

Distinct contributions of the caudate nucleus, rostral prefrontal cortex, and parietal cortex to the execution of instructed tasks

Andrea Stocco · Christian Lebiere ·
Randall C. O'Reilly · John R. Anderson

© Psychonomic Society, Inc. 2012

Abstract When we behave according to rules and instructions, our brains interpret abstract representations of what to do and transform them into actual behavior. In order to investigate the neural mechanisms behind this process, we devised an fMRI experiment that explicitly isolated rule interpretation from rule encoding and execution. Our results showed that a specific network of regions (including the left rostral prefrontal cortex, the caudate nucleus, and the bilateral posterior parietal cortices) is responsible for translating rules into executable form. An analysis of activation patterns across conditions revealed that the posterior parietal cortices represent a mental template for the task to perform, that the inferior parietal gyrus and the caudate nucleus are responsible for instantiating the template in the proper context, and that the left rostral prefrontal cortex integrates information across complex relationships.

Keywords Basal ganglia · Rostral prefrontal cortex · Parietal cortex · Cognitive control · fMRI · Rules · Learning · Instructions

Learning from instructions is such a common activity that its complexity can be easily underestimated. Instead, it represents a remarkable feat of human cognition. Consider, for

example, the case a scientist instructing a participant to perform the Stroop task. The scientist's instructions would probably sound like this: "Words will appear on the screen, one at the time, printed in different colors. For each word, you have to say out loud the name of the color the word is printed in, and ignore the word." To perform this task, participants need to create an internal "template" of the task to perform—that is, an internal representation that connects what features of the stimuli to attend to (the word's color, not its meaning), what cognitive operations to perform (identify the color name) and how to respond (say it out loud). When the task is executed, this template needs to be internally scanned and translated into actual behavior. This translation is a computationally sophisticated procedure where specific stimulus features (e.g., the color green) need to be put in specific placeholders within the template ("the words' color"), and abstract commands (e.g., "name the color") need to be transformed into sequences of basic mental operations ("retrieve from long-term memory the name of the color" and "say the color name out loud").

The process of reconfiguring one's own behavior from instructions is reminiscent of what computers do when executing a program written in a high-level programming language. The program provides a template for the operations that need to be performed, and when the program is interpreted, this template is transformed into actual operations of the underlying hardware, which are then applied to the input given by the user. This analogy is not without merit. Like programming computers makes them capable of performing any computational function, executing a task from a mental template gives us an unbounded flexibility of behavior. Thus, a characterization of the process of interpreting instructions is important for understanding the nature of the power and generality of human cognition at a computational level.

In summary, the capacity to carry out new tasks from instructions is a remarkable feature of human cognition that

A. Stocco (✉)
Institute for Learning and Brain Sciences and Department of
Psychology, University of Washington,
Portage Bay Bldg, Campus Box 357988, Seattle, WA 98915, USA
e-mail: stocco@uw.edu

C. Lebiere · J. R. Anderson
Department of Psychology, Carnegie Mellon University,
Pittsburgh, PA, USA

R. C. O'Reilly
Department of Psychology, University of Colorado,
Boulder, CO, USA

stills remains poorly understood. This article presents an experiment designed to identify the network regions involved in interpreting instructions and to provide preliminary characterizations of the specific computations supported by each region.

The cognitive neuroscience of learning from instructions

Studying the mechanism by which people learn to perform new tasks from instructions is experimentally challenging. In typical cognitive neuroscience experiments, participants are given instructions at the very beginning and then apply these instructions over a large number of consecutive trials. However, the more that the same task is practiced, the more it becomes automatic, and the less it depends on voluntary access to internal task representations, such as the “mental template” (Logan, 1988; Schneider & Chein, 2003; Schneider & Shiffrin, 1977). For this reason, traditional paradigms are not well suited to study the process by which instructions are interpreted. To overcome this difficulty, researchers have devised paradigms in which many different instructions (corresponding to different “tasks”) are given during the course of experiments (Brass, Wenke, Spengler, & Waszak, 2009; Cole, Bagic, Kass, & Schneider, 2010; Hartstra, Kühn, Verguts, & Brass, 2011; Ruge & Wolfensteller, 2010; Stocco, Lebiere, O’Reilly et al., 2010). In these paradigms, a new “task” is introduced by presenting its specific instructions on the screen, followed by a series of one or more stimuli on which the new task is applied. Thus, the execution of each task is divided into two consecutive phases: an initial phase in which the instructions are displayed on the screen and encoded (i.e., when the mental template is created), and a later phase in which the stimuli are displayed and the instructions are executed. To collect multiple observations during the experiment, paradigms employ a large number of comparable “tasks,” generated by systematically varying the structure and nature of its component operations. For instance, in the experiment by Hartstra et al. (2011), all of the tasks consisted of responding with fingerpresses to specific classes of visual stimuli, and different tasks were generated by varying the stimuli and the fingers used to respond. Thus, one task could be “respond with the index finger to cars, and with the middle finger to animals,” and a different task could be “respond with the middle finger to plants, and with the index finger to houses.” In our experiment, different tasks were generated by combining arithmetic operations into sets of three, such as “add 1 to x , divide y by 2, and sum the results” or “multiply x by 2, sum the result to y , and add 1 to the result.”

Regions involved in encoding instructions

In these types of paradigms, the brain regions involved in encoding instructions (i.e., creating the mental “template” to execute a specific task) can be inferred by comparing the brain activity during the encoding of novel instructions and during the encoding of practiced instructions—that is, instructions that have been already been encoded. Both Cole et al. (2010) and Hartstra et al. (2011) found that novel instructions activate regions in the lateral prefrontal cortex (LPFC) as well as parietal regions such as the inferior parietal sulcus. Cole et al. (2010) also reported that different parts of the LPFC were activated at different levels of practice. In particular, novel instructions elicited activity in the dorsal part of LPFC, while practiced instructions (e.g., instructions that had already been encoded) elicited activity in the more anterior rostral LPFC. These two regions have different cognitive functions: The rostral LPFC is responsible for integrating information in complex relationships (Bunge, Helskog, & Wendelken, 2009; Bunge, Wendelken, Badre, & Wagner, 2005; Wendelken, Nakhabenko, Donohue, Carter, & Bunge, 2008), while the dorsal LPFC is responsible for the maintenance of information in working memory (Cohen et al., 1997). On the basis of these established findings, Cole et al. (2010) interpreted their results in the following way: Initially, instructions are simply rehearsed in working memory, and these rehearsal operations increase the activation of dorsal LPFC. With practice, instructions are transformed into more specialized representations of the various terms and actions to be performed. This representation is what we have previously called a “mental template,” and what Ruge and Wolfensteller (2010) have named “pragmatic representations.” Cole et al. (2010) argued that, when such a template has already been established for a task, it can be accessed directly when encoding instructions. This access increases the activity of the rostral LPFC, but also reduces the need to rehearse instructions in verbal form, in turn decreasing the activity of the dorsal LPFC.

In summary, previous research has suggested that the lateral prefrontal and posterior parietal regions are actively engaged when encoding instructions for new tasks. Within the prefrontal cortex, both dorsal and rostral regions have been reported, likely reflecting the complementary functional processes of rehearsing instructions in working memory and developing an integrated, relational representation of the various components of the task.

Regions involved in interpreting instructions

The same paradigms can also be used to examine how new instructions are executed—that is, how the mental template

for a new task is applied to the stimuli. We will refer to this process as “interpretation,” to highlight its similarity to the process by which source code in a programming language is executed on a computer. As in the case of the source code, the interpretation process consists of translating abstract commands (e.g., “say the color out loud”) into the series of basic cognitive operations that can be performed by the biological hardware (“Retrieve from long-term memory the name of the color” and “Articulate the name”), and substituting variables or placeholders (“the word’s *color*”) with their proper stimulus values (e.g., the color green).

Tasks that have been practiced many times tend to become more automatic. When tasks are automatic, longer sequences of actions can be chunked together as a single procedure, thus requiring less effort and fewer accesses to the internal mental template (Chein & Schneider, 2005; Elio, 1986). For this reason, the brain regions involved in interpreting instructions can be observed by comparing the activity of the brain when novel instructions are applied to the experimental stimuli for the first time against the brain activity occurring during subsequent applications of the same instructions. Several authors have found that regions in the LPFC were more active during the execution of novel versus practiced instructions (Brass et al., 2009; Cole et al., 2010; Hartstra et al., 2011; Ruge & Wolfensteller, 2010). Both Ruge and Wolfensteller (2010) and Cole et al. (2010) reported that those regions in the LPFC that were active during the first execution of novel instructions were also located more anterior than those regions that were active for the execution of subsequent trials. In particular, Cole et al. (2010) found that the time courses of activation in the rostral and dorsal LPFC were the opposite of what was observed for the encoding of instructions. Thus, during the execution of novel instructions, rostral LPFC was more active than dorsal LPFC, but the dorsal LPFC was more active than the rostral LPFC when executing instructions that had been practiced before.

Ruge and Wolfensteller (2010) reported that, in addition to the lateral PFC, the striatum was also more active when instructions were executed for the first time than during subsequent executions. Furthermore, the degree of activation of the striatum correlated with shorter reaction times on subsequent trials, thus suggesting that this region mediates the acquisition of skills from instructions. This hypothesis agrees with the established connection between the striatum (and, more generally, the basal ganglia) and learning (Graybiel, 2005; Packard & Knowlton, 2002; Yin & Knowlton, 2006).

In summary, previous research has suggested that the process of executing novel instructions engages the same LPFC regions involved in encoding them. Research has also suggested that the execution of novel instructions also depends on the striatum, which is possibly involved in the transformation of instructed tasks into automatic procedures.

Limitations of the previous research

So far, previous studies have focused on comparing instructions that were encoded or executed for the first time against instructions that had been encoded or executed previously. While these comparisons have uncovered a number of regions that take part in the process of instruction interpretation, they suffer from a number of limitations.

The first limitation concerns the specific roles of different regions. In theory, the activation of a region in a neuroimaging experiment implies that the region either is *processing* specific information or is holding task-specific *representations*, or a mixture of the two. Let us take, for instance, the case of the rostral LPFC. Do neurons within the LPFC actually encode the internal template for the task, as suggested by its increased activity during the encoding of novel instructions? Or is LPFC involved in integrating information from the stimuli into the template itself? Or is it involved in both functions? These questions are difficult to answer when data analysis is confined to the “active” parts of the tasks—that is, when instructions are either being encoded or executed. Instead, processing and representing can be dissociated by comparing brain activity during the active parts of a task with brain activity during the delay periods following the encoding and execution of instructions. During the delay periods, and in particular during the delay that follows the encoding phase, the mental template for a task needs to be maintained, but no further processing needs to be done. The present study includes randomly varying delay intervals between the various phases of each trial (see Fig. 2 below). This allows for the estimation of brain activity associated with the maintenance of task representations.

