception than they did with control items that did not. According to these studies' authors, longer response times could be explained by the recruitment of inhibitory control mechanisms.

The three neuroimaging studies (Brault Foisy et al. 2015; Masson et al. 2014; Potvin et al. 2014) included in the present review under the category of overcoming misconceptions obtained results along the same lines, providing further evidence of an association between inhibitory control and scientific reasoning when students have to overcome misconceptions. At the neural level, inhibitory control refers to the capacity of a neural network to block the activation of another neural network, such as that in which a misconception is embedded, and which would otherwise lead to an inappropriate behavioral response (Hunter et al. 2011; Snyder et al. 2010).

Inspired by a preliminary study conducted by Dunbar, Fugelsang and Stein (2007), Masson et al. (2014) used fMRI, an electricity task, and a two-group design on undergraduate students to compare brain activations of experts ($n=11$; $M=22.1$ years; $\mathrm{SD}=3.5$) and novices ($n=12$; $M_{\mathrm{age}}=22.9$ years; $\mathrm{SD}=3.5$) in science, while evaluating the correctness of stimuli depicting simple electric circuits. Both groups were presented with images illustrating either *scientific circuits* (i.e., a bulb lights up when connected to a battery by two wires, thereby completing the circuit) or naïve, *misconceptual circuits* (i.e., a bulb lights up when connected to a battery by a single wire). Experts' answers were in accordance with the scientific concept of the electric circuit, whereas novices' answers were in accordance with the misconception. The authors' a priori hypothesis was that experts would show more activation than novices in brain areas involved in inhibitory control. On the basis of previous neuroscientific studies that examined tasks requiring inhibitory control, such as the Stroop (e.g., Bush et al. 1998) and Go/No-go tasks (e.g., Menon et al. 2001), they identified three such brain areas in their a priori hypothesis: the anterior cingulate cortex (ACC), the ventrolateral prefrontal cortex (VLPC), and the dorsolateral prefrontal cortex (DLPC).

Results of the contrast analysis *Experts > Novices* for misconceptual circuits conformed with the authors' a priori hypotheses by showing significantly greater relative activation of the ACC (Brodmann Area [BA] 32), left DLPC (BA 9), and left VLPC (BA 45). They reported that a likely interpretation of differential ACC activation, based on the neuroscientific literature (van Veen and Carter 2002; see also Botvinick 2007; Botvinick et al. 2001, 2004), was that experts had to monitor the conflict between two responses and detect the erroneous response. Similarly, the authors reported that the likely interpretation of differential activations of the lDLPC (Buchsbaum et al. 2005; see also Bush et al. 2006; Garavan et al. 2002; Menon et al. 2001; Monchi et al. 2001) and lVLPC (Buchsbaum et al. 2005; see also Badre and Wagner 2007; Casey et al. 1997; Gold et al. 2006; Levy and Wagner 2011) was that experts had to implement inhibitory control, a cerebral process distinct from conflict monitoring and error detection, which occurs after the brain has detected an error (Garavan et al. 2002; MacDonald et al. 2000; Menon et al. 2001). Based on these results, Masson et al. (2014) concluded that misconceptions and formal electricity knowledge likely still coexist in experts' neural networks, and that inhibitory control is necessary to avoid naïve answers.

Brault Foisy et al. (2015) used fMRI and a mechanics task in a similar two-group study design involving undergraduate students, to compare the cerebral activity of experts ($n=10$; $M_{\mathrm{age}}=22.3$ years; $\mathrm{SD}=2.4$) and novices ($n=19$; $M_{\mathrm{age}}=23.5$ years; $\mathrm{SD}=2.8$) in science, while evaluating the correctness of stimuli depicting two free-falling balls of different sizes. Both groups were presented with short films showing either a *scientific free fall* (i.e., both balls falling at the same speed) or a naïve, *misconceptual free fall* (i.e., largest ball falling faster). Experts' answers were in accordance with the scientific concept of free fall, whereas novices' answers were in accordance with the misconception. The authors' a priori hypothesis was identical to Masson et al.'s (2014), and experts were expected to activate the three same brain areas (ACC, DLPC, VLPC) more than novices.

Results of the contrast analysis *Experts > Novices* for the misconceptual free fall conformed, although not entirely, with the authors' a priori hypotheses, by showing significantly greater relative activation of the lDLPC (BA 46) and rVLPC (BA 47), but not of the ACC. They reported that the likely interpretation of differential activations of the DLPC and VLPC, based on the neuroscientific literature (Buchsbaum et al. 2005; see also Aron et al. 2003, 2004; Badre and Wagner 2007; Casey et al. 1997; Menon et al. 2001; Monchi et al. 2001), was that experts had to implement inhibitory control processes to avoid naïve answers. Moreover, Brault Foisy et al. proposed that the unexpected lack of relative activation of the ACC could be due to the fact that both experts and novices activated the ACC when evaluating misconceptual free falls. They suggested that novices could have activated the ACC because their brain possibly detected the presence of a conflict between their initial conception of free fall and the scientific conception previously learned in school. It thus appeared that the novices in this study may have been at a more advanced stage of conceptual change than expected, although they had not yet reached the stage of being able to overcome their initial conception to provide scientific answers. Similarly to Masson et al. (2014), Brault Foisy et al.'s general conclusion was that this cerebral activation pattern likely supports the hypothesis that experts' misconceptions about mechanics are not eradicated during formal learning. Instead, their misconceptions remain encoded in their neural networks where they coexist with scientific knowledge, and inhibitory control is required to provide correct answers.

Finally, in a somewhat different study by Potvin et al. (2014) using fMRI and a single group of undergraduates ($n = 22$; $M_{\text{age}} = 18.5$ years; SD = 0.7) considered to be novices in science, participants performed an electricity task in which they had to express their certainty level (i.e., sure or unsure) regarding the correctness of photographs of electric circuits. Some photographs depicted scientifically correct electric circuits, while others depicted incorrect electric circuits reflecting common misconceptions, such as "only one wire is enough to light up a bulb." The photographs for which participants claimed to be certain of their answers comprised a similar number of correct and incorrect electric circuits (45.3 vs. 54.7%, respectively). The photographs for which participants claimed to be uncertain of their answers also comprised a similar number of scientifically correct and incorrect electric circuits (50.7 vs. 49.3%, respectively). The authors' a priori hypothesis was that, due to internal lack of knowledge, novices would show more activation in the *Uncertainty* condition, compared to the *Certainty* condition, in brain areas involved in detecting uncertainty. On the basis of previous neuroscientific studies that examined this kind of uncertainty with tasks involving making predictions based on ambiguous rules (e.g., Volz et al. 2004) or judging written statements as being undecidable (e.g., Harris et al. 2008), Potvin et al. identified two such brain areas in their a priori hypothesis: the anterior cingulate cortex (ACC) and the dorsolateral prefrontal cortex (DLPC). The authors did not have any a priori hypothesis for the *Certainty* > *Uncertainty* comparison.

Results of the contrast analysis between *Uncertainty* > *Certainty* responses conformed with the authors' a priori hypothesis by yielding greater relative activation in the ACC (BA 24/32) and in the rDLPC (BA 8/9). They reported that the likely interpretation of ACC activation, based on the neuroscientific literature (van Veen and Carter 2002; see also Botvinick 2007; Botvinick et al. 2001, 2004; Bush et al. 2002; Eisenberger et al. 2003; van Duijvenvoorde et al. 2008), was the recruitment of processes related to conflict monitoring caused by competing conceptions. These competing conceptions were presumed to be scientific electricity knowledge learned in school versus intuitive misconceptions that continued to influence decision-making. Similarly, the authors reported that a likely interpretation of rDLPC activation could be the recruitment of the executive function of decision-making to resolve the conflict between two competing responses (Volz et al. 2005; see also Daniel et al. 2010; Harris et al. 2008; Hosseini et al. 2010; Stern et al. 2010). The opposite contrast analysis between *Certainty* > *Uncertainty* responses, for which they did not have any a priori hypotheses, yielded significantly stronger activations in posterior areas, such as the inferomedial occipital and inferior temporal gyri (BA 18/19/37), but not in the dorsal prefrontal areas. Based on task settings and on the neuroscientific literature (Fortin et al. 2001; Lamm et al. 2001; Waberski et al. 2008), they suggested that a possible interpretation of these activations was that scientific knowledge about electric circuits might be grounded in the visuospatial circuits of the brain. When certain about their answers, it is thus possible that participants automatically retrieve this scientific knowledge without recruitment of decision-making processes. Moreover, unlike the conclusion reached by the previously discussed Masson et al. (2014) and Brault Foisy et al. (2015) studies, Potvin et al. concluded that their results did not show inhibitory control, and suggested the possibility that the development of a certain level of expertise is required to record activations of inhibitory control areas when overcoming misconceptions.

Some common results and interpretations emerged from these three studies with respect to the neural correlates and processes involved in overcoming misconceptions. In two of the studies (Masson et al. 2014; Potvin et al. 2014), researchers observed ACC (BA 24/32) activation and, based on the task and on neuroscientific literature, suggested that this activation may have been associated with monitoring the conflict between competing conceptions and detecting the erroneous conception or answer. Two of the three studies (Brault Foisy et al. 2015; Masson et al. 2014) also interpreted VLPC and DLPC activations observed in experts' brains as a possible sign that inhibitory control was implemented to inhibit a task-relevant misconception. While Potvin et al.'s conclusions (2014) do not suggest the presence of inhibitory control, they explicitly hypothesized that participants likely did not possess enough expertise to implement inhibitory control mechanisms when overcoming misconceptions. Thus, the findings of these neuroimaging studies largely confirm the results of the cognitive psychology studies. They suggest that error detection, conflict monitoring, and particularly inhibitory control are important neural processes involved in overcoming misconceptions. Despite these common findings and similar conclusions, there were notable differences across the neural findings of these three studies, thus precluding a precise mapping of brain areas responsible for overcoming misconceptions. Most importantly, whereas Masson et al. (2014) and Brault Foisy et al. (2015) both found activations in the DLPC and VLPC, the respective activations occurred in different Brodmann Areas: left BA 9 and left BA 45 in Masson et al.'s study versus left BA 46 and right BA 47 in Brault Foisy et al.'s study.

These three studies also share certain limitations. Two of these limitations appear to be common among studies included in the three categories and will be examined in further detail in the general discussion of this review. These two limitations are (1) a problematic use of reverse inference to interpret neuroimaging findings and (2) small sample sizes ($n < 20$). However, the following discussion will concentrate on two limitations that appear to be specific to the three studies described above.

First, the three studies focused exclusively on undergraduates. Moreover, two of the studies focused on undergraduates who were considered to be science experts. Thus, it remains unclear whether error detection, conflict monitoring, and inhibitory control processes could be implemented in overcoming misconceptions among younger learners, such as elementary and secondary school students. Given the importance of brain development prior to adulthood (Kappel et al. 2015), adult neuroimaging findings cannot be readily generalized to children and adolescents. Indeed, because of developmental differences in both brain structure and function (e.g., Fair et al. 2009; Falk et al. 2013), the adult brain is generally considered non-representative of the younger brain (Henrich et al. 2010). In particular, neural structures and processes supporting executive functions, such as inhibitory control, are not fully developed in younger learners and continue to significantly strengthen throughout childhood and adolescence (Best and Miller 2010). Thus, it cannot be excluded that different mechanisms are recruited by younger learners to overcome misconceptions. Future neuroimaging studies could hence focus on younger learners to verify whether they recruit the same neural correlates and processes as adults when overcoming misconceptions. Moreover, focusing on younger learners appears necessary in future research, given the importance of gaining a better understanding of the learning mechanisms involved in overcoming misconceptions, as highlighted at the beginning of this section (diSessa 2017). Indeed, in comparison with adults, a greater proportion of younger learners possess common misconceptions (e.g., Cepni and Keles 2006; Potvin and Cyr 2017; Stavy et al. 2006), and formal science teaching in most Western countries starts in early elementary education.

Secondly, these three studies used an observational design and did not address the effect of instruction on brain activation patterns. An interesting question for future research would thus be whether different types of science instruction yield different patterns of change at the neural level. Previous research in non-scientific fields suggests that this might be the case; it also indicates that certain types of instruction could be more beneficial than others at the neural level, because they generate a brain activation pattern that the literature has previously associated with expertise in that particular domain. For example, in the domain of reading, it was observed by Yoncheva et al. (2010), using EEG, that two different types of artificial orthography training in young adults yielded different brain activation patterns. Training that consisted in directing participants' attention to grapheme–phoneme associations yielded more left-lateralized activation, which is associated in the neuroscientific literature with reading expertise (e.g., Shaywitz et al. 2002, 2007). In contrast, training that consisted in directing participants' attention to whole-word associations yielded more right-lateralized activation. In the area of arithmetics, it was observed by Delazer et al. (2005), using fMRI, that two different types of complex operations training in young adults yielded different brain activation patterns. Training that consisted of drills (i.e., learning the association between the operands and the result) yielded more activation in the left angular gyrus, an area associated in scientific literature with expertise in arithmetic fact retrieval (e.g., Grabner et al. 2009; Seghier 2013). In contrast, training that consisted in applying strategies yielded more activation in the precuneus. Thus, in the area of conceptual learning in science, future research could compare two different types of instruction to see if they result in different patterns of change at the neural level, and whether the brain activation pattern in novices nears the pattern observed in experts.

Causal reasoning

Another fundamental aspect of scientific literacy is *causal reasoning*, the mental process that consists in identifying causality—in other words, deciding whether and how one event (cause) brings about another event (effect) (Drouet 2012; McKay Illari et al. 2011). Causal reasoning can occur by *perceptual causality*, the direct perception of causality from observing two events, or *inferential causality*, the inference of causality from the observation of events combined with real-world knowledge (Roser et al. 2005). Causal perception is thus a more automatic process; the type of causal relationship is generally mechanical, meaning that it is based on the spatial or temporal contiguity of two events, such as two balls colliding (Ray and Schlottmann 2007; Scholl and

Nakayama 2002; Roser et al. 2005). Causal inference is a more complex process in which the reasoner must evaluate the extent to which two events have a cause-and-effect relationship. The type of causal relationship involved in causal inference is frequently covariation, that is, based on the observation of concomitant variations in two events, phenomena, or variables (Bullock et al. 1982; Cheng 1997; Novick and Cheng 2004; Roser et al. 2005). The recent literature (Lappi and Rusanen 2011; McKay Illari et al. 2011; Rusanen 2014) has pointed out that natural science instruction and learning routinely make use of causality, a long-established and central concept in scientific disciplines, but often lead to a range of interpretations. This tendency suggests the need for a better understanding of the tools underlying causality, including the cognitive mechanisms involved in perceiving and inferring causality.

