The Biological Basis of Rapid Instructed Task Learning

by

Michael William Cole

Bachelor of Arts, University of California, Berkeley, 2003

Submitted to the Graduate Faculty of

Arts and Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2009

UNIVERSITY OF PITTSBURGH

THE SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

Michael William Cole

It was defended on

June 3, 2009

and approved by

Raymond Cho, M.D., Psychiatry & Psychology

Marc Sommer, Ph.D., Neuroscience

Anthony Wagner, Ph.D., Neuroscience & Psychology, Stanford University

Mark Wheeler, Ph.D., Psychology

Chairperson: Julie Fiez, Ph.D., Neuroscience & Psychology

Dissertation Advisor: Walter Schneider, Ph.D., Psychology

THE BIOLOGICAL BASIS OF RAPID INSTRUCTED TASK LEARNING

Michael W. Cole, PhD

University of Pittsburgh, 2009

The uniquely human ability to rapidly learn novel tasks from instructions is extremely important in everyday life, and yet its evolutionary origin and basis in the brain remain mysteries. In order to address these gaps in scientific knowledge, comparative human-monkey studies were consulted to predict the human brain areas involved in rapid instructed task learning (RITL). These predictions were tested using functional MRI (fMRI), magnetoencephalography (MEG), and a novel cognitive paradigm developed to systematically investigate the neural basis of RITL for the first time. In accordance with cross-species neuroanatomical differences, anterior prefrontal cortex (aPFC), anterior temporal lobe (aTL), dorsolateral prefrontal cortex (DLPFC), and posterior parietal cortex (PPC) were found to be involved in RITL. DLPFC and PPC formed a network involved in loading individual task semantics into working memory, while aPFC and aTL formed a network involved in integrating semantics in preparation for task performance. Both networks supported novel task set formation, which occurred in a bottom-up manner (semantic loading, then integration), and practiced task set retrieval, which occurred in a top-down manner (integration retrieval, then semantic loading). These findings suggest that RITL relies upon semantic loading by DLPFC and PPC, but that aPFC and aTL support semantic integration both dynamically during RITL and from long-term memory after extensive practice. More broadly, the findings suggest RITL is enabled in humans via a combination of enhanced symbolic processing (language), enhanced working memory manipulation (aPFC), and enhanced integrated semantic representation (aTL).

The present document begins with a broad overview of RITL and related topics, such as its relation to animal cognition, other forms of learning, and cognitive control. These topics support several novel hypotheses regarding RITL and its likely basis in the brain. The fMRI study is then presented, verifying many of the hypotheses developed in the previous section. The MEG study is reported next, clarifying many of the questions about timing and causality suggested by the fMRI results. Finally, a general discussion integrates the results from both studies, expanding conclusions with an overview of brain connectivity findings, cross-species differences, and the role of neural hierarchies in RITL and cognition generally.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

In life, and in conducting the research reported here, I have had the distinct privilege of rapidly learning a great deal from some extraordinary instructors. Foremost among these instructors are my parents, whose lessons and support were essential for who I am today and ultimately what I accomplished here. Of course, they were not alone in these efforts, and I gladly thank my entire family (especially my aunt). My academic advisor Walter Schneider was instrumental in inspiring my interest in rapid task learning and my excitement for research generally. Feedback and support from my committee was essential to the quality of the work as well as to my endurance while conducting it. Anto Bagic and Robert Kass were especially adept instructors, whose support greatly improved data collection and analysis. Todd Braver provided wonderful conceptual and methodological feedback, as well as the immensely valuable means to expand upon this research in the future.

Of course, I would especially like to thank Bruna Martins, whose feedback and support during many long hours of work made life and this project of much higher quality.

Members of the Schneider lab, the UPMC MR Research Center, and the UPMC CABMSI were immensely helpful for data collection.

Finally, I would like to thank the broader academic community, especially the CNBC, the LRDC, and the CNUP, as well as the sources of my academic funding, an NSF Graduate Student Fellowship and an NSF IGERT Fellowship via the CNBC.

A man of clear ideas errs grievously if he imagines that whatever is seen confusedly does not exist; it belongs to him, when he meets with such a thing, to dispel the mist, and fix the outlines of the vague form which is looming through it.
– J.S. Mill, *Utilitarianism*

To study instructed learning is to study how the brain adapts itself to culture. It is to study the primary way in which humanity has distinguished itself among the creatures of the Earth.
– D. Noelle (1997)

1.0 GENERAL INTRODUCTION

Humans are rarely alone when learning something for the first time. Whether verbally, from text, or through imitation, we are uniquely capable of rapidly learning new tasks from instruction. This form of learning allows for rapid adaptation to novel situations and transfer of previously learned skills, increasing the cooperativity, productivity, and survivability of our species. The importance of rapid instructed task learning (RITL; pronounced "rittle") for both individuals and the species as a whole seems to be an excellent reason for there to be a large body of reasearch into its neural and psychological basis. Strangely, however, very little research has been conducted in this area. Here RITL is systematically characterized both behaviorally and neurally for the first time, making novel connections with research conducted in the related areas of memory, learning, and executive control.

The need for research in this area is not a new realization. It has been known for some time that the inability to adapt to novel situations is extremely debilitating for day-to-day life, and is perhaps the defining characteristic of disorders of both executive function and prefrontal cortex (PFC) (Rabbitt, 1997; Norman and Shallice, 2000). However, research in this area has eluded proper scientific investigation, perhaps because of methodological limitations. Indeed, according to Burgess (1997):

> The construct validity of a test whose purpose is to measure adaptation to novelty reduces with every testing. This is undoubtedly one of the key obstacles to our understanding of executive functions, and

> no doubt explains much of the "riddle of the frontal lobes" as Teuber (1964) called it: the purpose of non-routine processing is to make itself redundant.... The purest method of studying individuals' behavior in novel situations is to observe a large group on their first exposure and then attempt to extract general principles from their behavior.

The present work investigates situational novelty both within-subject and in a carefully controlled quantifiable manner for the first time, bringing us one step closer to solving the "riddle of the frontal lobes".

1.1 RITL AS A COGNITIVE PROCESS

> While I, no doubt, have learnt something extra about chick sexing, I was also reminded that man's abilities are limited only by his thinking.
> — R. D. Martin, as quoted by (Horsey, 2002)

Tasks consist of coordinated set of rules, or task sets, that specify how to act in a given context. Once basic skills and concepts are learned early in life, humans have the remarkable ability to rapidly learn novel tasks from instruction. The key capabilities that likely underlie RITL are the ability to integrate and recombine basic rules into novel task sets (generalizability) and the ability to communicate and understand the task sets of others (communicatability).

Tasks are considered here as points in a multi-dimensional state space of sets of rules, and RITL as transferring the specification of points in that space (Figure 1). Accordingly, a novel cognitive paradigm has been developed to traverse a portion of the task state space by permuting 12 rules such that 64 different task sets are available. A portion of task state space was used that consisted of relatively familiar rules, but in combinations that were quite likely novel for each participant so that RITL could be observed. Task instructions were given linguistically. Imitation

might have been used instead, but it is argued here that language is the most concise and accurate means of transferring task state specifications between individuals.

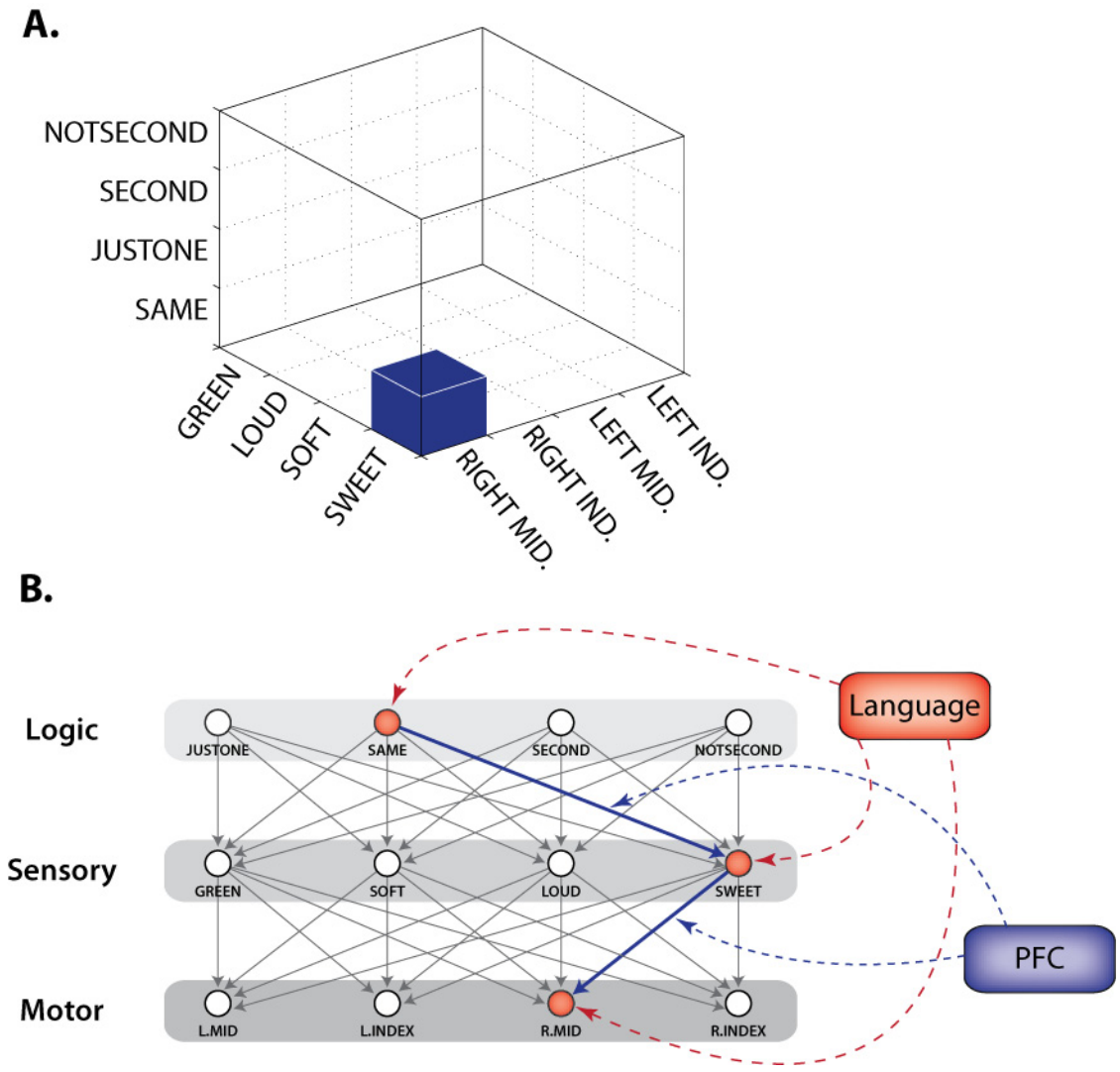

Figure 1 – Task specification in a task state space
A) The task state space for the PRO paradigm. Each task dimension is a spatial dimension in the Cartesian coordinate system. 64 (4 X 4 X 4) points exist in the three-dimensional grid. The depicted cube represents the task SAME – SWEET – RIGHT MID. ("If the answer to 'is it SWEET' is the SAME for both words, press the button with your RIGHT MIDDLE finger"). **B**) A depiction of the task search space in a neurally-inspired form. Different semantics are stored in different locations, with unit subpopulations representing each rule. Language can directly reference and activate these task rule semantics. PFC sustains these activations and integrates them by affecting connectivity between them.

Learning from feedback (i.e., 'trial-and-error' learning) is an alternative form of task learning. However, this form of learning can be extremely slow. For instance, Biederman &

Shiffar (1987) found that "chicken sexers" (who determine the sex of day-old chickens) in the poultry business learn their trade gradually by observing many thousand examples over several years. In contrast, by instructing the students using rules specifying what relationships to look for, Biederman & Shiffar were able to train them to perform as well as experts in a matter of minutes. This example demonstrates the tremendous power of rule-based instruction for speeding learning.

Several cognitive processes likely underlie RITL. First, there is a *task set formation* phase, in which task set components are interpreted from instructions (instruction interpretation) and integrated into a complete task set (working memory integration). Second, there is *task set maintenance*, in which the formed task set is maintained. Finally, there is *initial task set implementation*, in which a task set is applied to a stimulus (*decision-making*) for the first time. This research is the first to observe the brain processes underlying each of these processes to increase understanding of RITL.

Bovair & Kieras (1991) were one of the first to characterize the cognitive processes underlying RITL, through the development of a production-rule model of skill acquisition from text (e.g., reading a computer manual or a recipe to cook lasagna). In one of the only papers directly addressing this topic, Bovair & Kieras (1991) point out that 'procedure acquisition' from text is poorly understood[1] despite it taking up an estimated 75 percent of reading on the job (Sticht, 1977).

---

[1] A situation little remedied in the last 18 years; note that there were only 18 citations of Bovair & Kieras over that time period

Bovair & Kieras (1991) used the few findings in this area to develop a cognitive model of skill acquisition from text. They suggest that RITL occurs primarily during the procedure comprehension process, which consists of "generating the correct set of rules from the input propositions". These rules typically consist of a series of productions, or if-then statements, similar to those in Anderson's ACT models (Anderson, 1996). Bovair & Kieras go on to explain their model's various hypothesized processes for skill acquisition from text. They explain that the *initial reading comprehension* process is nearly identical to other forms of reading such as technical prose or narrative text. Like with these other forms of reading the text symbols are converted into semantic representations. However, their understanding of the next step, *procedure construction*, is "quite limited". It consists mainly of converting from the typically limited set of semantic representations from the text into fully specified declarative production rules. Thus, the core processes behind procedure construction from text is left to the imagination of the reader. The present research seeks to remedy this problem by providing some concrete observations of these core processes.

The experimental study of RITL provides a unique dilemma. On the one hand, RITL requires a novel task rule so the very first use of that rule can be observed. This requires a complex, arbitrary task rule very unlikely to have been experienced by the subjects before. On the other hand, the first experience with the rule must be repeated many times in order to attain proper statistical estimates of the associated behavior and brain activity. The difficulty of satisfying these two constraints may explain why there is so little research in this area to date.

The present research involves the development of a behavioral task, the permuted rule operations (PRO) paradigm, to satisfy these two constraints (see Figure 2 and section 2.4.3). The PRO paradigm combines a set of simple rule components in many different ways, creating

dozens of complex task rules certain to be novel to subjects. The present work used functional magnetic resonance imaging (fMRI) and magnetoencephalography (MEG) in combination with this new task paradigm to characterize the biological basis of RITL.

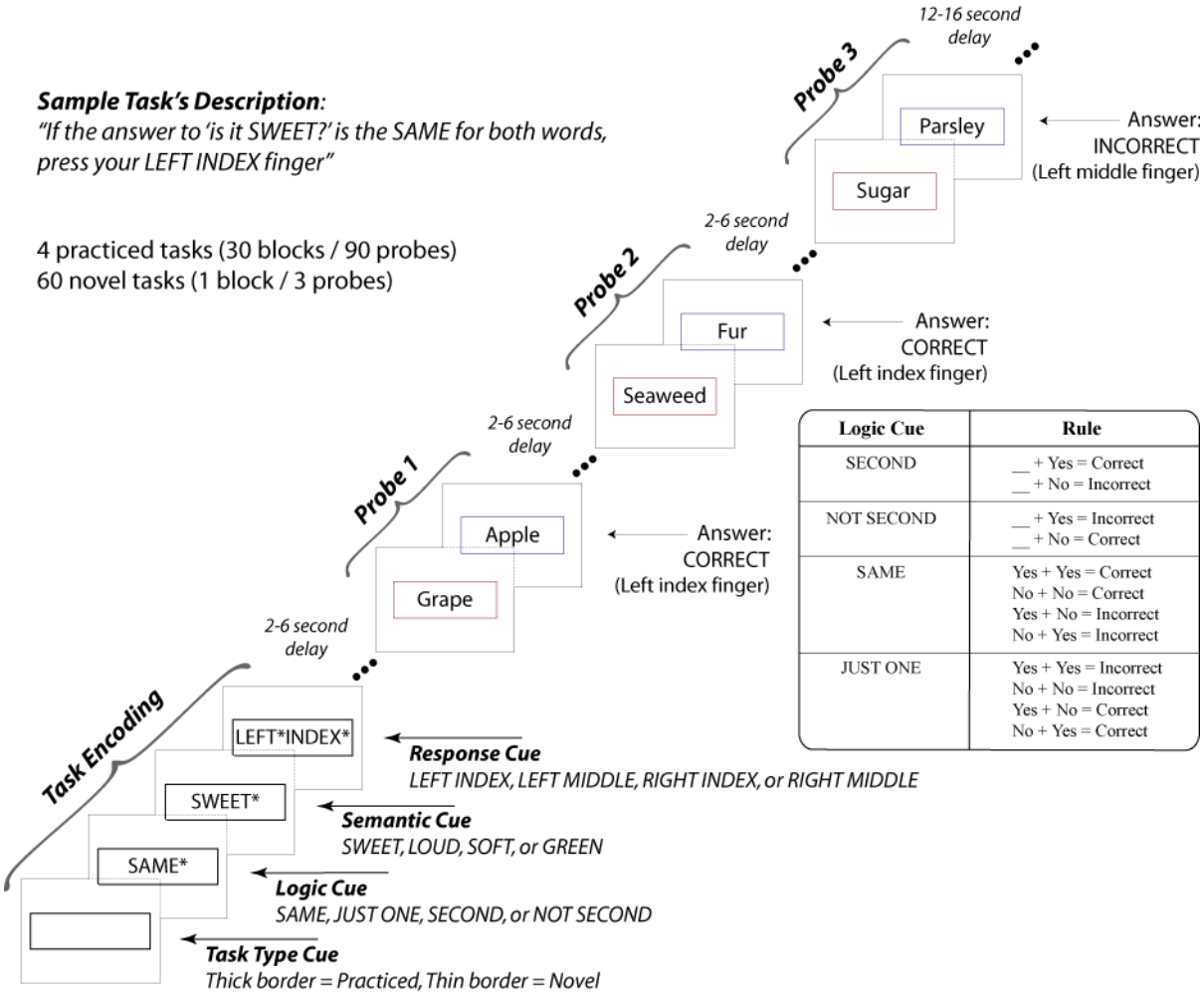

Figure 2 – The PRO cognitive paradigm
The permuted rule operations (PRO) paradigm involves permuting 12 task rules to create 64 unique tasks. Four of these tasks (counterbalanced across subjects) were practiced in a prior behavioral session, while the remaining 60 were completely novel. Importantly, the individual rules were equally practiced, while the task sets were not. Each task consisted of a logic, a sensory semantic, and a response rule, with four possible rules of each type. A block consisted of four encoding screens followed by three probes. Variable delays were present before probes (2-6 seconds) and between blocks (12-16 seconds).

1.2 RAPID LEARNING IN HUMANS AND NON-HUMANS

> Although we are made up of the same chemicals, with the same physiological reactions, we are very different from other animals. Just as gases can become liquids, which can become solids, phase shifts occur in evolution, shifts so large in their implications that it becomes almost impossible to think of them as having the same components.
> – M. Gazzaniga, *Human: The Science Behind What Makes Us Unique (2008)*

> Kanzi knows more about human language than humans know about bonobo language.
> – M. Gazzaniga, *Human: The Science Behind What Makes Us Unique (2008)*

There appears to be just three major ways for humans to learn tasks: exploration, conditioning, and RITL. The first two are shared most fully with our primate relatives.

Exploration is the most rudimentary of the three learning methods. It is a typically slow process by which many stimulus-response combinations are tried until the correct (i.e., reinforcing) set is learned via feedback from the environment. Tasks learned via exploration are not instructed, per se, though the subjects eventually learn to do the tasks. Even here not all animals are equal, as an abstract rule being reinforced for one species may not be conceptualized (and thus not reinforced) by another species.

Conditioning can allow some tasks to be learned very quickly via instruction. For instance, B. F. Skinner used operant conditioning to train a dog to open a step-base wastebasket on command in under 20 minutes (Skinner, 1951; Roddy, 1952). Figure 3 illustrates this process. Skinner used shaping and the method of successive approximation, wherein the animal is reinforced first liberally (for behaviors roughly similar to the desired one) and more conservatively over time (until only the desired behavior is reinforced). Operant conditioning is

thus limited to "shapable" behaviors that must come (at least in part) spontaneously from the subject so they can be reinforced.

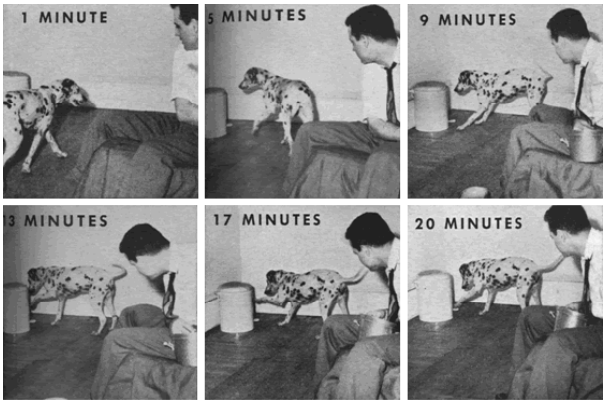

Figure 3 – B. F. Skinner trains a dog to open a wastebasket on command in 20 minutes
This is one of the fastest demonstrations of training an arbitrary behavior using operant conditioning. Figure from *LOOK* magazine (Roddy, 1952).

Most of the time conditioning is extremely slow. The worst case (typical of monkeys learning abstract cognitive tasks) involves a random search in which many instances of a task rule are repeated and "stick" only after hundreds of thousands of repetitions (e.g., Fuster and Jervey, 1982). Thus, conditioning of abstract rules with monkeys is in some sense akin to exploratory learning in humans. In contrast, simple "shapable" behaviors, such as learning to saccade on command, can be learned by monkeys in a matter of minutes since they involve reinforcing an innate reflex (fixating a bright spotlight).

While learning a task in 20 minutes is impressive, RITL can allow for task learning in seconds (Monsell, 1996). RITL is also not limited by "shapable" behaviors, but can be used to learn almost any sequence of thoughts or actions, especially when language is used. The advantages and limitations of RITL are explored in later sections.

RITL via explicit symbolic instruction is largely unique to humans. However, there is one notable exception to the human-only status of RITL. A bonobo chimpanzee named Kanzi learned to perform simple RITL (e.g., "put the apple on the soap") using symbolic and even English language instructions to implement tasks immediately after hearing instructions (Savage-Rumbaugh, Murphy et al., 1993; Lloyd, 2004). It was shown that Kanzi had attained RITL capabilities at the level of a 2½ year-old human.

Kanzi is a remarkable case not only because he was able to learn language but also because of the manner in which he learned it. When Kanzi was young he was occasionally brought into the training room while his wild-caught mother was being trained to use a symbol keyboard to refer to objects and actions. She was unable to learn these symbol-semantic associations, though Kanzi was able to learn some without explicit training. Once the experimenters realized this they began explicitly training Kanzi to form these associations. Amazingly, Kanzi not only learned the symbol-semantic associations for over 100 concepts, but also learned their English equivalents. Kanzi had learned rudimentary English by listening to the sounds the experimenters made during training.

Kanzi's language abilities were tested using RITL (Figure 4). An experimenter moved to a separate room (to avoid the possibility of non-verbal cues) and verbally gave Kanzi instructions to perform simple tasks in English. These tasks involved words, objects, and actions Kanzi knew, but in configurations he had never heard or performed before (e.g., "put your ball on the pine needles"). Kanzi was able to perform the 660 single-sentence instructions with 74% accuracy (in contrast to 65% for a 2 year-old human with the same sentences).

Kanzi's RITL ability, rather than his language ability, was probably the most remarkable aspect of this experiment. Kanzi was trained on symbolic representation using rote

memorization, yet was able to immediately and spontaneously perform RITL using these symbols. This suggests that the mechanisms of RITL were innate to Kanzi; that the brain structures essential to RITL had already evolved in our common ancestor with bonobo chimpanzees. The similarities (and differences) between human and monkey RITL behavior likely stem from cross-species similarities (and differences) in neuroanatomy.