A second limitation is that previous studies have only focused on within-phase comparisons—that is, between the encoding of novel and practiced instructions, or between the execution of novel and practiced instructions (Brass et al., 2009; Cole et al., 2010; Hartstra et al., 2011; Ruge & Wolfensteller, 2010). These “vertical” comparisons have some shortcomings. For instance, the transition from executing a task by interpreting a mental template to executing a task in an automatic fashion is likely a graded one, with regional activations decreasing continuously over multiple repetitions (Chein & Schneider, 2005). Because of this, the final results are inherently dependent on how much practice is given to participants. To overcome this difficulty, these analyses should be complemented by “horizontal” comparisons—that is, between the encoding and execution of novel instructions. Comparing the two phases with each other is more powerful than comparing novel versus practiced instructions, because no gradient effect hampers the power of the comparison. As long as a certain brain region is specific to the encoding or the execution phase, it will be picked up by this contrast. This principle is particularly helpful when studying the process of instruction

interpretation, because instructions cannot be interpreted during the encoding phase (since the stimuli have not been given to the participant yet).

It is important to note that, while this contrast is more powerful than the comparison of novel versus practiced instructions, it is not immune to problems. For instance, instructions to perform mathematical operations will increase activation in a number of known circuits that are specialized for number processing (Dehaene, Piazza, Pinel, & Cohen, 2003; Piazza, Izard, Pinel, Le Bihan, & Dehaene, 2004; Piazza, Pinel, Le Bihan, & Dehaene, 2007). In this case, number-processing circuits will be more active simply because they are specific to the stimuli used during the execution phase. Because each contrast is subject to limitations, this study takes into consideration the converging results from multiple contrasts to understand each region's functional role in interpreting instructions.

Experimental predictions

On the basis of neurophysiological and computational considerations, we hypothesized that the process of interpreting task instructions would be carried out by a network of regions comprising the striatum, the rostral LPFC, and the posterior parietal cortex. We further hypothesized that the rostral LPFC would be involved in resolving dependencies between the various operations in complex instructions, that the parietal cortex would be required to keep an internal representation of the task to be executed (the “mental template”), and that the caudate nucleus would be involved in integrating the various sources of information when the mental template needs to be transformed into actual behavior.

There is much evidence that the rostral LPFC, and in particular its left part, is involved when integrating pieces of information that are linked in complex relationships (Bunge et al., 2009; Bunge et al., 2005; Crone et al., 2009). As was noted above in the Stroop task example, executing a novel instructed task also requires solving complex relationships, like identifying the name of the color of the stimulus. With practice, this three-item chain of references can be solved seamlessly (e.g., Elio, 1986). But, before a task has become automatic, these relationships need to be resolved in a controlled and voluntary way, and the left rostral LPFC seems to be crucially involved in this process (Bunge et al., 2009). The precise mechanism by which the left rostral LPFC integrates complex relational information is still unknown. Many authors have suggested that the LPFC guides behavior by representing goals. They hypothesized that these goal representations exert a top-down bias on how the rest of the brain processes information, which ultimately directs the processing of information in the appropriate direction (Miller, 2000; Miller & Cohen, 2001). By generalizing this

principle, one can imagine that neurons in the rostral LPFC encode some sort of “link” between the current features of the attended stimulus and their role in the mental representation of the task (the “mental template”). In the Stroop task example, the rostral LPFC would provide the link between the “color” feature of the current stimulus (i.e., the color green) and its role in the instructions (“name the *color* of the word”). Thus, the rostral LPFC connects the current features of the stimulus with what we have called the “mental template.”

We propose that this mental template is represented in the posterior parietal cortices. In fact, many neuroimaging studies have found that this region is involved in building or modifying internal representations of a task. For example, when participants are solving algebra equations in their heads, the activity of the posterior parietal cortex is proportional to the number of manipulations performed over the equation. Let us consider the equation “ $2X + 5 = 21$.” To solve this equation, its original form must be updated after each solution step to keep track of its intermediate forms (e.g., “ $2X = 21 - 5$,” “ $2X = 16$,” and “ $X = 16/2$ ”). The number of such mental updates predicts the response magnitude of the parietal region (Danker & Anderson, 2007; Stocco & Anderson, 2008). This region is also recruited when an incomplete task is resumed after an interruption, and its activity depends on the complexity of the state of the task that is resumed. For instance, if participants are resuming a multicolumn addition, the activity of the parietal region is larger when they need to remember a previous carry (a more complex state) than when there is no carry (simpler state; Borst, Taatgen, Stocco, & van Rijn, 2010). In fact, in the cognitive architecture ACT-R (Anderson et al., 2004), the posterior parietal cortex has been identified with a computational module that holds “problem states”—for instance, intermediate representations of the task. In summary, converging evidence from different experiments has suggested that the posterior parietal cortex is involved in maintaining intermediate representations of complex tasks. These intermediate representations can be used to keep track of the current position in a series of mental operations (Stocco & Anderson, 2008) or to resume a task after an interruption (Borst, Taatgen et al., 2010; Borst, Taatgen, & van Rijn, 2010). What we have called the “mental template” of an instructed task closely resembles this type of task state or problem state representation, making the posterior parietal cortex an ideal candidate for its substrate.

Finally, the execution of a novel task from instructions requires a remarkable coordination between different regions. In the Stroop example, information about a stimulus's color (a perceptual feature) needs to be used as a cue to retrieve a color name, and the color name needs to be used as a vocal command. Because the acts of retrieving semantic information and programming vocal outputs depend on different brain circuits, successful performance on

the Stroop task requires a precise transmission of signals from one region to the other. While the LPFC is involved in resolving dependencies and the posterior parietal cortex is involved in holding a mental representation of the task, we believe that the coordination of signals between different regions is orchestrated by the basal ganglia, a set of interconnected nuclei located in the middle of the brain (Albin, Young, & Penney, 1989). A growing body of literature supports this conclusion. For instance, the basal ganglia are more active when memorizing an array of visual items that include distractors, as opposed to an array that contains only items to be memorized (McNab & Klingberg, 2008). This suggests that the basal ganglia play a role in deciding which contents are to be committed to working memory. In fact, a number of existing computational models have worked out, in detail, the biological mechanisms by which the basal ganglia control which signals gain access to PFC (Frank, Loughry, & O'Reilly, 2001; O'Reilly & Frank, 2006; Stocco, Lebiere, & Anderson, 2010). In a previous work, we have shown that, when augmented with a basal-ganglia-based routing mechanism, an artificial neural network can learn to perform new tasks when given explicit task representations—a process that is essentially similar to executing tasks from instructions (Stocco, Lebiere, & Anderson, 2010). For these reasons, we suggest that the caudate nucleus (the largest nucleus of the basal ganglia and the input station of the circuit) is involved in interpreting instructions.

Summary

Despite its pervasiveness in modern society, the process by which humans learn to perform new tasks from instructions is still poorly understood. Previous research has suggested that the rostral LPFC, the posterior parietal cortex, and the caudate nucleus play important roles in this process. We propose that the posterior parietal cortex holds a mental template for the task to be performed, that the LPFC resolves the relationship between the perceptual inputs and their roles in the mental template, and that the necessary coordination between these processing centers is mediated by the caudate nucleus. The next sections describe the results of an experiment that was designed to test these predictions.

Materials and method

The task

In the experimental paradigm that was used in this study, participants executed different instructions at every trial. Different instructions of comparable difficulty were generated by systematically varying the combinations of three

arithmetic operations, which were selected from a set of five. Furthermore, all of the instructions contained one binary (e.g., “multiply x by y ”) and two unary (e.g., “add 1 to x ”) operations. Table 1 illustrates the set of operations used in the experiment and provides some examples. The operations specified by the instructions were applied to two input numbers, indicated as x and y . For instance, one task could be “divide x by 3, multiply y by 2, and then multiply the two results with each other.”

As in previous instructed-learning paradigms (e.g., Cole et al., 2010), each trial was broken into three consecutive phases: (a) an *encoding* phase, in which the instructions were presented; (b) an *execution* phase, in which the two input numbers were unveiled and participants had to mentally execute the instructions; and (c) a *response* phase, in which participants indicated whether a given number was the result of executing the instructions. Both encoding and execution were introduced by a 2-s fixation and were self-paced; participants were instructed to press a key as soon as they had encoded the instructions or calculated the result. Encoding, execution, and response were separated by variable delays whose duration varied randomly between 2 and 8 s. The variable delays reduced the collinearity between the phases and allowed for a better estimation of the corresponding brain activity. Because the blank screen could potentially be very short and only 2 s were allowed for the response, participants were discouraged from extending either the encoding of the instructions or their mental calculation into the delay period. The time course of a sample trial is illustrated in Fig. 1.

The instructions for each task were presented as a string of letters and variables in prefix notation, where each operation was represented by one of the letters *A* through *E*. For example, the task “divide the first number by 3, add 1 to the second number, then multiply the two results” [e.g., $(x/3) \times (y + 1)$] was presented as “AExDy” (see Table 1). The prefix notation was chosen because it defines the proper order of application of operations unambiguously and ensures that all of the instructions can be expressed as a five-letter string, thus having the same visual complexity. Presenting task instructions in a special notation also made the encoding phase more difficult than the execution phase. This was important because one crucial step in the analysis consisted of identifying brain regions that were more active during the execution than during the encoding phase. Making the encoding phase behaviorally more difficult than the execution phase ensured that this analysis did not simply identify regions responding to task difficulty. To make sure that participants could read and understand the prefix notation, they were given extensive training the day before the experiment (see the Training Session section below).

During the execution phase, the values of x and y were presented with three randomly intermixed hash signs “#” (e.g., “#2##4”). This way, input values were also presented

Table 1 The five basic operations used in the experimental task

Letter	Operation Type	Notation	Meaning	Examples
A	Binary	Axy	$x \times y$	A(4, 2) = $4 \times 2 = 8$; A(2, 3) = $2 \times 3 = 6$
B	Binary	Bxy	x / y	B(8, 2) = $8 / 2 = 4$; B(6, 3) = $6 / 3 = 2$
C	Unary	Cx	$x \times 2$	C(4) = $4 \times 2 = 8$; C(3) = $3 \times 2 = 6$
D	Unary	Dx	$x + 1$	D(7) = $7 + 1 = 8$; D(3) = $3 + 1 = 4$
E	Unary	Ex	$x / 3$	E(9) = $9 / 3 = 3$; E(6) = $6 / 3 = 2$

Experimental problems were generated by randomly combining three different operations, with the requirement that one of the operations be binary

as five-character strings, thus matching the visual complexity of the instructions. All of the input numbers were in the 1–9 range and could be represented by a single character.