As with overcoming misconceptions, mechanisms involved in causal reasoning have been extensively studied in the field of cognitive development. Causal reasoning has been linked to cognitive processes related to attention, such as attentional shifting, and to memory, such as retrieval of prior knowledge (Bes et al. 2012; Bullock et al. 1982; Cheng 1997; Mendelson and Shultz 1976; Novick and Cheng 2004; Ray and Schlottmann 2007; Shultz et al. 1986). The four neuroimaging studies included in the present review under the category of causal reasoning (Blakemore et al. 2001; Fonlupt 2003; Fugelsang and Dunbar 2005; Fugelsang et al. 2005) obtained similar results, but they also suggest that different neural correlates may be recruited when causality is perceived versus when it is inferred.

First, Blakemore et al. (2001) used fMRI and a physics task to compare the brain activity of a single group of 8 college students (age range 20–25 years) who were presented with short movies depicting different types of visual events, including causal and non-causal events. The causal events comprised an elementary, billiard ball type of mechanical *causality* in which a ball rolled and collided with a second ball, with the second ball immediately moving upon impact. In the *non-causality* events, the first ball rolled beneath the second ball without colliding with it and without any movement of the second ball occurring. Thus, this task involved perceptual causality and a mechanical type of causal relationship. The contrast analysis *Causality* > *Non-causality* yielded significantly greater relative activation of the bilateral medial occipito-temporal area V5/MT, the bilateral superior temporal sulci (STS), and the left intraparietal sulcus (lIPS). These neural findings were entirely consistent with the authors' a priori hypotheses. This pattern of activation was thus interpreted as supporting the hypothesis that the elementary, billiard ball type of mechanical causality seems to be automatically processed by the visual system. More precisely, they interpreted (1) the V5/MT activation as being likely explained by engagement of motion information processing (Antal et al. 2004; Ueno et al. 2009; Zeki 2015), (2) the STS activation as being likely explained by engagement of processes related to the perception of complex non-biological motion that conveys significance (i.e., movement of inanimate objects, such as balls rolling) (Beauchamp et al. 2002, 2004; Redcay 2008), and (3) the lIPS activation as being likely due to processing of visuospatial relationships involving causal contingency in causal events, but not in non-causal events (Becker et al. 1999; Majerus et al. 2006; Ravizza et al. 2004). Blakemore et al. did not address the opposite contrast (i.e., *Non-causality* > *Causality*). Thus, they concluded that the visual system seems to work toward recovering the causal structure of the world, and may play a particularly important role in detecting causality in visual events.

Fonlupt (2003) used fMRI and the same billiard ball collision movies as in the previously described study by Blakemore et al. (2001) to compare brain activation in a single group of 10 college students (age range 20–25 years) in causal versus non-causal events. For each type of event, participants were instructed to make both a causal judgement (i.e., detect the presence or absence of causality) and a movement direction judgement (i.e., detect if the ball moved toward the right or left side of the screen). Despite implying a mechanical type of causal relationship, this task involved a conscious level of inference about the presence of causality. The contrast analysis *Causality detection* > *Direction detection* showed significantly greater relative activity of several foci in the medial dorsal part of the superior frontal cortex (dMPFC, BA 8/9), which was consistent with the authors' a priori hypothesis. Considering this finding, neuroscientific literature, and results obtained by Blakemore et al. (2001), Fonlupt interpreted this pattern of activation as likely suggesting that (1) the perception of causality is processed primarily by a visual module independently from higher-level processes, and that (2) output from this module is available to be read and interpreted by higher-order processes occurring in the dMPFC during the explicit search for causality and the inference of its presence or absence. More precisely, by comparing the foci of activation observed in the dMPFC with activation locations derived from the literature, the authors suggested that the pattern of activation observed in their study could be explained by the recruitment of processes such as externally focused attention (Banich et al. 2000; Dove et al. 2000), memory, e.g., working memory, recognition memory (Donaldson et al. 2001; Prabhakaran et al. 2000), general reasoning, e.g., inductive reasoning, relational integration (Christoff et al. 2001; Goel et al. 1997), and self-referential processes (Gusnard et al. 2001). In brief, they concluded that explicit determination of physical causality appears to involve comparison with a reference that could consist of participants' knowledge of physical laws governing collisions, memories of past events involving these laws, and the representation of complex interrelation between objects.

Fugelsang et al. (2005) also used fMRI and a billiard ball collision task to compare cerebral activity in a single group of 16 college students ($M_{\text{age}} = 26.8$ years) who were presented with short movies depicting three types of visual events. The first type of event was a *causal* event in which a ball rolled and collided with a second ball, with the second ball immediately moving upon impact. The two other types of events were non-causal and consisted of a *temporal gap* event in which the second ball moved with a short delay after impact, and of a *spatial gap* event in which the first ball stopped short of touching the second ball, but the second ball moved nevertheless. Thus, this task involved perceptual causality and a mechanical type of causal relationship. The contrast analysis comparing cerebral activity for the *causal* event with the conjunction of the two non-causal events (i.e., *Causal* > conjunction of *Temporal* and *Spatial gaps*) yielded significantly greater relative activity in the right inferior parietal lobule (rIPL, BA 39), the right inferior frontal gyrus (rIFG, BA 45), the middle and superior frontal gyri (MFG and SFG, BA 6/8), and the right superior parietal lobule (rSPL, BA 7).

These neural findings were entirely consistent with the authors' a priori hypotheses. Thus, this pattern of activation was interpreted as likely providing support for the hypothesis that perceiving causality from dynamic visual events recruits visual attention and executive processes. More precisely, they interpreted (1) the rIPL (BA 39) and rIFG (BA 45) activations as being likely explained by the recruitment of processes sustaining visual attention and vigilance (Buchsbaum et al. 2006; Singh-Curry and Husain 2009), and (2) MFG/SFG (BA 6/8) and rSPL (BA 7) activations as likely being due to recruitment of processes subserving visuospatial working memory (Babiloni et al. 2005; Cheng et al. 1995; Courtney et al. 1996). Also consistent with the authors' a priori hypotheses, the opposite contrast analysis (i.e., conjunction of *Temporal* and *Spatial gaps* > *Causal*) yielded significantly greater relative activity in parieto-occipital areas, such as the cuneus (BA 18) and the lingual gyrus (BA 19). These activations were thus interpreted as likely suggesting engagement of higher-order visuospatial information processing, but not engagement of attentional/executive processes (Fortin et al. 2001; Lamm et al. 2001; Waberski et al. 2008). In brief, they concluded that greater activation of regions in the prefrontal cortices when participants viewed causal events, compared with non-causal events, suggests that causal events may recruit more attentional/executive processes, thus resulting in more attentional resources devoted to such events.

Lastly, using fMRI and a task about pharmacology, Fugelsang and Dunbar (2005) compared cerebral activity in a single group of 14 college students (age range 18–31 years) who were presented with four visual event types: data either *consistent* or *inconsistent* with a *plausible* causal theory (e.g., a serotonin reuptake inhibitor [a type of antidepressant medication] improves mood) and data either *consistent* or *inconsistent* with an *implausible* causal theory (e.g., a topoisomerase inhibitor [a type of antibiotic medication] improves mood). Data were presented in combinations of cause (i.e., a colored pill representing the medication) and effect (i.e., mood variation). Thus, this task involved inferential causality and a covariation type of causal relationship. The contrast analysis *Consistent* > *Inconsistent* showed significantly greater relative activation of the left parahippocampal gyrus (lPHG) and the right precentral gyrus (BA 6) for both plausible and implausible theories, and significantly greater relative activation of the left caudate nucleus for implausible theories only. The contrast analysis *Inconsistent* > *Consistent* showed significantly greater relative activation of the ACC (BA 24/32), the precuneus (BA 7), and the lDLPC (BA 9) for plausible theories and no significantly greater relative activation for implausible theories.

These neural findings were consistent with the authors' a priori hypotheses only regarding lPHG and ACC activations. They had no a priori hypotheses on the other activations. Based on the task settings and on neuroscientific literature, they reported that a possible unified interpretation of the three activations observed in the contrast analysis *Consistent* > *Inconsistent* could be explained by the recruitment of processes related to memory and learning. More precisely, they interpreted the lPHG and caudate activations as being both likely explained by the engagement of processes related to encoding and binding of stimulus features to memory traces and the retrieval of information from declarative memory necessary to make causal inferences (Spaniol et al. 2009; see also Aminoff et al. 2013; Ekstrom and Bookheimer 2007; Grahn et al. 2008; Kohler et al. 1998; Monchi et al. 2006; Owen et al. 1996). The unexpected recruitment of the precentral gyrus (BA 6), which was coactivated with these regions, was interpreted as possibly reflecting the extent to which preparatory motor functions were occurring during the data accumulation phase of the task (Hanakawa et al. 2002; Lamm et al. 2001; Nobre et al. 1997; Tanaka et al. 2005).

Similarly, the authors reported that a possible unified account of the three activations observed in the contrast analysis *Inconsistent* > *Consistent* could be that data inconsistent with the plausibility of theories are likely processed by the brain as errors, and lead to reallocation of attentional resources away from the task. More precisely, they interpreted ACC (BA 24/32) activation as being likely explained by participants detecting erroneous data and monitoring conflict engendered by these data (van Veen and Carter 2002; see also Botvinick 2007; Botvinick et al. 2001, 2004). They also suggested that (1) lDLPC (BA 9) activation could be possibly explained by the active inhibition of attentional processes associated with the task (Buchsbaum et al. 2005; see also Bush et al. 2006; Garavan et al. 2002; Menon et al. 2001; Monchi et al. 2001), and that (2) precuneus (BA 7) activation could be possibly explained by the reorientation or reallocation of attentional resources away from the task (Cavanna and Trimble 2006; see also Bartolomeo et al. 2012; Corbetta et al. 1995; Mahayana et al. 2014; Simon et al. 2002). In brief, they concluded that conditions in which data are consistent with the plausibility of a theory may imply that participants are likely more apt to efficiently encode and retrieve information and, thus, likely more apt to learn. Conversely, they concluded that conditions in which data are inconsistent with the plausibility of a theory appear to recruit error processing and attentional reallocation mechanisms and, thus, do not seem to benefit learning.

With respect to overall findings, the four studies reviewed in this section did not find any neural commonalities between inference and perception of causality. Indeed, Fugelsang and Dunbar (2005) and Fonlupt (2003) examined causal inference, and their findings suggest that brain activations (e.g., lPHG, ldMPC) observed when inferring causality stem from the recruitment of processes related to attention, memory, and reasoning. The results of the two other studies, which examined causal perception (Blakemore et al. 2001; Fugelsang et al. 2005), suggest that brain activations (i.e., mainly visual system areas) observed when perceiving causality are different and stem from a more automatized processing of visuospatial information. Moreover, Fugelsang and Dunbar (2005) found more left-lateralized activations associated with inferring causality, whereas Blakemore et al. (2001) found more right-lateralized activations associated with perceiving causality. This discrepancy between neural correlates and processes recruited in causal inference and perception is consistent with the findings of several other studies in the literature (e.g., Roser et al. 2005; Schlottmann and Shanks 1992; Scholl and Nakayama 2002), and strengthens the idea that separate neural systems for causal perception and causal inference appear to exist in the human brain.

As a prime example of such studies, Roser et al. (2005) conducted a series of experiments on callosotomy (split-brain) patients ($n = 2$). In a first experiment, participants had to perceive the presence or absence of causality in the same billiard ball collision task as in the above-mentioned study by Fugelsang et al. (2005). When stimuli were shown in their left visual field, participants performed well and were able to identify spatial and temporal gap events as being non-causal events. However, when stimuli were shown in their right visual field, participants performed poorly and were not able to identify gap events as being non-causal events. Because left visual field stimuli are processed by the right hemisphere and vice versa, it was thus concluded that the right hemisphere was dominant in perceiving causality in this task. In a second experiment, the same participants were asked to infer causality in a series of movements of "switches" and their effect on a "lightbox" and then to decide whether a presented stimulus (i.e., the movement of one of the switches) was the reason the box lit up. Stimuli were once again shown in either their right or left visual field. Contrary to the results of the first experiment, participants performed well and were able to infer causality correctly when stimuli were shown in their right visual field, but not when they were shown in the left. It was thus concluded that the left hemisphere was dominant in inferring causality in this task. Thus, the human brain appears to have separate neural systems for causal perception and causal inference, with perception being more right-lateralized and inference being more left-lateralized.

As observed in the studies on overcoming misconceptions, the four studies on causal reasoning reviewed here have certain limitations. The two mentioned above (i.e., problematic reverse inference and small sample sizes) will be more closely examined in the general discussion of this review. First, the four studies focused on a single type of mechanical causality, namely, the collision between two bodies. However, the physical world includes a variety of types of mechanical causality, such as crushing, bending, pulling, shattering, and launching (Patterson and Barbey 2005). These are very common in the real world and underlie the learning of several concepts, theories, and laws in science curricula, such as physics (e.g., projectile motion, the behavior of a spring). Thus, it is possible that different neural correlates and processes are recruited when perceiving or inferring these other types of mechanical causality. Further studies are needed to achieve a more accurate picture of the neural correlates associated with other types of mechanical causality.

The second limitation of these causal reasoning studies is their use of a single type of perceptual modality (i.e., visual modality) to examine neural correlates recruited by the perception and inference of causality. However, there are other types of perceptual modalities that subtend the perception and inference of causality, such as touch and audition (Kording et al. 2007; McGann 2010). In a scientific context, auditory processing is involved, for instance, in several causal events that make characteristic sounds (Patterson and Barbey 2005), such as the tearing of a piece of paper, the cracking of a stick, reverberation, or the acceleration of a motor vehicle. Future neuroimaging research on causality in scientific contexts could thus focus on causal events involving perceptual modalities other than vision, to determine whether the neural correlates and processes recruited by such events are the same as those recruited by visual perception and inference of causality. Neural commonalities between modalities would suggest the presence of a unitary neural system subtending the perception and inference of physical causality.