Figure 4 – Kanzi the genius bonobo chimpanzee follows verbal instructions
Kanzi follows the verbal instructions to "Knife the sweet potatoes" on the first try, illustrating RITL performance in bonobo chimpanzees (Savage-Rumbaugh et al., 1993). Kanzi was 74% accurate for following verbal instructions. Figure from (Savage-Rumbaugh et al., 1993).

Much has been made of the role of PFC in higher-level cognition indicative of human thought such as RITL, especially the most anterior portion, the anterior PFC (aPFC) (Fuster, 2001). If aPFC (area 10) is involved in RITL, it is expected that aPFC development should reflect RITL abilities across species. Consistent with this hypothesis, Semendeferi et al. (2001) showed that aPFC developed tremendously (both in size and connectivity with other associative areas) between human and non-human primates. Importantly, among higher primates bonobo chimpanzees are second only to humans in aPFC development (i.e., total and relative size, and

connectivity). In conjunction with Kanzi's RITL abilities (and the lesser abilities demonstrated by gorillas and common chimpanzees: Williams, Brakke et al., 1997) this is highly suggestive that aPFC is tied to the neural basis of RITL. This also suggests that rudimentary RITL abilities may have evolved along with aPFC, which first appeared very early (at least 6-8 million years ago) in primates (Brakke and Savage-Rumbaugh, 1995). It is clear, however, that aPFC exploded in size (it is ~5 times larger in humans than in bonobo chimpanzees (Semendeferi et al., 2001); ~2 times larger if overall brain size differences are taken into account (Passingham, 2009)) and brain-wide connectivity with humans, along with a tremendous increase in RITL abilities.

Another part of the brain, the temporal lobe, is much larger (12X) for humans than monkeys (Passingham, 2009). Beyond simply being larger, it has been observed that the human temporal lobe is larger than predicted for a monkey with a human-sized brain (Rilling and Seligman, 2002). This finding suggests that the temporal lobe may be important for advanced human behaviors like RITL. Recent evidence suggests that the anterior temporal lobe (aTL) is especially important for storing conjunctive semantics, like those necessary for integrating rules during RITL (Rogalsky and Hickok, 2008; Taylor, Stamatakis et al., 2009). Accordingly, aTL is expected to play an important role in RITL.

Other cognitive control regions are likely to be both more highly developed in humans and also involved in RITL. For instance, posterior parietal cortex (PPC) is necessary for working memory (Quintana and Fuster, 1999; Cole and Schneider, 2007), which is essential for maintaining task components during RITL. A recent comparative functional study of macaque monkey and human posterior parietal cortex (PPC) found that an anterior portion of PPC is evolutionarily new in humans (Orban, Claeys et al., 2006). They found that this new subregion is sensitive to the perception of three-dimensional shapes. This is unlikely to be a function

necessary for RITL, but their finding points to the possibility of other portions of PPC (not tested in that comparative study) that might be specialized for processes necessary for RITL in humans, such as working memory (along with dorsolateral prefrontal cortex (DLFPC)) and language.

A recent morphometric study comparing human and common chimpanzees found many of the same regions mentioned above are significantly larger in humans than chimpanzees (Figure 5). The differences in regional brain size imply that these regions underwent extensive evolutionary change. I hypothesize that RITL was one of the major selectively advantageous traits leading to modern humans, predicting that the brain regions with the most change will be heavily involved in performing RITL.

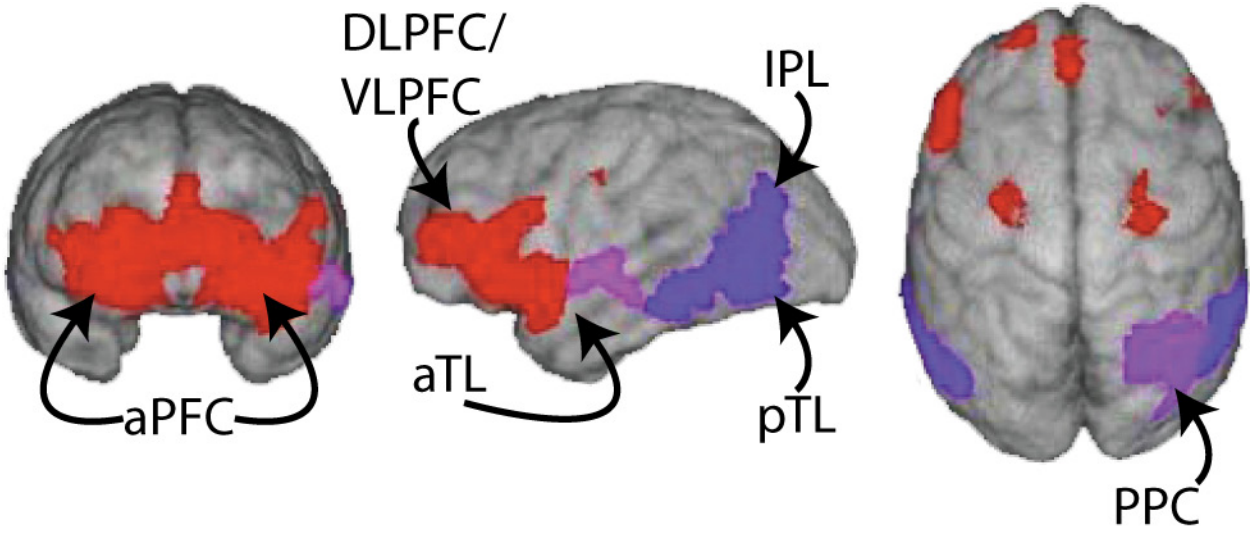

Figure 5 – Predicted RITL brain regions based on larger size in humans than chimpanzees
Avants et al. (2006) used spatial morphing between chimpanzee and human brains to identify the parts of the brain that significantly changed in size since the human-chimpanzee common ancestor. These are the brain regions that likely evolved the most in that time. I hypothesize that RITL was one of the major selectively advantageous traits leading to modern humans, predicting that the brain regions with the most change will be heavily involved in performing RITL.
The brain depicted above is a chimpanzee brain. Figure adapted from (Avants et al., 2006).
Regions: anterior prefrontal cortex (aPFC), dorsolateral/ventrolateral PFC (DLPFC/VLPFC), anterior temporal lobe (aTL), posterior TL (pTL), inferior parietal lobe (IPL), and posterior parietal cortex (PPC).

1.3 TASK LEARNING AS NAVIGATION THROUGH A STATE SPACE

> The use of punishments and rewards can at best be [only] a part of the teaching process. Roughly speaking, if the teacher has no other means of communicating to the pupil, the amount of information which can reach him does not exceed the total number of rewards and punishments applied. By the time a child has learnt to repeat "Casabianca" he would probably feel very sore indeed, if the text could only be discovered by a "Twenty Questions" technique, every "NO" taking the form of a blow. It is necessary therefore to have some other "unemotional" channels of communication. If these are available it is possible to teach a machine by punishments and rewards to obey orders given in some language, e.g., a symbolic language. These orders are to be transmitted through the "unemotional" channels. The use of this language will diminish greatly the number of punishments and rewards required.
> – A. Turing, in *Computing Machinery and Intelligence* (1950), in which the Turing Test was first proposed

Task learning can be conceptualized as a form of problem solving. Like problem solving, task learning is goal directed, involves subgoaling, and operator application (using known actions or rules in performing the task). Also, like problem solving, task learning can be conceptualized as a search through a problem/task state space: The participants start in one state and must search for the goal state (Newell and Simon, 1972), which is the correct task set for task performance (see Figure 1A).

RITL likely speeds task learning tremendously by paving a clear path through the search space of possible tasks. The other forms of task learning mentioned above (exploration and conditioning) are less efficient for navigating this task search space. Also relevant is the concept of immediate transfer (Bovair et al., 1991), which may allow slight modifications of previous solutions to cut through the task search space efficiently. More generally, each additional constraint on correct task performance provided (through detailed instruction, examples, feedback, etc.) is likely to aid subjects in finding the correct task set for proper task performance.

In light of the similarity of their component processes, previous research on problem solving may apply to RITL. For instance, symbolic instruction was shown to be more efficient than exploration for problem solving (see Anderson, 2000). It appears very likely that this will also be true for task learning. Note that including examples provides even further improvement beyond symbolic instruction for problem solving (Reed and Bolstad, 1991).

Evolutionarily, problem solving abilities have tracked with RITL. This supports the notion that common cognitive and neural substrates underlie both behaviors. Advanced problem solving requires working memory and subgoal processing. For instance, Kohler (1927) showed advanced problem solving in chimpanzees involving tool use. The goal, to get food beyond the chimp's reach, was solved by putting two sticks together to create a tool (a subgoal), and using this new tool to get the food.

RITL likely also requires subgoaling, as the act of task learning is taken on as a subgoal for the larger goal of performing the task correctly. The act of pulling together known rules and actions into a single task is similar to tool creation in that the final product is used to accomplish goals. It is likely that aPFC is primarily responsible for subgoal processing (i.e., integrating the task components in working memory) in the context of RITL (Braver and Bongiolatti, 2002).

1.3.1 The role of language in RITL

> The most powerful tool at the teacher's disposal is language. Utterances can present exercises, draw attention to salient aspects of a problem, praise success, and gently correct errors, but they can do more: Language allows for the direct communication of general and abstract knowledge from one person to the next.
> – D. Noelle (1997)

If possible tasks form a multidimensional state space, how do instructions specify points in that space? In other words, how do instructions both precisely and concisely specify what they are referring to (the procedural semantics)?

The 'task state space' framework described above predicts that the more directly a instructional method can refer to a set of task semantics, the more precise and concise (rapid) that method will be. Of the different forms of task learning, learning is most rapid via language or imitation. This suggests that language and imitation can more directly reference task semantics than exploratory learning or operant conditioning. Indeed, imitation involves a direct mapping between instructions and task semantics (body parts) via 'mirror neurons' (di Pellegrino, Fadiga et al., 1992; Goldenberg and Karnath, 2006), while language involves arbitrary mappings between instructions and task semantics (Hurford, 2001), unsupportable by 'mirror neurons' and limited only by the language used.

The extremely direct mapping between imitated instruction (others' body movements) and task semantics (self body movements) suggests that imitation is the most precise and concise means for task learning. However, imitation is severely limited in its ability to refer to non-embodied task semantics. For instance, visually checking for traffic before crossing a street can be readily learned via imitation (because of the associated body movements) but imitation alone cannot be used to learn to listen carefully for traffic coming around a blind corner. Neither can exploratory or operant conditioning be used here, as a single negative outcome could be your last. Task learning from language is the most viable option, since words ('listen for traffic') can refer directly to a behavior (see Figure 1B) hidden from imitation without the need for any (potentially life-ending) feedback. Note that this example can easily be extended to our

evolutionary past (e.g., listening for a lion on the savannah), with attending implications for evolutionary selective advantages.

Learning something as concrete as listening for traffic is just the tip of the iceberg for task learning from language. Indeed, unlike imitation, learning via language can refer to abstract and generalized concepts. These abstract concepts are not tied to the body, and can include logical relations and other broad generalizations. This opens up the 'task state space' in ways limited only by our (ever-expanding) language and our cognitive capacities.

One might argue that imitation *can* be used to learn abstract or non-embodied tasks, based on multiple instances. For example, in the first instance an instructor could look both ways before crossing, and stop for a car seen in the distance. For the next instance, the instructor may stop for no apparent reason, but then a car comes around the blind corner. The learner may deduce that the instructor was listening for the car, and can imitate that in the future.

The use of multiple instances and induction increases the usefulness of imitation (as well as exploratory learning and operant conditioning), but each additional instance necessary for the induction reduces the conciseness (and potentially the preciseness) of the learning method. As task instructions become more abstract, so many instances are eventually necessary to specify the logical induction that these forms of learning break down. Learning from language does not have this problem since it can refer directly to the abstract concept underlying the multiple instances. For example, the task 'lift the red items' is readily specified linguistically, but might require many trials via the other forms of task learning (e.g., lifting of a red dotted hat, lifting of a red dotted shirt ['lift red dotted clothes?'], lifting of a red striped shoe ['lift red clothes?'], lifting of a red striped ball ['ah, lift the red items']). In other words, many more tasks can be specified with

language than is either practical or possible with other forms of task learning. For this reason, language was used in the PRO paradigm for investigating RITL.

1.3.2 RITL as fundamental to language

It was demonstrated above that language is not necessary for RITL. However, it is possible that *RITL* is actually necessary or fundamental for *language*. As pointed out by Noelle (1997), "Most every speech act may be seen as an attempt to instruct another person – as an attempt to modify their behavior through the use of language".

Further, Noelle points out:

> The use of language for instruction may also form much of the foundation of explicit cognition. After evolving to instruct others, humans have evolved to instruct themselves. This may be the root of much of the systematic and controlled behavior exhibited by humans.

The current research will test this notion, checking if the same regions used to prepare a task from external instruction are also those regions used to prepare a task from memory. Future research will need to determine if these same brain regions are used for self-generated tasks as well.

1.3.3 Optimizing task learning

Expanding upon the notion of a 'task state space', one might ask how specifying tasks via instruction can be optimized from an information-theoretic perspective (Shannon, 1948; Cover and Thomas, 2006). This would entail the features that maximize mutual information between

the instructor's task set representation and the learner's task set representation following communication. In other words, what are the features that maximize the conciseness (speed) and preciseness (accuracy) of communication in the state space of possible tasks?

We have already identified one such feature: the ability for the instructor to directly refer to the target task state. The optimal means to achieving this feature would be to have a large database of symbolic-semantic associations that can be combined to form novel sets. While not completely optimal, human language fits this description. According to information theory, current languages could become more optimal by having greater orthogonality between symbols (fewer homophones/homonyms; e.g., 'flour' vs. 'flower') and semantics (fewer words with the same meaning; e.g., 'skinny' vs. 'thin').

Another feature of optimal task learning is maximal compression of information. Human language fits this description in that it can concisely refer to task rules that apply to a large variety of situations (i.e., abstract rules). As explained in the previous section, other forms of task learning typically require many learning instances to come to an abstract/generalizable task rule. For language, a single word in the right context can specify behavior over an infinite variety of situations.

Finally, optimal task learning requires a convention for converting multiple symbols into a coordinated set of task semantics. For imitation this is straightforward: the order of observed body movements is simply copied in the new behavior. This is less straightforward for language, since the specified task rules may be given in a different order than the task execution. However, human language has consistent syntax, which helps provide relational meaning between the semantics of the task instructions.

In summary, in order for RITL to be maximally flexible, it requires a large database of orthogonal symbolic-semantic associations, of varying abstraction, which can be combined and interpreted in arbitrary sets. The English language largely fits this description. In order to better approximate these optimal conditions, the PRO paradigm used a set of orthogonal symbolic-semantic associations based on English words to utilize previously learned associations (lexically optimal), rules at three levels of abstraction (highly compressed), and consistent within-subject presentation order for interpreting the symbols (syntactic consistency). This limited 'language' was optimized to communicate the broadest range of possible tasks in the (included) task state space as quickly and effectively as possible (see Figure 2).

1.3.4 Accounting for non-rapid instructed task learning

Everyday experience tells us that there are some tasks that can be instructed but are inevitably slow to learn. For instance, learning to play tennis is initially sped by learning the sport's rules via language, but quickly slows when one needs to actually hit the ball. Imitation is often used to improve performance of complex motor skills, but exploratory learning (with feedback enhanced by a trainer) typically takes over after some time. This suggests that the motor skill task state space is quite large, and that neither language nor imitation can fully specify the components of a complex motor task. This is likely because motor effectors have many free parameters, especially when implementing spatiotemporally dynamic processes like hitting a tennis ball. Only simple motor acts (single button presses using one of four possible fingers) are used here, making future research necessary to determine the limits of language and imitation in specifying complex motor tasks.

1.4 COMPUTATIONAL MODELS OF INSTRUCTED LEARNING

While there has been little experimental research on RITL, several groups have contributed important theoretical/computational advances in this area. Around the same time as the theoretical work of Bovair & Kieras (see section 1.1), Schneider & Oliver (1991) made further advances by implementing CAP2, a computational model that can learn from instruction (like a symbolic model) and also from experience (like a traditional connectionist model). Critically, unlike traditional connectionist models (which can take tens of thousands of trials to learn), CAP2 can learn in a more realistic number of trials (~20), closer to the number of trials required by humans (1 trial, 2 second reaction time). Note that these simple tasks involved learning how to respond to logic gates, which were certain to have been experienced before (though in a different context). A more complex logic gate problem took 10,835 trials to solve in a traditional connectionist model, and only ~300 trials for both humans and CAP2.

The key innovation of CAP2 was the separation of fast-learning/rule-based processes from slow-learning/associative processes. This separation has been characterized as mirroring the controlled and automatic processing stages in human skill acquisition (Schneider and Shiffrin, 1977; Shiffrin and Schneider, 1977). Importantly, this separation allows for a quick-to-learn yet fragile strategy (controlled processing, utilizing working memory) with limited practice, which can transition into a slow-to-learn yet robust strategy (automatic processing) after extensive practice. The present work hypothesizes a third stage, prior to regular controlled processing: RITL. This extremely early stage, in which the task rules are integrated for the very first time, likely utilizes controlled processing with additional processes involving instruction input and working memory integration.

The CAP2 model is revolutionary in its ability to learn tasks from instruction. However, the model has several limitations: 1) its ability to only learn sequences of actions, 2) the need to instruct the model non-linguistically, specifying the exact control signals necessary at each step, and 3) the assumption of complex connectionist structures without sufficient evidence from neuroscience (though this has been largely remedied recently, see Schneider and Chein, 2003; Chein and Schneider, 2005; Cole et al., 2007). A more recent computational model (Noelle and Cottrell, 1996; Noelle, 1997) demonstrated instructed learning in a network similar to CAP2, but with several of its limitations resolved. For instance, similar to humans, the Noelle model learns a language during early task learning, which is later used to provide task instructions to speed learning of novel tasks. Also, the Noelle model is not limited to sequences of actions, as it can learn novel categories from task instructions. Finally, while some connectionist structures are assumed in the Noelle model, fewer are assumed than CAP2.

A recently developed computational model (Rougier, Noelle et al., 2005) has several of the same features described above but in a more biologically realistic framework (see O'Reilly and Munakata, 2000). Rougier et al. created a model that could be trained to perform the Wisconsin Card Sort Task and Stroop task, using a large set of neuroscientific constraints based on key principles of PFC function. These well-established principles were, 1) active maintenance (tonic up and down states for neural populations), 2) adaptive updating (gated by dopamine/basal ganglia), and 3) PFC modulation of processing in other cortical areas. The model showed that these principles, along with Hebbian learning, were both sufficient and essential for (slow, >130000 trials) learning of task rules. Evidence for this came in part from lesioning the PFC module, which caused symptoms like those of PFC-damaged human patients (e.g., perseveration). Future research is necessary to build upon this model to implement rapid (single

trial) learning of tasks based on conjunctions of previously learned rules in a biologically realistic framework.

1.5 TASK PREPARATION AND THE NEURAL BASIS OF COGNITIVE CONTROL

Hundreds of studies have converged on a set of regions involved in cognitive control processes. These regions, which form a cognitive control network (Cole et al., 2007), include dorsolateral prefrontal cortex (DLPFC), anterior cingulate cortex / pre-supplementary motor area (ACC/pSMA), pre-motor cortex (PMC), anterior insula cortex (AIC), inferior frontal junction (IFJ), and posterior parietal cortex (PPC). The anterior portion of lateral prefrontal cortex (aPFC) has also been implicated in cognitive control processes. Cognitive control is defined here as the set of processes responsible for detecting (potentially before an overt error has been made) and correcting behavior (or lack of behavior) incompatible with the current task goals. This can include incompatible behavior due to *lack of* training with the correct task behavior, or incompatible behavior due to training *conflicting* with the correct task behavior.

A recent study showed that some of the control network regions maintain their activity over entire task blocks, while others are active only during trial events (Dosenbach, Visscher et al., 2006). A mixed blocked/event-related fMRI design was used to separate sustained from transient brain signals. Their results, which showed consistent sustained activity only in AIC and ACC/pSMA, are in sharp contrast with a large literature supporting DLPFC as a sustainer of information over time. Many studies have found sustained activity within DLPFC in many working memory contexts, both in human and non-human primates (Fuster, 2001; Curtis and D'Esposito, 2003).

It is likely that Dosenbach found sustained activity in AIC and ACC/pSMA because of their role in task-driven arousal. Portions of both regions have been implicated in emotional arousal and anxiety (Seeley, Menon et al., 2007), as well as in uncertainty during decision making (Grinband, Hirsch et al., 2006). These two regions may be essential for modulating the level of cognitive control based on the present level of anxiety, uncertainty, and/or difficulty.

PPC and PMC are highly functionally integrated in several domains. These domains extend from eye movement control (Curtis, Cole et al., 2005) to shifting attention (Wager, Jonides et al., 2004) to working memory (Constantinidis and Wang, 2004) to general perceptuo-motor integration (Wise, Boussaoud et al., 1997). PPC and PMC will likely interact during RITL, varying primarily during task implementation (where the regions will be most functionally relevant).

DLPFC appears to not only maintain information over time, but also to represent task rules and goals (Miller and Cohen, 2001). For instance, patients with prefrontal lesions have trouble with tasks involving simple rules such as the Stroop task (Vendrell, Junque et al., 1995) and the Wisconsin card sorting task (Milner, 1963). Further evidence comes from several studies of non-human primates, which found rule-like representations within DLPFC (Miller, Freedman et al., 2002). Asaad et al. (2000) found, for instance, that single neurons in lateral prefrontal cortex reflected the task rules rather than particular stimuli or responses.

aPFC (also known as fronto-polar cortex or Brodmann area 10) likely acts upon the representations within DLPFC, perhaps as a kind of second-order rule operator. Braver et al. (2002) used fMRI to compare a regular working memory task with a working memory task requiring subgoal processing (a concrete/abstract semantic judgment). The regular working memory task activated DLPFC as a function of working memory delay length, while only the

subgoal task activated aPFC. Similarly, De Pisapia et al. (2007) used mental arithmetic as a primary process and integration of a number from working memory into the arithmetic problem as a secondary process. This secondary process, which likely involved subgoaling, activated aPFC. Figure 5 illustrates the location of this activation.

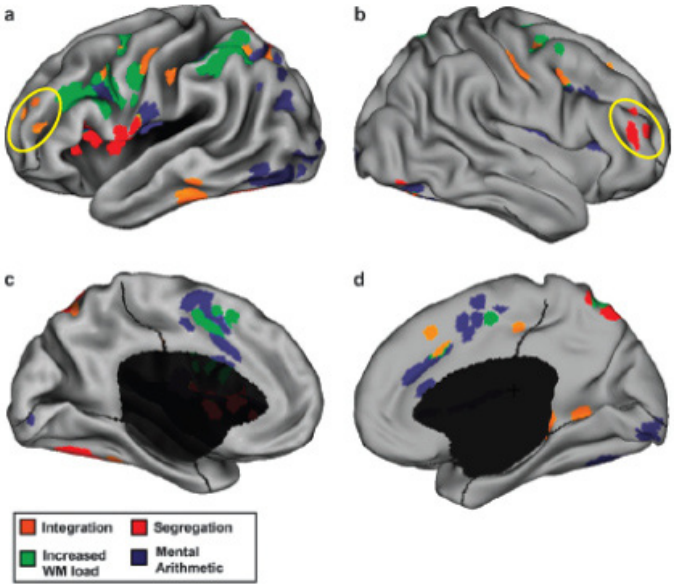

Figure 6 – aPFC is involved in integrating information in working memory
Modified figure from De Pisapia et al. (2007) illustrating aPFC activity (circled in yellow) during integration of information in working memory. Note that most of the control network has a role in this and other related processes as well.