Participants

A group of 21 Carnegie Mellon University undergraduates (18–29 years of age, mean = 22.4; 12 female, 9 male) participated in the experiment in exchange for monetary compensation. Informed consent was obtained from all participants prior to beginning the practice session. The study was reviewed by the Institutional Review Board at Carnegie Mellon University and conformed to both institutional and national guidelines on research ethics.

Upon collecting the neuroimaging data, one female participant was found to exhibit signs of neurological abnormality, and her data were consequently excluded from the analysis.

Training session

All participants underwent a training session the day before their scan. The training session was articulated in three

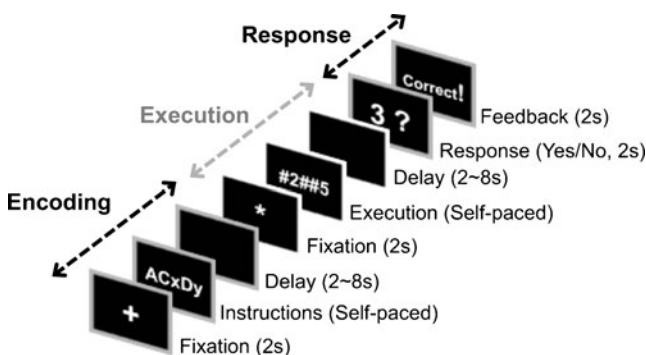


Fig. 1 Experiment design: Sample trial from the experiment, showing the three phases of encoding, execution, and response

consecutive parts. During the first part, participants familiarized themselves with the letter–operation associations. After we introduced the five operations and their associated letters, participants performed 10 single-operation tasks for each letter (e.g., “Axy” or “Cx”). The process was repeated until they were able to solve 10 consecutive problems for each letter without any mistakes. Whenever a mistake was made, the correct count for that operation was set back to zero, so that 10 more problems had to be solved. This procedure was analogous to the drop-out procedure that had been used to help participants memorize pair associates in studies of memory retrieval (Anderson & Reder, 1999). After every second error, participants were given an opportunity to restudy the five letter–operation associations.

The second part of the training introduced multioperation instructions. It consisted of 60 one- and two-letter problems divided into six blocks of 10 trials each. Each trial had the same structure as the experimental trials (see Fig. 1), consisting of consecutive encoding, execution, and response phases separated by randomly varying intervals. Thirty of the trials consisted of one-letter problems (six problems for each operations), while the remaining 30 consisted of two-letter problems (e.g., “AxCy” or “DBxy”; five problems for each combination of one binary and one unary operation). The two-letter tasks were meant to familiarize participants with the composition norm implied by the prefix notation. No trials had identical rules and input values.

The third and last part of the practice had the same duration and block structure of the experimental session—that is, 80 three-letter problems divided into four blocks of 20 each. All of the trials contained one of four three-operation instructions, each of which was practiced with 20 different combinations of values for x and y . This ensured that participants would eventually practice the specific operations and their order of application, and not simply form associations between a task and the final response. The four practiced instructions were different for each participant.

Experimental session

The experimental sessions took place in an fMRI scanner within 24 h of the practice session in order to minimize differences in retention. The experimental session consisted of 80 trials divided into four blocks of 20. Half of the trials were “practiced”—that is, they shared the same instructions used in the training session (10 problems for each of the four instructions), although with different input numbers. The other half consisted of “novel” instructions—that is, novel combinations of operations. These instructions had not been practiced before and were different from one another. Notice that in our experiment, all of the instructions were executed only once within a single trial, and all of the execution

phases were preceded by the corresponding phase. Thus, our design included a condition in which participants encoded “practiced” instructions (i.e., instructions they have encoded previously, which were absent in Ruge & Wolfensteller, 2010) and did not include a condition in which novel instructions were executed for a second or third time (as in Cole et al., 2010, and Hartstra et al., 2011).

Participants were required to complete as many problems as they could within a 45-min limit. This limit was imposed to contain the experimental session within the limit of a 1-h scan. Ten participants were able to complete all of the four experimental blocks (80 trials); eight participants completed three blocks. Due to additional time limitations at the scanning facility, two participants could only complete two blocks (40 trials). Even with only two blocks, we were able to collect a minimum of 15 observations per cell on all participants.

Imaging parameters

Functional data were collected using a Siemens Allegra 3T scanner using a gradient echoplanar pulse sequence with TR = 2,000 ms, TE = 30 ms, and a 72° flip angle. Each volume acquisition consisted of 34 oblique axial slices, each of which was 3.2 mm thick with 0-mm gap and contained 64×64 voxels with an in-plane resolution of 3.125×3.125 mm. A T2-weighted structural image was also acquired for each participant in the same space and with the same parameters as the functional images, but with an in-plane matrix of 256×256 voxels and a spatial resolution of 0.078×0.078 mm.

Image preprocessing and analysis

The data were analyzed using SPM8 (Wellcome Department of Imaging Neuroscience, www.fil.ion.ucl.ac.uk/spm). Images were corrected for differences in slice acquisition time, spatially realigned to the first image in the series, normalized to the Montreal Neurological Institute (MNI) ICBM 152 template, resampled to $2 \times 2 \times 2$ mm voxels, and finally smoothed with a $8 \times 8 \times 8$ -mm full-width-at-half-maximum Gaussian kernel to decrease spatial noise and to accommodate individual differences in anatomy. Statistical analyses were performed on individual and group data using the general linear model, as implemented in SPM8 (Friston, Ashburner, Kiebel, Nichols, & Penny, 2006). For individual participants, fixed-effects models that incorporated a high-pass filter with a cutoff of 128 s and an AR(1) correction for serial autocorrelation were used to estimate parameters. The models included one regressor for the response and one regressor for each of the four experimental conditions (novel instructions, practiced instructions, execution of novel instructions, and execution of practiced instructions). Errors and outlier trials were modeled separately and were not analyzed; outliers were

defined as trials whose reaction times for the encoding or the application phase were more than three standard deviations from the mean for that phase.

Group analyses were performed on the beta values obtained from the individual-participant analysis using a mixed-effects model in which Participants were treated as a random factor. To correct for multiple comparisons, a family-wise error (FWE) correction procedure based on Gaussian random-field theory (as implemented in SPM8) was applied, yielding a corrected Type I error probability of $p < .05$. In addition to a height threshold, a constraint was imposed that each identified cluster should include at least four contiguous voxels (32 cubic millimeters), corresponding to the volume of one voxel in the original native space (31.25 cubic millimeters).

Post-hoc analyses were performed by averaging the estimated mean beta values across all of the voxels within a single region of interest (ROI), using the procedure implemented in the MarsBaR toolbox for SPM (Brett, Anton, Valabregue, & Poline, 2002). The Bonferroni method was applied to correct for Type I errors when comparing the average beta values across multiple regions.

Results

Behavioral results

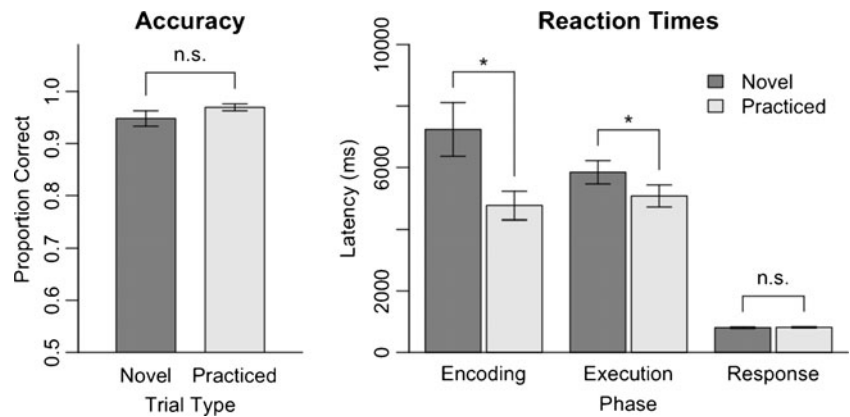
We first examined whether participants’ accuracy differed between the practiced and novel instructions. Since accuracy was a binary variable and participants’ accuracy was almost at ceiling levels ($M = .96$ proportion correct, $SD = .04$), the effect of practice on accuracy was analyzed with a mixed-effects logistic regression model in which participants were random factors. The model did not uncover any effect of practice on accuracy ($Z = 0.61$, $p = .54$).

On the other hand, practice did have a significant effect on the reaction times in both the encoding [$t(19) = 5.14$, $p < .0001$] and the execution [$t(19) = 4.78$, $p = .0001$] phases. In both cases, latencies were higher for the novel instructions. There was no effect of practice on the reaction times for the response phase [$t(19) = -1.28$, $p = .21$; see Fig. 2, right]. In summary, the behavioral data showed that participants were capable of successfully carrying out the task and maintaining high accuracy levels in both practiced and novel tasks, but that novel tasks took longer to be encoded and were executed faster than practiced ones. The difference in reaction times probably reflects the greater effort spent to create and access the internal templates for novel instructions.

Neuroimaging results

The regions significantly involved during the encoding and execution of novel instructions are depicted in Fig. 3. Only

Fig. 2 Behavioral results. (Left) Mean accuracy (\pm standard errors) for novel and practiced trials. (Right) Mean reaction times (\pm standard errors, in milliseconds) for novel and practiced trials, plotted separately for the three trial phases (encoding, execution, and response)



novel instructions were included in this figure because they were more representative of the types of processes that we were interested in investigating. A large network of regions were involved in both encoding the instructions and carrying them out (Fig. 3). Predictably, the common regions included bilateral frontal and parietal regions, which are commonly recruited by novel and effortful tasks (see Hill & Schneider, 2006, for a review), as well as visual areas in the occipital lobe.

Despite the large overlap, a number of regions were found to be preferentially active during one phase but not the other. The next sections will investigate these phase-specific differences.