Hypothesis generation

Hypothesis generation can be defined as a deductive reasoning process by which a theoretical proposition is formulated by a reasoner to explain or predict natural phenomena occurring in the environment (McGuire 1997). It is generally acknowledged (McGuire 1997; Zavrel and Sharpsteen 2016) that generating hypotheses is a critical ability among other skills related to scientific inquiry, the process by which new scientific knowledge is produced. Training science learners to generate hypotheses is usually part of the science curriculum (National Research Council [NRC] 2005). According to cognitivist models (Cooper and Yule 2013; Pleskac et al. 2007), the main cognitive processes involved in generating a hypothesis are the retrieval of relevant declarative knowledge from long-term memory, the inference of associations between this knowledge and cues from the environment, the updating of working memory with these representations, and the evaluation of the consequences of the action related to a hypothesis. Pleskac et al. (2007), however, underline that, because multiple cognitive processes appear to contribute to hypothesis generation, it is difficult to arrive at an exact understanding of how a hypothesis is generated by human cognition. The three neuroimaging studies (Lee and Kwon 2011, 2012; Kwon et al. 2009) on hypothesis generation included in this review obtained similar results, suggesting that the above-mentioned processes indeed appear to be involved when generating a hypothesis. Moreover, as highlighted below, these studies suggest that generating a hypothesis could have a more positive effect on learning and motivation than understanding a hypothesis. *Hypothesis understanding* in these studies comprises expository teaching about natural phenomena, in which students are provided with direct explanations of phenomena without formulating any hypotheses (Lee and Kwon 2011).

First, Kwon et al. (2009) used fMRI and an experimental two-group design to compare the change in cerebral activation patterns caused by two different training programs. Participants were female undergraduates (age range 20–25 years) who were randomly split between either of the instruction programs. The experimental group ($n=9$) underwent an eight-week program in which they were trained to generate their own hypotheses about various biological phenomena for 60 min per week. More precisely, the training session administered each week to the experimental group consisted of six steps: (1) observe a picture depicting a natural phenomena (e.g., sap oozing from a tomato stump), (2) generate a causal question with regard to the given phenomena (e.g., “Why did sap ooze from the stump?”), (3) analyze the question, (4) represent a phenomena similar to the given phenomena, which helped them, (5) find possible explanations for the given phenomena, and (6) construct a hypothesis (or explanation) about the given phenomena. Meanwhile, the control group ($n=9$) was given instructions on understanding hypotheses and was thus passively provided with correct explanations about the same phenomena rather than generating their own hypotheses. The cerebral activity of both groups was measured before and after the 2-month period, using a hypothesis generation visual task. In the task, participants were shown biological pictures depicting natural phenomena different from training, accompanied by a causal question (e.g., “Why is the monkey covered with white fur?”), and were asked to generate a hypothesis with regard to this question. In addition, immediately after the scanning session at both pretest and posttest, participants were asked to write down the hypotheses they had generated during the task, and these were scored.

With respect to behavioral data first, the experimental group showed substantial improvement between pretest ($M_{\text{score}}=2.33$, $\text{SD}=0.71$) and posttest ($M_{\text{score}}=4.59$, $\text{SD}=1.10$) scores, whereas the control group showed a more modest improvement between pretest ($M_{\text{score}}=2.00$, $\text{SD}=0.75$) and posttest ($M_{\text{score}}=2.26$, $\text{SD}=0.72$) scores. Comparative analysis of the two groups’ pretest and posttest scores was conducted using a one-way ANOVA with repeated measures. While the ANOVA showed no statistical difference between the two groups’ mean scores on the pretest ($p=0.195$), the mean score of the experimental group was significantly higher on the posttest ($p<0.001$). The authors thus concluded that participants in the experimental group performed better than those in the control group. However, they do not appear to have addressed the group (training/control) × time (pre/post) interaction using a two-way ANOVA. As discussed below, this type of statistical analysis is crucial when attempting to demonstrate significant training effects (Livelli et al. 2015; Sternberg 2008).

With respect to neuroimaging data, this study was mostly exploratory, and thus, the authors did not have precise a priori hypotheses with regard to expected brain activations. The *Post > Pre* contrast analysis for the control group, using a paired *t* test, yielded no significant difference in brain activation. The same *Post > Pre* (paired *t*-test) contrast analysis for the experimental group yielded significantly increased activation, notably in the bilateral superior frontal gyri (SFG, BA 6) and the left inferior frontal gyrus (lIFG, BA 9), as well as significantly decreased activation, notably in the left inferior parietal lobule (lIPL, BA 40) and right insula (BA 13). In addition, a one-way ANOVA was conducted to compare the brain activations of the two groups at posttest. The authors reported that the experimental group’s brain activity on the posttest was significantly higher than that of the control group in the same two above-mentioned regions (SFG, BA 6 and lIFG, BA 9) and significantly lower in the same two above-mentioned regions (lIPL, BA 40 and right insula, BA 13). Similar to behavioral data, a group (training/control) × time (pre/post) comparison of relative brain activation using a two-way ANOVA does not appear to have been conducted. Based on the task settings and on neuroscientific literature, the authors interpreted greater lIFG (BA 9) activation as being possibly due to the engagement of a greater working memory load necessary to generate hypotheses (Hirshorn and Thompson-Schill 2006; Kuperberg et al. 2006; Moss et al. 2005; Thompson-Schill et al. 1997). Similarly, SFG (BA 6) activation was explained as likely due to recruitment of higher-order inferential reasoning processes involved in generating hypotheses, which consist in evaluating information and establishing complex relationships among pieces of information (Christoff et al. 2001; Green et al. 2006; Kroger et al. 2002). Lesser activation of the right insula (BA 13) was explained as being possibly due to the experimental group's increased confidence in generating hypotheses following experimental training. This explanation was based on neuroscientific literature showing that the right insula is linked with combined cognitive and affective processing when adverse outcomes are expected, such as in the risky, uncertain decision-making behavior involved in hypothesis generation (Elliott et al. 2000; Paulus et al. 2003; Sawamoto et al. 2000). Finally, lesser activation of the lIPL (BA 40) was explained as being possibly due to the fact that participants in the experimental group more efficiently analyzed the visual stimuli after the training program than before it. This explanation was based on neuroscientific literature, showing that the lIPL is recruited by tasks assessing sustained visuospatial attention and vigilance (Adler et al. 2001; Coull et al. 1998; Foucher et al. 2004).

In brief, Kwon et al. (2009) found that hypothesis generation recruited processes similar to those described in cognitivist models as supporting hypothesis generation, such as working memory updating, inferential reasoning, and evaluation of the consequences (e.g., "risks") of making a hypothesis. In addition, they concluded that hypothesis-generating training had a more positive effect on learning than hypothesis-understanding instruction, because it induced (1) better task performance and (2) above-mentioned changes in neural activation patterns that suggest implementation of training-dependent brain plasticity mechanisms.

In a later study, Lee and Kwon (2011) used fMRI to compare the cerebral activity of a single group of 60 male participants ($M_{\text{age}} = 24.05$ years; age range 16–42 years) during completion of two types of tasks in biology. In both tasks, participants were shown visual stimuli representing biological phenomena (e.g., formation of dung balls by dung beetles). In the first task (hypothesis generating), they were presented with separate stimuli of the cause and effect of natural phenomena and were asked to generate a hypothesis (or explanation) on the biological process linking the cause to the effect. After producing a hypothesis, participants were shown the correct biological process to validate their hypothesis. In the second task (hypothesis understanding), participants did not generate any hypotheses. They were directly presented with the correct biological process linking the cause and effect of phenomena, which they only had to understand.

Similar to Kwon et al.'s (2009) study, the authors did not have precise a priori hypotheses with regard to expected brain activations. The contrast analysis *Hypothesis generating* > *Hypothesis understanding* revealed greater relative activity in a left hemisphere neural network notably comprising the lDLPC (BA 9/46), putamen, lPHG (BA 22), and left middle occipital gyrus (lMOG, BA 19). Based on the task and the literature, they suggested that the activation of the lDLPC (BA 9/46) could possibly be explained by the integration of abstract level causal relationships that are required to generate hypotheses from observing complex biological phenomena (Kwon et al. 2006). This process involves the establishment of complex relationships between pieces of information contained in working memory (Christoff et al. 2001; Green et al. 2006; Kroger et al. 2002). Similarly, activation of the putamen, an area of the mesolimbic system, was explained as being possibly due to the feeling of uncertainty felt while generating a hypothesis (Balleine et al. 2007; Chang et al. 2002; Sefcsik et al. 2009), or to the anticipation of the reward accompanying an accurate hypothesis (Haruno and Kawato 2006; Tanaka et al. 2006). Moreover, they emphasized that the putamen is part of the brain's reward system and, moreover, that its recruitment has been linked in the literature to academic achievement motivation (Berridge 1996; Berridge and Robinson 1998; Mizuno et al. 2008). Activation of the lPHG (BA 22) was explained as being likely due to the establishment of contextual associations between prior knowledge retrieved from declarative memory and informations processed from the task's stimuli (Spaniol et al. 2009; see also Aminoff et al. 2013; Ekstrom and Bookheimer 2007; Kohler et al. 1998; Owen et al. 1996). Finally, activation of the BA 19, a higher-order visuospatial processing area (Fortin et al. 2001; Lamm et al. 2001; Waberski et al. 2008), was explained as being possibly due to a more intense visual exploration of the stimuli by participants when they had to generate a hypothesis compared to when they had to understand it.

In brief, the findings of Lee and Kwon's study (2009) again suggested that generating a hypothesis appears to recruit processes similar to those described in cognitivist models of hypothesis generation, such as (1) inferring associations between externally derived cues and internally derived prior knowledge and (2) evaluating or anticipating the consequences (e.g., "reward") of making a hypothesis. In addition, because of the above-mentioned recruitment of the brain reward system, they suggested that students could be more interested in generating a hypothesis than in receiving passive instruction from which they have to understand a hypothesis.

Lastly, Lee and Kwon (2012) used fMRI to compare the cerebral activation patterns of two groups of high school students ($M_{age} = 16.71$ years; age range 16–17 years) who received either hypothesis-generating training ($n = 7$) or hypothesis-understanding instruction ($n = 7$) for 12 weeks. The hypothesis-generating group received twelve training sessions on biological hypothesis generation on different biological topics (e.g., rubber hand illusion) for 4 h each week. Each weekly training session comprised the same six steps as those used in the Kwon et al. (2009) study previously described. Meanwhile, the control group received the same type of hypothesis-understanding instruction as in the Kwon et al. (2009) study. Both groups were scanned before and after the 12-week period using both a hypothesis-generating visual task and a hypothesis-understanding visual task. In the hypothesis-generating task, participants were presented with biological pictures showing natural phenomena different from training. Pictures were accompanied by a causal question (e.g., "What causes a dung ball to appear?") and participants were asked to generate a hypothesis explaining the phenomena in question. They were then shown the correct explanation and were required to indicate, by clicking on one of the two mouse buttons, whether or not their hypothesis was in agreement with the correct explanation. Data obtained from participants' mouse clicks, although subjective, were considered as behavioral data representing accuracy during the hypothesis-generating task. In the hypothesis-understanding task, participants were passively presented with the correct explanation of the phenomena and only had to understand it. They were then required to indicate, by clicking on one of the two mouse buttons, whether or not they fully understood the correct explanation. Data obtained from participants' mouse clicks were considered as representing accuracy during the hypothesis-understanding task. In addition, immediately after the scanning session at both pretest and posttest, both groups were administered a paper and pencil test to measure their general scientific hypothesis generation ability (Lee 2009).

With respect to behavioral data first, this study conducted a two-way ANOVA (group [training/control] × time [pre/post]) on both types of behavioral data (i.e., accuracy during the task and hypothesis-generating ability). For accuracy, the authors reported a nonsignificant interaction effect ($p = 0.441$) between group and time, suggesting that variation in the means of the two groups' accuracy scores over the 12-week treatment period was not significantly different depending on the type of treatment received. For the hypothesis-generating ability, they reported a significant main effect of group ($p = 0.004$), but a nonsignificant interaction effect ($p = 0.623$) between group and time. Because the mean scores of the two groups' hypothesis-generating ability were not statistically different at pretest ($p = 0.124$), the authors interpreted these findings as suggesting participants of the trained group showed significantly higher ability improvement than those in the control group due to the 12-week training program. However, this interpretation appears questionable because of the absence of a significant interaction effect between group and time. Such an interaction, as previously stated and as discussed below, is critical when attempting to demonstrate significant training effects.

With respect to neuroimaging data, the authors used a one-way ANOVA at posttest to compare experimental versus control group cerebral activity during hypothesis generation and hypothesis-understanding tasks. In addition to the one-way ANOVA, they conducted paired *t*-tests to determine the regions in which a group showed a relative increase or decrease in activity compared with the other group from pretest to posttest. Similarly to Kwon et al.'s (2009) study, a two-way ANOVA (group × time) with cerebral activity as a dependant variable was not conducted. Moreover, because two unrelated groups were compared, it seems unclear why and how paired rather than independent *t* tests were used. Lastly, authors had precise hypotheses with regard to expected brain activations, based on the findings of previous studies, such as Kwon et al. (2009) and Lee and Kwon (2011). Neural findings confirmed their hypotheses.

For the hypothesis generation task, it was found that the experimental group's neural activity increased significantly more from pretest to posttest, compared with the control group, in the left PHG, left DLPC, left superior temporal gyrus (lSTG), and left putamen. Conversely, the experimental group's neural activity decreased significantly more from pretest to posttest, compared with the control group, in the left middle occipital gyrus (lMOG) and right lingual gyrus. On the other hand, for the hypothesis-understanding task, it was found that the control group's neural activity increased significantly more from pretest to posttest, compared with the experimental group, in the right middle frontal gyrus (rMFG) and the right precuneus. Conversely, the control group's neural activity decreased significantly more from pretest to posttest, compared with the experimental group, in the left corpus callosum.