A subgoal process involving the integration of task rules was hypothesized to occur during RITL task instruction. In light of the above evidence, it is likely that aPFC will be involved in this task rule integration process.

Very little work has investigated these regions as part of a network for cognitive control. The present research seeks to advance our understanding of this network, especially the roles of DLPFC and aPFC, in the context of RITL.

1.5.1 Changes in the control network with learning

A wide variety of studies have demonstrated that the control network's activity drops with practice (see Chein et al., 2005). This is likely because performance approaches automaticity with practice (Schneider et al., 1977), which allows performance to occur via direct links between the involved perceptuo-motor regions rather than links mediated by the control network (Schneider et al., 2003).

It is clear that the control network is involved early in task performance, but it is unclear what happens during initial loading of the task rules and during the first several trials applying those rules. It is likely that the same system is involved in the first several trials as well as during later (non-automatic) trials. However, there is probably an additional process, the integration of rule components into a single task representation, during the first several trials. This additional process likely involves aPFC.

The present research seeks to demonstrate that task learning typically takes place in three stages. These include RITL, controlled performance, and automatic performance. The transition between controlled and automatic performance is relatively well characterized (Chein et al., 2005). The present research seeks to characterize the transition between RITL and controlled task performance. Further research will be necessary to fully characterize the transition points between these stages, what prompts these transitions, and the mechanisms behind each of them.

1.5.2 Control network functions during task instruction

In two of only a handful of studies explicitly looking at task instruction, Bunge et al. (2003) and Brass et al. (2004) suggest a distinction between instruction interpretation and (practiced) task rule loading.

Bunge et al. used 'instruction synonyms', two cues (one visual and one verbal) for each task rule, in order to measure inter-modality differences in instruction interpretation. They found that portions of aPFC, AIC, Broca's area, IFJ, PPC, and the hippocampal formation were more active for visual relative to verbal task instructions. This difference could be due to an innate disparity in how tasks are loaded based on modality of input.

However, a more likely interpretation is that subjects converted the visual instructions into verbal instructions, which added an additional cognitive process for the visual cues. This suggests that task rule loading (including RITL) is easier the closer the task instructions are to the internal representations subjects use to encode task rules. This also suggests that any task loading study should minimize this difference (using verbal cues) in order to cleanly measure the task loading process.

The little that is known about the neural basis of task rule loading implicates IFJ, DLPFC, and aPFC as primary players in the process, with other control network regions implicated as secondary players. Bunge et al. found that left IFJ (labeled 'VLPFC' in the paper) and DLPFC showed sensitivity to rule complexity (independent of instruction interpretation) during task loading. IFJ additionally showed sensitivity to rule complexity during task maintenance.

Bunge et al. also found that aPFC was sensitive to the 'non-match' relative to the 'match' rule (in comparing two complex visual stimuli) during both task loading and maintenance, suggesting the subgoal process of negation recruited aPFC. This in turn suggests that each

additional production step in the presently proposed experiments will increase activation within aPFC during task loading, maintenance, and execution.

Similar to Bunge et al., Brass et al. separated instruction interpretation from task loading by using 'instruction synonyms'. Subjects were presented with same-modality task cues that differed slightly in form but not meaning, while other cues differed in both meaning and form.

Brass et al. found that IFJ, VLPFC, and PPC were sensitive to task loading relative to instruction interpretation. In contrast, ACC/pSMA, PMC, and fusiform gyrus were sensitive to instruction interpretation, suggesting they have a role in evaluating the instruction stimuli but not in loading the task rules.

These studies suggest key roles for VLPFC/DLPFC, IFJ, and PPC in task loading and maintenance. They also suggest that instruction interpretation processes are distinct from task loading processes, and that both processes involve control network regions.

The present research includes VLPFC/DLPFC, IFJ, and PPC as regions of interest and will determine if they are involved in RITL. Further, the present research minimizes instruction interpretation processes by using verbal cues during task instruction similar or identical to verbalization used by subjects to encode the task rules.

1.5.3 The internal organization of prefrontal cortex

Investigating the nature of processing and representation within PFC has garnered a great deal of recent excitement. One line of research has emphasized the representation of abstract rules within PFC (Asaad et al., 2000; Wallis and Miller, 2003), while another has emphasized a hierarchical representation structure within PFC (Koechlin, Basso et al., 1999; Fuster, 2001; Braver et al., 2002; Badre and D'Esposito, 2007; Botvinick, 2008). Still another has supported

the role of PFC in active process-driven representations (with permanent representations only in posterior cortex), while an emerging view supports more permanent representations within PFC itself (Wood and Grafman, 2003). The present research may provide some evidence for one or several of these views by demonstrating if part of PFC stores task representations (i.e., is active during practiced task recall) and if PFC involved a hierarchy (i.e., one PFC region organizes the activity of a 'lower' PFC region during RITL).

Despite the variety of different views of processing and representation within PFC, there are several key features that are generally agreed upon. First, PFC likely maintains task goals and the means to achieve those goals (i.e., task sets) (Miller et al., 2001). Second, representations within PFC are generally more abstract than other areas (Asaad et al., 2000; Fuster, 2001; Miller et al., 2002), possibly with a further caudal-to-rostral hierarchy of abstractness *within* PFC (Badre et al., 2007; Botvinick, 2008). Finally, representations within PFC implement task sets by biasing activity in other regions in more posterior cortex, including a rostral-to-caudal biasing cascade within PFC (Koechlin, Ody et al., 2003).

When considering the need for the coordination of abstract information (rules) during RITL, PFC's involvement in rule representation and goal-driven hierarchical control suggests that hierarchical interactions within PFC are likely to be essential for RITL. In particular, it may be that aPFC interacts with DLPFC in a hierarchical manner during RITL, allowing for rapid integration of task rules in working memory.

Hypotheses are the first murmurings of reason in the darkness of the unknown.
– S. Ramon y Cajal, *Advice for a Young Investigator*

Table 1– General experimental questions with corresponding hypotheses

Questions	**Hypotheses**
What regions are involved in RITL?	The same regions that are most different between humans and monkeys: aPFC, temporal lobe, and language regions. Also: The cognitive control network and relevant domain-specific semantic regions.
How are task sets formed during RITL task encoding?	Task set formation occurs via language circuits mapping to semantic regions, whose contents are loaded into DLPFC (working memory) and bound into rules by aPFC.
How does task preparation change with practice?	Task preparation shifts from loading and integrating rule components to simply loading pre-integrated task sets from long-term memory (speeding task preparation).
When do task preparation processes occur?	*For RITL*: Semantic working memory loading occurs early (as each word is comprehended) and then the contents of working memory are integrated into a new task set. *For practiced tasks*: The first instruction cues long-term memory retrieval of the pre-integrated task set, which then triggers semantic working memory loading (see Section 2.1).

2.0 THE NEURAL BASIS OF RAPID INSTRUCTED TASK LEARNING[2]

2.1 INTRODUCTION

Playing a new game, following a recipe, or learning a new task at work may seem like relatively ordinary instances of day-to-day life, and yet these skills exemplify one of the greatest mysteries of human cognition: Individuals can be instructed to perform an almost infinite variety of novel tasks with a high degree of proficiency, even with very little practice (Monsell, 1996; Braver and Barch, 2006). This phenomenon, which can be referred to as rapid instructed task learning (RITL), has eluded scientific investigation despite its central role in executive control functions uniquely characteristic of and extremely important to human beings. The inability to adapt to novel situations is extremely debilitating for day-to-day life, and is perhaps the defining characteristic of disorders of both executive function and prefrontal cortex (PFC). The present research uses novel situations as a window into the inner workings of executive control in humans.

---

[2] This section focuses on an fMRI dataset collected for investigation of RITL. The section is a modified version of a journal article manuscript yet to be submitted for publication. In accordance with the CNUP dissertation guidelines, the formatting of this section is similar to that of the article. Co-author Walter Schneider had an advisory role in this work.

Several lines of research have supported the central role of situational novelty in executive control and PFC function. Schneider & Shiffrin (1977) performed early work characterizing executive control using varied mappings between stimulus and response to make tasks non-routine. Norman & Shallice (2000) found that capacity limitations, a key characteristic of executive control, were most prominent in non-routine (relative to routine) tasks. Rabbitt (1997) found that individual differences during novel task performance (rather than after some practice) were the most strongly predictive of fluid intelligence and other relevant outcome variables. Burgess (quoted by Rabbitt (1997)) commented that, "[a task is] only novel once and it is only on initial presentation that we can be sure it works as a diagnostic tool".

Further, according to Burgess (1997):

> The construct validity of a test whose purpose is to measure adaptation to novelty reduces with every testing. This is undoubtedly one of the key obstacles to our understanding of executive functions, and no doubt explains much of the "riddle of the frontal lobes" as Teuber (1964) called it: the purpose of non-routine processing is to make itself redundant.

The present research combines functional (fMRI) with a new cognitive paradigm that resolves this failure of construct validity to investigate the neural basis of task novelty and bring us one step closer to solving the "riddle of the frontal lobes".

As Burgess points out, the experimental study of RITL provides a unique dilemma. First, investigating RITL requires a novel task rule so the very first use of that rule can be observed. This requires a complex, arbitrary task rule very unlikely to have been experienced by the subjects before. Second, the first experience with the rule must be repeated many times in order to attain proper statistical estimates of the associated behavior and brain activity. Third, a control condition is necessary to differentiate RITL from the particular task rules and stimuli used. The

difficulty of satisfying these three constraints may explain why there is so little research in this area to date.

We developed a new cognitive paradigm, the permuted rule operations (PRO) paradigm, to satisfy these three constraints. This new paradigm permutes a set of simple rule components in many different ways, creating dozens of complex task sets certain to be novel to subjects. The PRO paradigm, which makes it feasible to investigate RITL experimentally, is illustrated in Figure 1.

The paradigm involves reading three instructions, each representing a task rule, that together specify a task set for how to perform semantic judgements on pairs of words. Task sets are either novel (never seen before) or practiced (performed 90 times in a previous session). Each task block begins with a cue indicating if the upcoming task is novel or practiced, followed by three instruction screens. Each set of instruction screens indicates a logic gate, a sensory semantic judgement, and a motor response. Together these instructions fully specify a task set. Since there are four possible rules for each of these three rule categories, there were 4 X 4 X 4 = 64 different task sets tested. Four of them were practiced (counterbalanced across subjects), making available 60 different tasks to investigate RITL.

The PRO paradigm makes it possible to test several hypotheses of how neurocognitive processes are similar and different for novel and practiced tasks. First, since multiple task instructions need to be maintained between encoding and probes, it was expected that brain regions supporting working memory (WM) would be involved during task preparation (Figure 2A & 2B). Second, since each instruction has a specific meaning, the semantics of the rules (logic, sensory, and motor) were expected to load into WM via the neural representations corresponding to each semantic domain. Third, since the task rule types are initially semantically

independent, it was expected that semantic integration would be a crucial component in the formation of an implementable task set.

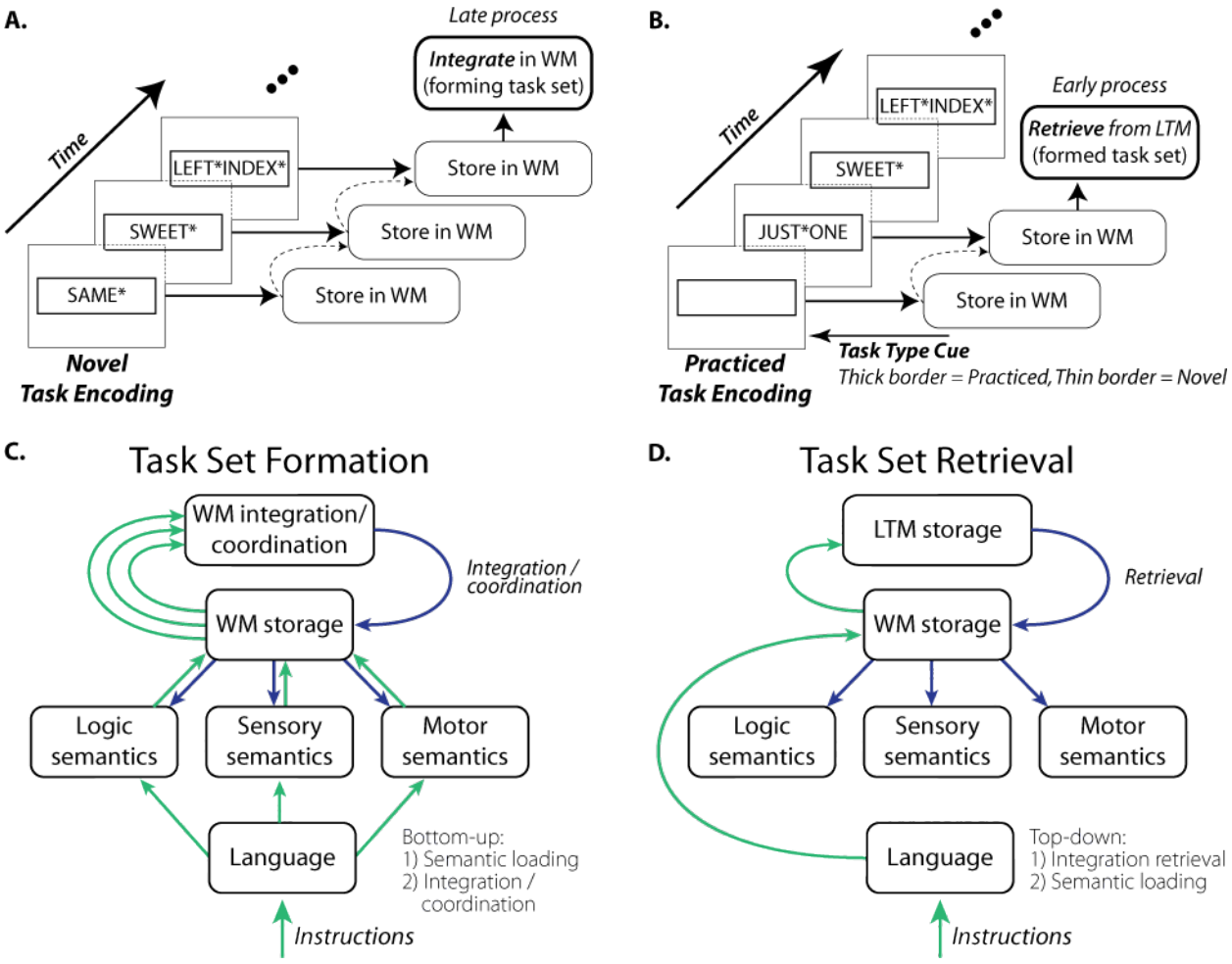

Figure 7 – Task preparation hypotheses
A) We hypothesized that a process, which can be called *task set formation*, leads to the formation of novel task set neural representations from instructions. Since this process requires all relevant instructions to proceed, we predicted that this process would occur after the final encoding screen. **B**) In contrast, we hypothesized that practiced tasks can be loaded from long-term memory (LTM) early in the encoding period. Subjects only had four practiced tasks to choose from, and so were likely able to recall the entire task set as soon as they knew it was a practiced task (first encoding screen) and they knew one of the task rules (second encoding screen). **C**) A hypothesized cognitive process model of novel task set formation that may map onto discrete brain regions. First, language processes activate the relevant (unintegrated) semantics and drive the loading of the semantics into WM storage. Later, the unintegrated semantics are loaded into a WM integration/coordination module to be integrated/coordinated as a procedurally unified task set. **D**) A cognitive process model of practiced task set retrieval. First, language input is loaded into WM storage, which triggers a procedurally integrated LTM representation (formed during the previous practice session). Later, this retrieved task set is loaded into WM storage and the relevant semantics are loaded for task performance.

Research with human and non-human primates has demonstrated that dorsolateral prefrontal cortex (DLPFC) is essential for WM processes (Quintana et al., 1999; Curtis et al., 2003), suggesting that DLPFC activity will reflect WM processes during RITL. Other research suggests roles for posterior parietal cortex (PPC), inferior frontal junction (IFJ), and pre-motor cortex (PMC) as well. Indeed, it has been recently suggested that these regions (along with DLPFC) form a cortical network for maintaining and manipulating WM in the context of cognitive control tasks (Cole et al., 2007). We expected this network to support WM in the context of RITL.

Coordination of the contents of WM, such as WM integration, involves a distinct set of regions from WM maintenance alone. Recent theories of PFC organization suggest that lateral PFC is organized in a hierarchy (Fuster, 2001; Botvinick, 2008), with WM processes becoming more abstract from posterior to anterior portions of PFC. Some have suggested that anterior regions receive input and send output primarily with more posterior PFC regions, providing higher-level transformations on the contents of WM in lower portions of the hierarchy (Ramnani and Owen, 2004). Recent research has supported the role of anterior PFC (aPFC) along the lateral portion of Brodmann's area (BA) 10 in higher-level control and coordination of information (Badre, 2008), including WM integration (Reynolds, McDermott et al., 2006; De Pisapia et al., 2007). Accordingly, aPFC was expected to be involved in WM integration/coordination processes during RITL.

Task instructions were given as three rules with independent semantics associated with each. Recent evidence suggests that even when considered abstractly (i.e., with words and in imagined situations), semantics activate the same brain areas that would be involved in processing that information if it were directly experienced (Wheeler, Petersen et al., 2000;

Goldberg, Perfetti et al., 2006b). For instance, the semantic space for color semantics has been localized to visual cortex, loudness to auditory cortex, sweetness to orbitofrontal cortex, and softness to somatosensory cortex (Goldberg, Perfetti et al., 2006a). However, another line of research suggests that more abstract semantics, especially those involving integration of multiple modalities, are represented in the anterior temporal lobe (aTL) (Rogers, Hocking et al., 2006; Rogalsky et al., 2008). Like aPFC, aTL is thought to sit atop a hierarchy of abstractness (Ungerleider and Haxby, 1994; Taylor et al., 2009), though aTL's hierarchy is thought to be organized by represenation, while aPFC's hierarchy is thought to be organized by process. In the present study we expected the unintegrated semantic task rules to be represented in the areas that would be involved in instantiating those semantics (e.g., occipital lobe for color rules, PMC/M1 for motor rules). In contrast, we expected the integrated task representations to be represented in aTL.

Previous research on RITL has primarily involved theoretical work, usually grounded in comptuational modeling. Most symbolic computational models have implemented RITL simply as a means of directly translating instructions into the form the model is 'thinking', which arguably amounts to little more than a roundabout way to program a computer (but see Kieras and Bovair, 1986). In contrast, the more neurally plausible connectionist computational models might provide some insight into how the human brain performs RITL. The vast majority of connectionist computational modeling has focused on slow (10000+ trials) instructed learning, rather than RITL. However, Schneider & Oliver (1991) as well as Noelle (1997) were able to construct wholly-connectionist models that accomplished RITL by separating a fast rule-based WM system from a typical connectionist distributed slow learning system. Reflecting these models, the present work hypothosizes that a WM system located in PFC, as well as sematic

representations separated by semantic domain, will be differentially involved in RITL and practiced task preparation.

Two specific neurocognitive processes are postulated here: *task set formation* and *task set retrieval*. Task set formation involves converting instructions to an implementable task set, while task set retrieval involves loading a previously formed task set from long-term memory (LTM). Task set formation occurs only during novel task learning, and involves loading of semantic instructions into WM followed by a late WM integration process (Figure 7A & Figure 7C). In contrast, task set retrieval occurs only during practiced task encoding, and involves retrieval of an integrated task set from LTM followed by a late semantic loading process (Figure 7B & Figure 7D). It was hypothesized that brain regions implementing these neurocognitive processes would be dissociated by the differential timing implicit in these processes.

The hypothesized differences in timing were between early (during or just after the encoding period) or late (the first probe) task preparation. Early semantic WM loading (for novel tasks) was likely to occur just after the encoding period, since it involves all task information (see Figure 7A). In contrast, task set retrieval (practiced tasks) was likely to occur early in the encoding period, since it required only a small portion of task information (a LTM cue; see Figure 7B).

The prediction that the first probe of each block would include late task preparation processes is substantiated primarily by behavioral evidence. Specifically, a long line of task switching research has shown that tasks can be prepared for if given enough time, but that there is always some behavioral cost on the first trial no matter how much preparation takes place (Monsell, 2003). This suggested that some task preparation would occur during the encoding period, and also that some task preparation would occur during the first probe.

It is perhaps strange that a behavior so important for daily life, and so central to executive control and PFC functions generally, would have been researched so little. This lack of research may be due to the difficulties of observing task novelty repeatedly while controlling for other factors such as stimulus novelty. The PRO paradigm overcomes these difficulties, allowing the systematic investigation of RITL for the first time. We hypothesized that contrasting novel to practiced tasks would allow dissociation of the neurocognitive processes underlying task set formation (for novel tasks) and task set retrieval (for practiced tasks), providing unprecidented understanding of the processes underlying executive control and PFC processes during task learning and task preparation.

2.2 RESULTS

2.2.1 Individuals can rapidly learn complex novel tasks from instruction

We used the PRO paradigm to statistically test whether subjects can rapidly and accurately learn numerous novel tasks. Accuracy across 15 subjects on the very first performance of the 64 novel tasks was 92.2% ± 1.0% (mean ± standard error). This high accuracy demonstrates that individuals are capable of accurately forming and applying complex novel task sets within seconds.

Overall accuracy in the PRO paradigm was 92.6%. Accuracy for the novel tasks (across all probes, not just the first) was 92.0% ± 1.0%, and accuracy for the practiced tasks was 93.3% ± 0.9%. We used mixed-effect repeated-measures ANOVAs (subjects as a random effect) to test accuracy for novel relative to practiced probes. The overall difference in accuracy between novel

and practiced probes was not statistically significant F(1,14)=3.04, P=0.10. However, both the main effect of probe number (F(2,26)=3.77, P=0.036) and the probe number by condition interaction (F(2,26)=3.78, P=0.036) were significant.

Follow-up comparisons between all possible pairs of probes (first, second, third) across the conditions (novel, practiced) revealed only one significant difference: Including only the first probe of each block, the difference between novel (92.2% ± 1.0%) and practiced (94.3% ± 0.9%) accuracy was statistically significant F(1,14)=10.63, P=0.0057.

The overall mean reaction time (RT) for all probes was 1323 ms. The mean RT for novel probes was 1338 ms ± 70 ms, while the mean RT for practiced probes was 1308 ms ± 69 ms. This 30 ms difference was statistically significant: F(1,14)=11.62, P=0.0042. The main effect of probe number was marginally significant (F(2,26)=2.66, P=0.089), while the probe number by condition interaction was not significant (F(2,26)=0.032, P=0.97).

Follow-up comparisons between all possible pairs of probes (first, second, third) across the conditions (novel, practiced) revealed a marginally significant task switching effect. Collapsing between novel and practiced conditions for probe 1 (1336 ms ± 72 ms) and probe 2 (1295 ms ± 67 ms): F(1,14)=4.44, P=0.054. This effect was approximately identical for novel and practiced probes. For novel probe 1 (1344 ms ± 71 ms) vs. novel probe 2 (1302 ± 68 ms): F(1,14)=4.56, P=0.051. For practiced probe 1 (1328 ms ± 73 ms) vs. practiced probe 2 (1287 ms ± 67 ms): F(1,14)=4.56, P=0.051. There were no other statistically significant RT effects except for novel probe 3 (1329 ms ± 70 ms) vs. practiced probe 3 (1290 ms ± 70 ms): F(1,14)=14.23, P=0.0021. This indicates that, after a reduced RT on probe 2, novel (but not practiced) probe 3 increased in RT again.

Note that effects, such as task switching costs, that are measured by RT may have been weakened by having two stimuli (and only one response) per probe. This is because any slowing while processing the first stimulus may have had little (if any) effect on the time to react to the second stimulus, due to a fixed delay between the two stimuli. This fixed delay may have acted as a buffer, allowing any processes with increased duration to complete prior to onset of the second stimulus.