Regions more active during the encoding than during the execution phase

Eighteen regions were found to be more active during encoding than during the execution of novel instructions (Fig. 4 and Table 2). These regions include early (e.g., BA 17) and late (e.g., left and right fusiform gyrus) visual processing areas, likely because the instructions were visually examined longer than the inputs. The list also included the bilateral hippocampus and medial temporal regions. Increased activation in these regions is associated with the creation of new memories (Nadel, Samsonovich, Ryan, & Moscovitch, 2000; Squire et al., 1992), an activity that also can be expected in the process of creating new representations for novel tasks.

The orbitofrontal cortex, the posterior cingulate cortex, and the right parahippocampal gyrus were also more active during encoding than during the execution of instructions. These three regions belong to the so-called default-mode network, a set of brain regions that share common functional characteristics, such as showing low-frequency correlations in resting-state fMRI and deactivation when participants are actively engaged in tasks (Fox et al., 2005). The functional role of this network is still debated, but authors generally agree that its regions are associated with an information-processing mode that is “self-centered”—that is, focused on internal representations rather than external stimuli (Buckner, Andrews-Hanna, & Schacter, 2008).

The recruitment of the default-mode network when encoding instructions is noteworthy and deserves further investigation. Two hypotheses can explain our finding. One possibility is that the process of encoding instructions shares some characteristics of the conditions in which the default-mode network is active. For instance, the creation of mental templates for new tasks is an entirely internal process, it concerns one’s own future mental states, and it has to do with stimuli that are yet unseen. Thus, this hypothesis states that encoding instructions is a mode of information processing similar to those that characterize the default-mode network.

An alternative account is that the default-mode network is in fact deactivated during the process of encoding instructions, but to a lesser extent than during their execution. This possibility is

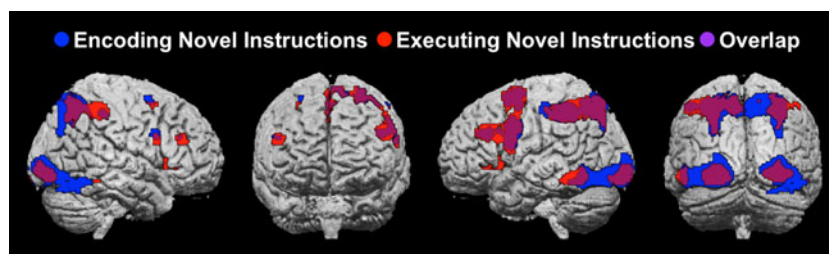
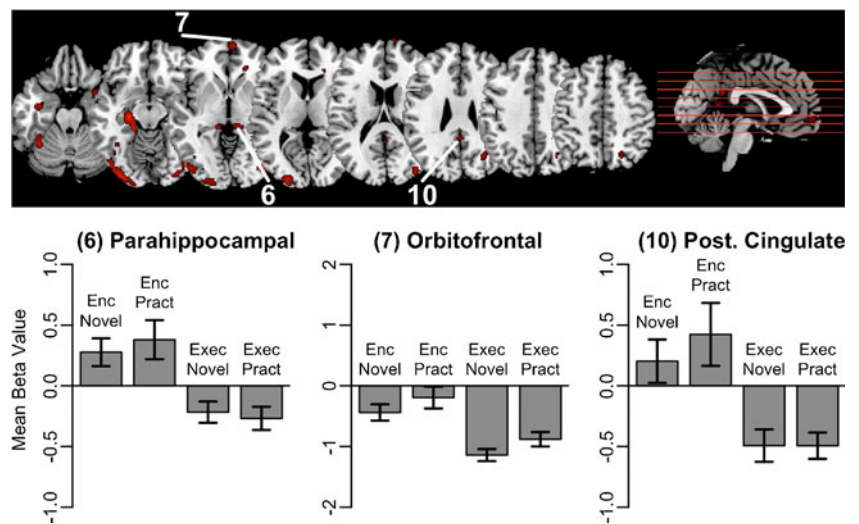


Fig. 3 Imaging results. The regions significantly engaged ($p < .05$, family-wise error corrected) in the instruction phase (red in electronic version), the execution phase (blue in electronic version), or both

(purple in electronic version) of novel tasks, overlaid over the MNI single-participant template

Fig. 4 Regions more active during the encoding than during the execution of novel instructions ($p < .05$, family-wise error corrected). The region numbers refer to the indexes in Table 2. (Top) Multislice view of the regions, overlaid on the MNI template. (Bottom) Follow-up analysis, showing the mean values of the regressors for the four experimental conditions. In the labels, “Enc” stands for “encoding,” “Exec” for “execution,” and “Pract” for “practiced”; see Fig. 1 for a depiction of the different experimental phases



consistent with the fact that regions in the default-mode network typically show deactivation when participants perform cognitive tasks (Greicius, Krasnow, Reiss, & Menon, 2003).

These two hypotheses can be distinguished by examining each region’s beta values for the encoding-phase and the execution-phase regressors. The first hypothesis predicts that these values will be positive and significantly greater than zero for the encoding phases (implying active recruitment of the network) and negative for the execution phases. The second

hypothesis, on the other hand, predicts that the beta values will be negative and significantly smaller than zero (implying deactivation) across all phases. An examination of the regions’ beta values uncovered an interesting dissociation: While the orbitofrontal cortex showed consistently negative values for all of the four task-related phases, the posterior cingulate and the right parahippocampal gyrus exhibited positive responses during the encoding phases and negative responses during execution. One-tailed t tests showed that the activity of the right

Table 2 Regions more active during the Instruction phase than during the execution phase of novel tasks ($p < .05$, family-wise error corrected)

Index	MNI (Peak)	Anatomical Location	Brodmann Areas (% of ROI volume)	Size (Voxels)	Peak Intensity
1	30, 10, -16	Right parahippocampal gyrus	BA 34 (7.14 %); BA 28 (5.95 %); BA 47 (3.57 %); BA 13 (2.38 %); BA 38 (2.38 %)	84	8.34
2	48, -70, -10	Right fusiform gyrus	BA 37 (13.47 %); BA 19 (12.36 %); BA 18 (2.84 %)	809	11.78
3	-24, -22, -12	Left hippocampus	BA 35 (6.79 %); BA 28 (4.91 %); BA 30 (2.64 %); BA 27 (1.51 %)	265	12.56
4	-36, -52, -22	Left fusiform	BA 37 (26.36 %); BA 20 (0.91 %)	110	9.20
5	-32, -90, -16	Left inferior occipital gyrus	BA 18 (14.61 %); BA 19 (7.00 %); BA 17 (1.94 %)	671	12.97
6	14, -32, -4	Right parahippocampal gyrus	BA 27 (4.46 %); BA 30 (2.68 %); BA 35 (1.79 %); BA 28 (0.89 %)	112	9.11
7	4, 62, -4	Right and left inferior orbitofrontal cortex	BA 10 (36.19 %); BA 11 (4.76 %)	105	8.89
8	20, 36, -2	White matter	–	40	7.81
9	16, 50, 2	Right superior middle frontal gyrus	BA 10 (30.00 %)	10	7.54
10	2, -52, 12	Posterior cingulate	BA 29 (9.01 %); BA 31 (8.11 %); BA 30 (8.11 %); BA 23 (7.21 %)	111	8.03
11	32, -60, 22	Right middle occipital gyrus	BA 19 (9.09 %); BA 39 (2.39 %); BA 7 (0.48 %)	209	8.73
12	-26, -68, 14	Left middle occipital gyrus	BA 18 (4.76 %); BA 31 (4.76 %)	21	8.62
13	10, 70, 14	Right medial superior frontal gyrus	BA 10 (22.22 %)	36	8.78
14	38, -6, 28	Right postcentral gyrus	BA 6 (27.91 %)	43	8.43
15	24, 6, 34	White matter	–	16	8.12
16	22, -20, 34	White matter	–	33	9.51
17	-44, -16, 36	Left postcentral gyrus	–	13	7.28
18	20, -66, 50	Right precuneus	BA 7 (65.22 %)	23	7.39

Anatomical labels were obtained from the Automated Anatomical Labeling atlas (Tzourio-Mazoyer et al., 2002).

parahippocampal gyrus was significantly greater than zero during the encoding of both novel and practiced instructions [$t(19) > 2.44, p < .013$] and that the activity of the precuneus was significantly greater than zero during the encoding of novel instructions [$t(19) = 1.68, p = .05$]. In contrast, the beta values were negative for the execution of novel and practiced instructions across these three regions. The bottom part of Fig. 4 illustrates the results of this analysis.

Regions more active during the execution of novel than of practiced instructions

As a first step in identifying the regions involved in interpreting instructions, we proceeded by identifying regions that were more active during the execution of novel instructions than during the execution of instructions that had been practiced in advance. This analysis was a replication of similar analyses conducted by Hartstra et al. (2011), Cole et al. (2010), and Ruge and Wolfensteller (2010) and could be used to observe the regions involved in instruction interpretation.

Three regions were more active during the execution of novel than of practiced instructions: the left rostral LPFC and two posterior parietal regions located between the superior occipital gyrus and the posterior part of the superior parietal gyrus. These clusters are shown in Fig. 5 and are described in Table 3. Because the underlying component operations were equal (in terms of number and complexity) in novel and practiced tasks, these regions' increased activity cannot be attributed to any difference in the underlying operations that

were executed and should be attributed to the additional mental effort for interpreting the mental template for the new task.

These results are broadly consistent with our hypothesized framework, according to which both prefrontal and parietal cortices are involved in interpreting novel instructions. However, this analysis cannot test a more specific prediction of our framework—that is, that the rostral LPFC is involved in the online resolution of complex relationships between the terms of instructions, while the posterior parietal cortices maintain the internal mental template of the task to be performed. To test this specific prediction, one needs to compare the activity of these regions during the active parts of a task against their activity during the delay phases. It seems reasonable to assume that a region that is involved in representing task instructions should be active during the delay, because these representations need to be voluntarily maintained until the execution begins. In contrast, a region that is only involved in transforming instructions does not need to become active until the execution phase starts.

To verify this hypothesis, we created a second set of first-level statistical models, which we will refer to as the “augmented models.” The augmented models contained the same regressors as the original models, but included four additional regressors capturing the delay periods that followed the encoding and the execution of novel and practiced instructions (see Fig. 1). These additional regressors were included to capture any metabolic activity that might be related to maintaining information in memory.