For the hypothesis-generating task, the pattern of comparative change in cerebral activity, and its accompanying interpretations, presents major similarities with findings from previously described studies by Lee and Kwon (2011) and Kwon et al. (2009). Eloquent examples of such similarities are the greater comparative activations in the lPHG, lDLPC, and left putamen observed in the experimental group. The comparative lPHG activity increase in the experimental group was explained by Lee and Kwon (2012) as being likely due to the creation of contextual associations between informations derived from the task's stimuli and prior knowledge stored in declarative memory (Spaniol et al. 2009; see also Aminoff et al. 2013; Ekstrom and Bookheimer 2007; Kohler et al. 1998; Owen et al. 1996; Yue et al. 2007). The comparative lDLPC activity increase in the experimental group was explained as being likely due to the engagement of information integration processes consisting of establishing complex relationships between pieces of information (Christoff et al. 2001; Green et al. 2006; Kroger et al. 2002; Kwon et al. 2009). This finding was explained as being likely caused by the 12-week training program that engaged these processes in the experimental group, whereas these processes were less recruited in the control group. The comparative putamen activity increase in the experimental group was explained as being likely due to the uncertainty felt while generating a hypothesis or to the engagement of the brain reward system (Balleine et al. 2007; Chang et al. 2002; Mizuno et al. 2008; Sefcsik et al. 2009). This finding was explained as being likely caused by experimental group participants being more motivated because they could seek explanations actively, whereas control group participants were accustomed to having the explanations passively provided to them.

Lastly, Lee and Kwon's (2012) study also presents similar findings to Lee and Kwon (2011) on the important role of higher-order visuospatial processing areas (i.e., left MOG, lingual gyrus) during hypothesis generation. The comparative decrease in these regions in the experimental group was explained as being likely due to the neural efficiency hypothesis (Dunst et al. 2014; Haier et al. 1992). Participants who were trained in hypothesis generation for 12 weeks processed task-dependent visual information more efficiently. Conversely, participants who received the hypothesis-understanding treatment were less efficient at processing task-dependent visual information because of the few opportunities given to them to do so during treatment. Thus, participants in the experimental group needed substantially less recruitment of brain areas subserving higher-order visuospatial processing during the hypothesis generation task.

For the hypothesis-understanding task, the authors interpreted the pattern of comparative change in cerebral activity as pointing toward a greater working memory load in the control group. For example, the comparative rMFG activity increase in the control group was explained as being likely due to increased requirements of information maintenance in working memory (Olesen et al. 2004). The comparative right precuneal increase in the control group was explained as being likely due to increased requirements of temporary storage of basic visuospatial information in working memory (Smith and Jiondes 1997). These findings were explained as being likely caused by control group participants being constantly pushed to memorize the correct explanations of the presented natural phenomena during the hypothesis-understanding instruction.

In summary, when considered as a whole, the neuroimaging findings from these three studies suggest the recruitment of mostly left-lateralized neural correlates in frontal (e.g., superior and inferior frontal gyri), temporal (e.g., parahippocampal gyrus), and basal ganglia (e.g., putamen) areas when a hypothesis is generated. Moreover, these activations are interpreted by the authors as subtending mental processes similar to those described in cognitivist models as supporting hypothesis generation—namely, retrieval of prior knowledge from long-term memory, inference of associations between prior knowledge and stimuli from the task, updating of working memory, and evaluation of the consequences of making a hypothesis (e.g., being right and receiving a reward, such as positive feedback, or the risk of being wrong and receiving negative feedback).

In addition to appearing to confirm the recruitment of mental processes similar to those suggested by cognitivist models, the findings of the three neuroimaging studies reviewed here suggest that hypothesis generation training may have a more positive effect on learning and motivation than hypothesis understanding. An example of such findings is the behavioral data gathered by Kwon et al. (2009) and Lee and Kwon (2012). In both studies, researchers found that participants trained in generating hypotheses appeared to perform better on written tests measuring the ability to generate scientific hypotheses than participants who received instruction in hypothesis understanding. Although the tests represented a near transfer type of task (Barnard and Jacobs 2007) that measured learning in a context closely related to training, these results can nevertheless be interpreted as indicating a positive effect on learning due to the critical role of hypothesis generation in scientific inquiry (McGuire 1997; Zavrel and Sharpsteen 2016).

Another example of a possible positive effect on learning are the results of the study by Lee and Kwon (2012), showing that, at posttest, trained participants exhibited a relative decrease in activity in higher visuospatial processing areas compared with control participants. This was interpreted as the trained group becoming more neurally efficient at processing sufficient visual information from the task's stimuli. Neural efficiency is a well-established concept in neuroscientific studies on human intelligence. It has been observed in a variety of studies employing different neurophysiological measurement methods and a broad range of cognitive task demands (Dunst et al. 2014; Neubauer and Fink 2009; Neubauer et al. 2002). Neurally efficient individuals use a more limited group of neural circuits and/or fewer neurons, resulting in lower total cortical activation and greater efficiency when engaged in performing cognitively demanding tasks (Haier et al. 1988, 1992). Lastly, as already pointed out, Lee and Kwon (2011, 2012) explained the greater relative putamen recruitment in hypothesis generation as likely being due to greater engagement of motivation processes. This appears to be in line with findings from previous behavioral studies comparing emotional changes in hypothesis-generating vs. hypothesis-understanding conditions (Lee and Kwon 2008;

Teixeira-Dias et al. 2005; van Zee 2000). In these studies, participants in hypothesis-generating conditions generally expressed more positive emotions, such as higher interest and motivation.

As discussed earlier in this review, the three studies on hypothesis generation have certain limitations in common. One of these has already been briefly mentioned and concerns statistical data analysis conducted in the two-group design studies (i.e., Kwon et al. 2009; Lee and Kwon 2012). These two studies evaluated the effect on a dependent variable (e.g., cerebral activity, task accuracy) of a mix of one between-subject factor (i.e., group: trained vs. control) and one within-subject factor (i.e., time: pretest vs. posttest). Thus, a two-way mixed ANOVA would have appeared to be the most appropriate type of statistical data analysis in this case to demonstrate a significant training effect. More precisely, a significant interaction effect between group and time would have made it possible to infer that the variation in the dependent variable over the treatment period (time) differed depending on the type of treatment received (group) (Hinkle et al. 2003; Howell 2002; Seltman 2015). In addition to being largely documented in the analysis of behavioral data (e.g., Seltman 2015), the use of a two-way ANOVA is also well documented for the analysis of neuroimaging data (Auffermann et al. 2001; Chen et al. 2013; Henson and Penny 2005; Ward and Chen 2006). However, in the Kwon et al. (2009) and Lee and Kwon (2012) studies, a two-way mixed ANOVA was not conducted, and the interaction effect was addressed using another form of statistical analysis. For example, Kwon et al. (2009) used paired *t*-tests to compare the pre and posttest cerebral activity of each group and a one-way ANOVA to compare the cerebral activity of the two groups at posttest. This combination of analyses appears less appropriate than a two-way mixed ANOVA, for several reasons. First, the within-subject and between-subject factors were considered separately, making it impossible to test for the interaction of their effects (Auffermann et al. 2001). In addition, the one-way ANOVA did not consider the pretest cerebral activity of the two groups, which may have revealed comparative differences. Lastly, this combination of *t*-tests and one-way ANOVA involved analyzing certain data twice, potentially increasing the risk of committing a type I error (Seltman 2015).

Another significant limitation is that none of the two-group design studies in which one group received hypothesis generation training (Lee and Kwon 2012; Kwon et al. 2009) used a delayed posttest to compare the cerebral activity of the two groups after a supplementary period of no training. It is therefore unclear whether differences in brain activation patterns of the two groups observed immediately after training would still be present a few weeks later. A future study could examine the delayed effect of hypothesis generation training on cerebral activity by comparing the cerebral activity of previously trained participants and control participants during a hypothesis-generating task completed several weeks after training. A delayed comparison would seem especially relevant given results obtained by Kwon et al. (2009) and Lee and Kwon (2012), which suggest the occurrence of neural changes related to learning and motivational benefits in the trained group, such as greater neural efficiency and greater engagement of the brain reward system.

General discussion

Thus far, the present review has examined the conclusions of various studies regarding neural correlates and processes for each type of cognitive task, highlighting the limitations specific to each category. This section will discuss convergent findings and limitations across all three scientific reasoning tasks.

Convergent findings

Firstly, recruitment of the lateral prefrontal areas subtending diverse executive functions (i.e., inhibitory control, decision-making, and working memory updating) was observed for all three types of cognitive tasks. This observation seems to support the idea that executive functions play a role in scientific reasoning and, potentially, in scientific literacy and science learning achievement in school. Cognitive studies have previously found associations between science learning achievement and (1) working memory updating (Bahar and Hansell 2000; Danili and Reid 2004; Gathercole et al. 2004; Jarvis and Gathercole 2003; Rhodes et al. 2014, 2016; St Clair-Thompson and Gathercole 2006; St Clair-Thompson et al. 2012); (2) attention set-shifting (Latzman et al. 2010); and (3) planning (Rhodes et al. 2014).

With respect to pedagogical implications, one interesting hypothesis arising from these findings is the possibility that executive function-based teaching strategies improve scientific reasoning ability, and might therefore also improve overall academic performance in scientific disciplines. This hypothesis has been tested by previous empirical research in literacy and mathematics in two contexts: far transfer situations and near transfer situations (Sala and Gobet 2017a, b). A far transfer situation involves applying previous learning to dissimilar situations, such as contexts and tasks different from those in which the original learning event took place (Barnard and Jacobs 2007; Laker 1990). A significant number of empirical studies were conducted to test for far transfer between executive function-based training and academic achievement in mathematics (e.g., Henry et al. 2014; Holmes et al. 2009; Karbach et al. 2015; Kroesbergen et al. 2014) and literacy (e.g., Lee 2014; Loosli et al.

2012; Nevo and Breznitz 2014; Studer-Luethi et al. 2016). For example, St Clair-Thompson, Stevens, Hunt, and Bolder (2010) examined the effectiveness of computerized working memory training (e.g., digit recall, block recall, listening recall) among 5- to 8-year-olds. Although training led to an improvement in working memory skills, no improvement was observed on standardized tests for mathematics and reading, either immediately following training or 5 months later. Recent meta-syntheses (Cragg and Gilmore 2014) and meta-analyses (Melby-Lervåg et al. 2016; Sala and Gobet 2017b; Simons et al. 2016) have concluded that, in far transfer situations, there is not enough evidence to date to support the idea that executive function-based training enhances academic performance in mathematics and literacy.

A near transfer situation involves applying previous learning to similar situations, such as contexts and tasks closely related to those in which the original learning event took place (Barnard and Jacobs 2007; Laker 1990). Several empirical studies have been conducted to test for near transfer between a training task comprising executive function-based training strategies and a closely related testing task. These studies have largely focused on mathematical abilities in the fields of logic (e.g., Houdé et al. 2001; Houde and Moutier 1996, 1999; Moutier et al. 2002) and geometry (e.g., Babai et al. 2015). Their findings generally indicate that executive function-based training strategies have a positive effect on learning. For example, Babai et al. (2015) tested the effects of warnings on performance in a geometry task where participants (sixth graders aged 11–12) had to determine which of two shapes had the larger perimeter. *Warnings* are a teaching strategy intended to activate inhibitory control processes and consist in explicitly informing participants about the appealing but erroneous reasoning strategies that they have to inhibit. The warning intervention was administered to an experimental group in training trials similar though not identical to the test trials. Warnings resulted in an improvement in the accuracy of responses in comparison with the control group.

Thus, the literature suggests that executive function-based training strategies may have a beneficial effect on learning when used in near transfer but not far transfer situations. However, in the domain of science, as pointed out by Sala and Gobet's recent meta-analysis (2017a, b), no peer-reviewed journal appears to have published studies on the effects of executive function-based training strategies on science achievement, either in far transfer or near transfer situations. This subject would therefore be an interesting area for future empirical research to examine, particularly with near transfer situations, which appear to be most promising.

Another interesting finding arising from the present review is that other brain areas seem to be associated with scientific reasoning abilities across more than one type of task. A noteworthy example is the parahippocampal gyrus, or the medial temporal lobe in general, which was found to be recruited when participants performed causal reasoning and hypothesis generation. This observation suggests an association between declarative memory processes (i.e., encoding, consolidation, and retrieval) and scientific reasoning; it also reinforces previous evidence from cognitive and developmental research showing the profound influence of prior knowledge and experience on scientific reasoning (Shah et al. 2017). In other words, scientific reasoning in a given task always draws on prior knowledge, including facts, concepts, theories, laws, and models. These conclusions, along with the massive amount of information covered in science curricula and the prevalence of misconceptions, all point to one straightforward pedagogical implication: one way to improve students' scientific reasoning could be to provide better guidance on what knowledge should be accessed and for what purpose during a learning task in science. This hypothesis echoes recent meta-analytic literature, which suggests that instructional guidance is beneficial to the development of scientific knowledge and reasoning skills and, moreover, that what constitutes the best type of instructional guidance in science depends on students' age (Alfieri et al. 2011; D'Angelo et al. 2014; Carolan et al. 2014; Furtak et al. 2012; Lazonder and Harmsen 2016).

As an eloquent example, the meta-analysis completed by Lazonder and Harmsen (2016) synthesized the results of 72 studies to compare the effectiveness of different types of instructional guidance on science learning outcomes—namely, scientific knowledge and reasoning abilities, as assessed through posttests, criterion tasks, interviews, or questionnaires. The different types of instructional guidance ranged in specificity. More specific types of guidance consisted of precise *scaffolds* and *explanations* with regard to the knowledge and reasoning skills relevant to the task. Less specific or more minimal types of guidance consisted of *prompts*, or reminders to perform an action, and *process constraints*, which restricted the comprehensiveness of the learning task. Results showed that receiving guidance (all types combined) had a significant positive overall effect ($d=0.50$) on learning outcomes compared with receiving no guidance at all. In addition, results showed that less specific types of guidance appeared to be more beneficial for older students ($d=0.94$), aged 15–22, than for younger students ($d=0.78$), aged 5–12. Conversely, more specific types of guidance appeared to be more beneficial for younger students ($d=3.62$) than for older students ($d=0.70$). The authors explained these results as being likely due to the fact that younger students have less content knowledge and less familiarity with scientific reasoning skills, therefore requiring more specific instructional guidance.