2.2.2 Task set formation and task set retrieval rely on distinct cortical networks during encoding

Task set formation (novel tasks) was likely to occur late in encoding, since it requires all task information (see Figure 7A). In contrast, task set retrieval (practiced tasks) was likely to occur early in encoding, since it required only a small portion of task information (a LTM cue; see Figure 7B). Accordingly, no shape was assumed when modeling the BOLD signal, allowing early and late encoding activity to be distinguished.

We identified regions likely to show the predicted patterns using a Condition X Time interaction analysis (novel & practiced by 11 time points), which tests for differences in BOLD response shape across conditions (Figure 8A; Table 2). Eighteen regions were identified ($P<0.05$, whole-brain FDR corrected, PFC-only FWE corrected). Regions were colored (red = novel > practiced, blue = practiced > novel) based on statistical differences either early (8 seconds after encoding onset) or late (14 seconds after encoding onset). All regions were statistically different between novel and practiced conditions at 14 seconds ($P<0.05$, mean $P=0.001$), with only aPFC, aTL, and PCC greater for practiced than novel. This corresponds to neural activity during the delay period just after the instructions (see section 2.4.7). Only aPFC, aTL, and PCC were

statistically different (practiced > novel) between novel and practiced conditions at 8 seconds (mean P=0.001, overall mean P=0.35). This corresponds to neural activity early in the encoding period.

Table 2 – Novel vs. practiced encoding activity voxel clusters

Cluster labels	Hemi-sphere	Voxels (3.2 mm^3)	Talairach coord: x	Talairach coord: y	Talairach coord: z	Brodmann areas
IPL	Left	51	-52	-27.2	25.7	40
PCC	Both	40	-2	-46.8	29.7	31
M1/S1/S2	Left	39	-18.3	-29.3	66.7	4, 3, 5
aPFC	Left	31	-22.3	48.1	18.5	10
DLPFC	Right	25	29.6	26.5	35.3	9
Fusiform/Cerebellum	Right	24	25.7	-37.8	-19.7	20, 36
ACC	Left	24	-11.5	1.4	28.7	24
M1/S1	Right	24	41.7	-25.4	58	3, 4
VLPFC	Left	23	-46.8	29.3	-4.2	47, 45
BG (Medial globus pallidus & putamen)	Left	16	-17	-2.8	-7	–
aTL	Right	15	51.1	-2.5	-24.8	20, 21
IFJ	Right	15	27.8	0.5	32	9, 6
PMC	Right	15	22.9	-13.4	47.5	6
VLPFC	Right	14	48.9	25.5	-0.7	47, 45
PMC	Left	14	-32	-15.5	60.5	6
Cuneus	Left	13	-10.2	-62.6	5.6	18, 19

Hierarchical clustering was used to identify distinct brain networks involved in task encoding based on BOLD response shape (Figure 8B). The clustering of the regions was based on the Euclidean distance between the novel, practiced, and novel - practiced time series for each region. Thus, the clustering reflects each region's response over time, as well as the relative difference between novel and practiced responses over time.

The regions formed three distinct clusters. Example time series, as well as the average time series of each cluster, are plotted in Figure 8C. The first cluster includes regions showing a long-tailed extended response for novel tasks. The second cluster includes regions showing greater early activity for practiced tasks (consistent with task set retrieval). The third cluster includes regions showing greater late activity for novel tasks (consistent with task set formation).

This final cluster also included regions with relatively flat, extended responses, possibly involved in maintaining information in WM.

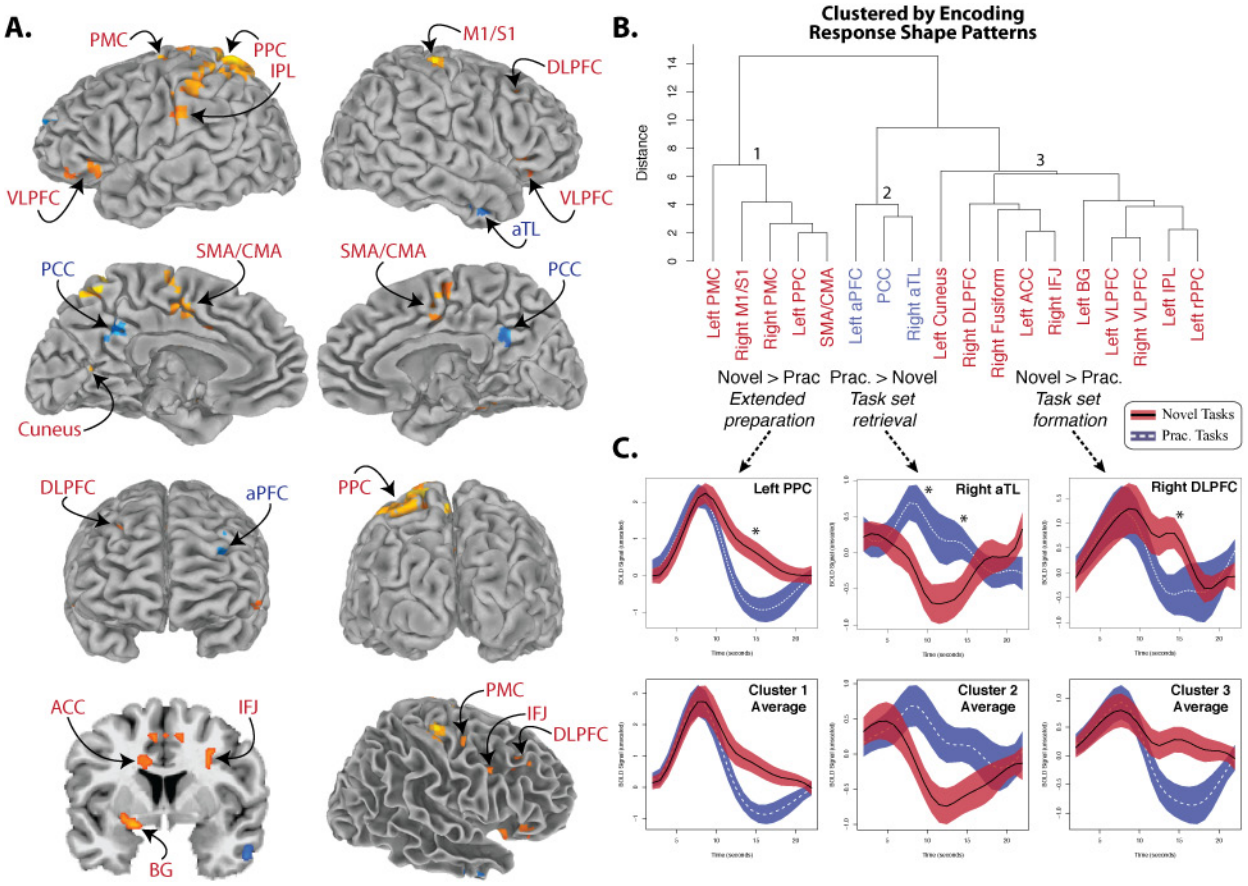

Figure 8 – Encoding Activity Clusters Into Three Networks
Blood oxygen level dependent (BOLD) responses to novel and practiced task encodings were measured with functional MRI (fMRI), and statistically modeled with no assumption of response shape. **A**) We performed a Condition X Time interaction to identify the brain regions with statistical differences in reponse shape across conditions. This allowed us to identify regions showing early increases for practiced tasks (task set retrieval regions) and regions showing late increases for novel tasks (task set formation regions). Red regions have statistically significant time points larger for novel tasks, while blue regions have time points larger for practiced tasks. **B**) The regions' response shapes clustered into three distinct networks. Each branch of the hierarchical cluster plot was investigated for identifiable within-cluster similarities and between-cluster differences. The first three clusters were consistent internally and differentiated externally. **C**) The hemodynamic responses for novel (red) and practiced (blue) encoding periods across the three clusters. An example region is depicted on top, with the average response across regions in each cluster on the bottom. Cluster 1 showed a similar response between novel and practiced encoding, but with a long tail on the end for novel encoding. The regions in Cluster 1 are consistent with task set semantic loading (see **Figure 7**). Cluster 2 was consistent with the hypothesized role of task set retrieval, showing greater (early) activity for practiced encoding.

Cluster 3 included regions consistent with task set formation, which showed greater (late) activity for novel encoding. Other regions in Cluster 3 showed a flat extended response (as depicted in the Cluster 3 average), perhaps reflecting maintenance in WM.

2.2.3 aPFC is dissociated from other regions active during first probes

Probe responses were modeled, like encoding responses, with no assumption of BOLD response shape. However, since there was no hypothesis regarding differential BOLD response shape (only amplitude) the mean BOLD responses were contrasted between novel and practiced probes (Figure 9A; Table 3).

Six spatially clustered voxel zones were identified ($P<0.05$, FWE corrected). Two of these zones (one on each hemisphere) were quite extensive, including Broca's area (BA 44; on the left), PMC, M1, S1, A1, IPL, and possibly other nearby regions. Notably, these regions are known to have sensory and motor functions, suggesting that these clusters primarily reflect the activation of sensory-motor semantic representations.

Similar to the encoding analysis, regions were clustered based on their activity patterns (Figure 9B). However, clustering occurred here based on mean activity patterns for novel and practiced conditions (rather than the time series themselves). Further, because of the hypothesis that encoding and first probes would best reflect task preparation processes that differed between novel and practiced tasks, only encoding and first probe responses were included in the cluster analysis. The analysis separated regions into two clusters. Only aPFC was in the first cluster, while all five other regions were in the second cluster.

The two clusters differed between novel and practiced tasks across encoding and first probes (Figure 9C). Confirming this, both clusters showed a Task Type X Task Period interaction (cluster 1: $F(1,14)=10.23$, $P=0.0006$; cluster 2: $F(1,59)=13.02$, $P=0.0006$).

Importantly, these interactions were in opposite directions. This double dissociation (Condition X Cluster interaction) was highly statistically significant (F(1,42)=23.41, P=4.58e-09).

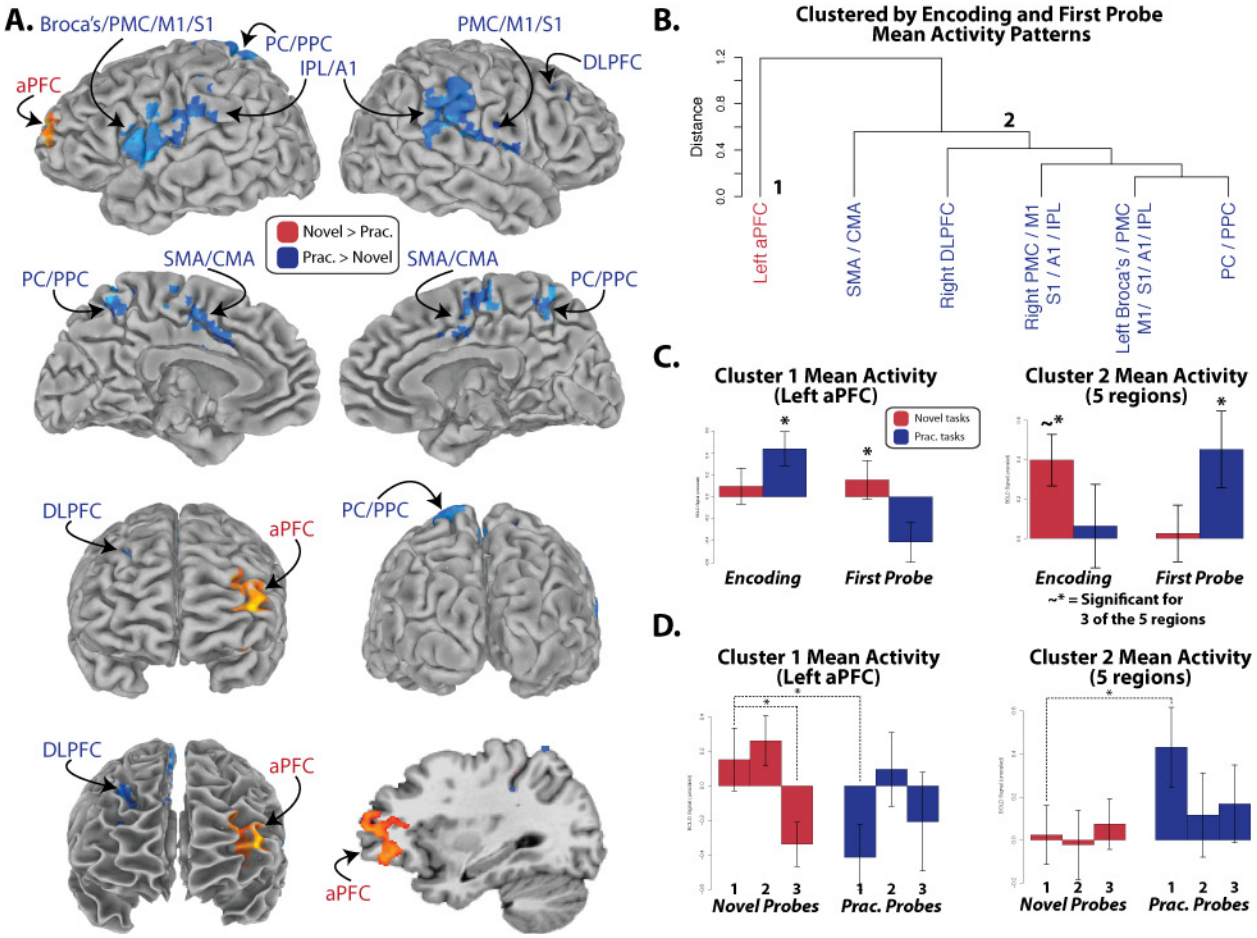

Figure 9– A Double Dissociation Between aPFC And Other Probe-Responsive Regions
Activity patterns across conditions were investigated for regions differentially active for novel and practiced probes. **A**) The main effect of novel vs. practiced probes (P<0.05, FWE corrected) is depicted. **B**) The task set formation and retrieval hypotheses were primarily relevant to encoding and first probe time periods, and so regions were clustered based on their activity patterns across them. All of the regions clustered tightly, except for aPFC. **C**) Mean BOLD signal is plotted for novel and practiced tasks across encoding and first probe time periods. There was a highly statistically significant double dissociation between the two clusters (Condition X Cluster interaction; F(1,42)=23.41, P=4.58e-09). Cluster 1 showed a Task Type X Time Period interaction (F(1,14)=10.23, P=0.0006). Cluster 2 also showed a Task Type X Time Period interaction (F(1,59)=13.02, P=0.0006). Cluster 2 regions SMA/CMA, PC/PCC, and right DLPFC (for multiple time point, not mean activity) were more active for novel than practiced encoding. Activity in Cluster 1 (aPFC) is consistent with integrated task set representation, while activity in Cluster 2 is consistent with task set semantics loading (see Figure 2). **D**) Mean BOLD signals are plotted across all three probes for novel and practiced tasks. These results suggest a unique role for these regions in process the first probes.

Cluster 1 (aPFC) was more active for practiced than novel tasks during encoding, suggesting a role in task set retrieval (early practiced task preparation activity, consistent with Figure 7B & 6D). In contrast, Cluster 2 was more active during novel than practiced tasks during encoding (mean response $P<0.05$ for 3 of the 5 regions). This suggests a role in semantic loading as the instructions are read, understood, and loaded into working memory (early novel task preparation activity, consistent with Figure 7A & 6C).

Table 3 – Novel vs. practiced probe activity voxel clusters
The contrast was defined by all novel probes vs. all practiced probes ($P<0.05$, FWE corrected). The right aTL region was defined by novel probe 1 vs. novel probe 3 ($P<0.05$, FWE corrected).

Cluster labels	Hemi-sphere	Voxels ($3.2\ mm^3$)	Talairach coord: x	Talairach coord: y	Talairach coord: z	Brodmann areas
Broca's / PMC M1/ S1/ A1/ IPL	Left	318	-51.2	-19.9	20.3	44, 6, 4, 3, 13, 22, 41, 43, 40
aTL	Right	244	51	-4.8	-16	20, 21, 38, 22
PC/PPC	Both	240	-8	-52.2	58	7
PMC / M1 S1 / A1 / IPL	Right	228	57.9	-24.4	21.4	40, 6, 22, 41, 42, 21
SMA / CMA	Both	204	-0.9	-7.6	47.9	6, 24
aPFC	Left	148	-29.4	46.6	11.5	10, 11
DLPFC	Right	113	25.5	28.7	29.7	9

Cluster 1 (aPFC) was also more active for novel than practiced tasks during first probes, suggesting a role in task set integration/coordination (late novel task preparation activity, consistent with Figure 7A & Figure 7C). In contrast, Cluster 2 was more active for practiced than novel tasks during first probes. This suggests a role in loading of an integrated/coordinated task set from long-term memory (late practiced task preparation activity, consistent with Figure 7D).

The clusters were analyzed across all three probes in order to confirm that encoding and first probes were unique among the task periods. Figure 9D illustrates enhanced activity in cluster 1 (aPFC) for novel probe 1 vs. novel probe 3 and vs. practiced probe 1. In constrast,

cluster 2 was more active for practiced than novel first probes. Though practiced first probes were not consistently higher than subsequent practiced probes, the mean was nontheless much higher for first than subsequent practiced probes, suggesting something was different about the first probes in these regions.

2.2.4 aPFC performs task set integration/coordination and DLPFC performs task set semantic loading

aPFC (cluster 1) and DLPFC (cluster 2) were differentially active for both encoding and first probes. These regions' activity patterns were investigated further because of existing hypotheses regarding a PFC hierarchy, and their likely involvement in WM storage and integration/coordination processes expected to be involved in RITL (Figure 7).

Since the previous analysis found specific differences for first probes, another analysis focusing on novel vs. practiced first probes was performed. This statistical contrast was restricted to lateral PFC (inferior, middle, superior, and lateral orbitofrontal gyri), with a statistical contrast thresholded at $P<0.05$. Multiple comparisons were corrected with FWE for aPFC and confirmed to be a subset of the probe main effect map for DLPFC (DLPFC would not have survived the FWE correction).

The first probe PFC map was included along with the two other novel vs. practiced contrasts reported above in a conjunction map (Figure 10A). Within the map, each contrast and their spatial overlap was assigned a unique color. Both aPFC and DLPFC contain a portion that overlapped between the three contrasts (depicted in white).

The aPFC and DLPFC regions defined by the first probe contrast were investigated further. First, the double dissociation found in the previous analysis between the two clusters was

specifically confirmed for left aPFC and right DLPFC (Region X Condition interaction; $F(3,42)=11.0$, $P=0.000019$). The dissociation between the regions was significant even when including all encoding and probe conditions in the Region X Condition interaction test ($F(7,98)=5.10$, $P=0.000057$).

Figure 10B illustrates activity in left aPFC across all encoding and probe conditions. This region showed activity largely consistent with coordinating and/or representing integrated task sets, since it is active later for novel than practiced tasks (see Figure 7). aPFC's activity hovered at baseline during novel encoding (significantly less than practiced encoding), rose for novel probe 1 (significantly more than practiced probe 1 and novel probe 3), stayed high for novel probe 2 (significantly higher activity than novel probe 3), and dropped significantly for novel probe 3 (relative to the first two novel probes). In contrast, aPFC was active during practiced encoding (relative to novel encoding and practiced probe 1), dropped for practiced probe 1 (relative to practiced encoding and novel probe 1), rose again for practiced probe 2 (relative to practiced probe 1, but *not* relative to novel probe 2), and hovered around baseline for practiced probe 3.

Figure 10C illustrates activity in right DLPFC across all encoding and probe conditions. This region showed activity largely consistent with semantic working memory, since it is active earlier for novel than practiced tasks (see Figure 7). The region's activity was high for novel encoding (significant relative to novel probe 1 and practiced encoding (see right side of figure for encoding difference)), dropped for novel probe 1 (relative to novel encoding and practiced probe 1), rose again for novel probe 2 (relative to novel probes 1 and 3, but *not* relative to practiced probe 2), and dropped again for novel probe 3 (relative to novel probe 2). In contrast, right DLPFC's activity was low for practiced encoding (relative to novel encoding), rose for

practiced probe 2 (relative to novel probe 1), and hovered around baseline for practiced probes 2 and 3.

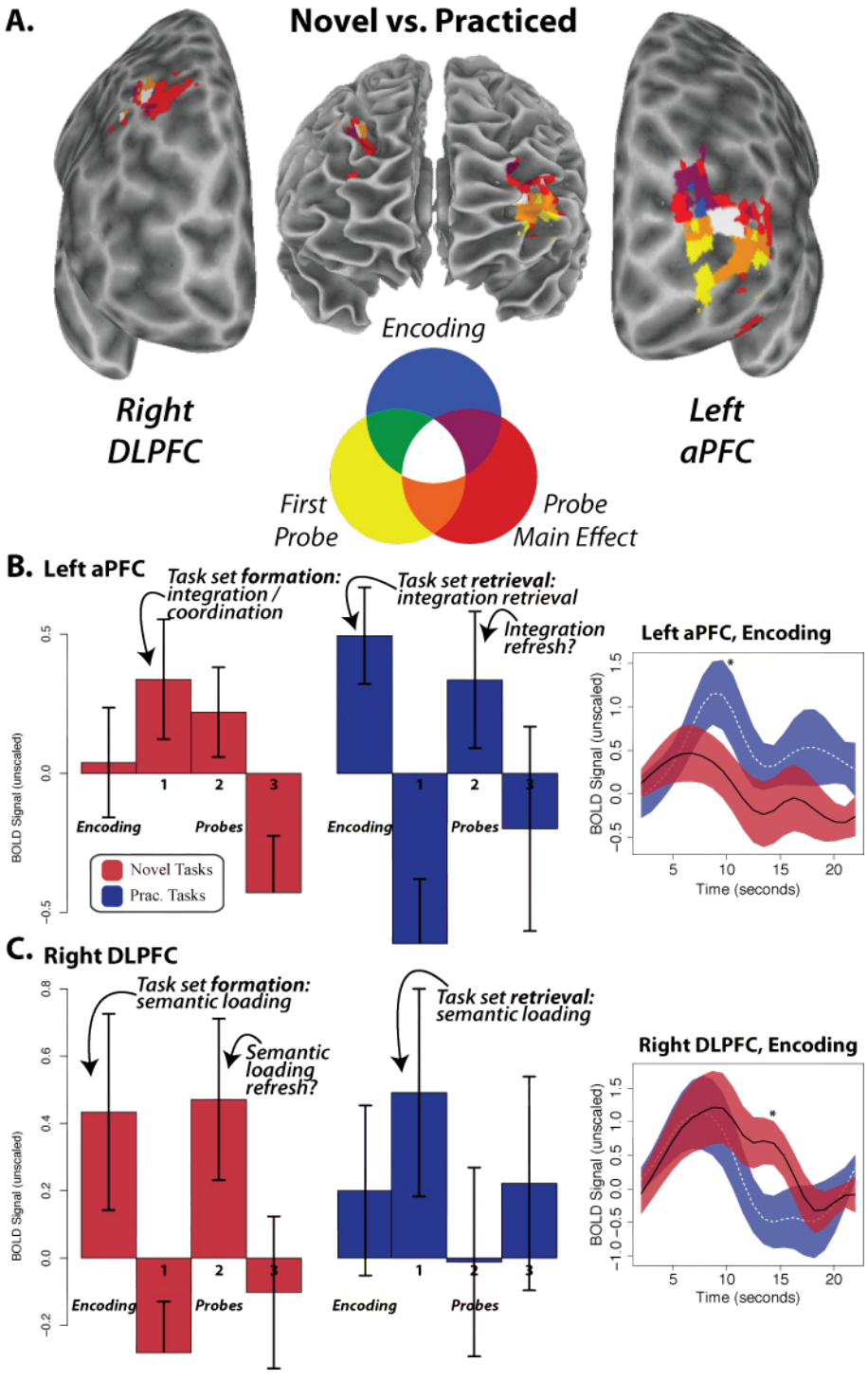

Figure 10 – aPFC Performs Task Set Integration and DLPFC Performs Task Set Semantic Loading
Differences across novel and practiced first probes were used to localize PFC regions for further investigation. **A**) The first probe contrast was plotted along with encoding and probe main effect

contrasts on white matter and inflated surfaces. The color legend indicates the overlap between the statistically significant voxels across the analyses. Both aPFC and DLPFC contain a portion that overlapped between the three contrasts (depicted in white). Note that right DLPFC was probably centered on the middle frontal gyrus, but showed up on either side of the gyrus due to anatomical differences between the template and these subjects. **B**) Activity in aPFC across all encoding and probe conditions are depicted, along with the time series during encoding. As indicated, activity during the first probe for the novel condition is consistent with task rule integration/coordination during task set formation. For the practiced condition, activity during the encoding period is consistent with loading procedurally integrated/coordinated task sets from LTM. Both of these situations are consistent with representing integrated task sets. **C**) Activity in DLPFC is depicted in the same way as aPFC above. DLPFC and aPFC show essentially opposite activity patterns. As indicated, activity during novel encoding is consistent with loading task set semantics into WM. For the practiced condition, activity during the first probe is also consistent with loading task set semantics into WM. Unexpectedly, both aPFC and DLPFC have an increased response for the second probes (relative to first probes, *not* relative to second probes in the control condition) when they were active during encoding. These increases may reflect the need to 'refresh' WM when the relevant information was loaded early. Further research is necessary to confirm this possibility. The double dissociation was statistically significant between aPFC and DLPFC ($F(3,42)=11.0$, $P=0.000019$).