The beta values for the two delay periods that followed the encoding of either novel or practiced instructions were extracted

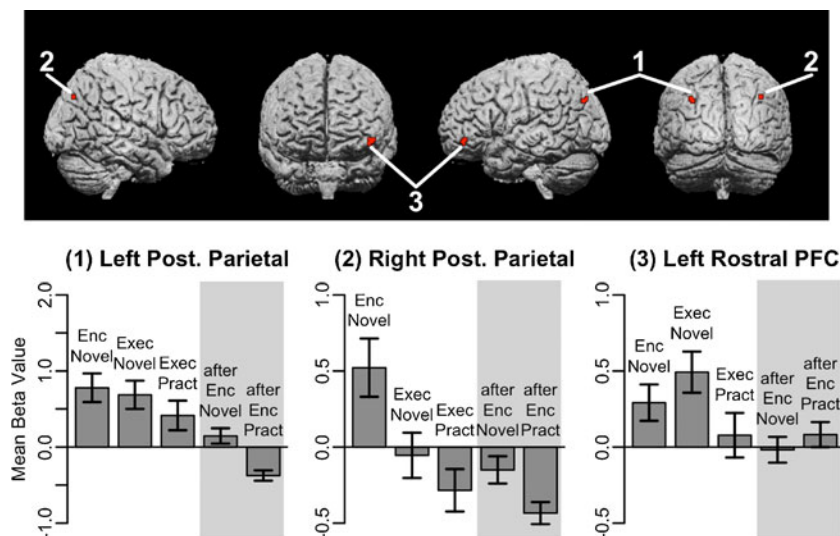


Fig. 5 Regions more active during the execution of novel instructions than during the execution of practiced instructions ($p < .05$, family-wise error corrected). The region numbers refer to the indexes in Table 3. (Top) Location of the regions on the MNI template. (Bottom) Follow-up analysis, showing the mean values of the regressors for five experimental conditions of interest. In each of the three plots, the three vertical bars on

the left represent the beta values estimated from the regular single-participant models, and the two columns on the right (over a gray background) represent the beta values estimated from the augmented models, which included additional regressors for the delay intervals. In the labels, “Enc” stands for “encoding,” “Exec” for “execution,” and “Pract” for “practiced”; see Fig. 1 for a depiction of the different experimental phases

Table 3 Regions more active during the execution of novel instructions than during the execution of practiced ones ($p < .05$, family-wise error corrected)

Index	MNI (Peak)	Anatomical Location	Brodman Areas (% of ROI Volume)	Size (Voxels)	Peak Intensity (t Value)
1	-44, 48, -8	Left middle orbitofrontal	BA 11 (5.88 %)	17	7.54
2	-32, -78, 36	Left posterior parietal	BA 19 (63.64 %)	11	8.35
3	38, -76, 40	Right posterior parietal	BA 19 (66.67 %)	6	7.42

Anatomical labels were obtained from the Automated Anatomical Labeling atlas (Tzourio-Mazoyer et al., 2002).

and calculated for each of the three regions; their mean values are shown in Fig. 5, represented by the two rightmost bars in each region's plot, drawn against a gray area (the three leftmost bars represent beta values for active parts of the task, extracted from the original first-level models). Pairwise comparisons revealed that the left posterior parietal cortex was significantly more active after encoding novel instructions than after encoding practiced ones (corrected $p < .001$). In contrast, the difference between the two delay periods was not significant for the right posterior parietal region (corrected $p > .37$) and was only marginally significant for the rostral LPFC (corrected $p > .06$). Another significant difference is that, while the left rostral LPFC was more active during the execution than the encoding of novel instructions (corrected $p < .04$), the left parietal region was not (corrected $p = 1$) and the right parietal regions was, in fact, significantly more active during the encoding than during the execution phase (corrected $p < .04$). An ANOVA revealed that the ROI-by-contrast interaction was significant [$F(2, 38) = 14.49$, $p < .001$].

This pattern of results is consistent with our predictions, and suggests that the rostral PFC does not represent the task-specific rules per se, but might play a role in integrating information about the current stimuli within the mental template for a task.

In the case of our experiment, such a template would likely be used to represent the order and hierarchy in which the three mental operations should be applied, and possibly the current step in the processing order. This template is, in turn, represented bilaterally in the posterior parietal cortices. To further verify this hypothesis, we performed a regression-based analysis of those voxels whose difference in activity between the execution of novel and practiced instructions correlated significantly with the mean behavioral difference in the reaction times needed for carrying out novel versus practiced instructions (see Fig. 2). We expected that the practice-related decrease in reaction times during the execution phase would be mostly due to reduced effort to link the working memory contents with their proper role in the mental template. We reasoned that, as a task gets more and more familiar, this process should become more and more automatic, resulting in a corresponding reduction of the activity of the rostral PFC. On the other hand, we did not expect a similarly strong correlation in the case of the posterior parietal cortex. These regions do become less active with practice, but, since a mental template is still needed to execute instructions, we

believe that their reduction in activity would be less correlated with the reduction in reaction times than would be the case for the rostral LPFC. Thus, a regression analysis identifying voxels whose decrease in activity during the execution phase matched the corresponding decrease in reaction times should identify the rostral LPFC, but not the posterior parietal cortex.

This regression analysis identified four regions, which are shown in Fig. 6 and described in Table 4. As expected, one of these clusters included all of the voxels of rostral LPFC region previously identified, thus confirming this region's involvement in the integration of information for the execution of novel tasks. The remaining three clusters included two regions in the left dorsolateral PFC and one region in the medial temporal lobe. The two dorsolateral prefrontal regions have been previously identified as being involved in working memory (Cohen et al., 1997) and are located in proximity to the dorsolateral region also reported by Cole et al. (2010). The left middle temporal cluster corresponds to the region that Donohue, Wendelken, Crone, and Bunge (2005) identified as being involved in the long-term representations of abstract rules. Taken together, these results suggest that, with practice, participants become faster at executing instructed tasks because they rely less on retrieving the instructions and rehearsing them in working memory. These results are consistent with Cole et al.'s (2010) interpretation of their own findings. Consistent with our prediction, the reduction in reaction times was correlated with a decrease in activity in the prefrontal regions, but not in the posterior parietal regions that hold the template.

In summary, the rostral LPFC and the left and right posterior parietal cortices are specifically more involved in the execution of novel versus practiced instructions. An analysis of their activation patterns suggests that the two parietal regions are involved in representing an internal model or template of the task to be performed, while the rostral LPFC is involved in resolving the internal relationships between individual operations at the moment of execution.

Regions more active during execution than during encoding of novel instructions

As we explained in the introduction, a more detailed picture of the mechanisms underpinning instruction interpretation

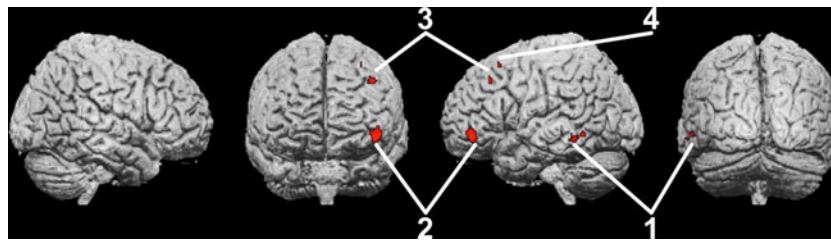


Fig. 6 Locations of the four regions whose difference in activation between the execution of novel and practiced instructions reliably predicted the corresponding difference in reaction times between the

execution of novel versus practiced instructions ($p < .05$, family-wise error corrected). The region numbers refer to the indexes in Table 4

can be gathered by identifying brain regions that were more active during the execution than during the encoding of novel instructions. This analysis identified two regions, the superior part of the left inferior parietal gyrus (BA 40) and the right head of the caudate nucleus (Fig. 7 and Table 5). As we noted above, this comparison is limited because it identifies regions involved in *any* of the mental processes that are specific to the execution phase, and not only those that support instruction interpretation. In this particular case, the inferior parietal gyrus and the caudate nucleus might simply be involved in the execution of mathematical operations, which do not occur during the encoding phase. In fact, both regions have previously been found to be active in mathematical problem solving (Stocco & Anderson, 2008). If this were the case, however, the same regions would also come up when comparing brain activity during the execution of practiced instructions against the encoding of novel instructions, or even against the encoding of practiced instructions. Neither of these contrasts, however, yielded any significant voxels.

In summary, the involvement of these two regions is specific to the execution of novel tasks but does not seem to relate to either the nature of the operations being executed (otherwise, the same regions would also be more active for practiced tasks) or to a mere effect of novelty (otherwise, the same regions would be similarly active during novel instructions). We propose, therefore, that the left inferior parietal gyrus and the right caudate nucleus are part of the network that interprets instructions.

The identification of the caudate nucleus is consistent with our proposed framework. The identification of an

additional parietal region comes, however, was a surprise. To gather insight into the specific functions of these regions, we analyzed the data from the augmented single-participant models. An analysis of the activity of the caudate nucleus and the parietal areas during the delay periods could point out whether these regions are involved in processing information or in holding some form of representations.

The results showed that a significant difference exists between the caudate and the parietal regions. Specifically, the parietal region is sensitive to novelty during the delay periods, as evidenced by its higher activation after novel than after practiced instructions [$t(19) > 8.22$, $p < .0001$, Bonferroni corrected], while the caudate is not [$t(19) = 1.54$, $p > .28$; Fig. 5, bottom]. The interaction between ROI and delay type (after novel vs. after practiced instructions) was also significant [$F(1, 19) = 67.84$, $p < .0001$, Bonferroni corrected]. This difference is consistent with the parietal region being engaged in the internal representation of the instructions, while the right caudate is more likely involved in the process of transforming these representations into actual behaviors.

Discussion

This study has identified a network of regions that jointly contribute to the execution of instructions for novel tasks. This network includes rostral prefrontal and posterior cortices and the basal ganglia. Although further research on the functional specialization of these regions needs to be pursued, a number of preliminary conclusions can be drawn from the data.

Table 4 Regions whose difference in activation between the execution of novel versus practice instructions reliably predicted the corresponding difference in reaction times ($p < .05$, family-wise error corrected)

Index	MNI (Peak)	Peak Anatomical Location	Brodmann Areas (% of ROI Volume)	Size (Voxels)	Peak Intensity
1	-58, -54, -4	Left middle temporal gyrus	BA 37 (3.92 %)	51	7.90
2	-46, 46, -6	Left inferior orbitofrontal cortex	BA 10 (4.60 %); BA 11 (2.30 %); BA 47 (1.15 %)	87	9.24
3	-42, 30, 42	Left middle frontal gyrus	BA 9 (42.86 %); BA 8 (23.81 %)	21	8.26
4	-30, 22, 56	Left superior frontal gyrus	BA 8 (100.00 %)	5	7.76

Anatomical labels were obtained from the Automated Anatomical Labeling atlas (Tzourio-Mazoyer, et al., 2002).