Several authors have also proposed that simply carrying out science learning tasks, with minimal or no guidance, could impose excessive demands on novice learners' working memory, leaving too little capacity to recruit long-term memory processes during the task (Hushman and Marley 2015; Kirschner et al. 2006; Paas et al. 2003; Sweller 1999; Tuovinen and Sweller 1999). With more specific instructional guidance, novice learners receive more support in terms of relevant knowledge and reasoning skills, thus decreasing their working memory load and increasing the capacity to recruit long-term memory processes, such as encoding new knowledge (Mayer 2004; Hushman and Marley 2015; Shulman and Keisler 1966). This mechanism, postulated as subtending the beneficial behavioral effects of more specific instructional guidance on novice learners, does not yet appear to have been addressed using neuroimaging. It could thus represent an interesting lead for future research, which could examine the combined neural and behavioral effects of different types of instructional guidance on novice learners. In accordance with Lazonder and Harmsen's (2016) results discussed above, it would be expected that more, rather than less, specific guidance would benefit younger, more novice learners, such as elementary school students. Providing these students with specific guidance could lead, on the one hand, to lesser relative activation in working memory areas and greater relative activation in long-term memory areas and, on the other hand, to a positive correlation between greater neural activity in long-term memory areas and better posttest behavioral results.

Common limitations

With regard to common limitations among the studies included in this review, one that applies to most of these studies is the use of potentially problematic reverse inferences. *Reverse inference* is the inference of the engagement of a specific cognitive process based on brain activation observed during a task (Aguirre 2003; Poldrack 2006; Hutzler 2013). Reverse inference is a frequent practice in the interpretation of neuroimaging results and is used most often in cases where the activation of a brain region that was not part of the a priori research hypotheses occurs (Poldrack 2011). Pinal and Nathan (2017) point out that reverse inference is widely criticized, and that some researchers have gone as far as suggesting that it be removed from the cognitive neuroscience tool kit. However, reverse inference is not problematic per se if there is a high likelihood of engagement of the particular cognitive process (C) inferred from the observed pattern of brain activity (A) and the task settings (T). This likelihood can be framed as the conditional probability $P\,(C|A \cap T)$ (Hutzler 2013; Poldrack 2011; Sarter et al. 1996). According to this formula, the two conditions that would increase conditional probability, thereby increasing the validity of a reverse inference, are functional specificity and task specificity. *Functional specificity* is the extent to which a brain region is specialized for a single cognitive process (Kanwisher 2010)—in other words, the specificity of activation of a brain region. As emphasized by Poldrack (2006, 2011) and Pinal and Nathan (2017), several brain regions, especially those in the heteromodal associative cortex, can be activated by a wide range of cognitive processes. Inferring the engagement of a specific cognitive process from the activation of one of these regions would therefore be a weak form of reverse inference and consequently diminish the validity of the results' interpretation. Likewise, the fact that smaller brain regions (e.g., BA 19) are more selective of cognitive processes than larger brain regions (e.g., medial temporal lobe) weakens the validity of reverse inference based on the activation of a larger region (Poldrack 2006). *Task specificity* refers to the extent to which task settings isolate a particular cognitive process. As Hutzler (2013) explains, higher task specificity reduces the probability that the observed activation of a brain region could be explained by other cognitive processes implemented in that region. Thus, reverse inference in the case of tasks involving several cognitive processes could also be unreliable.

The studies in all three categories of the present review provide examples of problematic instances of reverse inference. For example, Potvin et al. (2014, p. 9) interpreted the "bilateral extended activations in the posterior region of the brain, beginning at the inferior occipital gyrus and ending at the angular gyrus and the superior parietal lobule," for which there was no a priori hypothesis, as likely being the result of "visuospatial processing." This inference is problematic mostly with regard to functional specificity, because the observed activation appears to encompass a large region of the brain. Similarly, Fugelsang and Dunbar (2005, p. 1210) interpreted the observed DLPC activation, for which there was no a priori hypothesis, as being possibly explained by "the active inhibition of the attentional processes associated with the task." Once again, the problem with this reverse inference relates primarily to functional specificity, as the DLPC is part of the frontal heteromodal associative cortex, and several cognitive processes are implemented in this region. As a last example, Kwon et al. (2009, p. 394) interpreted the activation of the DLPC, for which there was no precise a priori hypothesis, as being likely explained by the "integration of abstract level causal relation structure […] to generate biological hypotheses from complex biological phenomena." The issues with this reverse inference mirror those outlined above regarding the Fugelsang and Dunbar (2005) study.

Another significant limitation in several of the studies discussed in this review is the use of small sample sizes. For example, Masson et al. (2014) used two groups of 11 and 12 participants, respectively, Blakemore et al. (2001) used a single group of 8 participants, and Lee and Kwon (2012) used two groups of 7 participants each. As will be explained below, small sample sizes are associated with decreased statistical power, and greater risk of committing a type II error and missing true activations. Using methods such as simulations (Desmond and Glover 2002) and resampling (Murphy and Garavan 2004), fMRI-specialized statisticians have predicted power in fMRI studies and produced power curves to summarize their results. Their work is often referenced in the neuroimaging community with respect to power and sample size issues (Hayasaka et al. 2007). To reach the acceptable 80% threshold of statistical power at conservative alpha values (e.g., $\alpha < 0.001$) commonly used in fMRI studies, these statisticians recommend using sampling groups of about 20 participants. It is suggested that once the sample size exceeds 20, fMRI activation maps generally cease to show significant supplementary activations for a given task. For example, there would be little to no benefit in scanning a sample of 50 participants in terms of significant observed supplementary activations. Conversely, scanning smaller samples ($n < 20$) for a given task yields activation maps showing fewer significant activations. Smaller samples, such as those used in most of the studies discussed in this review, could therefore result in less complete activation maps, increasing the risk of missing true activations and thereby committing a type II error.

Moreover, lower statistical power could lead to another problem affecting the reliability of findings, namely, a lower positive predictive value (Button et al. 2013; Heston and King 2017; Nichols et al. 2016). The positive predictive value (PPV) is the probability that a significant research finding reflects a true effect, or, in other words, that the finding is a true positive. The PPV depends on factors such as the a priori probability of the finding being true, the statistical power of the study, and the level of statistical significance (Button et al. 2013; Nichols et al. 2016). Through their meta-analysis, Button et al. (2013) demonstrate that low statistical power appears to be endemic in the field of neuroscientific research and argue that the major implication of a lower associated PPV is the decreased likelihood that any nominally significant neuroimaging finding actually reflects a true effect.

In summary, while the potential problems associated with the use of reverse inference and small sample sizes do not invalidate the findings of neuroimaging studies, such as those discussed in this review, they are nevertheless worth keeping in mind.

Conclusion

The objective of this review was to synthesize and classify neuroimaging studies that examined brain activation during three types of scientific reasoning tasks: overcoming misconceptions, causal reasoning, and hypothesis generation. The findings across all three tasks suggest converging evidence for the recruitment of (1) lateral prefrontal areas, likely caused by the engagement of executive function processes, and (2) medial temporal areas, likely caused by the engagement of declarative memory processes. The main hypotheses with respect to pedagogical implications that can be derived from these findings, in combination with previous literature, are that scientific reasoning skills and scientific literacy could be improved with the use of (1) executive function-based teaching strategies applied in near transfer situations and (2) better instructional guidance during scientific learning activities regarding what knowledge to use and why. These hypotheses need to be tested and refined in future research. Most of the studies discussed in this review also present two notable limitations: a problematic use of reverse inference, which could undermine the interpretation of results, and small sample sizes, which could diminish the completeness and reliability of findings.

References

Adler CM, Sax KW, Holland SK, Schmithorst V, Rosenberg L, Strakowski SM (2001) Changes in neuronal activation with increasing attention demand in healthy volunteers: an fMRI study. Synapse 42(4):266–272

Aguirre GK (2003) Functional imaging in behavioral neurology and cognitive neuropsychology. In: Feinberg TE, Farah MJ (eds) Behavioral neurology and neuropsychology, 2nd edn. McGraw-Hill, New York, pp 85–96

Alfieri L, Brooks PJ, Aldrich NJ, Tenenbaum HR (2011) Does discovery-based instruction enhance learning? J Educ Psychol 103:1–18. https://doi.org/10.1037/a0021017

Aminoff EM, Kveraga K, Bar M (2013) The role of the parahippocampal cortex in cognition. Trends Cogn Sci 17(8):379–390. https://doi.org/10.1016/j.tics.2013.06.009

Antal A, Nitsche MA, Kruse W, Kincses TZ, Hoffmann KP, Paulus W (2004) Direct current stimulation over V5 enhances visuomotor coordination by improving motion perception in humans. J Cogn Neurosci 16(4):521–527

Aron AR, Fletcher PC, Bullmore ET, Sahakian BJ, Robbins TW (2003) Stop-signal inhibition disrupted by damage to right inferior frontal gyrus in humans. Nat Neurosci 6:115–116. https://doi.org/10.1038/nn1003

Aron AR, Robbins TW, Poldrack RA (2004) Inhibition and the right inferior frontal cortex. Trends Cogn Sci 8(4):170–177

Auffermann WF, Ngan SC, Sarkar S, Yacoub E, Hu X (2001) Nonadditive two-way ANOVA for event-related fMRI data analysis. NeuroImage 14:406–416. https://doi.org/10.1006/nimg.2001.0809

Babai R, Amsterdamer A (2008) The persistence of solid and liquid naive conceptions: a reaction time study. J Sci Educ Technol 17:553–559

Babai R, Levyadun T, Stavy R, Tirosh D (2006) Intuitive rules in science and mathematics: a reaction time study. Int J Math Educ Sci Technol 37(8):913–924. https://doi.org/10.1080/00207390600794958

Babai R, Sekal R, Stavy R (2010) Persistence of the intuitive conception of living things in adolescence. J Sci Educ Technol 19:20–26

Babai R, Shalev E, Stavy R (2015) A warning intervention improves students’ ability to overcome intuitive interference. ZDM 47:735–745

Babiloni C, Ferretti A, Del Gratta C, Carducci F, Vecchi F, Romani GL, Rossini PM (2005) Human cortical responses during one-bit delayed-response tasks: an fMRI study. Brain Res Bull 65:383–390. https://doi.org/10.1016/j.brainresbull.2005.01.013

Badre D, Wagner AD (2007) Left ventrolateral prefrontal cortex and the cognitive control of memory. Neuropsychologia 45:2883–2901

Bahar M, Hansell M (2000) The relationship between some psychological factors and their effects on the performance of grid questions and word association tests. Educ Psychol 20:349–364. https://doi.org/10.1080/713663739

Balleine BW, Delgado MR, Hikosaka O (2007) The role of the dorsal striatum in reward and decision-making. J Neurosci 27(31):8161–8165

Banich M, Milham M, Atchley R, Cohen N, Webb A, Wszalek T, Kramer A, Liang Z, Barad V, Gullett D, Shah C (2000) Prefrontal cortex play a predominant role in imposing an attentional 'set': evidence from Fmri. Cogn Brain Res 10:1–9

Barnard JK, Jacobs RL (2007, February) The effects of a near versus far transfer of training approach on trainees' confidence to coach related and unrelated tasks. Paper presented at the international research conference in the Americas of the Academy of Human Resource Development, Indianapolis, IN, USA

Bartolomeo P, deSchotten MT, Chica AB (2012) Brain networks of visuospatial attention and their disruption in visual neglect. Front Hum Neurosci. https://doi.org/10.3389/fnhum.2012.00110

Beauchamp MS, Lee KE, Haxby JV, Martin A (2002) Parallel visual motion processing streams for manipulable objects and human movements. Neuron 34:149–159

Beauchamp MS, Argall BD, Bodurka J, Duyn JH, Martin A (2004) Unraveling multisensory integration: patchy organization within human STS multisensory cortex. Nat Neurosci 7:1190–1192. https://doi.org/10.1038/nn1333

Becker JT, MacAndrew DK, Fiez JA (1999) A comment on the functional localization of the phonological storage subsystem of working memory. Brain Cogn 41:27–38

Berg EA (1948) A simple objective technique for measuring flexibility in thinking. J Gen Psychol 39:15–22. https://doi.org/10.1080/00221309.1948.9918159

Berridge KC (1996) Food reward: brain substrates of wanting and liking. Neurosci Biobehav Rev 20(1):1–25

Berridge KC, Robinson TE (1998) What is the role of dopamine in reward: hedonic impact, reward learning, or incentive salience? Brain Res Rev 28:309–369

Bes B, Sloman S, Lucas CG, Raufaste E (2012) Non-Bayesian inference: causal structure trumps correlation. Cogn Sci 36(7):1178–1203

Best JR, Miller PH (2010) A developmental perspective on executive function. Child Dev 81(6):1641–1660. https://doi.org/10.1111/j.1467-8624.2010.01499.x

Blakemore S-J, Fonlupt P, Pachot-Clouard M, Darmon C, Boyer P, Meltzoff AN, Segebarth C, Decety J (2001) How the brain perceives causality: an event related fMRI study. NeuroReport 12(17):3741–3746

Botvinick M (2007) Conflict monitoring and decision-making: reconciling two perspectives on anterior cingulate function. Cogn Affect Behav Neurosci 7(4):356–366

Botvinick M, Braver T, Barch D, Carter C, Cohen J (2001) Conflict monitoring and cognitive control. Psychol Rev 108:625–652

Botvinick M, Cohen JD, Carter CS (2004) Conflict monitoring and anterior cingulate cortex: an update. Trends Cogn Sci 8(12):539–546

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

Brickhouse NW, Ebert-May D, Wier BA (1989) Scientific literacy: perspectives of school administrators, teachers, students, and scientists from an urban mid-Atlantic community. In: Champagne AB, Lovitts BE, Callinger BJ (eds) This year in school science. Scientific literacy. AAAS, Washington, pp 157–176