2.2.5 aTL forms a functional network with aPFC

In the first functional cluster analysis (see Figure 8B), aPFC appeared to form a functional network with aTL (in addition to PCC). However, aTL was not present in the probe main effect map (see Figure 9A), and so its functional similarity to aPFC was not tested in this broader context. It is possible that aTL matched aPFC's activity pattern, but the large variance during practiced probes reduced the statistical significance of the probe main effect contrast in this region (see standard error bars in Figure 9D and Figure 10B). Since aPFC showed a large difference between novel probe 1 and novel probe 3 (see Figure 10B) , we used this contrast to specifically investigate the possibility of right aTL activity patterns matching those of aPFC.

The statistical analysis was performed over the whole brain ($P<0.05$, FWE corrected). Figure 11A illustrates the large portion of right aTL significantly more active for novel probe 1 relative to novel probe 3. Figure 11B illustrates the same region projected onto the surface, so its

spatial extent is illustrated despite sulcal activity and anatomical variability between the template's gray matter and those of the 15 subjects.

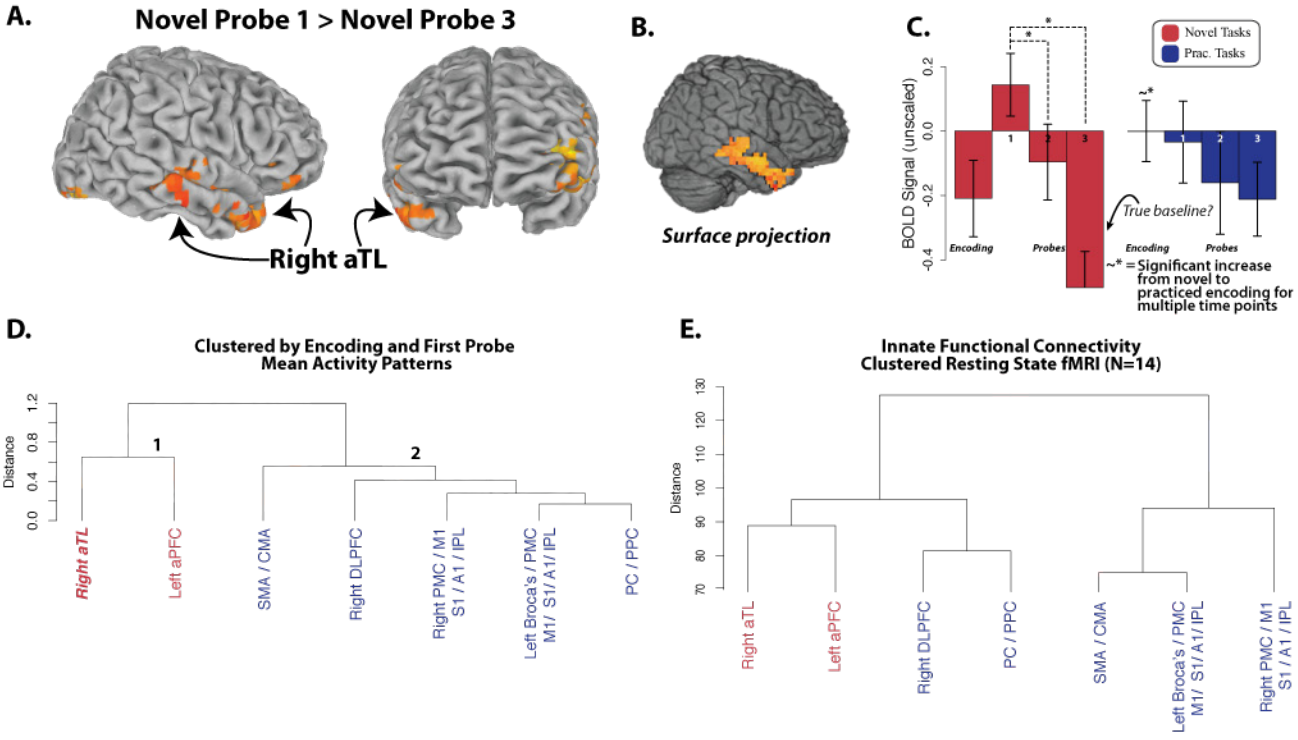

Figure 11 – aTL Forms A Functional Network With aPFC
aTL was included with aPFC in the first functional cluster analysis (**Figure 8**B). However, the functional similary between aTL and aPFC across time periods was not assessed because aTL activity was not significantly different for novel and practiced probes (**Figure 9**B). It is possible that aTL matched aPFC's activity pattern, but the large variance during practiced probes reduced the statistical significance of the probe main effect contrast in this region. Since aPFC showed a large difference between novel probe 1 and novel probe 3, we used this contrast to specifically investigate the possibility of right aTL activity patterns matching those of aPFC. **A**) Right aTL, as defined by novel probe 1 vs. novel probe 3, is depicted on a Talairach template brain surface. **B**) The aTL activity is projected onto the surface to see activity buried in the sulci and not registering perfectly with the template surface. **C**) aTL activity across all encoding and probe conditions is depicted. Note that aTL (and possibly aPFC) is part of the 'default mode network' (Raichle, MacLeod et al., 2001), which is known to be active during rest periods. This may have shifted the baseline upward, such that apparent negative activity such as during novel probe 3 may actually be decreases to baseline. **D**) The functional cluster analysis in **Figure 9**B was rerun with right aTL included. aTL clustered with aPFC, as expected. Like aPFC, aTL was functionally dissociated from Cluster 2 ($F(3,42)=4.87$, $P=0.005$). **E**) The regions were included in a functional cluster analysis using preprocessed fMRI data when subjects were at rest (10 minutes). The rest data was collected prior to task performance for 14 of the 15 subjects. The innate functional connectivity pattern is similar to the task-induced clustering pattern of these regions. Right DLPFC and PC/PPC are separated from Cluster 2, perhaps reflecting their role in cognitive control and the involvement of the other regions in semantic representation. In turn, the task-

induced functional similarity of DLPFC and PC/PPC with the semantic regions likely reflects WM control of the semantic regions by DLPFC and PC/PPC.

Right aTL's activity across all encoding and probe conditions is illustrated in Figure 6C. Similar to left aPFC, right aTL was more active for practiced than novel probes (for multiple time points, though the mean activity was not quite significant), was more active for novel probe 1 (relative to novel probes 2 and 3), and dropped down for novel probe 3 (relative to all other conditions).

The functional cluster analysis run on the probe main effect regions (Figure 9B) was rerun with this right aTL region included. As expected from its activity pattern, right aTL clustered with left aPFC (Figure 11D). The dissociation between the clusters (with aTL included in cluster 1) including all conditions was highly statistically significant ($F(7,98)=5.92$, $P=0.000009$). This dissociation remained significant when aPFC was removed from cluster 1, making it a test for a dissociation between aTL and cluster 2 ($F(7,98)=2.98$, $P=0.007$). This dissociation was also significant when only encoding and probe 1 conditions were included ($F(3,42)=4.87$, $P=0.005$).

An additional functional cluster analysis was run on resting state data in order to determine the functional connectivity of these regions outside the context of the tasks (Figure 11E). The cluster analysis verified that aPFC and aTL form a functional network even outside the context of the PRO paradigm. Unlike the task-induced activity patterns, DLPFC and PC/PPC clustered separately from semantic regions (CMA/SMA and the two lateral voxel clusters). This task-induced change in connectivity may reflect the loading of information from the semantic regions into WM storage (DLPFC and PC/PPC). A separate correlation analysis showed that

correlations between all ROIs were statistically greater than 0 (mean R=0.42, T(13)=8.71, P=8.67e-07). See the Materials and Methods section for details.

2.2.6 Semantic domain activity patterns

A standard fMRI block analysis (assuming a canonical hemodynamic response shape) was used to identify the locations of semantic processing (Figure 12). Motor semantics were localized by contrasting blocks with a right-hand response to blocks with a left-hand response ($P<0.05$, FWE corrected). Logic semantics were localized by contrasting blocks using complex logic rules (SAME and JUSTONE) with blocks using simple logic rules (SECOND and NOTSECOND) ($P<0.05$, FWE corrected). Each form of sensory semantics (gustatory, auditory, tactile, and visual) was localized by contrasting blocks using that sensory domain (e.g., auditory) with blocks using the other sensory domains (e.g., gustatory + tactile + visual). These maps were thresholded at $P<0.05$, uncorrected for multiple comparisons. No voxels survived thresholding using either FWE or FDR multiple comparison corrections ($P<0.05$). In order to (at least partially) account for multiple comparisons, the analyses were restricted to the approximate anatomical areas in which each domain would be expected to activate for perceptual experiences. Accordingly, the sensory rule for color was restricted to visual cortex, loudness to auditory cortex, softness to somatosensory cortex, and sweetness to gustatory cortex. Further, these regions have previously been shown to respond to perceptual knowledge retrieval from linguistic stimuli (Goldberg et al., 2006b).

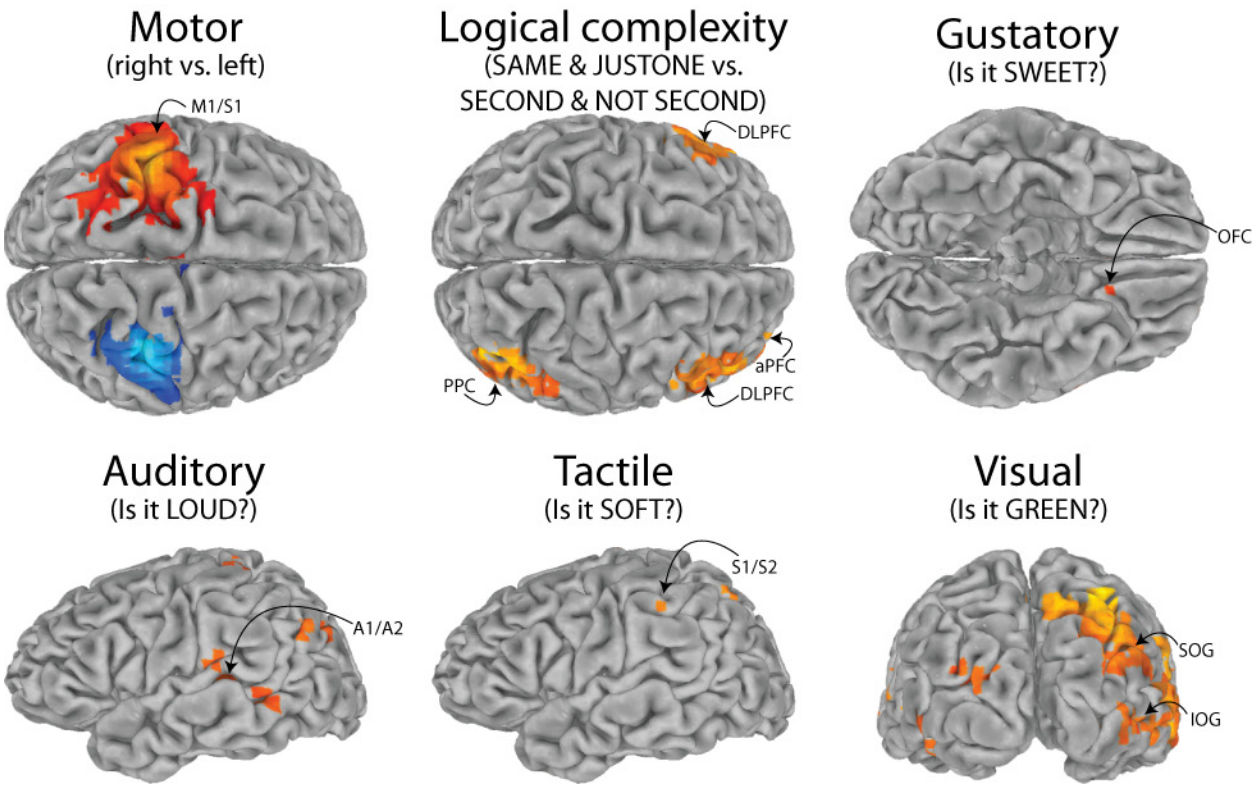

Figure 12 – Semantic Domain Activity Patterns
Six statistical contrasts were used to determine the likely locations of semantic domain-specific activity. Motor activity was determined by contrasting blocks with a right-hand response to blocks with a left-hand response. The logical complexity effect was observed by contrasting SAME and JUSTONE (both requiring integration of both probe stimuli) with SECOND and NOT SECOND (both involving only the second probe stimulus). Both of the preceding contrasts were corrected for multiple comparisons using FWE correction ($P<0.05$, FWE corrected). Multiple comparison correction was not used for the sensory semantic contrasts ($P<0.05$, uncorrected), due to the lower statistical power for these contrasts (~1/2 of the first two contrasts). However, each contrast was restricted to the approximate anatomical vicinity expected to show the effect based on previous work (Goldberg et al., 2006b). The tentative locations of each sensory semantic activity pattern are depicted. Gustatory activated a portion of orbito-frontal cortex (OFC), auditory A1/A2, tactile S1/S2, and visual superior occipital gyrus (SOG) and inferior occipital gyrus (IOG).

2.3 DISCUSSION

The results indicate the involvement of the expected WM maintenance, WM integration/coordination, and semantic representation brain regions in RITL. Importantly, these brain regions were strongly dissociated across novel and practiced tasks (Figures 8, 9, & 10).

Further, these dissociations are largely consistent with the hypothesized neurocognitive processes, suggesting a mapping between those processes and specific brain regions (Figure 13).

Statistically significant dissociations were present between a network including aPFC and aTL and a network including DLPFC, PPC, and a set of regions likely representing linguistic and sensory-motor semantic information. Both networks were active for novel and practiced tasks, but they were differentially active for encoding and first probe periods across the two task types.

The DLPFC/PPC network was more active early (encoding period) for novel tasks, likely due to loading semantic information into WM during (and just after) reading the instructions. This network was more active late (first probe) for practiced tasks, also likely due to the loading of semantic information into WM. However, this loading was likely induced in a top-down manner (rather than bottom-up from instructions) by aPFC/aTL retrieving the integrated/coordinated task set from LTM.

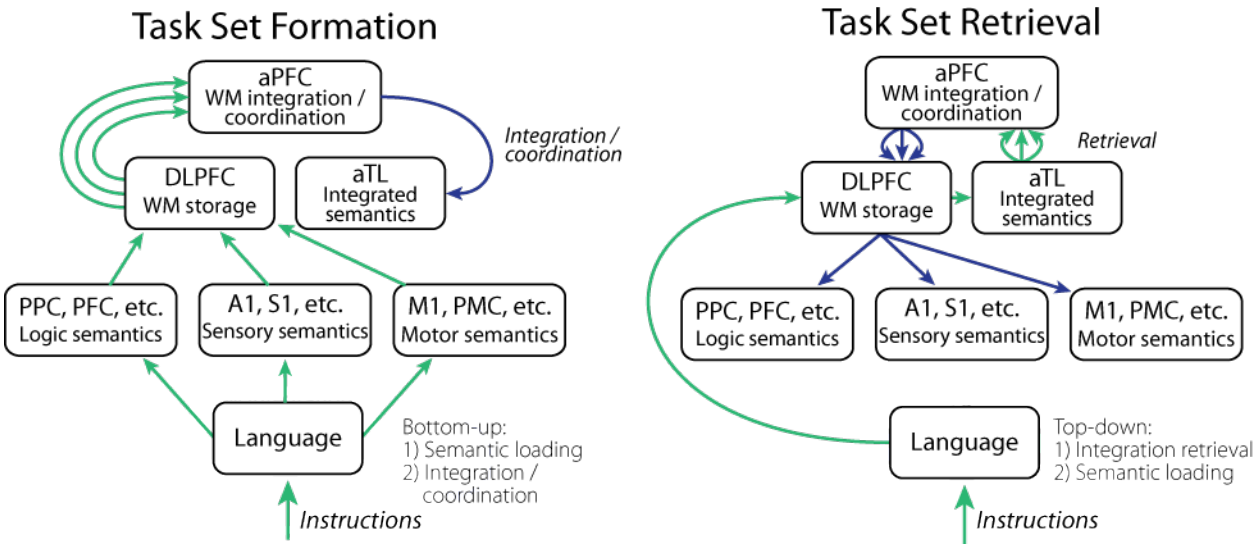

Figure 13 – Mapping Neural Structures to Functions for Novel and Practiced Task Preparation
The brain regions showing activity consistent with the hypothesized neurocognitive processes are paired with those processes. **A)** The results are largely consistent with the neurocognitive processes outlined in **Figure 7**C. However, the addition of an 'integrated semantics' representation region was necessary (aTL). Further, rather than loading back into the WM storage region, the integrated task set appears to be implemented via aTL (perhaps being controlled by aPFC). These changes are based on aTL's activity late in novel task preparation (first probe) and

the region's known role in integrated semantics (Taylor et al., 2009). aPFC may integrate the task rules into a task set in preparation for task implementation. Alternatively, aPFC may implement some other form of hierarchical control of the task set (i.e., task coordination), possibly by sequentially activating the task rules in order to implement the task procedures for the first time. The procedures may be implemented by lower-level regions after some practice. **B)** The results are also largely consistent with the neurocognitive processes outlined in **Figure 7**D. However, the 'integrated semantics' representation region (aTL) appears to be necessary here as well. This is likely because aTL contains the LTM store for integrated semantic representations. aPFC was also active along with aTL, perhaps since LTM retrieval involves reactivation of neural activity at the time of encoding (which had involved aPFC) and because aPFC has no direct access to semantic regions (Ramnani et al., 2004). aPFC could then load the integrated task into DLPFC (and/or coordinate the rule representations in DLPFC) for task performance. Whereas DLPFC appears to not be involved in novel first probe task performance, it was clearly involved for practiced first probes. This may reflect a shift from aPFC control to lower-level DLPFC control for highly practiced tasks. More specifically, DLPFC may have developed (via practice and consolidation) a long-term representation of the practiced tasks, but aTL and aPFC are necessary to fully retrieve that integrated task and reduce interference within DLPFC between the unintegrated and integrated semantic representations.

The aPFC/aTL network was more active late (first probe) for novel tasks, possibly due to aPFC's role in integrating and coordinating information in WM (De Pisapia et al., 2007) and aTL's role in representing integrated/conjunctive semantics (Taylor et al., 2009). The network was more active early (encoding period) for practiced tasks, also probably due to its role in representing integrated information. However, this activity was likely cued by the first instruction screen (which was associated with a single practiced task during training) and retrieved from LTM.

The roles of semantic WM maintenance for the DLPFC/PPC network and integrating/coordinating WM for the aPFC/aTL network are largely consistent with the hypothesized task set formation (Figure 13A) and task set retrieval (Figure 13B) processes. There were several inconsistencies with the original hypotheses (Figure 7), however.

First, it was unclear *a priori* what role aTL would play in RITL and practiced task performance. aTL is known to sit at the top of a semantic hierarchy, representing integrated multimodal and abstract semantics, such as during crossmodal integration of object features

(Taylor et al., 2009) and during semantic (as opposed to syntactic) sentence processing (Rogalsky et al., 2008). This, in conjunction with the region's activity patterns here, suggests that aTL's involvement is related to representing task sets formed from multiple semantic domains. Given that the lateral temporal lobes store complex (associative and autobiographical) consolidated LTM representations (Miyashita, Kameyama et al., 1998; Lah, Lee et al., 2008), it also seems likely that aTL stores LTM representations of integrated task sets.

Another unexpected result involved the lack of aPFC/aTL activity late in practiced task preparation. If aTL represents integrated task sets, it seems that its activity would be necessary for practiced task performance (i.e., during the first probe). However, it is also possible that repeated task performance shifted task control from the higher-level aPFC/aTL to the lower-level DLPFC/PPC. Such a shift would parallel the shift seen from controlled (moderately practiced) to automatic (highly practiced) task performance between DLPFC/PPC and sensory-motor regions (Chein et al., 2005). Under this interpretation, repeated exposure to a task builds an integrated representation within DLPFC, which only needs to be activated by aPFC/aTL before it can implement the task set on its own. This is consistent with the frequent observation of DLPFC/PPC (rather than aPFC/aTL) for even very complex (moderately practiced) tasks (Chein et al., 2005; Toro, Fox et al., 2008).

The lack of DLPFC activity during novel task implementation was also unexpected. However, if an integrated task set representation had not yet developed within DLPFC, it may have been impossible for DLPFC to coordinate the unintegrated semantic represenations necessary for task performance. Rather, a region (such as aTL) both capable of representing integrated semantics and having connectivity to lower-level semantic regions (Grabowski, Damasio et al., 2001; Pobric, Jefferies et al., 2007) would be necessary to coordinate task

activity. aPFC, which is well-poised to control activity in a task-relevant manner (Burgess, Dumontheil et al., 2007), has connectivity with aTL (see Figure 11B) (Petrides and Pandya, 2007) but no direct access to lower-level semantic regions (Ramnani et al., 2004). This suggests that aPFC implements task set formation by integrating/coordinating semantic representations loaded into DLPFC and implementing them via aTL.

In addition to these unexpected results, it was unclear *a priori* whether late task preparation processes would occur during or just after the encoding period. As expected, some early preparatory processes, such as WM semantic loading for novel tasks (see Figure 8C), occurred late in the encoding period. However, based on correspondence between the cognitive model and known functions of the regions active during the first probes (see Figure 13), all hypothesized late preparatory processes appear to have occurred during the first probes. There are several possible reasons for this. First, the role of aPFC/aTL may be to coordinate the implementation of task rules during the first probe. Second, the processing of a stimulus may have been necessary to fully specify a recently loaded task set, since task set formation and task set retrieval may have involved a lack of task set detail. This second possibility is compatible with findings in task switching showing that there is a switch cost even after extensive preparation time (Monsell, 2003).

Another unexpected result involved increased activity of some regions during second (relative to first) probes. Specifically, aPFC was more active for second than first probes for practiced tasks (Figure 10B), while DLPFC was more active for second than first probes for novel tasks (Figure 10C). There are a number of possible explanations for these activity patterns, but the most parisimonious may involve a common mechanism for both regions. It is possible that since both regions were active early (encoding period) for the condition showing this effect,

that WM needed to be refreshed by the time the second probe occurred. It is also possible that a 'stop' signal (Dosenbach et al., 2006) was necessary after the regions were no longer necessary. Yet another possibility is that this effect is illusory, given that there was no difference in activity between novel and practiced second probe activity for either region. Futher research is necessary to determine the exact nature of this phenomenon.