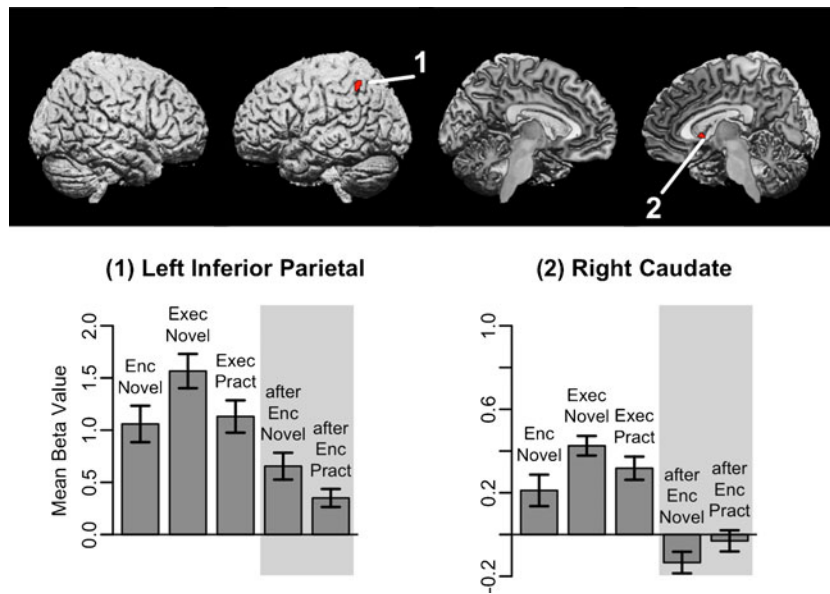


Fig. 7 Regions more active during the execution than during the encoding of novel instructions ($p < .05$, family-wise error corrected). The region numbers refer to the indexes in Table 5. (Top) Location of the regions on the MNI template. (Bottom) Follow-up analysis, showing the mean values of the regressors for five experimental conditions of interest. In each of the two plots, the three vertical bars on the left represent the beta values estimated from the regular single-participant

models, and the two columns on the right (over a gray background) represent the beta values estimated from the augmented models, which included additional regressors for the delay intervals. In the labels, “Enc” stands for “encoding,” “Exec” for “execution,” and “Pract” for “practiced”; see Fig. 1 for a depiction of the different experimental phases

Role of the rostral LPFC

A huge amount of literature supports the role of the rostral LPFC in handling and representing complex rules, whether these rules are learned from instructions (Cole et al., 2010; Hartstra et al., 2011) or by trial and error (Anderson, Betts, Ferris, & Fincham, 2011; Strange, Henson, Friston, & Dolan, 2001). Various authors have suggested that more frontal regions are capable of representing more complex and abstract information in virtue of their more anterior positions, which puts them at the top of the hierarchy of prefrontal connections (Badre, 2008; Koehlin, Ody, & Kouneiher, 2003). In fact, there is substantial experimental evidence that the rostral LPFC is specialized for integrating information across complex relationships (Bunge et al., 2009, 2005; Crone et al., 2009; Wendelken et al., 2008). For instance, it comes online when people have to judge the similarity of relationships between pairs of words (Bunge et al., 2005; Wendelken et al., 2008). Consistent with our findings, this functional specialization seems to hold for

the left part more than for the right (Bunge et al., 2009). The execution of novel instructions also requires integrating pieces of information that are linked in a complex relationship. In particular, when novel instructions are executed for the first time, certain aspects of the stimuli need to be connected with their role in the mental template for the task. In our experiment, intermediate results in working memory (e.g., “ $X = 4$ ”) needed to be passed as input arguments to arithmetic functions (e.g., “2 times X ”), and, in turn, their results (“8”) needed to be used to update the contents of working memory (e.g., “ $X = 8$ ”).

The process of resolving these complex dependencies would be especially needed during instruction execution, when the unveiled stimuli need to be linked to the abstract representation of the operations to perform. This is consistent with our findings that the rostral LPFC is not significantly more active for novel versus practiced instructions during the delay periods that follow the encoding phase (i.e., periods in which the mental template needs to be maintained). It is also consistent with the significant ROI-by-

Table 5 Regions more active during the execution than during the encoding of novel instructions

Index	MNI (Peak)	Peak Anatomical Location	Brodmann Area (% of ROI volume)	Size (Voxels)	Peak Intensity
1	12, 14, 0	Right head of the caudate nucleus	–	5	7.33
2	–44, –52, 52	Left inferior parietal gyrus	BA 40 (65.00 %)	20	7.34

Anatomical labels were obtained from the Automated Anatomical Labeling atlas (Tzourio-Mazoyer, et al., 2002).

contrast interaction, whereby the rostral LPFC, but not the parietal regions, is significantly more responsive to the execution than to the instruction phase.

On the other hand, the integration of information into complex relationships is also needed, at least in part, during the encoding of novel instructions. This is because, in our paradigm, the various component operations are linked together so that the results of one operation are the input values for another. For example, when encoding the instructions “divide the first number by 3, add 1 to the second number, then multiply the two results,” the results of first two operations (i.e., dividing by 3 and adding 1) are the arguments for the third (i.e., multiplication). Because a proper representation of a task requires keeping track of how the various variables are passed from one operation to the other, the rostral LPFC is at least partially needed during the encoding phase, thus explaining why this region was not identified when contrasting the execution against the encoding of novel instructions. Thus, the hypothesis that the rostral LPFC integrates relationships across different representations fits with our experiment results.

Role of the posterior parietal cortices

Like the rostral LPFC, the left and right posterior parietal cortices were significantly more active during the execution of novel than of practiced instructions. These regions were not, however, more active during the execution phase than during the instruction phase. This pattern implies that the posterior parietal cortices support cognitive processes that are at least partially shared by both the encoding and the execution of novel instructions. This is consistent with the hypothesis that the posterior parietal cortices represent the “mental template” for a task. This mental template needs to be built during the encoding phase and updated during the execution phase. With practice, the mental template becomes easier to build and update (explaining the difference between novel and practiced instructions), but not necessarily more active during execution than during encoding.

Further comparisons showed that the left posterior parietal region is more active in the delay period that follows the encoding of novel instructions (as opposed to the delay following the re-encoding of practiced instructions) and that both parietal regions were more active during the encoding than during the execution of novel instructions (albeit the difference was not significant for the left region). This pattern suggests that the two posterior parietal regions are involved in actively maintaining at least some components of the instructions during the delay periods. Their role in maintaining complex task representations has been confirmed by other research. The two parietal regions lie in close proximity to the two predefined ROIs that correspond to the imaginal module in ACT-R, a cognitive architecture

that has been successfully used to predict patterns of brain activation in neuroimaging studies (Anderson, Fincham, Qin, & Stocco, 2008). The imaginal module is specialized for holding so-called problem state representations—that is, representations of the intermediate states of the task being performed. Tasks that stress these limits (because, e.g., they require switching back and forth between different representations) elicit strong imaging responses in the two areas identified by our analysis (Borst et al., 2010a; Borst, Taatgen, & van Rijn, 2011; Stocco & Anderson, 2008). In the case of our experiment, the problem state that needs to be maintained is the initial model of the task, as specified by the instructions (explaining the regions’ activity elicited by novel instructions). This internal model needs to be accessed and, possibly, updated more frequently in the case of novel tasks than in the case of practiced tasks, thus explaining the behavior of the posterior parietal cortices in our experiment. It also needs to be accessed when rehearsing during the delays between encoding and execution. As predicted by the ACT-R architecture, the activity of the posterior parietal cortices reflects the mental effort spent on creating, updating, and inspecting this mental model.

Roles of the caudate nucleus and the inferior parietal cortex

The inferior parietal gyrus and the right caudate were significantly more active during the execution than during the encoding of novel instructions. At the same time, these regions were not significantly more active during the execution of novel than of practiced instructions (albeit they both showed a trend in this direction; see Fig. 7), and the difference between executing novel and practiced instructions did not correlate significantly with the corresponding difference in reaction times. Thus, the inferior parietal gyrus and the caudate nucleus must contribute some computation that is, to some extent, required even when tasks are practiced. In our framework, the caudate nucleus is involved in coordinating the transmission of information between different cortical areas when no established pathways exist. For instance, during the execution of our mathematical tasks, intermediate results in working memory (e.g., the number “3”) are used as inputs to new operations (e.g., “the double of 3”), resulting in updated values (e.g., the number “6”). This transfer of information corresponds to the transmission of signals between the posterior parietal cortex (the mental template) and the dorsolateral PFC (working memory), and their coordination is mediated by the basal ganglia (Stocco, Lebiere, & Anderson, 2010). As we discussed above, the rostral PFC aids this process when the connections between the role of a particular input (“ $X = 3$ ”) and a particular place within the template (“double of X ”) are novel and unfamiliar. But the underlying transmission of information still occurs, to a certain extent, after the instructions have been

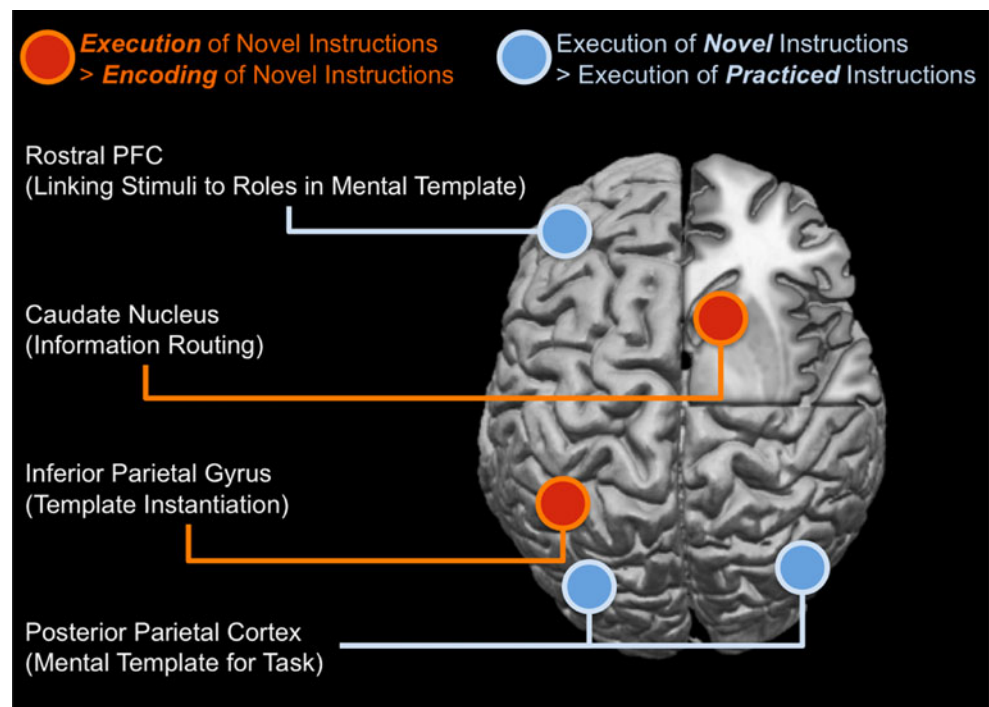
practiced, thus explaining why the caudate nucleus was not identified when comparing the execution of novel and practiced instructions.