Buchsbaum BR, Greer S, Chang WL, Berman KF (2005) Meta-analysis of neuroimaging studies of the wisconsin card-sorting task and component processes. Hum Brain Mapp 25:35–45. https://doi.org/10.1016/j.neulet.2006.05.063

Buchsbaum MS, Buchsbaum BR, Chokron S, Tang C, Wei T-C, Bynea W (2006) Thalamocortical circuits: fMRI assessment of the pulvinar and medial dorsal nucleus in normal volunteers. Neurosci Lett 404:282–287. https://doi.org/10.1016/j.neulet.2006.05.063

Bullock M, Gelman R, Baillargeon R (1982) The development of causal reasoning. In: Friedman W (ed) The developmental psychology of time. Academic Press, New York, pp 209–254

Bush G, Paul J, Whalen PJ, Rosen B, Jenike MA, McInerney SC, Rauch SL (1998) The counting Stroop: an interference task specialized for functional neuroimaging—validation study with functional MRI. Hum Brain Mapp 6:270–282

Bush G, Vogt BA, Holmes J, Dale AM, Greve D, Jenike MA, Rosen BR (2002) Dorsal anterior cingulate cortex: a role in reward-based decision making. Proc Natl Acad Sci USA 99:523–528. https://doi.org/10.1073/pnas.012470999

Bush G, Whalen PG, Shin LM, Rauch SL (2006) The counting Stroop: a cognitive interference task. Nat Protoc 1(1):230–233. https://doi.org/10.1038/nprot.2006.35

Button KS, Ioannidis JPA, Mokrysz C, Nosek BA, Flint J, Robinson ESJ, Munafo MR (2013) Power failure: why small sample size undermines the reliability of neuroscience. Nat Rev Neurosci 14(5):365–376. https://doi.org/10.1038/nrn3475

Carolan TF, Hutchins SD, Wickens CD, Cumming JM (2014) Costs and benefits of more learner freedom: meta-analyses of exploratory and learner control training methods. Hum Factors 56:999–1014. https://doi.org/10.1177/0018720813517710

Casey BJ, Trainor RJ, Orendi JL, Schubert AB, Nystrom LE, Giedd JN, Castellanos FX, Haxby JV, Noll DC, Cohen JD, Forman SD (1997) A developmental functional MRI study of prefrontal activation during performance of a go-no-go task. J Cogn Neurosci 9:835–847

Cavanna AE, Trimble MR (2006) The precuneus: a review of its functional anatomy and behavioural correlates. Brain 129:564–583. https://doi.org/10.1093/brain/awl004

Cepni S, Keles E (2006) Turkish students' conceptions about the simple electric circuits. Int J Sci Math Educ 4(2):269–291

Chang J-Y, Chen L, Luo F, Shi L-H, Woodward DJ (2002) Neuronal responses in the frontal cortico-basal ganglia system during delayed matching-to-sample task: ensemble recording in freely moving rats. Exp Brain Res 142:67–80. https://doi.org/10.1007/s00221-001-0918-3

Chen G, Saad ZS, Britton JC, Pine DS, Cox RW (2013) Linear mixed-effects modeling approach to FMRI group analysis. Neuroimage 73:176–190. https://doi.org/10.1016/j.neuroimage.2013.01.047

Cheng PW (1997) From covariation to causation: a causal power theory. Psychol Rev 104:367–405

Cheng K, Fujita H, Kanno I, Miura S, Tanaka K (1995) Human cortical regions activated by wide-field visual motion: an H2150 PET study. J Neurophysiol 74(1):413–427

Christoff K, Prabhakaran V, Dorfman J, Zhao Z, Kroger JK, Holyoak KJ, Gabrieli JDE (2001) Rostrolateral prefrontal cortex involvement in relational integration during reasoning. NeuroImage 14:1136–1149. https://doi.org/10.1006/nimg.2001.0922

Cooper RP, Yule P (2013) Decision making. In: Cooper RP (ed) Modelling high-level cognitive processes. Psychology Press, London, pp 223–268

Corbetta M, Shulman GL, Miezin FM, Petersen SE (1995) Superior parietal cortex activation during spatial attention shifts and visual feature conjunction. Science 270:802–805

Coull JT, Frackowiak RSJ, Frith CD (1998) Monitoring for target objects: activation of right frontal and parietal cortices with increasing time on task. Neuropsychologia 36(12):1325–1334

Courtney SM, Ungerleider LG, Keil K, Haxby JV (1996) Object and spatial visual working memory activate separate neural systems in human cortex. Cereb Cortex 6:39–49

Cragg L, Gilmore C (2014) Skills underlying mathematics: the role of executive function in the development of mathematics proficiency. Trends Neurosci Educ 3:63–68

D'Angelo C, Rutstein D, Harris C, Bernard R, Borokhovski E, Haertel G (2014) Simulations for STEM learning: systematic review and meta-analysis. SRI International, Menlo Park

Daniel R, Wagner G, Koch K, Reichenbach JR, Sauer H, Schlösser RG (2010) Assessing the neural basis of uncertainty in perceptual category learning through varying levels of distortion. J Cogn Neurosci 23:1781–1793. https://doi.org/10.1162/jocn.2010.21541

Danili E, Reid N (2004) Some strategies to improve performance in school chemistry based on two cognitive factors. Res Sci Technol Educ 22:203–223

Delazer M, Ischebeck A, Domahs F, Zamarian L, Koppelstaetter F, Siedentopf CM (2005) Learning by strategies and learning by drill-evidence from an fMRI study. Neuroimage 25(3):838–849. https://doi.org/10.1016/j.neuroimage.2004.12.009

Dempster FN (1995) Interference and inhibition in cognition: an historical perspective. In: Dempster FN, Brainerd CJ (eds) Interference and inhibition in cognition, pp 3–26. https://doi.org/10.1016/b978-012208930-5/50002-7

Dempster FN, Corkill AJ (1999) Interference and inhibition in cognition and behavior: unifying themes for educational psychology. Educ Psychol Rev 11(1):1–88

Desmond JE, Glover GH (2002) Estimating sample size in functional MRI (fMRI) neuroimaging studies: statistical power analyses. J Neurosci Methods 118(2):115–128

Diamond A (2013) Executive functions. Annu Rev Psychol 64:135–168. https://doi.org/10.1146/annurev-psych-113011-143750

diSessa AA (2017) Conceptual change in a microcosm: comparative learning analysis of a learning event. Hum Dev 60:1–37. https://doi.org/10.1159/000469693

Donaldson D, Petersen S, Ollinger J, Buckner R (2001) Dissociating state and item components of recognition memory using fMRI. Neuroimage 13:129–142

Dove A, Pollman S, Schubert T, Wiggins C, von Cramon D (2000) Prefrontal cortex activation in task switching: an event-related fMRI Study. Cogn Brain Res 9:103–109

Dragos V, Mih V (2015) Scientific literacy in school. Procedia Soc Behav Sci 209:167–172. https://doi.org/10.1016/j.sbspro.2015.11.273

Drouet I (2012) Causes, probabilités, inferences [Causes, probabilities, inferences]. Vuibert, Paris

Dunbar KN, Fugelsang JA, Stein C (2007) Do naïve theories ever go away? Using brain and behavior to understand changes in concepts. In: Lovett MC, Shah P (eds) Thinking with data: 33rd Carnegie symposium on cognition. Lawrence Erlbaum, Mahwah, pp 193–206

Dunst B, Benedek M, Jauk E, Bergner S, Koschutnig K, Sommer M, Ischebeck A, Spinath B, Arendasy M, Bühner M, Freudenthaler H (2014) Neural efficiency as a function of task demands. Intelligence 42:22–30

Eisenberger NI, Lieberman MD, Williams KD (2003) Does rejection hurt? An fMRI study of social exclusion. Science 302:290–292. https://doi.org/10.1126/science.1089134

Ekstrom AD, Bookheimer SY (2007) Spatial and temporal episodic memory retrieval recruit dissociable functional networks in the human brain. Learn Mem 14:645–654

Elliott R, Dolan RJ, Frith CD (2000) Dissociable functions in the middle and lateral orbitofrontal cortex: evidence from human neuroimaging studies. Cereb Cortex 10:308–317

Fair DA, Cohen AL, Power JD, Dosenbach NUF, Church JA, Miezin FM, Schlaggar BL, Petersen SE (2009) Functional brain networks develop from a ''local to distributed'' organization. PLoS Comput Biol 5(5):e1000381. https://doi.org/10.1371/journal.pcbi.1000381

Falk EB, Hyde LW, Mitchell C, Faul J, Gonzalez R, Heitzeg MM, Keating DP, Langa KM, Martz ME, Maslowsky J, Morrison FJ (2013) What is a representative brain? Neuroscience meets population science. Proc Natl Acad Sci USA 110(44):17615–17622. https://doi.org/10.1073/pnas.1310134110

Fonlupt P (2003) Perception and judgement of physical causality involve different brain structures. Cogn Brain Res 17:248–254. https://doi.org/10.1016/S0926-6410(03)001125

Fortin A, Ptito A, Faubert J, Ptito M (2001) Cortical areas mediating stereopsis in the human brain: a PETstudy. NeuroReport 13(67):895–898

Foucher JR, Otzenberger H, Gounot D (2004) Where arousal meets attention: a simultaneous fMRI and EEG recording study. NeuroImage 22:688–697

Friedman NP, 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

Fugelsang JA, Dunbar KN (2005) Brain-based mechanisms underlying complex causal thinking. Neuropsychologia 43:1204–1213. https://doi.org/10.1016/j.neuropsychologia.2004.10.012

Fugelsang JA, Roser ME, Corballis PM, Gazzaniga MS, Dunbar KN (2005) Brain mechanisms underlying perceptual causality. Cogn Brain Res 24:41–47. https://doi.org/10.1016/j.cogbrainres.2004.12.001

Furtak EM, Seidel T, Iverson H, Briggs DC (2012) Experimental and quasiexperimental studies of inquiry-based science teaching: a meta-analysis. Rev Educ Res 82:300–329. https://doi.org/10.3102/0034654312457206

Garavan H, Ross TJ, Murphy KR, Roche AP, Stein EA (2002) Dissociable executive functions in the dynamic control of behavior: inhibition, error detection, and correction. NeuroImage 17:1820–1829. https://doi.org/10.1006/nimg.2002.1326

Gathercole SE, Pickering SJ, Knight C, Stegmann Z (2004) Working memory skills and educational attainment: evidence from national curriculum assessments at 7 and 14 years of age. Appl Cogn Psychol 18:1–16

Goel V, Gold B, Kapur S, Houle S (1997) The seat of reason? An imaging study of deductive and inductive reasoning. NeuroReport 8:1305–1310

Gold BT, Balota DA, Jones SJ, Powell DK, Smith CD, Andersen AH (2006) Dissociation of automatic and strategic lexical-semantics: functional magnetic resonance imaging evidence for differing roles of multiple frontotemporal regions. J Neurosci 26:6523–6532

Grabner RH, Ischebeck A, Reishofer G, Koschutnig K, Delazer M, Ebner F (2009) Fact learning in complex arithmetic and figural–spatial tasks: the role of the angular gyrus and its relation to mathematical competence. Hum Brain Mapp 30(9):2936–2952. https://doi.org/10.1002/hbm.20720

Grahn JA, Parkinson JA, Owen AM (2008) The cognitive functions of the caudate nucleus. Prog Neurobiol 86:141–155. https://doi.org/10.1016/j.pneurobio.2008.09.004

Green AE, Fugelsang JA, Kraemer DJM, Shamosh NA, Dunbar KN (2006) Frontopolar cortex mediates abstract integration in analogy. Brain Res 1096:125–137. https://doi.org/10.1016/j.brainres.2006.04.024

Greenhalgh T, Peacock R (2005) Effectiveness and efficiency of search methods in systematic reviews of complex evidence: audit of primary sources. BMJ 331:1064–1065. https://doi.org/10.1136/bmj.38636.593461.68

Gusnard D, Akbudak E, Shulman G, Raichle M (2001) Medial prefrontal cortex and self-referential mental activity: relation to a default mode of brain function. Proc Natl Acad Sci 98:4259–4264

Haier RJ, Siegel BV, Nuechterlein KH, Hazlett E, Wu JC, Paek J, Browning HL, Buchsbaum MS (1988) Cortical glucose metabolic rate correlates of abstract reasoning and attention studied with positron emission tomography. Intelligence 12:199–217

Haier RJ, Siegel BV Jr, MacLachlan A, Soderling E, Lottenberg S, Buchsbaum MS (1992) Regional glucose metabolic changes after learning a complex visuospatial/motor task: a positron emission tomographic study. Brain Res 570:134–143

Hanakawa T, Honda M, Sawamoto N, Okada T, Yonekura Y, Fukuyama H, Shibasaki H (2002) The role of rostral brodmann area 6 in mental operation tasks: an integrative neuroimaging approach. Cereb Cortex 12:1157–1170

Harris S, Sheth SA, Cohen MS (2008) Functional neuroimaging of belief, disbelief, and uncertainty. Ann Neurol 63:141–147. https://doi.org/10.1002/ana.21301

Haruno M, Kawato M (2006) Different neural correlates of reward expectation and reward expectation error in the putamen and caudate nucleus during stimulus-action-reward association learning. J Neurophysiol 95:948–959

Hayasaka S, Peiffer AM, Hugenschmidt CE, Laurienti PJ (2007) Power and sample size calculation for neuroimaging studies by non-central random field theory. Neuroimage 37(3):721–730

Henrich J, Heine SJ, Norenzayan A (2010) The weirdest people in the world. Behav Brain Sci 33:61–135. https://doi.org/10.1017/S0140525X0999152X

Henry LA, Messer DJ, Nash G (2014) Testing for near and far transfer effects with a short, face-to-face adaptive working memory training intervention in typical children. Infant Child Dev 23:84–103. https://doi.org/10.1002/icd.1816

Henson RNA, Penny WD (2005) ANOVAs and SPM (Technical report). Wellcome Department of Imaging Neuroscience, London, UK

Heston TF, King JM (2017) Predictive power of statistical significance. World J Methodol 7(4):112–116. https://doi.org/10.5662/wjm.v7.i4.112

Hinkle DE, Wiersma W, Jurs SG (2003) Applied statistics for the behavioral sciences, 5th edn. Houfton Mifflin, Boston