The drop in activity during the novel task third probe (see Figure 9B, Figure 9C & Figure 10C) was also unexpected. This drop in activity may reflect an end of task setup processes, which may have continued through the second probe (e.g., for aPFC; see Figure 5B). After that, only task implementation regions may have been active. There were a number of regions equally active for novel and practiced probes relative to baseline (see Figure 14), and some of these continued to be active for novel third probes. For instance, a left fusiform region was equally active across the three probes (see section 2.4.7). It is also possible that many more regions continued to be involved during novel third probes, but that the baseline was not accurate. This is almost certainly true for right aTL (see Figure 11C), which is part of the default mode network (DMN), known to be active during 'baseline' rest periods (Raichle et al., 2001; Uddin, Clare Kelly et al., 2009). This suggests that direct comparisons of activity across conditions are more accurate than comparisons to resting 'baseline'.

The present account of aPFC's role in novel and practiced task preparation does not specify wheather the region represents the integrated task set in WM (representational account), or if it controls initial task implementation and task set retrieval from LTM (control account). It is also possible that aPFC has both roles, as the act of coordinating rules and semantics may *create* integrated task set representations in WM. Further research is necessary to distinguish between these three possibilities.

These three views (representation, control, or control + representation) are consistent with a variety of theories proposing a PFC hierarchy of function. It has been suggested that aPFC sits atop a rostro-caudal gradient of function, and contains representations of domain-general WM (e.g., D'Esposito, Postle et al., 1999), relational complexity (e.g., Bunge, Wendelken et al., 2005), and/or abstract concepts, events, or procedures (e.g., Koechlin et al., 1999; Koechlin et al., 2003) (see Badre, 2008 for review). It has also been suggested, sometimes in relation to the representational theories listed above, that aPFC controls actions/procedures extended in time as the top of a hierarchy of action (Fuster, 2001; Botvinick, 2008).

The 'integrated representation' account appears to be compatible with the majority of existing theories of aPFC function, as it involves domain-general WM (multiple semantic domains are involved), relational complexity (the task sets involve complex integrations among rules and between rules and stimuli), and abstraction (operations based on task rules during task set preparation likely involve more abstract concepts/actions than the task rules themselves). In contrast, the 'task set coordination' account appears most compatible with the theory that aPFC controls actions/procedures extended in time. Further research should focus on spatial patterns of activity within aPFC (predicted by the 'integration' account to reflect unique task representations) and causal relations between aPFC and more posterior regions (predicted by the 'coordination' account to reflect rostral-to-caudal control processes) is necessary. Additionally, if a future study were to include a variable number of rules across tasks, the 'coordination' account would predict a strong correlation between reaction times and aPFC activity for novel first probes. It is also possible that both accounts will be supported by these tests, implying some form of 'integration + coordination' account (e.g., integrating representations is necessary for coordination).

The relationship between aPFC and DLPFC is also relevant to theories of a PFC hierarchy. According to these theories, feedforward/bottom-up processes should arrive in DLPFC (lower-level) before aPFC (higher-level), while feedback/top-down processing should involve aPFC before DLPFC. Given that RITL *must* be bottom-up (since the task set is formed from external instructional stimuli), the activation of DLPFC before aPFC during novel tasks (Figure 10 & Figure 13) strongly supports the hierarchical view of aPFC-DLPFC relations. Further support for this hierarchical relationship comes from the task set retrieval framework, which is compatible with the notion that aPFC influences DLPFC in a top-down manner following LTM retrieval of practiced task sets. Finally, while aPFC and DLPFC were dissociated by this study, they were nonetheless functionally connected (see section 2.2.5; at least R=0.42) and likely interacted strongly during the short time between the end of encoding (novel: DLPFC active; practiced: aPFC active) and the beginning of the first probe (novel: aPFC active; practiced: DLPFC active).

The network status of aPFC/aTL and DLPFC/PPC was determined here in three ways. First, network status was identified based on the similarity of functional activity patterns across conditions (see Figure 8B & Figure 11D). Second, innate functional connectivity was used to determine network status outside task performance. Finally, anatomical and functional connectivity studies in non-human primates have shown strong and direct connectivity between aPFC and aTL (Petrides et al., 2007) and between DLPFC and PPC (Quintana et al., 1999). It was also shown in non-human primates that both DLPFC and PPC contain populations of cells with sustained activity related to sensory/semantic information as well as motor information (Quintana and Fuster, 1992), suggesting that these WM regions form temporary functional

networks with semantic regions during task preparation. The shift in functional connectivity between rest (Figure 11E) and task preparation (Figure 11D) supports this view.

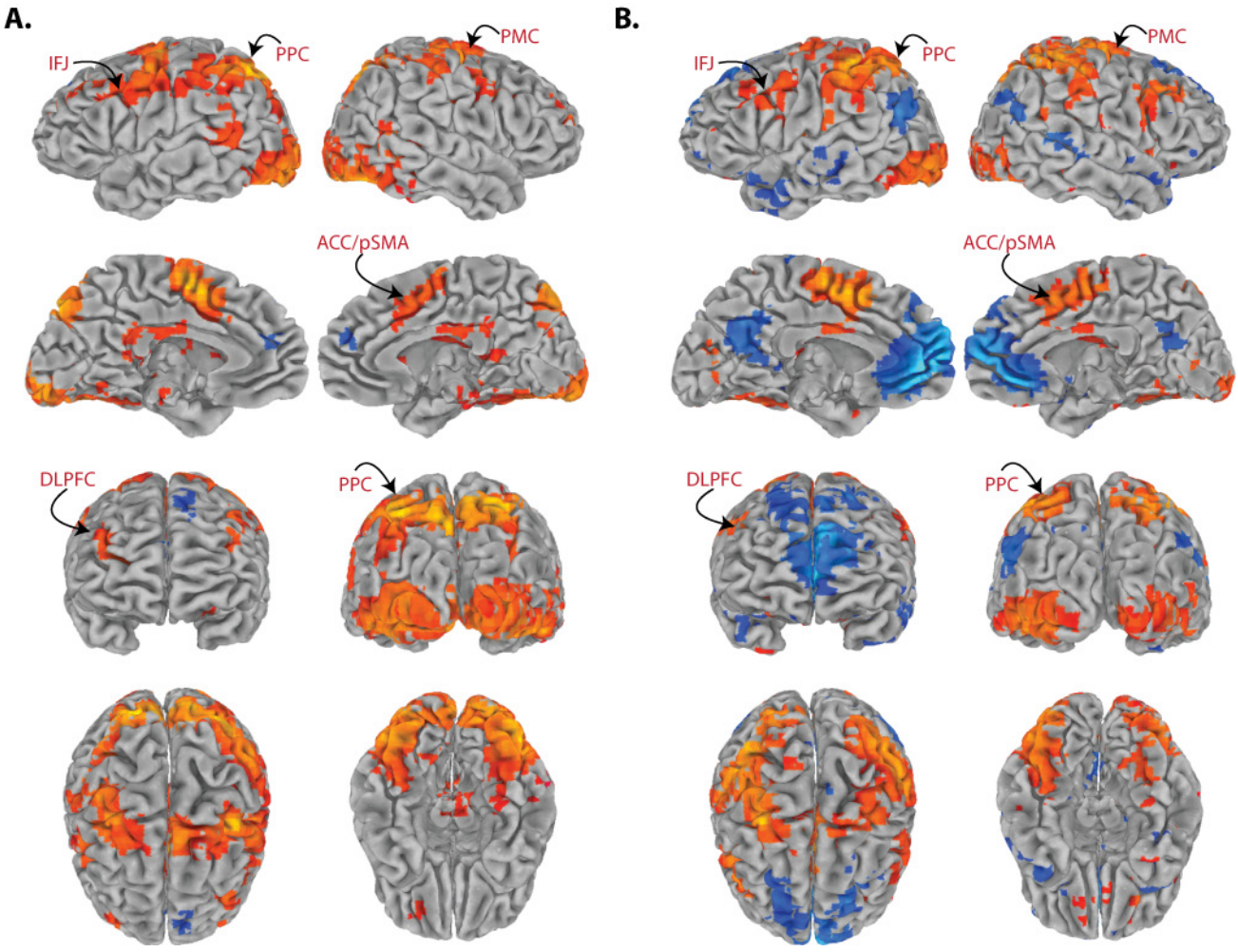

Figure 14– Encoding and Probe Activity vs. Rest
Average BOLD signals for (**A**) all encoding periods (novel & practiced) and (**B**) all probes (novel & practiced) relative to baseline ($P<0.05$, FWE corrected). Cognitive control network regions (Cole et al., 2007) are labeled. Left anterior insula cortex, another cognitive control network region, is not depicted but was also active for both conditions. The entire cognitive control network was active relative to baseline for both conditions. Note that the default-mode network (Uddin et al., 2009) shows negative activity (blue) relative to rest, suggesting that these regions have shifted baselines and should be compared across conditions rather than with the baseline.

Conclusions drawn from the findings reported here are likely generalizable to other RITL situations, since many tasks were used to support the findings. The use of logic, sensory, and

motor rules may seem obscure at first, but they are actually fundamental categories of task performance. For instance, nearly all cognitive tasks involve a motor component in order to obtain behavioral measures. Similarly, sensory information is typically used to select between different motor responses. Finally, even simple WM match to sample or lexical decision tasks involve logic rules (e.g., 'if it visually *matches*, press button 1, otherwise button 2'). It is also the case that rapid learning from instruction must involve multiple instructions and that they must be integrated. Otherwise, the task 'instruction' would simply be a LTM cue for a task learned previously. RITL relies upon the ability to generalize previously learned rules to new contexts via novel task set integration (i.e., task set formation).

RITL is not the only means of learning a new task. Other methods (shared with other animals) include exploratory (i.e., trial-and-error) and operant conditioning (B. Klein, 2008). Exploratory learning is very slow and involves pseudo-randomly trying behaviors until a particular behavior is reinforced and learned. In contrast, operant conditioning can be faster since it can involve instruction (called 'shaping'). While exploratory learning can take hours, operant conditioning can take minutes. Neither is as fast as RITL, however, which we show here can take seconds. Note that immitation, which can also take just seconds (Jackson, Meltzoff et al., 2006), can be considered another form of RITL.

Future research will be important for clarifying several aspects of the neural basis of RITL. First, separate localization of the domain-specific semantic regions would clarify which exact regions are involved for each type of task rule. Second, methods with higher temporal resolution (EEG/MEG) could test the directions of information flow proposed in Figure 13. Finally, biologically plausible computational modeling could test if the proposed process model

is adequate for implementing RITL, likely clarifying and proposing more accurate mechanisms for rapid learning.

Understanding RITL has implications for understanding a wide range of behaviors not only because so many everyday behaviors utilize RITL, but also because rapidly learning a task from instructions is a key aspect of nearly every study of cognition. Cognitive studies typically begin with a practice session in which instructions are given and subjects implement RITL in preparation for the main experimental session. Understanding RITL and the transition to steady-state task performance will be essential for characterizing everything from Stroop to lexical decision to a variety of WM and LTM tasks. The present research suggests that the transition from task set formation to task set retrieval is a key event in learning to efficiently perform a new task. Future research will be important for clarifying the timing and nature of this transition, as well as the causal dynamics between the aPFC/aTL and DLPFC/PPC networks during both processes.

2.4 MATERIALS AND METHODS

2.4.1 Participants

We included 15 right-handed subjects (8 male, 7 female), aged 19 to 29 (mean age 22) in the study. These subjects were recruited from the University of Pittsburgh and surrounding area. Subjects were excluded if they had any medical, neurological, or psychiatric illness, any contraindications for MRI scans, were non-native English speakers, or were left-handed. All subjects gave informed consent.

We included 30 subjects (10 male, 20 female) in the stimuli selection study. In the stimuli normalizing study, we included 54 subjects (21 male, 33 female). These subjects were recruited from the University of Pittsburgh and surrounding area.

2.4.2 Stimuli selection and normalization tasks

For stimulus selection, subjects were given a sheet of paper and asked to write down as many words as they could think of in each of four semantic categories. The categories included sweet, loud, soft, and green objects.

All words collected during the stimuli selection study were used in the stimuli normalization study. In this and subsequent studies, experimental stimuli were presented with E-Prime software (Schneider, Eschman et al., 2002). Subjects were shown each word on a screen, and were asked to indicate (at separate times for each category) if a given word was considered sweet, loud, soft, or green. The subjects' responses were collected as button presses (button 1 = yes, button 2 = no).

Words were included in the full study if they were considered within at least one category over 75% of the time and also considered in at least one other category under 25% of the time. During the full study, a given word would be included with exactly two category questions; one in which the the answer is 'yes' over 75% of the time and one in which the answer is 'yes' less than 25% of the time. For instance, since the word "butterscotch" was considered sweet 93% of the time, green 0%, loud 7%, and soft 27%, the word would have been included in the full study when subjects were asked 'is it sweet?' and also when they were asked 'is it green?' (it could have been used for 'is it loud?' instead, but not for 'is it soft?'). Including each word in a 'yes' context and a 'no' context (50% of the time for each) ensured that subjects could not simply

memorize the correct answer for a given word. Further, this ensured that subjects could not infer the current task's semantic question based on the most salient semantic category across the stimuli. Exactly 45 stimuli were included per semantic category, making the total stimulus count 180. Each stimulus was presented exactly 8 times (50% in a 'yes' context, 50% in a 'no' context) in the full study, across both the behavioral (50%) and scanning (50%) sessions.

The average word length for all stimuli included in the full study was 6.58 characters, with a standard deviation of 2.05. The average word length varied across categories from 5.80 (minimum) to 7.29 (maximum) characters. Thus, the average word length for each semantic category was within one standard deviation of all other semantic categories.

2.4.3 PRO task paradigm

The full study consisted of performing the permuted rule operations (PRO) task paradigm, which was developed for the first time here. The PRO paradigm combines a set of simple rule components in many different ways, creating dozens of complex task sets certain to be novel to subjects. The PRO paradigm, which makes it feasible to investigate RITL experimentally, is illustrated in Figure 2.

Four sensory semantic ('sweet', 'loud', 'soft', 'green'), four logic ('same', 'just one', 'second', 'not second'), and four motor response ('left index', 'left middle', 'right index', 'right middle') rules were used in the PRO paradigm. Each semantic judgment task consisted of one rule from each of these categories, allowing the creation of 4 X 4 X 4 = 64 distinct tasks by permuting the possible rules. Of these tasks, four (counterbalanced across subjects) were practiced (30 blocks, 90 probes each) during a two-hour behavioral session one to seven days prior to the fMRI session. These 'practiced' tasks were chosen for each subject such that each

rule was included in exactly one of the four tasks, ensuring that all rules were equally practiced. During the fMRI session, half of the blocks consisted of the practiced tasks and half of novel tasks (never seen before by the subjects). Novel and practiced blocks were randomly interleaved, with the constraint that exactly 6 blocks of each type occur within every run. With 10 runs total per subject, each novel task was presented in one block and each practiced task was presented in 15 blocks.

The semantic rules consist of questions (e.g., 'is it sweet?'). The logic rules consist of logic gates that specify how to respond based on the two semantic judgments for each probe. Specifically, 'same' is like the AND (more precisely: the XNOR, or biconditional) logic gate, 'just one' is the XOR logic gate, 'second' is a match logic gate (ignoring the first stimulus and only responding to the second), and 'not second' is the NOT logic gate (also ignoring the first stimulus, but negating the 'second' rule). The motor response rules consist of associations between decision outcomes and motor responses. Specifically, each response rule indicates the finger that must be pressed if the logic rule is 'correct', while also indicating the finger that must be pressed if the logic rule is 'incorrect'. The task instructions make explicit reference to the motor response for a 'correct' outcome, while subjects know (from the practice session) to use the other finger on the same hand for an 'incorrect' outcome.

Figure 2 illustrates an example task block, which includes encoding and three probes. Each block began with a task type cue (1 second), indicating if the upcoming task is novel (thin border) or practiced (thick border). Next was a series of three instruction screens (presented for 0.8 seconds with a 0.2 second delay between each). The order of these instructions was consistent for each subject but counterbalanced across subjects. Asterisks filled in extra spaces in each instruction screen to control for differences in visual stimulation across task rules. A series

of three probes was then shown, each one consisting of two word stimuli (presented for 0.8 seconds each with a 0.2 second delay between them), with a 2, 4, or 6 second delay (with each delay duration equally frequent) before each probe. The first stimulus of each probe included a thin red border, while the second stimulus included a thin blue border. Subjects were instructed to respond only after the second stimulus had appeared. A long delay of 12, 14, or 16 seconds (with each delay duration equally frequent) occurred at the end of each block. Each delay period longer than 0.2 seconds included a central fixation cross, with a slightly thicker fixation cross for the final block delay (to facilitate subjects entering a rest state). The delays were chosen for each block such that the block lasted exactly 36 seconds (with equal frequency for all delay durations).

2.4.4 MRI data collection

Image acquisition was carried out on a 3T Siemens Trio MRI scanner. Thirty-eight transaxial slices were acquired every 2000 ms **(**FOV: 210 mm, TE: 30 ms, Flip angle: 90°, voxel dimensions: 3.2 X 3.2 X 3.2 mm), with a total of 216 echo-planar imaging (EPI) volumes collected per run. Three-dimensional anatomical MP-RAGE images and T2 structural in-plane images were collected for each subject prior to fMRI data collection. One rest fMRI run (10 minutes, 300 EPI volumes) was collected just before the ten fMRI PRO task runs for 14 of the 15 subjects.

2.4.5 Behavioral analysis

All behavioral analyses were carried out in R (R Development Core Team, 2009). Accuracy and reaction time statistical tests were modeled in mixed-effect repeated-measures ANOVAs (subjects as a random effect). All non-switch blocks (1.2% of the probes) were removed prior to statistical analysis (see below for more details).

2.4.6 fMRI analysis

Preprocessing and analysis were carried out using AFNI (Cox, 1996). Preprocessing consisted of standard slice timing correction and motion correction and spatial smoothing (6mm FWHM). Freesurfer (Desikan, Ségonne et al., 2006) anatomical segmentations of the subjects' skull-stripped brain volumes were used as masks during AFNI preprocessing.

The preprocessed data were analyzed with a rapid event-related general linear model (GLM) (Serences, 2004) (11 regressors, encompassing 22 seconds per event). No hemodynamic response shape was assumed, as is typical of deconvolution analysis. Events included novel encoding and probes and practiced encoding and probes.

Since all novel tasks necessarily involve task switching, a lack of task switching in the control (practiced task) condition is undesirable. The probability of a task repeat was small in the present design (1.6%; only 1.2% actually repeated). Nonetheless, the few task repeats among the practiced task blocks were removed from analysis to control for any task switching effects. Incorrect probes were also removed from analyses by modeling them with a separate set of regressors (to reduce possible contamination of the correct probe responses).

The resulting statistical maps were fit to a template brain in Talairach space (Talairach, 1988) and were analyzed using group ANOVAs (with subjects as a random effect). The first ANOVA estimated mean across-subject blood oxygen level dependent (BOLD) responses to novel and practiced encoding events. Condition (novel & practiced) and time (11 time points) factors were estimated, and the Condition X Time interaction estimated for each voxel. The resulting map illustrates the statistical reliability of BOLD response shape differences between conditions.

The second group ANOVA estimated BOLD responses to probe events. The three probe events per block were estimated separately in order to test hypotheses regarding response differences across probes. In particular, we expected differences between the first novel probe (i.e., the very first time the subject performs a task) and subsequent probes. The ANOVA included task type (novel or practiced), probe number (1, 2, or 3), and subject (as a random effect) factors. The main effect of task type is reported, as well as novel vs. practiced tasks for the first probes only.

Multiple comparisons were corrected for the statistical maps using false discovery rate (FDR) (Genovese, Lazar et al., 2002) or family-wise error (FWE) cluster size thresholding (Forman, Cohen et al., 1995). FWE cluster thresholds were estimated using Monte Carlo simulations (AFNI's AlphaSim). The whole-brain FWE simulation was restricted to an anatomical mask of all gray matter (including both cortical and subcortical structures) and used a cluster-defining threshold of $P<0.05$, resulting in a 0.05 FWE-corrected cluster threshold of 89 voxels. The group anatomical mask was created from individual-subject Freesurfer anatomical segmentations, added across subjects after transforming each subject's anatomical mask to fit the Talairach template. Since we specifically hypothesized that PFC activity should be differential

across our planned comparisons, we determined an FWE cluster threshold using a Freesurfer anatomical mask restricted to lateral PFC (including the gray matter in lateral PFC: superior frontal, middle frontal, inferior frontal, and lateral orbitofrontal gyri). This FWE simulation used a cluster-defining threshold of $P<0.05$, resulting in a 0.05 FWE-corrected cluster threshold of 28 voxels.

FDR was used for the encoding period analysis in order to include small regions (e.g., those smaller than 89 voxels). The whole-brain FWE cluster threshold was used to correct for multiple comparisons for all other statistical maps. The hypothesis-driven lateral PFC-restricted FWE cluster threshold was used for all statistical maps, and the resulting PFC clusters were added to the whole-brain thresholded statistical maps for presentation in the figures.

Regions with different BOLD response shapes between encoding conditions (identified using a Condition X Time interaction analysis) had their group average time series plotted and time points with large amplitude differences across the conditions were identified. Post-hoc t-tests were used to verify if these differences were statistically significant ($P<0.05$). Two time points were tested across all regions: 4 TRs from encoding onset (early) and 7 TRs from encoding onset (late). Time series were interpolated from 2-second to 1-second temporal resolution using cubic spline interpolation for presentation in the figures.

Hierarchical cluster dendrograms were produced using R (R Development Core Team, 2009) based on the 'complete' linkage algorithm applied to the Euclidean distance between the regions' functional activity patterns. For the encoding period dendrogram (Figure 8B), the GLM-estimated time series for novel, practiced, and novel minus practiced were concatenated prior to hierarchical cluster analysis. This analysis reflects the similarity of BOLD response shape between regions for each condition and the differences between the conditions. For the probe

main effect dendrogram (Figure 9B & Figure 11D), the means of novel and practiced conditions across encoding and first probe periods were input as patterns to match across regions.

2.4.7 fMRI timing validation

Several regions identified by the fMRI analysis showed multiple BOLD responses time-locked to encoding events. Since deconvolution analysis typically involves a single BOLD response to each stimulus, we performed additional analyses to ensure the secondary response was not due to inaccurate separation of serially presented events. First, we identified a highly active region in left fusiform gyrus (identified in a statistical map of encoding activity vs. baseline ($P<0.005$, FDR corrected) and consistent with a region known to respond to word stimuli (Cohen, Lehéricy et al., 2002)). We determined that the region was responsive to all estimated events and that single BOLD responses in this region were well separated by the deconvolution analysis. Second, we performed a simulation using the task event timing (identical to that used in the fMRI study), convolving it with the canonical SPM hemodynamic response function, and adding 5% Gaussian noise (Figure 15A & B). We then analyzed the simulated data using deconvolution analysis (identically to how the fMRI data was analyzed), replicating the results in the left fusiform region. Finally, we tested many possible activity timing patterns in the simulation to reveal the only one that showed two BOLD responses similar to those seen in, e.g., right DLPFC (Figure 15C). This timing suggests that regions with doubled responses were active early in encoding and also just before the first probe.

The only other activity pattern that looked like this had early encoding activity and greater activity during the first probe than the two subsequent probes. Here the second encoding response was due to estimation of all three probes with the same regressors, which left the

additional activity in the first probe to be included in the encoding regressors. When the three probes were estimated separately, however, the illusory secondary response was removed. These simulations illustrate the importance of separately estimating the three probes, as was done in the fMRI GLM analysis.