Like the caudate nucleus, the inferior parietal gyrus has been suggested to play a role in coupling different pieces of information. For instance, Corbetta and Shulman (2002) noted that the inferior parietal region modulates the attentional processes needed to properly link visual stimuli with motor responses, and Collette et al. (2005) observed that this region is more active when two tasks need to be performed concurrently (more interference in coupling information) than when the same cognitive operations are integrated in a single task (less interference when coupling information). Unlike the caudate nucleus, however, the inferior parietal gyrus shows signs of also being involved in maintaining the mental template during delay periods (Fig. 7). A possible interpretation of this behavior is that this region mediates the transition between the static mental template (e.g., “the double of X ”) and its current instantiation (the “double of 3”).

A functional network of regions for interpreting instructions

Overall, our experimental results are consistent with our initial framework and suggest a preliminary model for a network of regions that mediates instruction interpretations—that is, the transformation of abstract mental representations of a task into actual behaviors. This network, together with a preliminary characterization of the distinct contributions of each region, is represented in Fig. 8.

Fig. 8 A proposed framework for the distinct contributions of the different regions in the instruction interpretation network



The process of encoding instructions

In addition to the regions involved in executing instructions, we also identified a network of 18 regions that was more active when encoding than when executing novel instructions. This network included brain regions involved in memory formation, like the hippocampus (see Table 2 and Fig. 4). Interestingly, three of these regions (the orbitofrontal cortex, the right parahippocampal gyrus, and the posterior cingulate) belong to the so-called default-mode network (Buckner et al., 2008).

A follow-up analysis of the regions’ beta values revealed that the parahippocampal regions and the posterior cingulate were, in fact, actively recruited during the encoding of novel instructions (as indicated by beta values significantly greater than zero). The default-mode network has been previously associated with the creation of autobiographical and prospective memories (Spreng & Grady, 2010), as well as with “self-centered” and “internal” processing of information (Buckner et al., 2008). Each of these functions can, in principle, also play a role in encoding instructions, and the exact relationship between the process of encoding instructions and the activity of the default-mode network remains an open question for future research.

Present limitations

Although our interpretation of the experimental results is internally consistent and is compatible with the

existing literature, a number of limitations need to be acknowledged.

A first caveat concerns the role of the caudate nucleus. In our proposed framework, the caudate nucleus is involved in coordinating how information is passed between other processing centers. However, the caudate nucleus is also involved in learning and skill acquisition (Knowlton, Mangels, & Squire, 1996; Packard & Knowlton, 2002; Yin & Knowlton, 2006). Thus, it is conceivable that its activity reflects the ongoing learning of the task, which is obviously more likely to happen when the task is entirely novel. Although this is possible in theory, one must note that, in our experiment, the activity in the caudate nucleus was not significantly smaller during the execution of practiced instructions. Furthermore, its reduction in activity while executing practiced instructions (as opposed to novel instructions) did not correlate significantly with the corresponding reduction of reaction times.

A second confound is that the caudate is also involved in the inhibition of practiced routines (Casey et al., 1997). Therefore, it is possible that its activation during the execution of novel tasks was due to the concurrent inhibition of practiced routines. Inhibition of practiced responses is speculated to initiate in the right inferior frontal gyrus (IFG) and to reach the right subthalamic nucleus, passing through the right caudate (Frank, Samanta, Moustafa, & Sherman, 2007; Rubia, Smith, Brammer, & Taylor, 2003). In the context of our experiments, increased caudate activity during the execution of novel instructions would be needed to overcome the procedures that have already been learned for practiced instructions. Thus, this account can easily explain why the right (and not the left) caudate was identified by the analysis. On the other hand, this account would predict a similarly significant pattern of activation for the right IFG, which was not found in any of our analyses.

A third limitation concerns our interpretation of the rostral LPFC's function. In addition to integrating complex relationships and representing complex rules, the rostral PFC is involved in prospective memory (Burgess, Scott, & Frith, 2003; Costa et al., 2011; Reynolds, West, & Braver, 2009; Volle, Gonen-Yaacovi, de Lacy Costello, Gilbert, & Burgess, 2011). Prospective memory plays an important role in our experimental design because instructions are encoded at the beginning of a trial but executed only at a later time (Fig. 1). Therefore, it is theoretically possible that the activity of the rostral PFC during the execution of novel tasks reflects the retrieval of procedures that participants have previously encoded in preparation for the forthcoming execution phase. However, the hypothesis that the lateral rostral PFC integrates information across complex relationships has been supported by a number of neuroimaging studies in which rostral PFC activity occurred *after* the presentation of stimuli, and thus could not be an effect of prospective memory (Anderson et al., 2011; Bunge et al., 2005;

Christoff et al., 2001). Instead, it is possible that memory for prospective actions is encoded as an internal rule that is rehearsed until the moment of future application.

The generality of our conclusions is also constrained by some limitations of the present study. First, this research was designed with a focus on one of the computational problems posed by the process of understanding instructions—that is, how instructions are instantiated into actual behaviors. Thus, no factors were explicitly manipulated that would give us insight into the mechanism by which instructions are encoded, or by which practice takes place. For instance, instructions were not varied in complexity, and practice was not varied parametrically.

Also, in our study the instructions were presented in an artificial notation, making the encoding phase behaviorally more difficult (i.e., slower) than the execution phase. While this notation let us correctly interpret the comparisons between the execution and the encoding of novel instructions, it also yielded some drawbacks. In particular, the use of an artificial notation likely increased participants' reliance on working memory during both novel and practiced instructions. As a consequence, the dorsolateral prefrontal cortex was not found among the regions more active during the execution of novel versus practiced instructions. Even more dramatically, no reliable difference was found when comparing the encoding of novel and practiced instructions, despite the existence of a significant behavioral difference (Fig. 2). In summary, the use of a prefix notation prevented us from observing the dynamics of working memory during instruction interpretation, and our proposed framework (Fig. 8) should probably be revised in the future to include the role of the dorsolateral prefrontal cortex (Cole et al., 2010).

Finally, all of the tasks that were used in our experimental paradigm were composed of similar arithmetic operations. This choice was made so as not to introduce additional factors in our search for the neural underpinnings of rule instantiation. However, it is possible to envision tasks made of operations of a different nature—for example, operations consisting of associated retrievals or mental imagery manipulations. The use of different operations would provide a potential means to further investigate the nature of abstract rule representations in the parietal (and possibly prefrontal) cortices.

These present limitations represent an exciting opportunity for future research, where the processes by which humans can rapidly communicate how to perform new tasks will be further studied and better understood.

Author note This research was made possible by Award No. FA9550-08-1-0404 from the Air Force Office of Scientific Research (AFOSR) to J.R.A., R.C.O., and C.L.; by support from the Army Research Laboratory's Robotics Collaborative Technology Alliance to C.L.; and by a special award from the Brain Imaging Research Center (BIRC) of Pittsburgh to A.S.