Hirshorn EA, Thompson-Schill SL (2006) Role of the left inferior frontal gyrus in covert word retrieval: neural correlates of switching during verbal fluency. Neuropsychologia 44:2547–2557

Holbrook J, Rannikmae M (2009) The meaning of scientific literacy. Int J Environ Sci Educ 4(3):275–288

Holmes J, Gathercole SE, Dunning DL (2009) Adaptive training leads to sustained enhancement of poor working memory in children. Dev Sci 12(4):F9–F15. https://doi.org/10.1111/j.1467-7687.2009.00848.x

Hosseini SM, Rostami M, Yomogida Y, Takahashi M, Tsukiura T, Kawashima R (2010) Aging and decision making under uncertainty: behavioral and neural evidence for the preservation of decision making in the absence of learning in old age. Neuroimage 52:1514–1520. https://doi.org/10.1016/j.neuroimage.2010.05.008

Houdé O, Borst G (2014) Measuring inhibitory control in children and adults: brain imaging and mental chronometry. Front Psychol 5(616):1–7. https://doi.org/10.3389/fpsyg.2014.00616

Houde O, Moutier S (1996) Deductive reasoning and experimental inhibition training: the case of the matching bias. Curr Psychol Cogn 15:409–434

Houde O, Moutier S (1999) Deductive reasoning and experimental inhibition training: the case of the matching bias. New data and reply to Girotto. Curr Psychol Cogn 18:75–85

Houdé O, Zago L, Crivello F, Moutier S, Pineau A, Mazoyer B, Tzourio-Mazoyer N (2001) Access to deductive logic depends on a right ventromedial prefrontal area devoted to emotion and feeling: evidence from a training paradigm. NeuroImage 14:1486–1492. https://doi.org/10.1006/nimg.2001.0930

Howell DC (2002) Statistical methods for psychology, 5th edn. Duxbury, Pacific Grove

Hunter SK, Kisley MA, McCarthy L, Freedman R, Rossi RG (2011) Diminished cerebral inhibition in neonates associated with risk factors for schizophrenia: parental psychosis, maternal depression, and nicotine use. Schizophr Bull 37(6):1200–1208

Hushman CJ, Marley SC (2015) Guided instruction improves elementary student learning and self-efficacy in science. J Educ Res 108(5):371–381. https://doi.org/10.1080/00220671.2014.899958

Hutzler F (2013) Reverse inference is not a fallacy per se: cognitive processes can be inferred from functional imaging data. NeuroImage. https://doi.org/10.1016/j.neuroimage.2012.12.075

Jarvis HL, Gathercole SE (2003) Verbal and nonverbal working memory and achievements on national curriculum tests at 11 and 14 years of age. Educ Child Psychol 20:123–140

Kanwisher N (2010) Functional specificity in the human brain: a window into the functional architecture of the mind. Proc Natl Acad Sci 107(25):11163–11170

Kappel V, Lorenz RC, Streifling M, Renneberg B, Lehmkuhl U, Ströhle A, Salbach-Andrae H, Beck A (2015) Effect of brain structure and function on reward anticipation in children and adults with attention deficit hyperactivity disorder combined subtype. Soc Cogn Affect Neurosci 10:945–951. https://doi.org/10.1093/scan/nsu135

Karbach J, Strobach T, Schubert T (2015) Adaptive working memory training benefits reading, but not mathematics in middle childhood. Child Neuropsychol 21:285–301. https://doi.org/10.1080/09297049.2014.899336

Kelemen D, Rosset E (2009) The human function compunction: teleological explanation in adults. Cognition 111(1):138–143

Kelemen D, Rottman J, Seston R (2013) Professional physical scientists display tenacious teleological tendencies: purpose-based reasoning as a cognitive default. J Exp Psychol 142(4):1074–1083. https://doi.org/10.1037/a0030399

Kirschner PA, Sweller J, Clark RE (2006) Why minimal guidance during instruction does not work: an analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. Educ Psychol 41:75–86. https://doi.org/10.1207/s15326985ep4102_1

Kohler S, Black SE, Sinden M, Szekely C, Kidron D, Parker JL, Foster JK, Moscovitch M, Wincour G, Szalai JP, Bronskill MJ (1998) Memory impairments associated with hippocampal versus parahippocampal gyrus atrophy: an MR volumetry study in Alzheimer's disease. Neuropsychologia 25(8):901–914

Kording KP, Beierholm U, Ma WJ, Quartz S, Tenenbaum JB, Shams L (2007) Causal inference in multisensory perception. PLoS ONE 2(9):e943. https://doi.org/10.1371/journal.pone.0000943

Kroesbergen EH, van't Noordende JE, Kolkman ME (2014) Training working memory in kindergarten children: effects on working memory and early numeracy. Child Neuropsychol 20(1):23–37. https://doi.org/10.1080/09297049.2012.736483

Kroger JK, Sabb FW, Fales CL, Bookheimer SY, Cohen MS, Holyoak KJ (2002) Recruitement of anterior dorsolateral prefrontal cortex in human reasoning: a parametric study of relational complexity. Cereb Cortex 12:477–485

Kuperberg GR, Lakshmanan BM, Caplan DN, Holcomb PJ (2006) Making sense of discourse: an fMRI study of causal inferencing across sentences. Neuroimage 33:343–361

Kwon Y-J, Lawson AE (2000) Linking brain growth with the development of scientific reasoning ability and conceptual change during adolescence. J Res Sci Teach 37:44–62

Kwon Y, Jeong J, Park Y (2006) Roles of abductive reasoning and prior belief in children's generation of hypotheses about pendulum motion. Sci Educ 15:643–656

Kwon Y-J, Lee J-K, Shin D-H, Jeong J-S (2009) Changes in brain activation induced by the training of hypothesis generation skills: an fMRI study. Brain Cogn 69:391–397

Laetsch WM (1987) A basis for better understanding of science. In: Evered D, O'Connor M (eds) Communicating science to the public. Wiley, London, pp 1–10

Laker DR (1990) Dual dimensionality of training transfer. Hum Resour Dev Q 1(3):209–224

Lamm C, Windischberger C, Leodolter U, Moser E, Bauer H (2001) Evidence for premotor cortex activity during dynamic visuospatial imagery from single-trial functional magnetic resonance imaging and event-related slow cortical potentials. NeuroImage 14:268–283. https://doi.org/10.1006/nimg.2001.0850

Lappi O, Rusanen AM (2011) Turing machines and causal mechanisms in cognitive sciences. In: McKay Illari P, Russo F, Williamson J (eds) Causality in the sciences. Oxford University Press, Oxford, pp 224–239

Latzman RD, Elkovitch N, Young J, Clark LA (2010) The contribution of executive functioning to academic achievement among male adolescents. J Clin Exp Neuropsychol 32(5):455–462. https://doi.org/10.1080/13803390903164363

Laugksch RC (1998) Scientific literacy: a conceptual overview. Sci Educ 84(1):71–94. https://doi.org/10.1002/(SICI)1098-237X(200001)

Lawson AE (1978) The development and validation of a classroom test of formal reasoning. J Res Sci Teach 15(1):11–24

Lazonder AM, Harmsen R (2016) Meta-analysis of inquiry-based learning: effects of guidance. Rev Educ Res 86(3):681–718. https://doi.org/10.3102/0034654315627366

Lee JK (2009) Dissociation of the brain activation network associated with hypothesis-generating and hypothesis-understanding in biology learning: evidence from an fMRI study (Unpublished doctoral dissertation). Korea National University of Education, Cheongwon

Lee SE (2014) The impact of working memory training on third grade students' reading fluency and reading comprehension performance (Doctoral dissertation). Southern Illinois University, Carbondale, IL

Lee JK, Kwon YJ (2008) Types of emotion during the hypothesis-generating and hypothesis-understanding process on the biological phenomena. Second Educ Res 56(3):1–36

Lee J-K, Kwon Y-J (2011) Why traditional expository teaching–learning approaches may founder? An experimental examination of neural networks in biology learning. J Biol Educ 45(2):83–92. https://doi.org/10.1080/00219266.2010.548874

Lee J-K, Kwon Y-J (2012) Learning-related changes in adolescents' neural networks during hypothesis-generating and hypothesis-understanding training. Sci Educ 21:1–31. https://doi.org/10.1007/s11191-010-9313-4

Levy BJ, Wagner AD (2011) Cognitive control and right ventrolateral prefrontal cortex: reflexive reorienting, motor inhibition, and action updating. Ann N Y Acad Sci 1224(1):40–62. https://doi.org/10.1111/j.1749-6632.2011.05958.x

Livelli A, Orofino GC, Calcagno A, Farenga M, Penoncelli D, Guastavigna M, Carosella S, Caramello P, Pia L (2015) Evaluation of a cognitive rehabilitation protocol in HIV patients with associated neurocognitive disorders: efficacy and stability over time. Front Cogn Neurosci 9(306):1–10. https://doi.org/10.3389/fnbeh.2015.00306

Loosli SV, Buschkuehl M, Perrig WJ, Jaeggi SM (2012) Working memory training improves reading processes in typically developing children. Child Neuropsychol 18:62–78. https://doi.org/10.1080/09297049.2011.575772

MacDonald AW, Cohen JD, Stenger VA, Carter CS (2000) Dissociating the role of dorsolateral prefrontal and anterior cingulate cortex in cognitive control. Science 288:1835–1838

Mahayana IT, Tcheang L, Chen C-Y, Juan C-H, Muggleton NG (2014) The precuneus and visuospatial attention in near and far space: a transcranial magnetic stimulation study. Brain Stimul 7:673–679. https://doi.org/10.1016/j.brs.2014.06.012

Majerus S, Poncelet M, Van der Linden M, Albouy G, Salmon E, Sterpenich V, Vandewalle G, Collette F, Maquet P (2006) The left intraparietal sulcus and verbal short-term memory: focus of attention or serial order? NeuroImage 32:880–891. https://doi.org/10.1016/j.neuroimage.2006.03.048

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 Educ 8(1):37–48

Mateen FJ, Oh J, Tergas AI, Bhayani NH, Kamdar BB (2013) Titles versus titles and abstracts for initial screening of articles for systematic reviews. Clin Epidemiol 5:89–95. https://doi.org/10.2147/CLEP.S43118

Mayer R (2004) Should there be a three-strike rule against pure discovery learning? The case for guided methods of instruction. Am Psychol 59:14–19. https://doi.org/10.1037/0003-066X.59.1.14

McGann M (2010) Perceptual modalities: modes of presentation or modes of interaction? J Conscious Stud 17(1–2):72–94

McGuire WJ (1997) Creative hypothesis generating in psychology: some useful heuristics. Annu Rev Psychol 48:1–30

McKay Illari P, Russo F, Williamson J (2011) Causality in the sciences. Oxford University Press, New York

Melby-Lervåg M, Redick TS, Hulme C (2016) Working memory training does not improve performance on measures of intelligence or other measures of "far-transfer": evidence from a meta-analytic review. Perspect Psychol Sci 11:512–534. https://doi.org/10.1177/1745691616635612

Mendelson R, Shultz TR (1976) Covariation and temporal contiguity as principles of causal inference in young children. J Exp Child Psychol 22(3):408–412

Menon V, Adleman NE, White CD, Glover GH, Reiss AL (2001) Error-related brain activation during a Go/No Go response inhibition task. Hum Brain Mapp 12(3):131–143

Mizuno K, Tanaka M, Ishii A, Tanabe HC, Onoe H, Sadato N, Watanabe Y (2008) The neural basis of academic achievement motivation. NeuroImage 42:369–378

Monchi O, Petrides M, Petre V, Worsley K, Dagher A (2001) Wisconsin card sorting revisited: distinct neural circuits participating in different stages of the task identified by event-related functional magnetic resonance imaging. J Neurosci 21(19):7733–7741

Monchi O, Petrides M, Strafella AP, Worsley KJ, Doyon J (2006) Functional role of the basal ganglia in the planning and execution of actions. Ann Neurol 59:257–264. https://doi.org/10.1002/ana.2074

Moss HE, Abdallah S, Fletcher P, Bright P, Pilgrim L, Acres K, Tyler LK (2005) Selecting among competing alternatives: selecting and retrieval in the left inferior frontal gyrus. Cereb Cortex 15:1723–1735

Moutier S, Angeard N, Houdé O (2002) Deductive reasoning and matching-bias inhibition training: evidence from a debiasing paradigm. Think Reason 8:205–224

Murphy K, Garavan H (2004) An empirical investigation into the number of subjects required for an event-related fMRI study. NeuroImage 22(2):879–885

National Research Council (2005) America's lab report: investigations in high school science. National Academies Press, Washington

Neubauer AC, Fink A (2009) Intelligence and neural efficiency: measures of brain activation versus measures of functional connectivity in the brain. Intelligence 37:223–229. https://doi.org/10.1016/j.intell.2008.10.008

Neubauer AC, Fink A, Schrausser DG (2002) Intelligence and neural efficiency: the influence of task content and sex on the brain–IQ relationship. Intelligence 30(6):515–536

Nevo E, Breznitz Z (2014) Effects of working memory and reading acceleration training on improving working memory abilities and reading skills among third graders. Child Neuropsychol 20:752–765. https://doi.org/10.1080/09297049.2013.863272

Nichols TE, Das S, Eickhoff SB, Evans AC, Glatard T, Hanke M, Kriegeskorte N, Milham MP, Poldrack RA, Poline JB, Proal E (2016) Best practices in data analysis and sharing in neuroimaging using MRI (Report No. bioRxiv). https://doi.org/10.1101/054262

Nobre AC, Sebestyen GN, Gitelman DR, Mesulam MM, Frackowiak RSJ, Frith CD (1997) Functional localization of the system for visuospatial attention using positron emission tomography. Brain 120:515–533

Novick LR, Cheng PW (2004) Assessing interactive causal influence. Psychol Rev 111(2):455–485. https://doi.org/10.1037/0033-295X.111.2.455

OECD (2007) Understanding the brain: the birth of a learning science. OECD Editions, Paris

OECD (2017) Education at a glance 2017: OECD indicators. OECD Publishing, Paris. https://doi.org/10.1787/eag-2017-en