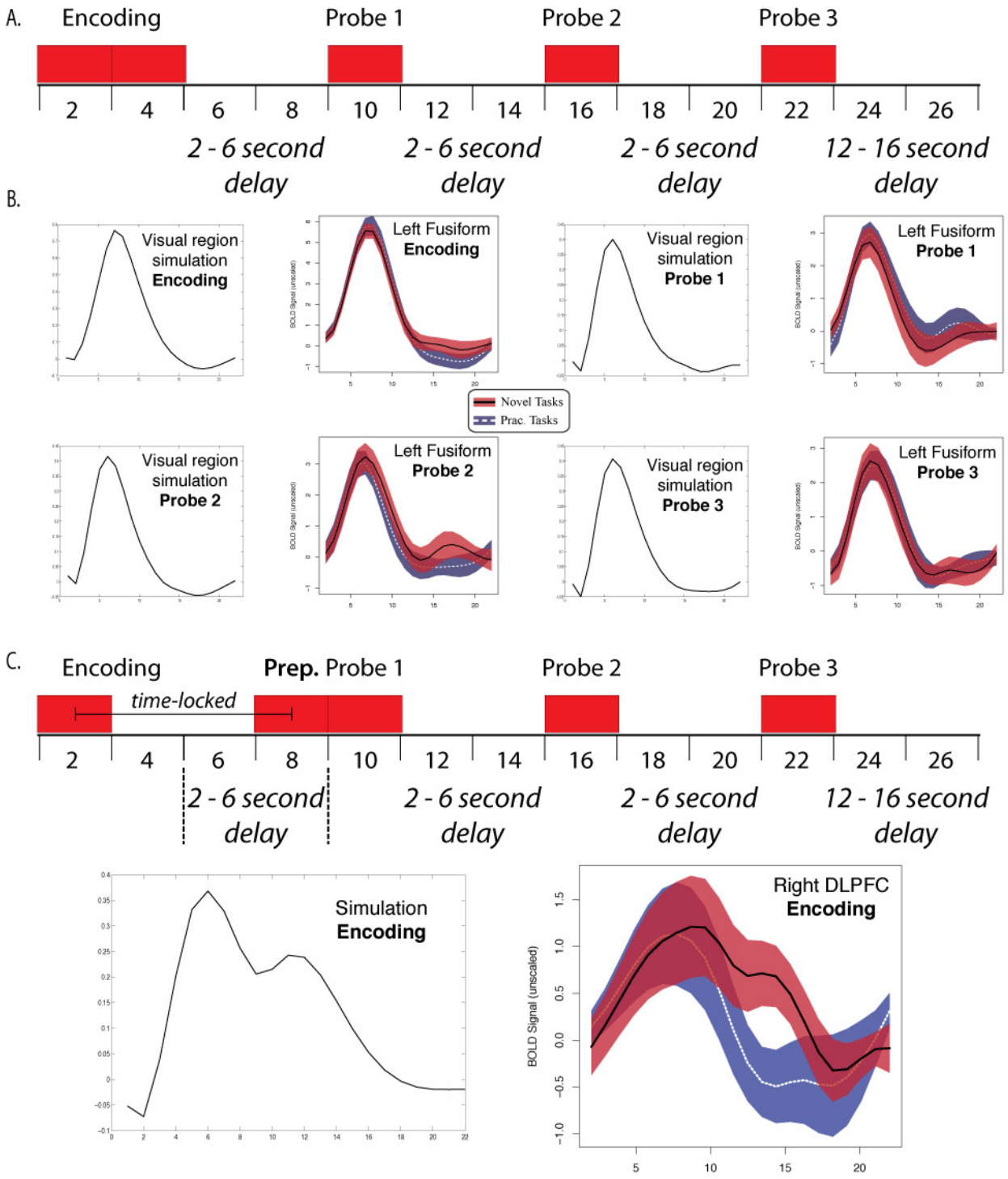

Figure 15– fMRI Timing Validation and Simulations
A) Visually evoked BOLD signal was simulated in Matlab by convolving a canonical hemodynamic response function with the timing of the visual stimuli (depicted). The exact same timing was used as during the fMRI experiment. **B**) Visual activity (from the fMRI study) in a left fusiform region was verified to separate accurately between encoding and the three probes. This separation, using an identical statistical model as was used on the fMRI data, was replicated with the simulated data. **C**) The simulation method validated in part B was used to simulate the doubled response seen in right DLPFC (among other

regions). This simulated activity pattern was identified by checking many activity timing patterns for this kind of response shape. The necessary timing to produce this result is depicted on the top.

2.4.8 Resting state functional connectivity analysis

Freesurfer (Desikan et al., 2006) was used with each subject's MP-RAGE to segment gray matter, white matter, and ventricle voxels. Following this step, fMRI preprocessing methods as implemented in AFNI (Cox, 1996) were used. Standard slice timing correction and motion correction procedures were used. As is standard with resting state functional connectivity MRI (rs-fcMRI) analyses (Shehzad, Kelly et al., 2009), time series were then extracted from white matter and ventricle voxels (with some distance from the gray matter in order to ensure no overlap), and linear regression was used to remove any correlations between the gray matter voxels and these nuisance covariates (as well as the motion parameters). The global signal was not regressed out from the gray matter voxels in order to avoid artifactual anti-correlations (Murphy, Birn et al., 2008). As is typical of rs-fcMRI, temporal filtering was applied with $0.009 < f(Hz) < 0.08$ in order to remove any remaining cardiac and respiratory effects not removed by the regression. The functional cluster dendrogram (Figure 11E) was made in an identical manner as the other dendrograms, but using the preprocessed resting fMRI data as input.

3.0 RAPID INSTRUCTED TASK LEARNING: INITIAL FINDINGS ON THE PRECISE TIMING OF PROCESSES[3]

3.1 INTRODUCTION

A major agenda within cognitive neuroscience is mapping brain structures to cognitive functions. This becomes difficult, however, for complex behaviors that involve multiple functions distributed across multiple brain networks. The need to characterize the dynamics and coordination of these networks requires both high spatial resolution to identify the networks' nodes and high temporal resolution to determine their interactions. Using magnetoencephalography (MEG), and informed by recent findings with functional MRI (fMRI), we characterized the precise timing and interactions between several cortical networks during rapid instructed task learning (RITL).

Methods suitable for localizing brain activity, such as fMRI and lesion studies, have been quite effective at discovering the brain's structure-function relationships. However, a given function rarely maps to a single structure. Rather, brain functions tend to be distributed, with functions emerging from dynamic network interactions. Much is learned by localizing a function

[3] This section reports an experiment performed using MEG. Co-authors Walter Schneider, Anto Bagic, and Robert Kass had advisory roles in this work.

to a network, yet much is left unexplained. A stronger focus on characterizing network dynamics is clearly necessary to attain a mechanistic understanding of brain function. The present study focuses on characterizing the network dynamics between regions involved in RITL in order to promote a mechanistic understanding of RITL and high-level cognition generally.

3.1.1 Measuring directed connectivity

Directed connectivity methods estimate the amount one brain region affects another. There are several methods for estimating these directed influences. I will focus here on Granger causality (GC).

In Clive Granger's 2003 Nobel Prize acceptance speech he described the basic assumptions behind his causality metric as the following: 1) Causes occur before their effects, and 2) The cause contains information about the effect that is unique, and is in no other variable (see Granger, 1969). Applied to neuroscience, region A is said to Granger-cause region B if activity in region A can be used to uniquely predict activity in region B.

Proper neuroscientific application of GC requires several kinds of information about the neural activity taking place. First, all major causal sources should be included within the statistical model in order to infer direct versus indirect directed connectivity. Neuroimaging is well suited for this since it can typically obtain brain-wide activity estimates very quickly. Second, high spatial resolution must be obtained in order to ensure multiple causal sources are not collapsed within the same time series. fMRI is well suited for this purpose, since it can obtain up to one millimeter spatial resolution. Third and finally, high temporal resolution is necessary to characterize the interactions occurring between the regions of interest, which typically occur on the order of 10 to 500 milliseconds (Schmolesky, Wang et al., 1998). MEG is well suited for this

purpose because of its ability to obtain data at under one millisecond per observation (MEG system can typically acquire data at up to 5 kHz). Together, fMRI and MEG promise to complement each other by providing high spatial *and* temporal resolution. Indeed, techniques have been developed for using fMRI to aid in the spatial location of fast MEG signals (Dale and Halgren, 2001), effectively combining these complimentary methods.

Granger causality has recently been applied to fMRI data. Goebel et al. (2003) showed that fMRI's poor temporal resolution limited GC to estimating dominant influences between only two regions at a time (also see Roebroeck, Formisano et al., 2005). In other words, fMRI cannot be used to distinguish between bidirectional and unidirectional influences, or bidirectional and non-directional influences. If, for instance, region A influences region B and region B influences region A equally in return, then there would be no dominant influence and an fMRI GC measure would indicate no interaction between the regions. Due to this limitation fMRI can only be used in the bivariate case, such that direct interactions can only be inferred circuitously.

These major limitations in the utility of GC with fMRI all stem from fMRI's poor temporal resolution. MEG has much higher temporal resolution, and so does not suffer from the same issues as fMRI. This method has other issues, however, such as confounding of spatial localization due to the "inverse problem", which is also shared with electroencephalography (EEG). MEG measures the magnetic fields produced by synchronous neuronal populations which, unlike the electrical fields measured by EEG, pass through the skull and scalp undistorted (Hämäläinen and Hari, 2004). Thus, though MEG shares spatial localization issues with EEG, it is less of a problem due to the nature of magnetic versus electric signals.

Estimating directed influences between brain regions is important for furthering understanding of RITL. This information can likely distinguish between several hypothetical models of RITL function.

3.2 MATERIALS AND METHODS

3.2.1 Participants

We included ten right-handed subjects (4 male, 6 female), aged 18 to 30 (mean age 23.5) in the MEG study. These subjects were recruited from the University of Pittsburgh and surrounding area. Subjects were excluded if they had any medical, neurological, or psychiatric illness, any contraindications for MEG or MRI scans, were non-native English speakers, or were left-handed. All subjects gave informed consent.

3.2.2 Cognitive paradigm

The cognitive paradigm used in the MEG study was very similar to the paradigm used in the fMRI-only study (see section 2.4.3). There were several changes made, however. First, the amount of practice per subject was doubled. The actual time practicing was similar (~2 hours), but the duration of delays between blocks was shortened to allow for 60 blocks (180 probes) for each of the 4 practiced tasks.

Second, the delays prior to probes were shortened to take advantage of the higher temporal resolution afforded by MEG. Delays varied from 2 to 6 second in the fMRI version,

and this was shortened to 2 to 3 seconds (in 500 ms intervals) for MEG. Stimulus duration remained the same (800 ms), with 150 to 250 ms (in 50 ms intervals) delays following each one. Each block ended with a 6 to 7 second (in 500 ms intervals) delay.

Third, rather than have novel and practiced task blocks randomly intermixed (as in the fMRI study), these block types were placed together in epochs. The order of epochs was counterbalanced across runs (one novel task epoch and one practiced task epoch were presented per run). A 2 second 'epoch cue' indicated if the upcoming blocks were novel or practiced. Additionally, as in the previous version, the border around the instructions for each block was thicker for practiced than novel tasks.

Fourth, no task repeats were allowed. The previous version had allowed occasional task repeats (~1.2% of the time). The present design restricted the practiced tasks such that the upcoming task could not have been just performed.

Fifth, it was possible (though unlikely) in the previous design for the same stimuli to occur with the same practiced task repeatedly. This could confound novel task learning ('generativity') with novel stimulus-task pairing ('stimulus generalization'). Note that this was less likely for SAME and JUSTONE rules since the *pairs* of words were typically unique for a given practiced task. In order to correct for this subtle confound, all stimulus pairings were ensured to be unique for practiced tasks including the SAME and JUSTONE rules. For practiced tasks including the SECOND and NOT SECOND rules, the second stimulus (the first one is ignored by subjects) was only allowed to be presented once for each of the rules. These corrections took into account the presentation history during the practice session and also during the neuroimaging session.

All of these modifications were used for the MEG-subject fMRI data collection as well. However, the 2-6 second delays just before probes were retained from the previous design, as well as the 12-16 second inter-block delays. Further, only 5 runs of data were collected (rather than the previous 10) along with the structural MRI images. Note that only results from the MEG data are reported here.

3.2.3 MEG data collection

Data acquisition was carried out with a Vectorview (Neuromag-Elekta Oy, Helsinki, Finland) 306-sensor MEG system. To monitor eye movements and eye blinks, two bipolar electrode pairs were used for vertical and horizontal electrooculogram (EOG) recordings. These were placed above and below the left eye for vertical EOG and laterally to each eye for horizontal EOG. Two electrocardiogram (ECG) electrodes were placed on each subject's chest in order to record heart and breathing signals. A ground electrode was placed behind each subject's right ear, and reference electrode was placed behind each subject's left ear. Four head position indicator (HPI) coils were attached onto each subject's scalp. Next, the 3D locations of three cardinal landmarks (nasion, left and right periauriculars), HPI coils, and additional points on the scalp were digitized using a Fastrak system (Polhemus, Colchester, VT) to allow subsequent registration of the MEG data with each subject's structural MRI.

Each subject was then seated in a magnetically shielded room in an upright position and his or her head was positioned inside the MEG helmet. Before recording, current was passed in the HPI coils in order to determine the position of each subject's head relative to the MEG sensor array. After every scanning run (10 runs, 5 minutes each), recording was stopped and an additional HPI reading was taken. The MEG data were acquired and saved continuously at a

sampling rate of 1000 Hz with a recording lowpass filter set to 325 Hz.

3.2.4 MRI data collection

Image acquisition was carried out on a 3T Siemens Trio MRI scanner. Thirty-eight transaxial slices were acquired every 2000 ms **(**FOV: 210 mm, TE: 30 ms, Flip angle: 90°, Slice thickness: 3.2 mm), with a total of 216 EPI volumes collected per run. Five fMRI runs were performed (not reported here). Two three-dimensional anatomical MP-RAGE (voxel volume: 1.0 mm^3) images and one T2 structural in-plane image were collected for each subject.

3.2.5 MEG data analysis

MNE software (Hämäläinen and Ilmoniemi, 1994) was used for most of the MEG analyses. Several preprocessing steps were carried out prior to full analysis. First, the two structural MRI MPRAGE images were averaged (following motion correction) and converted to three-dimensional anatomical surfaces using Freesurfer (Desikan et al., 2006) (see Appendix A for details). Next, signal-space separation (SSS) (Taulu, Simola et al., 2004; Taulu, Simola et al., 2005) as implemented by MaxFilter software (Neuromag-Elekta Oy, Helsinki, Finland) was used to remove artifacts from signals outside the head. SSS was used with the temporal extension (Taulu and Hari, 2008) using a 60-second moving window. The MaxMove feature in MaxFilter was used to bring each subject's 10 runs into alignment with his/her first run. Next, using a custom MATLAB (2007a; The MathWorks, Natick, MA) program, EOG and ECG signals were regressed out of the MEG signals to reduce eye blink (and cardiac) artifacts (Wallstrom, Kass et al., 2004) using a 60-second moving window. Next, in MNE software, a high-resolution three-

dimension MRI-based reconstruction of each subject's head was used to align the cardinal points between the MRI and each subject's first run of MEG data. This alignment was later applied while calculating the forward solution between the cortical surface and the MEG sensors. The data were bandpass filtered from 1 to 40 Hz prior to MNE analysis (no filtering was applied for the GC analyses).

In MNE software, the data were averaged by condition in 800 ms (stimulus duration) windows for all but the last encoding stimulus. The window for this last encoding stimulus included the following delay period because of hypothesized WM processes during this period. Any trials with MEG artifacts exceeding 200 pico-Teslas (pT) were removed prior to averaging. The forward solution for each run was then computed for each run's average and then the forward solutions were averaged together into a grand average. The covariance matrices for each run, which estimate the level of noise in the data, were computed using part of the rest period data between blocks (1000 time points from 4 to 3 seconds prior to the first encoding stimulus). This same time period was used to baseline the data. The inverse solution was then computed using the grand mean covariance matrix and the grand mean forward solution.

Using the grand mean data, the inverse solution was used to localize the estimated current (minimum norm estimate (MNE)) derived from the magnetic fields measured with the MEG sensors. Depth-weighting was used to help counter the localization bias toward the superficial surface (Lin, Witzel et al., 2006). Further, because most of the MEG signal is thought to derive from pyramidal neurons and because pyramidal neurons are normal to the cortical surface, current orientation was loosely constrained to surface-normal (constraint value of 0.4) (Lin, Belliveau et al., 2005). These MNE values, located across the cortical surface, were then divided by the grand mean noise estimate to produce a dynamic statistical parametric map (dSPM) (Dale,

Liu et al., 2000) for each condition. The dSPM value is a signal-to-noise estimate, and has an associated F-value and p-value for evaluating statistical confidence.

For group analysis, each subject's dSPM data were projected onto the cortical surface and morphed (Fischl, Sereno et al., 1999b) to a common Talairach (Talairach, 1988) template surface. All subjects' data were averaged on this common template surface. These group averaged dSPM values were also subtracted between conditions and projected back to the surface for condition contrast analyses.

3.2.6 Granger causality analysis

Multivariate GC was implemented using the Causal Connectivity Toolbox (Seth, 2005) in MATLAB. Prior to running the GC analysis, four regions of interests (ROIs) were selected from the set of dSPM contrasts (see Figure 16), based on dissociations found in a previous fMRI study (see section 2.0). These regions included: right dorsolateral prefrontal cortex (DLPFC), left posterior parietal cortex (PPC), left anterior prefrontal cortex (aPFC), and right anterior temporal lobe (aTL). The regions were identified on a Talairach brain surface, and morphed to each subject's individual space (see Section 1.01(a)(i)Appendix A). MNE software was used to apply the inverse transformation for each subject to attain time series for each ROI for every time point collected (10 runs per subject). These time series were then imported into MATLAB for GC analysis.

Multivariate GC, in which all ROIs are included in the same model, was used to rule out indirect influences among the ROIs. Note, however, that other regions not included in the model may have played intermediary roles between the ROIs. Each trial of each condition was modeled separately in order to account for inter-trial and inter-subject differences in timing (though the

same lag was assumed). The Akaike information criterion (Akaike, 1976) was used to find the optimal model order by estimating it for all trials and averaging the estimates. The GC magnitude was estimated for all trials separately as the log F-ratio (Geweke, 1982), and these estimates were included in several statistical tests. First, t-tests were used to check for GC magnitudes relative to 0. Second, t-tests (paired by subject) compared the directions of influence between each pair of regions to determine if there is a dominant influence between them (i.e., if one direction is significantly larger than the other). Finally, t-tests (paired by subject) compared the GC magnitudes across conditions (novel vs. practiced tasks) for each time period.

Stationarity, the notion that causal influences do not change during the duration of the time series, was not assumed here. Rather, the GC magnitude was assumed to accurately summarize the dominant linear (stationary and non-stationary) Granger-causal influences within the time series, and inter-trial variance was used to produce p-values based on t-tests. The p-values based on the t-tests were used rather than the p-value associated with each GC test since the GC tests' p-values assume stationarity.

3.2.7 Limitations of the analyses due to eye artifacts

The present MEG analyses were performed on only five of the ten subjects due to eye blink artifacts. These artifacts have high amplitude, primarily affect prefrontal cortex and temporal pole, and occur late in (or just after) stimulus presentation (Croft and Barry, 2000). Several of the main ROIs here are in prefrontal cortex and near temporal pole (aPFC and aTL), and activity in these regions is likely to occur late or just after stimulus presentation because they involve high-level cognitive processes. This 'perfect storm' of overlap between the artifacts and the signals of interest was problematic for the planned analyses.

Many existing eye artifact removal methods were tried with the present dataset. These methods included principle components analysis (PCA), independent components analysis (ICA) (Li and Principe, 2006), PCA-enhanced ICA (Okada, Jung et al., 2007), adaptive filtering (He, Wilson et al., 2004), wavelet-enhanced ICA (wICA) (Castellanos and Makarov, 2006; Inuso, La Foresta et al., 2007; Kumar, Arumuganathan et al., 2008), and EOG regression (Wallstrom et al., 2004). The most effective method with these data was EOG regression. This conclusion was based on visual identification of the eye artifacts (which were larger than brain-originating signals and tended to have a distinct temporal profile) in the raw data of several frontal and temporal pole channels both before and after each method was applied.

Importantly, Wallstrom et al. (2004) came to the same conclusion using a more systematic approach involving electroencephalography (EEG) and simulations. One of their concerns was distortion of the data by the ICA and PCA methods. The second-best method found here, wICA, was specifically developed in response to Wallstrom et al. and reduces ICA distortions by only removing the frequencies contaminated by the artifact (Castellanos et al., 2006). The major issue with applying wICA to the present study's data was the variability in its effectiveness across subjects.

EOG regression was used for the presently reported analyses. However, some of the subjects' EOG data were of lower quality, possibly because of poor conductance of the EOG electrodes during data collection. These subjects (N = 5) were not used in the present analyses, since EOG regression left many of the eye artifacts in these cases.

3.3 RESULTS

3.3.1 Novel vs. practiced activity contrasts

The dSPM estimates for the encoding and first probe periods were contrasted across novel and practiced conditions (Figure 16). The maps were thresholded at $P<0.05$. Time points were chosen for their ability to summarize the differences seen for each time period, and also for illustration of the ROIs identified in the previous fMRI study (see Figure 9 & Figure 11).

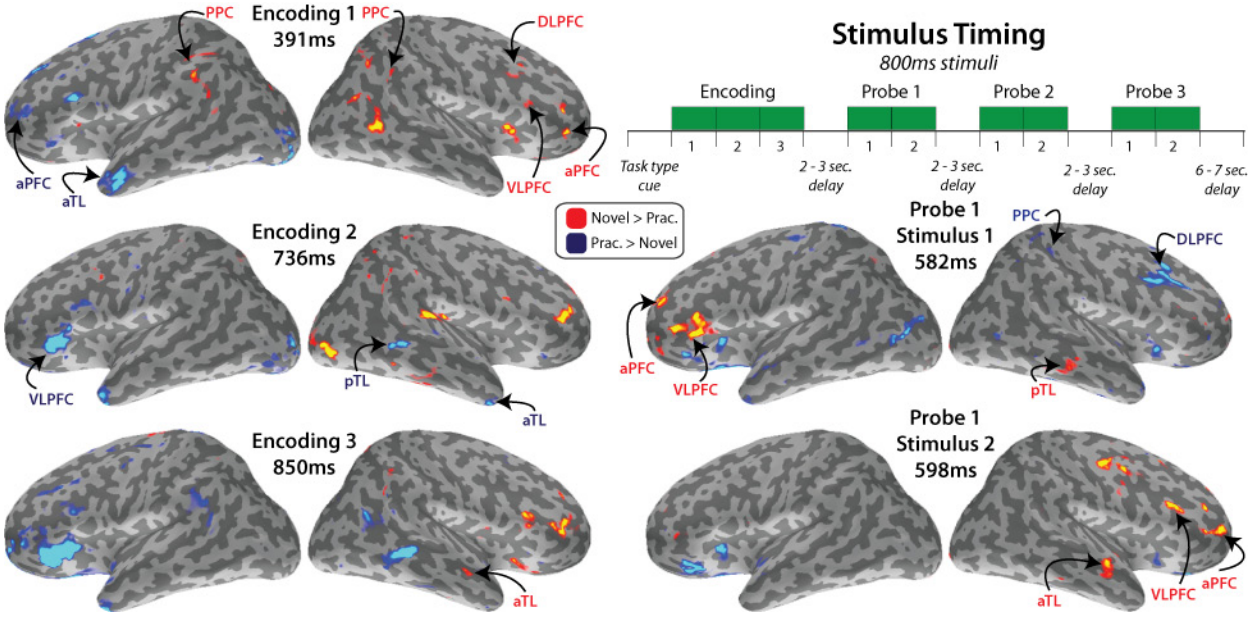

Figure 16 – The MEG results replicate the fMRI results with more precise timing
The normalized electric current estimates (dSPMs) were contrasted between novel and practiced tasks across five time periods thought to be involved in task preparation. The activity patterns largely replicated previous findings with fMRI. The more precise timing afforded by MEG reveals that right PFC (aPFC/DLPFC) and PPC are more active for novel than practiced throughout the encoding, but switch to be more active for practiced for the first part of the first probe. The opposite pattern occurred for left PFC (aPFC/VLPFC) and temporal lobe (aTL/pTL). The dSPM difference maps are depicted on an inflated cortical surface.