References

- Albin, R. L., Young, A. B., & Penney, J. B. (1989). The functional anatomy of basal ganglia disorders. *Trends in Neurosciences*, *12*(10), 366–375.
- Anderson, J. R., Betts, S., Ferris, J. L., & Fincham, J. M. (2011). Cognitive and metacognitive activity in mathematical problem solving: Prefrontal and parietal patterns. *Cognitive, Affective, & Behavioral Neuroscience*, *11*(1), 52–67.
- Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review*, *111*(4), 1036–1060.
- Anderson, J. R., Fincham, J. M., Qin, Y., & Stocco, A. (2008). A central circuit of the mind. *Trends in Cognitive Sciences*, *12*(4), 136–143.
- Anderson, J. R., & Reder, L. M. (1999). The fan effect: New results and new theories. *Journal of Experimental Psychology: General*, *128*, 186–197.
- Badre, D. (2008). Cognitive control, hierarchy, and the rostro-caudal organization of the frontal lobes. *Trends in Cognitive Sciences*, *12*(5), 193–200.
- Borst, J. P., Taatgen, N. A., Stocco, A., & van Rijn, H. (2010). The neural correlates of problem states: Testing fMRI predictions of a computational model of multitasking. *PLoS ONE*, *5*(9), e12966. doi:10.1371/journal.pone.0012966
- Borst, J. P., Taatgen, N. A., & van Rijn, H. (2010). The problem state: A cognitive bottleneck in multitasking. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(2), 363–382.
- Borst, J. P., Taatgen, N. A., & van Rijn, H. (2011). Using a symbolic process model as input for model-based fMRI analysis: Locating the neural correlates of problem state replacements. *NeuroImage*, *58*(1), 137–147.
- Brass, M., Wenke, D., Spengler, S., & Waszak, F. (2009). Neural correlates of overcoming interference from instructed and implemented stimulus–response associations. *Journal of Neuroscience*, *29*(6), 1766–1772.
- Brett, M., Anton, J.-L., Valabregue, R., & Poline, J.-B. (2002, June). *Region of interest analysis using an SPM toolbox*. Paper presented at the 8th International Conference on Functional Mapping of the Human Brain, Sendai, Japan.
- Buckner, R. L., Andrews-Hanna, J. R., & Schacter, D. L. (2008). The brain's default network: Anatomy, function, and relevance to disease. *Annals of the New York Academy of Sciences*, *1124*, 1–38.
- Bunge, S. A., Helsen, E. H., & Wendelken, C. (2009). Left, but not right, rostrolateral prefrontal cortex meets a stringent test of the relational integration hypothesis. *NeuroImage*, *46*(1), 338–342.
- Bunge, S. A., Wendelken, C., Badre, D., & Wagner, A. D. (2005). Analogical reasoning and prefrontal cortex: Evidence for separable retrieval and integration mechanisms. *Cerebral Cortex*, *15*(3), 239–249.
- Burgess, P. W., Scott, S. K., & Frith, C. D. (2003). The role of the rostral frontal cortex (area 10) in prospective memory: A lateral versus medial dissociation. *Neuropsychologia*, *41*(8), 906–918.
- Casey, B. J., Castellanos, F. X., Giedd, J. N., Marsh, W. L., Hamburger, S. D., Schubert, A. B., . . . Rapoport, J. L. (1997). Implication of right frontostriatal circuitry in response inhibition and attention-deficit/hyperactivity disorder. *Journal of the American Academy of Child and Adolescent Psychiatry*, *36*(3), 374–383.
- Chein, J. M., & Schneider, W. (2005). Neuroimaging studies of practice-related change: fMRI and meta-analytic evidence of a domain-general control network for learning. *Cognitive Brain Research*, *25*(3), 607–623.
- Christoff, K., Prabhakaran, V., Dorfman, J., Zhao, Z., Kroger, J. K., Holyoak, K. J., & Gabrieli, J. D. E. (2001). Rostrolateral prefrontal cortex involvement in relational integration during reasoning. *NeuroImage*, *14*(5), 1136–1149. doi:10.1006/nimg.2001.0922
- Cohen, J. D., Perlstein, W. M., Braver, T. S., Nystrom, L. E., Noll, D. C., Jonides, J., & Smith, E. E. (1997). Temporal dynamics of brain activation during a working memory task. *Nature*, *386*(6625), 604–608. doi:10.1038/386604a0
- Cole, M. W., Bagic, A., Kass, R., & Schneider, W. (2010). Prefrontal dynamics underlying rapid instructed task learning reverse with practice. *Journal of Neuroscience*, *30*(42), 14245–14254.
- Collette, F., Olivier, L., Van der Linden, M., Laureys, S., Delfiore, G., Luxen, A., & Salmon, E. (2005). Involvement of both prefrontal and inferior parietal cortex in dual-task performance. *Cognitive Brain Research*, *24*(2), 237–251. doi:10.1016/j.cogbrainres.2005.01.023
- Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience*, *3*(3), 201–215.
- Costa, A., Oliveri, M., Barban, F., Torriero, S., Salerno, S., Lo Gerfo, E., . . . Carlesimo, G. A. (2011). Keeping memory for intentions: A cTBS investigation of the frontopolar cortex. *Cerebral Cortex*, *21*, 2696–2703. doi:10.1093/cercor/bhr052
- Crone, E. A., Wendelken, C., van Leijenhorst, L., Honomichl, R. D., Christoff, K., & Bunge, S. A. (2009). Neurocognitive development of relational reasoning. *Developmental Science*, *12*(1), 55–66.
- Danker, J. F., & Anderson, J. R. (2007). The roles of prefrontal and posterior parietal cortex in algebra problem solving: A case of using cognitive modeling to inform neuroimaging data. *NeuroImage*, *35*(3), 1365–1377.
- Dehaene, S., Piazza, M., Pinel, P., & Cohen, L. (2003). Three parietal circuits for number processing. *Cognitive Neuropsychology*, *20*(3), 487–506.
- Donohue, S. E., Wendelken, C., Crone, E. A., & Bunge, S. A. (2005). Retrieving rules for behavior from long-term memory. *NeuroImage*, *26*(4), 1140–1149.
- Elio, R. (1986). Representation of similar well-learned cognitive procedures. *Cognitive Science*, *10*(1), 41–73.
- Fox, M. D., Snyder, A. Z., Vincent, J. L., Corbetta, M., Van Essen, D. C., & Raichle, M. E. (2005). The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proceedings of the National Academy of Sciences*, *102*(27), 9673–9678.
- Frank, M. J., Loughry, B., & O'Reilly, R. C. (2001). Interactions between frontal cortex and basal ganglia in working memory: A computational model. *Cognitive, Affective, & Behavioral Neuroscience*, *1*(2), 137–160. doi:10.3758/CABN.1.2.137
- Frank, M. J., Samanta, J., Moustafa, A. A., & Sherman, S. J. (2007). Hold your horses: Impulsivity, deep brain stimulation, and medication in parkinsonism. *Science*, *318*(5854), 1309–1312.
- Friston, K. J., Ashburner, J. T., Kiebel, S. J., Nichols, T. E., & Penny, W. D. (2006). *Statistical parametric mapping: The analysis of functional brain images*. Amsterdam, The Netherlands: Elsevier.
- Graybiel, A. M. (2005). The basal ganglia: Learning new tricks and loving it. *Current Opinion in Neurobiology*, *15*(6), 638–644.
- Greicius, M. D., Krasnow, B., Reiss, A. L., & Menon, V. (2003). Functional connectivity in the resting brain: A network analysis of the default mode hypothesis. *Proceedings of the National Academy of Sciences*, *100*(1), 253–258.
- Hartstra, E., Kühn, S., Verguts, T., & Brass, M. (2011). The implementation of verbal instructions: An fMRI study. *Human Brain Mapping*, *32*, 1811–1824. doi:10.1002/hbm.21152
- Hill, N. M., & Schneider, W. (2006). Brain changes in the development of expertise: Neurological evidence on skill-based adaptations. In K. A. Ericsson, N. Charness, P. Feltovich, & R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance* (pp. 653–682). New York, NY: Cambridge University Press.
- Knowlton, B. J., Mangels, J. A., & Squire, L. R. (1996). A neostriatal habit learning system in humans. *Science*, *273*(5280), 1399–1402.

- Koechlin, E., Ody, C., & Kouneiher, F. (2003). The architecture of cognitive control in the human prefrontal cortex. *Science*, *302*(5648), 1181–1185. doi:10.1126/science.1088545
- Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review*, *95*(4), 492–527. doi:10.1037/0033-295X.95.4.492
- McNab, F., & Klingberg, T. (2008). Prefrontal cortex and basal ganglia control access to working memory. *Nature Neuroscience*, *11*(1), 103–107. doi:10.1038/nn2024
- Miller, E. K. (2000). The prefrontal cortex and cognitive control. *Nature Reviews Neuroscience*, *1*(1), 59–65.
- Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, *24*, 167–202. doi:10.1146/annurev.neuro.24.1.167
- Nadel, L., Samsonovich, A., Ryan, L., & Moscovitch, M. (2000). Multiple trace theory of human memory: Computational, neuroimaging, and neuropsychological results. *Hippocampus*, *10*(4), 352–368.
- O'Reilly, R. C., & Frank, M. J. (2006). Making working memory work: A computational model of learning in the prefrontal cortex and basal ganglia. *Neural Computation*, *18*(2), 283–328.
- Packard, M. G., & Knowlton, B. J. (2002). Learning and memory functions of the basal ganglia. *Annual Review of Neuroscience*, *25*(1), 563–593.
- Piazza, M., Izard, V., Pinel, P., Le Bihan, D., & Dehaene, S. (2004). Tuning curves for approximate numerosity in the human intraparietal sulcus. *Neuron*, *44*(3), 547–555. doi:10.1016/j.neuron.2004.10.014
- Piazza, M., Pinel, P., Le Bihan, D., & Dehaene, S. (2007). A magnitude code common to numerosities and number symbols in human intraparietal cortex. *Neuron*, *53*(2), 293–305. doi:10.1016/j.neuron.2006.11.022
- Reynolds, J. R., West, R., & Braver, T. (2009). Distinct neural circuits support transient and sustained processes in prospective memory and working memory. *Cerebral Cortex*, *19*(5), 1208–1221.
- Rubia, K., Smith, A. B., Brammer, M. J., & Taylor, E. (2003). Right inferior prefrontal cortex mediates response inhibition while mesial prefrontal cortex is responsible for error detection. *NeuroImage*, *20*(1), 351–358.
- Ruge, H., & Wolfensteller, U. (2010). Rapid formation of pragmatic rule representations in the human brain during instruction-based learning. *Cerebral Cortex*, *20*(7), 1656–1667.
- Schneider, W., & Chein, J. M. (2003). Controlled & automatic processing: From mechanisms to biology. *Cognitive Science*, *27*, 525–559.
- Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: 1. Detection, search, and attention. *Psychological Review*, *84*(1), 1–66.
- Spreng, R. N., & Grady, C. L. (2010). Patterns of brain activity supporting autobiographical memory, prospection, and theory of mind, and their relationship to the default mode network. *Journal of Cognitive Neuroscience*, *22*(6), 1112–1123.
- Squire, L. R., Ojemann, J. G., Miezín, F. M., Petersen, S. E., Videen, T. O., & Raichle, M. E. (1992). Activation of the hippocampus in normal humans: A functional anatomical study of memory. *Proceedings of the National Academy of Sciences*, *89*(5), 1837–1841.
- Stocco, A., & Anderson, J. R. (2008). Endogenous control and task representation: An fMRI study in algebraic problem-solving. *Journal of Cognitive Neuroscience*, *20*(7), 1300–1314.
- Stocco, A., Lebiere, C., & Anderson, J. R. (2010). Conditional routing of information to the cortex: A model of the basal ganglia's role in cognitive coordination. *Psychological Review*, *117*(2), 541–574.
- Stocco, A., Lebiere, C., O'Reilly, R. C., & Anderson, J. R. (2010). The role of the anterior prefrontal–basal ganglia circuit as a biological instruction interpreter. *Frontiers in Artificial Intelligence and Applications*, *221*, 153–162.
- Strange, B. A., Henson, R. N., Friston, K. J., & Dolan, R. J. (2001). Anterior prefrontal cortex mediates rule learning in humans. *Cerebral Cortex*, *11*(11), 1040–1046.
- Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., . . . Joliot, M. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage*, *15*(1), 273–289. doi:10.1006/nimg.2001.0978
- Volle, E., Gonen-Yaacovi, G., de Lacy Costello, A., Gilbert, S. J., & Burgess, P. W. (2011). The role of rostral prefrontal cortex in prospective memory: A voxel-based lesion study. *Neuropsychologia*, *49*, 2185–2198. doi:10.1016/j.neuropsychologia.2011.02.045
- Wendelken, C., Nakhachenko, D., Donohue, S. E., Carter, C. S., & Bunge, S. A. (2008). “Brain is to thought as stomach is to ??”: Investigating the role of rostralateral prefrontal cortex in relational reasoning. *Journal of Cognitive Neuroscience*, *20*(4), 682–693. doi:10.1162/jocn.2008.20055
- Yin, H. H., & Knowlton, B. J. (2006). The role of the basal ganglia in habit formation. *Nature Reviews Neuroscience*, *7*(6), 464–476.