Ogunkola BJ (2013) Scientific literacy: conceptual overview, importance and strategies for improvement. J Educ Soc Res 3(1):265–274. https://doi.org/10.5901/jesr.2013.v3n1p265

Olesen PJ, Westerberg H, Klingberg T (2004) Increased prefrontal and parietal activity after training of working memory. Nat Neurosci 7:75–79

Owen AM, Milner B, Petrides M, Evans AC (1996) A specific role for the right parahippocampal gyrus in the retrieval of object-location: a positron emission tomography study. J Cogn Neurosci 8(6):588–602

Paas F, Renkl A, Sweller J (2003) Cognitive load theory and instructional design: recent developments. Educ Psychol 38:1–4

Patterson R, Barbey AK (2005) A multiple systems approach to causal reasoning. In: Grafman J, Krueger F (eds) Neural basis of belief systems. Psychology Press, New York, pp 43–70

Paulus MP, Rogalsky C, Simmons A, Feinstein JS, Stein MB (2003) Increased activation in the right insula during risk-taking decision making is related to harm avoidance and neuroticism. Neuroimage 19:1439–1448

Pinal GD, Nathan MJ (2017) Two kinds of reverse inference in cognitive neuroscience. In: Leefmann J, Hildt E (eds) The human sciences after the decade of the brain. Academic Press, London, pp 121–139

Pleskac TJ, Dougherty MR, Busemeyer J, Risekamp J, Tenenbaum J (2007) Cognitive decision theory: developing models of real-world decision behavior. Proc Annu Meet Cogn Sci Soc USA 29(29):39–40. https://escholarship.org/uc/item/2hh7462x

Poldrack RA (2006) Can cognitive processes be inferred from neuroimaging data? Trends Cogn Sci 10(2):59–63

Poldrack RA (2011) Inferring mental states from neuroimaging data: from reverse inference to large-scale decoding. Neuron 72(5):692–697. https://doi.org/10.1016/j.neuron.2011.11.001

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. J Res Sci Teach 54(9):1121–1142. https://doi.org/10.1002/tea.21396

Potvin P, Turmel E, 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. Front Hum Neurosci. https://doi.org/10.3389/fnhum.2014.00014

Prabhakaran V, Narayanan K, Zhao Z, Gabrieli J (2000) Integration of diverse information in working memory within the frontal lobe. Nat Neurosci 3:85–90

Ravizza SM, Delgado MR, Chein JM, Becker JT, Fiez JA (2004) Functional dissociations within the inferior parietal cortex in verbal working memory. NeuroImage 22:562–573. https://doi.org/10.1016/j.neuroimage.2004.01.039

Ray E, Schlottmann A (2007) The perception of social and mechanical causality in young children with ASD. Res Autism Spectr Disord 1:266–280. https://doi.org/10.1016/j.rasd.2006.11.002

Redcay E (2008) The superior temporal sulcus performs a common function for social and speech perception: implications for the emergence of autism. Neurosci Biobehav Rev 32(1):123–142. https://doi.org/10.1016/j.neubiorv.2007.06.004

Rhodes SM, Booth JN, Campbell LE, Blythe RA, Wheate NJ, Delibegovic M (2014) Evidence for a role of executive functions in learning biology. Infant Child Dev 23(1):67–83. https://doi.org/10.1002/icd.1823

Rhodes SM, Booth JN, Palmer LE, Blythe RA, Delibegovic M, Wheate NJ (2016) Executive functions predict conceptual learning of Science. Br J Dev Psychol 34:261–275

Roser ME, Fugelsang JA, Dunbar KN, Corballis PM, Gannaziga MS (2005) Dissociating processes supporting causal perception and causal inference in the brain. Neuropsychology 19(5):591–602

Rusanen A-M (2014) Towards to an explanation for conceptual change: a mechanistic alternative. Sci Educ 23(7):1413–1425. https://doi.org/10.1007/s11191-013-9656-8

Sala G, Gobet F (2017a) Does far transfer exist? Negative evidence from chess, music, and working memory training. Curr Dir Psychol Sci 26(6):515–520. https://doi.org/10.1177/0963721417712760

Sala G, Gobet F (2017b) Working memory training in typically developing children: a meta-analysis of the available evidence. Dev Psychol 53(4):671–685. https://doi.org/10.1037/dev0000265

Sarter M, Berntson GG, Cacioppo JT (1996) Brain imaging and cognitive neuroscience. Toward strong inference in attributing function to structure. Am Psychol 51(1):13–21

Sawamoto N, Honda M, Okada T, Hanakawa T, Kanda M, Fukuyama H, Konishi J, Shibasaki H (2000) Expectation of pain enhances responses to nonpainful somatosensory stimulation in the anterior cingulate cortex and parietal operculum/posterior insula: an event-related functional magnetic resonance imaging Study. J Neurosci 20(19):7438–7445

Schlottmann A, Shanks DR (1992) Evidence for a distinction between judged and perceived causality. Q J Exp Psychol Hum Exp Psychol 44(A):321–342

Scholl BJ, Nakayama K (2002) Causal capture: contextual effects on the perception of collision events. Psychol Sci 13(6):493–498

Sefcsik T, Nemeth D, Janacsek K, Hoffmann I, Scialabba J, Klivenyi P, Gergely GA, Haden G, Vecsei L (2009) The role of the putamen in cognitive functions—a case study. Learn Percept 1(2):215–227. https://doi.org/10.1556/LP.1.2009.2

Seghier ML (2013) The angular gyrus: multiple functions and multiple subdivisions. Neuroscientist 19(1):43–61. https://doi.org/10.1177/1073858412440596

Seltman AJ (2015) Experimental design and analysis. http://www.stat.cmu.edu/~hseltman/309/Book/Book.pdf

Shah P, Michal A, Ibrahim A, Rhodes R, Rodriguez F (2017) What makes everyday scientific reasoning so shallenging?

Psychol Learn Motiv 66:251–299. https://doi.org/10.1016/bs.plm.2016.11.006

Shaywitz BA, Shaywitz SE, Pugh KR, Mencl WE, Fulbright RK, Skudlarski P, Constable RT, Marchione KE, Fletcher JM, Lyon GR, Gore JC (2002) Disruption of posterior brain systems for reading in children with developmental dyslexia. Biol Psychiatry 52:101–110

Shaywitz BA, Skudlarski P, Holahan JM, Marchione KE, Constable RT, Fulbright RK, Zelterman D, Lacadie C, Shaywitz SE (2007) Age-related changes in reading systems of dyslexic children. Ann Neurol 61:363–370

Shtulman A, Harrington K (2015) Tensions between science and intuition across the lifespan. Top Cogn Sci 8:118–137. https://doi.org/10.1111/tops.12174

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

Shulman L, Keisler E (1966) Learning by discovery: a critical appraisal. Rand McNally, Chicago

Shultz TR, Altmann E, Asselin J (1986) Judging causal priority. Br J Dev Psychol 4:67–74

Simon O, Mangin J-F, Cohen L, Le Bihan D, Dehaene S (2002) Topographical layout of hand, eye, calculation, and language-related areas in the human parietal lobe. Neuron 33:475–487

Simons DJ, Boot WR, Charness N, Gathercole SE, Chabris CF, Hambrick DZ, Stine-Morrow EAL (2016) Do "brain-training" programs work? Psychol Sci Public Interest 17:103–186. https://doi.org/10.1177/1529100616661983

Singh-Curry V, Husain M (2009) The functional role of the inferior parietal lobe in the dorsal and ventral stream dichotomy. Neuropsychologia 47:1434–1448. https://doi.org/10.1016/j.neuropsychologia.2008.11.033

Smith EE, Jiondes J (1997) Working memory: a view from neuroimaging. Cogn Psychol 1:5–42

Snyder HR, Hutchison N, Nyhus E, Curran T, Banich MT, O'Reilly RC, Munakata Y (2010) Neural inhibition enables selection during language processing. Proc Natl Acad Sci USA 107(38):16483–16488

Spaniol J, Davidson PSR, Kim ASN, Han H, Moscovitch M, Grady CL (2009) Event-related fMRI studies of episodic encoding and retrieval: meta-analyses using activation likelihood estimation. Neuropsychologia 47:1765–1779. https://doi.org/10.1016/j.neuropsychologia.2009.02.028

St Clair-Thompson HL, Gathercole SE (2006) Executive functions and achievement in school: shifting, updating, inhibition and working memory. Q J Exp Psychol 59(4):745–759. https://doi.org/10.1080/17470210500162854

St Clair-Thompson HL, Stevens R, Hunt A, Bolder E (2010) Improving children's working memory and classroom performance. Educ Psychol 30(2):203–219. https://doi.org/10.1080/01443410903509259

St Clair-Thompson HL, Overton T, Bugler M (2012) Mental capacity and working memory in chemistry: algorithmic versus open-ended problem solving. Chem Educ Res Pract 13(4):484–489

Stavy R, Babai R (2010) Overcoming intuitive interference in mathematics: insights from behavioral, brain imaging and intervention studies. ZDM 42:621–633. https://doi.org/10.1007/s11858-010-0251-z

Stavy R, Babai R, Tsamir P, Tirosh D, Lai-Lin F, McRobbie C (2006) Are intuitive rules universal? Int J Sci Math Educ 4:417–436

Stern ER, Gonzalez R, Welsh RC, Taylor SF (2010) Updating beliefs for a decision: neural correlates of uncertainty and underconfidence. J Neurosci 30:8032–8041. https://doi.org/10.1523/JNEUROSCI.4729-09.2010

Sternberg RJ (2008) 'g', g's, or jeez: which is the best model for developing abilities, competencies, and expertise? In: Kyllonen PC, Roberts RD, Stankov L (eds) Extending intelligence: enhancement and new constructs. Taylor & Francis, London, pp 225–266

Studer-Luethi B, Bauer C, Perrig WJ (2016) Working memory training in children: effectiveness depends on temperament. Mem Cogn 44:171–186. https://doi.org/10.3758/s13421-015-0548-9

Sweller J (1999) Instructional design in technical areas. ACER Press, Camberwell

Tanaka SC, Honda M, Sadato N (2005) Modality-specific cognitive function of medial and lateral human Brodmann area 6. J Neurosci 25(2):496–501. https://doi.org/10.1523/JNEUROSCI.4324-04.2005

Tanaka SC, Samejima K, Okada G, Ueda K, Okamoto Y, Yamawaki S, Doya K (2006) Brain mechanism of reward prediction under predictable and unpredictable environmental dynamics. Neural Netw 19:1233–1241

Teixeira-Dias JJC, Pedrosa de Jesus MH, Neri de Souza FN, Watts M (2005) Teaching for quality learning in chemistry. Int J Sci Educ 27(9):1123–1137. https://doi.org/10.1080/09500690500102813

The Royal Society (2011) Neuroscience: implications for education and lifelong learning. The Royal Society, London

Thompson-Schill SL, D'Esposito M, Aguirre GK, Farah MJ (1997) Role of left inferior prefrontal cortex in retrieval of semantic knowledge: a reevaluation. Proc Natl Acad Sci USA 94:14792–14797

Treagust DF, Duit R (2008) Conceptual change: a discussion of theoretical, methodological and practical challenges for science education. Cult Sci Edu 3(2):297–328

Tuovinen JE, Sweller J (1999) A comparison of cognitive load associated with discovery learning and worked examples. J Educ Psychol 91:334–341

Ueno A, Abe N, Suzuki M, Shigemune Y, Hirayama K, Mori E, Tashiro M, Itoh M, Fujii T (2009) Reactivation of medial temporal lobe and human V5/MT + during the retrieval of motion information: a PET study. Brain Res 1285:127–134

UNESCO (2010) UNESCO science report 2010: the current status of science around the world. http://unesdoc.unesco.org/images/0018/001899/189958e.pdf

UNESCO (2013) Educational neuroscience: more problems than promise?. UNESCO Bangkok, Bangkok

van Duijvenvoorde AC, Zanolie K, Rombouts SA, Raijmakers ME, Crone EA (2008) Evaluating the negative or valuing the positive? Neural mechanisms supporting feedback-based learning across development. J Neurosci 28(38):9495–9503. https://doi.org/10.1523/JNEUROSCI.1485-08.2008

van Veen V, Carter CS (2002) The anterior cingulate as a conflict monitor: fMRI and ERP studies. Physiol Behav 77:477–482

van Zee EH (2000) Analysis of a student-generated inquiry discussion. Int J Sci Educ 22(2):115–142. https://doi.org/10.1080/095006900289912

Volz KG, Schubotz RI, von Cramon DY (2004) Why am I unsure? Internal and external attributions of uncertainty dissociated by fMRI. Neuroimage 21:848–857. https://doi.org/10.1016/j.neuroimage.2003.10.028

Volz KG, Schubotz RI, von Cramon DY (2005) Variants of uncertainty in decision-making and their neural correlates. Brain Res Bull 67:403–412. https://doi.org/10.1016/j.brainresbull.2005.06.011

Waberski TD, Gobbele R, Lamberty K, Buchner H, Marshall JC, Fink GR (2008) Timing of visuo-spatial information processing: electrical source imaging related to line bisection judgements. Neuropsychologia 46:1201–1210. https://doi.org/10.1016/j.neuropsychologia.2007.10.024

Wandersee JH, Mintzes JJ, Novak JD (1994) Research on alternative conceptions in science. In: Gabel DL (ed) Handbook of research on science teaching and learning. MacMillan, New York, pp 177–210

Ward BD, Chen G (2006) Analysis of variance for fMRI data (Technical report). https://afni.nimh.nih.gov/afni/doc/manual/ANOVAm.pdf

Yoncheva YN, Blau VC, Maurer U, McCandliss BD (2010) Attentional focus during learning impacts N170 ERP responses to an artificial script. Dev Neuropsychol 35(4):423–445. https://doi.org/10.1080/87565641.2010.480918

Yue X, Vessel EA, Biederman I (2007) The neural basis of scene preferences. NeuroReport 18:525–529

Zavrel E, Sharpsteen E (2016) How the television show “MythBusters” communicates the scientific method. Phys Teach 54:228–232. https://doi.org/10.1119/1.4944364

Zeki S (2015) Area V5-a microcosm of the visual brain. Front Integr Neurosci. https://doi.org/10.3389/fnint.2015.0002