Left aPFC was identified as more active for practiced than novel during the first encoding period. This pattern was present throughout encoding, but switched for the first probe, with left aPFC being more active for novel than practiced. Left VLPFC also showed this pattern.

In contrast, right DLPFC was more active for novel than practiced during encoding, while it was more active for practiced than novel for the first probe. Consistent with the interpretation that right DLPFC is involved in task preparation, the region was only more active for practiced than novel for the first probe's first stimulus. Right (not left) aPFC showed a similar pattern as right DLPFC, but was also more active for novel during the first probe's second stimulus.

The location of activity in the temporal lobe was more variable here than the fMRI study. Left aTL was initially (encoding 1) more active for practiced than novel, which is consistent with the interpretation that this region is involved in task set retrieval, but it showed little more differential activity after this. Right aTL at the temporal pole was more active for practiced for the second encoding stimulus. Interestingly, right aTL was already more active for novel during the third encoding stimulus, suggesting task component integration may begin at that time. This region was also more active for novel than practiced during the first probe's second stimulus.

Right posterior temporal lobe (pTL) appears to be more consistent with the previous fMRI study, possibly as a result of mislocalization with MEG. This region was more active for practiced during encoding, but more active for novel during the first probe's first stimulus.

3.3.2 Novel vs. practiced Granger causality

Multivariate GC was used to estimate directed influences between four ROIs during task performance. GC estimates across both novel and practiced tasks were compared to 0, and dominant influences were determined from differences in GC magnitude between directions (Figure 17A). Further, GC estimates for novel and practiced tasks GC were compared to each other (Figure 17B).

It was found that all connections were statistically greater than 0 ($P<0.05$) except for aTL with PPC during the first probe. These connections were all bidirectional, which is consistent with anatomical studies showing that cortical regions are typically connected bidirectionally (Maunsell and van Essen, 1983; Pandya and Yeterian, 1985; Fuster, 2006). However, some of the directions were statistically greater than others. aPFC dominantly influenced DLPFC for the practiced condition in both time periods. aPFC also dominantly influenced PPC in the same way, but also during the first probe for novel tasks. Importantly, aTL dominantly influenced aPFC during encoding for the practiced condition, suggesting that aTL stores the pre-integrated task set and loads it into aPFC from long-term memory.

Changes in GC between novel and practiced tasks were estimated for the post-encoding delay period, revealing increased influences from PPC ($P=0.02$) and DLPFC ($P=0.054$) to aPFC for novel tasks. In contrast, aTL ($P=0.03$) was found to increase its influence on aPFC during practiced tasks, further supporting the loading of task sets from long-term memory.

Changes in GC were also estimated for the first probe's first stimulus. It was found that DLPFC ($P=0.03$) increased its influence on aPFC for novel tasks. In contrast, aTL ($P=0.076$) increased its influence on DLPFC for practiced tasks, suggesting that both aPFC and aTL load the task set into DLPFC during the first probe's first stimulus.

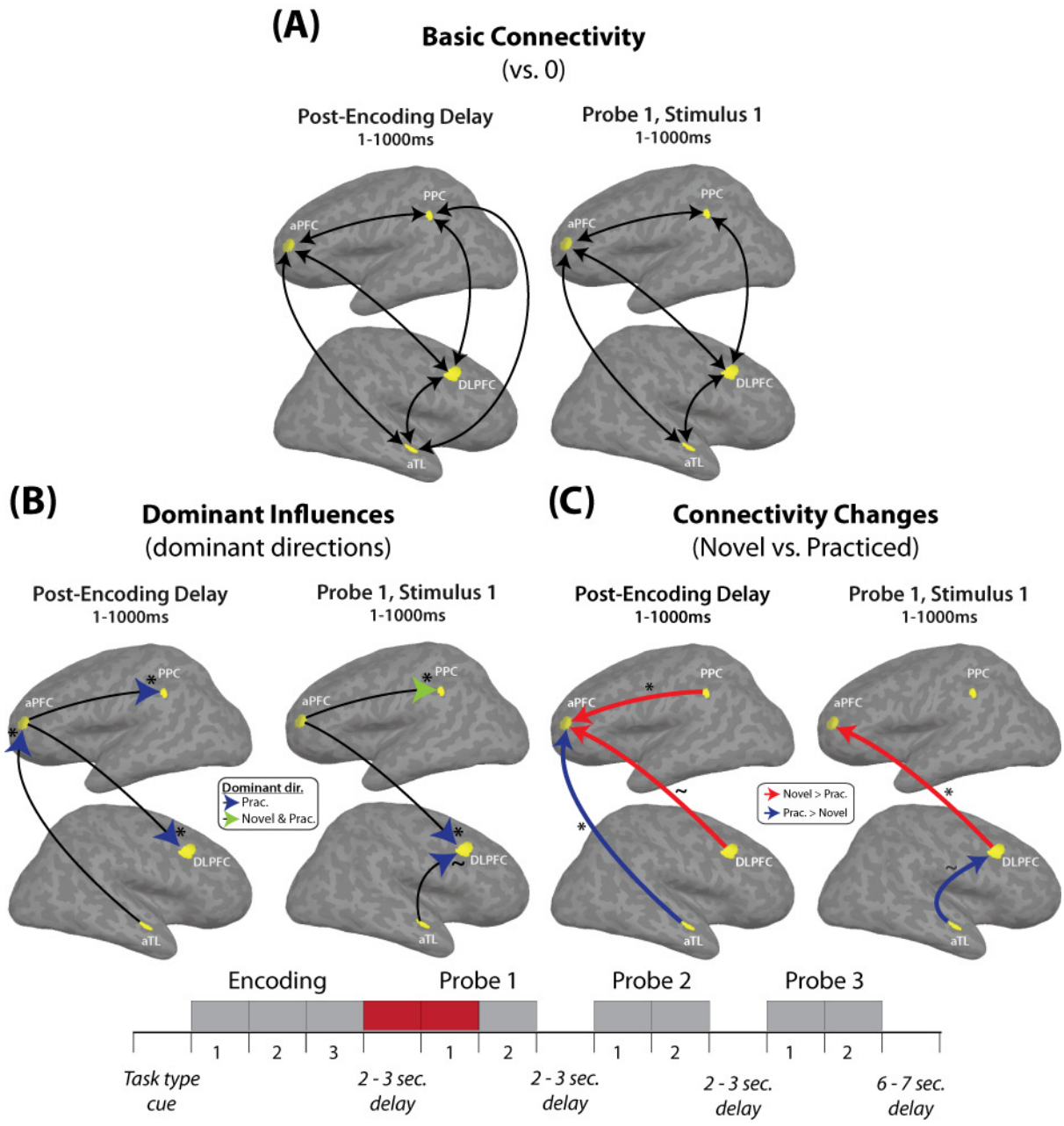

Figure 17 – Granger causality reveals directed influences between regions of interest
Multivariate GC was used to determine the directed influences among ROIs during time periods thought to involve task preparation. Of special interest were time periods likely to show transfer of information between aPFC/aTL and DLPFC/PPC, as a previous fMRI study revealed these two networks but not how they interact. These time periods included the first second of the post-encoding delay, and the first probe's first stimulus. **A**) The black arrows indicate GC relative to 0 for both novel and practiced conditions. Notice that all regions show bidirectional connectivity, possibly reflecting complementary feedforward and feedback interactions for network settling. **B)** The blue arrows indicate the dominant influence between pairs of brain regions for the practiced condition, while the green arrow indicates a dominant influence for both novel and practiced conditions. **C**) The GC magnitudes were compared between novel and practiced tasks. Red arrows indicate novel > practiced, while blue arrows indicate practiced > novel.
~ = P<0.10, * = P<0.05

3.4 DISCUSSION

A previous fMRI study found that aPFC and aTL form a network and DLPFC and PPC form a network (see Figure 11 & Figure 13), but not how these two networks interact. The present study used MEG to not only replicate the previous findings with higher temporal resolution (see Figure 16), but also to determine that the networks interact during the post-encoding delay and during the first probe's first stimulus (see Figure 17). This narrow window in time (3-4 seconds), between the end of task instruction and prior to the first response, is apparently long enough to bring a task from markings on a screen to complex thoughts and actions in a human being.

Multivariate GC was used to test the previously predicted directions of influence between the ROIs. All predicted connections that were tested were verified, with stronger verification (based on significant changes in connectivity) for some connections (Figure 18). This stronger verification was found for DLPFC → aPFC for novel tasks, and aTL → aPFC for practiced tasks. These findings support the hypothesis that DLPFC loads semantic information from instructions into aPFC during novel tasks, but that this semantic information is loaded into aPFC from long-term memory stored in aTL for practiced tasks. Further, an unexpected connection from aTL to DLPFC significantly increased during practiced tasks, suggesting aTL can load task sets from long-term memory directly into DLPFC for working memory maintenance.

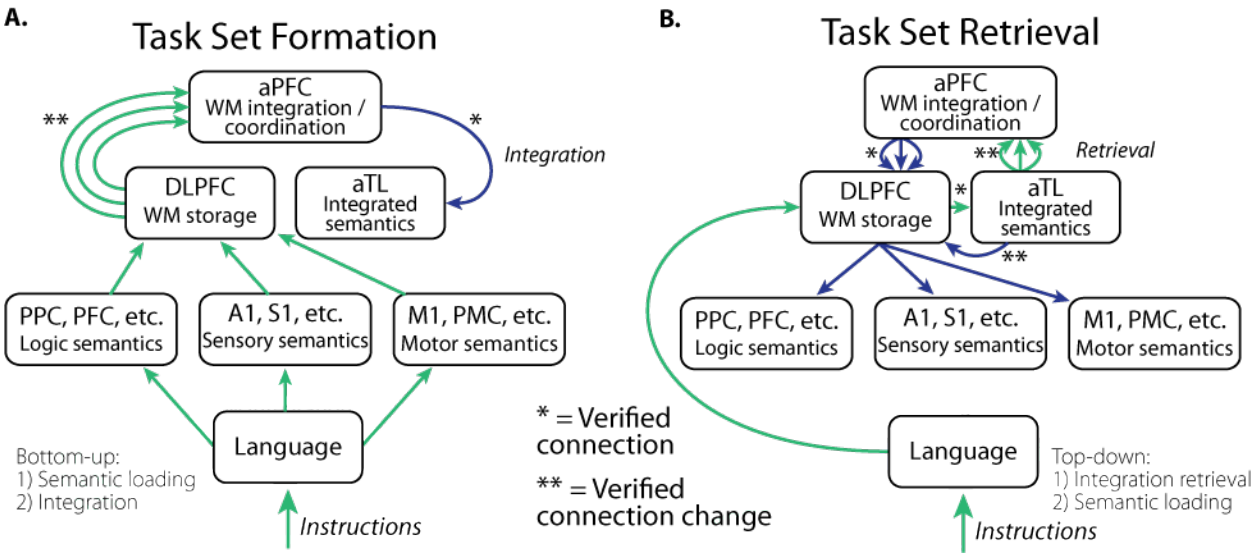

Figure 18 – Mapping neural dynamics to cognitive functions
All tested connections were verified. A stronger verification, based on a significant change in connection strength in the predicted direction, was found for DLPFC → aPFC (for novel tasks) and aTL → aPFC (for practiced tasks). An unexpected influence from aTL to DLPFC was discovered, likely reflecting the unity of aPFC and aTL as they exert top-down control on DLPFC.

Further research is necessary to test for the predicted interactions between the language, semantic, and cognitive control regions. Such research might be tricky, since the current version of the PRO paradigm provides less-than-ideal statistical power for each semantic category. However, even the data presented here may be sufficient to characterize these network interactions, since the current study is preliminary and the included set of analyses focused solely on interactions between the cognitive control regions (DLPFC, PPC, aPFC, and aTL). I plan to analyze a larger set of connections for the complete version of this study.

4.0 GENERAL DISCUSSION

> The transmission and alteration of humanity's vast stores of cultural information have been the key to *Homo sapiens'* rise to dominance – the reason why the world isn't run by chimpanzees.
> – A. Ehrlich, *The Dominant Animal*

Increased understanding of RITL can shed light on basic human cognition, disorders of cognition, education, artificial intelligence, and the nature of hierarchy in the human brain, to name just a few. This list may seem too broad for something so seemingly obscure and so little researched. Upon careful examination, however, it is clear that RITL is not obscure but rather a fundamental human trait, almost certainly instrumental in our rise to become the dominant species on the planet. Further, it is clear that RITL is so poorly understood not for lack of trying (see section 1.4)(also, Monsell, 1996; Burgess, 1997) but rather, as explained below, because of difficulties overcoming limitations inherent to the very nature of the scientific method.

Every cognitively healthy human being can perform RITL, with the vast majority of 3+ year-olds performing it better than the very best non-human (Savage-Rumbaugh et al., 1993). Further, it has been argued that RITL is necessary for cultural transmission of knowledge and behaviors between individuals (Noelle, 1997), and that cultural transmission is essential to the stature humanity has attained in the world (Ehrlich, 2008). Thus, RITL may be so fundamental that it is a defining trait of *Homo sapiens*. Research on RITL and the cognitive processes underlying it, such as imitation, language, and working memory, is clearly of great importance.

There has been little empirical research on RITL likely because of the difficulty in satisfying the scientific method's need for replication and experimental control. If a single observation were sufficient for drawing strong scientific conclusions, then observing an individual's behavior on the first trial of a new task would be sufficient for understanding RITL (see Burgess, 1997). This approach is not sufficient, however, because we wish to attain statistical confidence and generalize our findings using multiple observations, in addition to ruling out other lurking variables like stimulus novelty using a proper control condition. The PRO cognitive paradigm (see Figure 2) satisfies both the need for replication (within-subject) as well as experimental control, ruling out most lurking variables.

Neuroimaging was used in conjunction with this new cognitive paradigm to better understand RITL. The findings suggest RITL is made possible in humans via a combination of language, enhanced WM manipulation (aPFC), and enhanced integrated semantic representation (aTL). These findings and their implications are explored in the following sections.

4.1 SUMMARY OF FINDINGS

> There are no small problems. Problems that appear small are large problems that are not understood.
> – S. Ramon y Cajal, *Advice for a Young Investigator*

I have presented several novel findings here. The various small contributions I have made to this colossal problem will be discussed below.

Behaviorally, I showed that individuals can learn very complex tasks virtually instantly upon being instructed. Interestingly, though there were consistent behavioral differences between

novel and practiced first tasks, these differences were quite small (2% accuracy, 30ms reaction time) given the large difference in practice between them (0 vs. 90 trials). This small difference is likely due to a ceiling effect. Importantly, however, it suggests that differences between the novel and practiced conditions were unlikely to be due to differences in accuracy or reaction time. Several recent studies have emphasized the potentially confounding nature of reaction time (Grinband et al., 2006; Yarkoni, Barch et al., 2009) and accuracy effects (Brown and Braver, 2005) for neuroimaging studies, which are likely avoided here.

I demonstrated a series of effects with fMRI. First, predicted differences in BOLD response shape were verified, supporting the separation between the postulated neurocognitive task preparation processes: *task set formation* and *task set retrieval*. However, some preparation was also expected to occur during the first probe. The pattern of activity across encoding and first probes suggested that the late encoding period activity was actually an *early* part of task set formation. Thus, the encoding period involved semantic WM loading (DLPFC, PPC, and semantic regions), followed by WM integration (aPFC and aTL) during the first probe. In contrast, aPFC and aTL were active early (during the encoding period) for practiced tasks, while DLPFC, PPC, and semantic regions were active (late) during the first probe. This double dissociation between the DLPFC/PPC and aPFC/aTL networks suggests that the first is involved in WM semantic loading and the second is involved in WM integration.

I also presented preliminary findings on precise temporal estimation of RITL processes using MEG. First, the MEG findings support the fMRI results, showing a similar dissociation between brain regions across encoding and first probes. However, the precise timing of MEG provides some novel insights as well. For instance, right PFC and PPC were active for all three

encoding screens for novel tasks, supporting the hypothesis that these regions are involved in loading each (unintegrated) instruction into WM (see Figure 7A).

In contrast, it was unclear from the fMRI results if during practiced tasks left PFC responded to each encoding stimulus or simply used the first encoding stimulus to retrieve the task from memory. The MEG results suggest that the first encoding stimulus initiates the retrieval process, but that all encoding stimuli contribute to that process. Further, multivariate GC revealed that the retrieval process continues until just after the final encoding stimulus, and that transfer from the aPFC/aTL network to the DLPFC/PPC network may not occur until the start of the first probe.

Taken together, the MEG results 1) solidify the fMRI results by replicating them, 2) increase the certainty of the timing inferred from the fMRI time series, 3) expand knowledge of the exact timing of neural activity during each cognitive event, and 4) verify network interactions predicted by the patterns of fMRI activity.

4.2 THE NATURE OF CONNECTIVITY AND NETWORK STATUS

A classic debate within neuroscience has been between *distribution* (originally supported by Flourens in 1824 and Lashley in 1929) and *localization* (originally supported by Bouillaud in 1825 and Broca in 1861) of function within cortex (see Bennett and Hacker, 2008). Modern electrophysiological and neuroimaging methods have firmly established the truth of *both* these approaches (Fuster, 2006). Instead of 'equipotentiality' or 'phrenology', function maps to brain structure via networks and their components. The future of neuroscience is in characterizing networks.

It is strange, given the central role of networks to neuroscientific progress, that the concept 'brain network' is so poorly defined (Horwitz, 2003). The core of the problem lies in the concept of connectivity. If we say a 'network' is any set of connected brain regions, we can safely call the entire brain a network and be done with the whole enterprise. Indeed, graph theorists have shown that the brain is a small-world network, meaning that any point in the brain can be reached from any other point via very few connections (Bullmore and Sporns, 2009). How then might a network be objectively defined such that the concept remains useful?

I propose the use of a dual approach to network definition. We are ultimately interested in network mechanisms, and so a *dynamic* network can be defined in a particular functional context. This definition could include, for instance, DLPFC with visual cortex when visual information is maintained in WM, and DLPFC with auditory cortex when auditory information is maintained. However, this definition would (rightly, in some sense) include primary visual cortex in the same network as primary motor cortex in a visuo-motor task. In order to capture the typical functional relationships between brain regions *static* networks can be defined as those networks with strong anatomical connectivity, coactivation across many contexts, and strong functional connectivity in a task-neutral context. Together these definitions can capture both typical and context-dependent network mechanisms.

This dual approach to network definition was used in the present research. First, co-activation of aPFC/aIT and DLPFC/PPC/etc. were established. This could reflect their status as networks, but not by necessity (see Stephan, 2004). A cluster analysis (see Figure 8) verified similar BOLD response shapes across conditions, suggesting high functional connectivity. Differences in this functional connectivity between the regions suggested that these two sets of regions formed two *dynamic* networks. A second cluster analysis, demonstrating similar

responses across conditions, confirmed this possibility (see Figure 11D). Resting state functional connectivity MRI was used to test for *static* network status. Clustering the low-frequency resting state fMRI data revealed that aPFC/aTL forms a static network, while DLPFC/PPC/etc. does not. Rather, DLPFC/PPC form a static network while the other (task semantic) regions were only connected to DLPFC/PPC in a dynamic task-driven context.

Characterizing these two networks using the dual approach proposed here allows for a deeper understanding not only of network dynamics, but also for two forms of prediction: one that is context independent (e.g., DLPFC and PPC are likely to interact across many contexts) and one that is context dependent (e.g., DLPFC and auditory cortex are likely to interact in auditory WM contexts).

4.3 PROCESSES UNDERLYING RESTING STATE CONNECTIVITY[4]

Resting state correlations in fMRI signals were used to assess functional connectivity in order to rule out connectivity patterns tied to any specific task (see 2.4.8). The fluctuations underlying correlations likely originate from spontaneous slow-wave electrical activity across cortex. Beggs & Plenz (2003) showed that such spontaneous electrical activity occurs across cortex in rats at rest. Beggs & Plenz characterized these burst in activity as 'neuronal circuit avalanches', which quickly traverse neuronal circuits periodically. Golanov et al. (1994) had previously shown these same bursts of activity, and that they are followed by increases in regional cerebral blood flow lasting ~12 seconds. Such increases in blood flow are known to increase the blood oxygen level

[4] Most of this text is from Cole & Schneider, 2007

dependent (BOLD) signal, as measured by fMRI (Ogawa, Tank et al., 1992). Further, Golanov et al. found that these spontaneous waves of electrical activity and the corresponding blood flow changes (R=0.94 between these events) occurred at ~0.1 Hz, which is approximately the same rate as resting state BOLD fluctuations in humans (Cordes, Haughton et al., 2001).

These slow BOLD fluctuations are correlated between functionally connected regions (Xiong, Parsons et al., 1999) likely because when two regions are anatomically connected with significant synaptic weights these connections do not 'turn off' when not being used for task performance. Instead, any spontaneous neural firing in one region will likely cause an increase in action potential propagation across even long-distance axonal connections. Thus, once potential confounding signals are dealt with (see section 2.4.8), resting state BOLD correlations likely reflect functional connectivity as utilized by task performance. Supporting this conclusion, Cordes et al. (2000) found very similar patterns across cortex when comparing fMRI functional activity (using motor, language, and visual tasks) and cortex-wide resting state correlations. BOLD correlations between well-known functionally and anatomically connected regions (e.g., LIP and FEF) have been found in anaesthetized monkeys as well (Vincent, Patel et al., 2007).

4.4 HIGH GLOBAL CONNECTIVITY FOR APFC AND ATL

> Little is known about the functions of the brain, but we can perceive that as the intellectual powers become highly developed, the various parts of the brain must be connected by very intricate channels of the freest intercommunication.
> – C. Darwin, *The Descent of Man*

The functional role of aPFC and aTL in WM integration during task set formation (see Figure 11) suggests that these regions may have high connectivity throughout the brain. Verifying this

hypothesis would help solidify the conclusions drawn from the fMRI RITL study (see section 2.0). As part of a separate project to assess global brain connectivity (GBC) across the entire brain, I developed a novel method for measuring GBC using resting state functional connectivity[5]. This new analysis provides a good opportunity to test the hypothesis that aPFC and aTL have among the highest GBC.

The GBC method provides a whole-brain map of average correlation values across all gray-matter voxels, characterizing the degree to which each voxel is functionally connected to all other gray-matter voxels (see Figure 19A). Further methodological detail is available in Section 1.01(a)(i)Appendix B. In short, we found that both aPFC and aTL were among the top 10% of all gray matter voxels in terms of global connectivity (Figure 19C). Importantly, bilateral aPFC was also in the top 5% of globally connected voxels, firmly establishing aPFC as among the most highly globally connected regions in the brain (Figure 19B).

This finding of high GBC for aPFC was not without precedent. Also supporting this conclusion, human postmortem anatomical studies have shown that aPFC has among the largest dendrites (and associated spines) of any tested brain region (Jacobs, Driscoll et al., 1997; Jacobs, Schall et al., 2001). Such highly distributed connectivity suggests aPFC may be important for integrating information from across the brain during RITL and other complex cognitive tasks (Fuster, 2004; Badre, 2008; Botvinick, 2008; Wendelken, Nakhabenko et al., 2008).

---

[5] The manuscript for this method is in preparation. An early version of the GBC method was published previously (see Cole & Schneider, 2007). My collaborator for this new version of the method, Sudhir Pathak, helped with methodological details, while Walter Schneider played an advisory role.

The finding of high GBC for aPFC and aTL strongly suggests that these regions are involved in kinds of RITL other than those tested here. For instance, children must perform RITL all the time with tasks much more simple than those used here (e.g., tying shoe laces). It could be that brain regions different than those responsive to the PRO paradigm can integrate simple task information. However, the high global connectivity of aPFC and aTL suggests that these regions are likely involved in integrating task information across many domains, simple or not. Of course, further research is necessary to fully verify this conjecture.

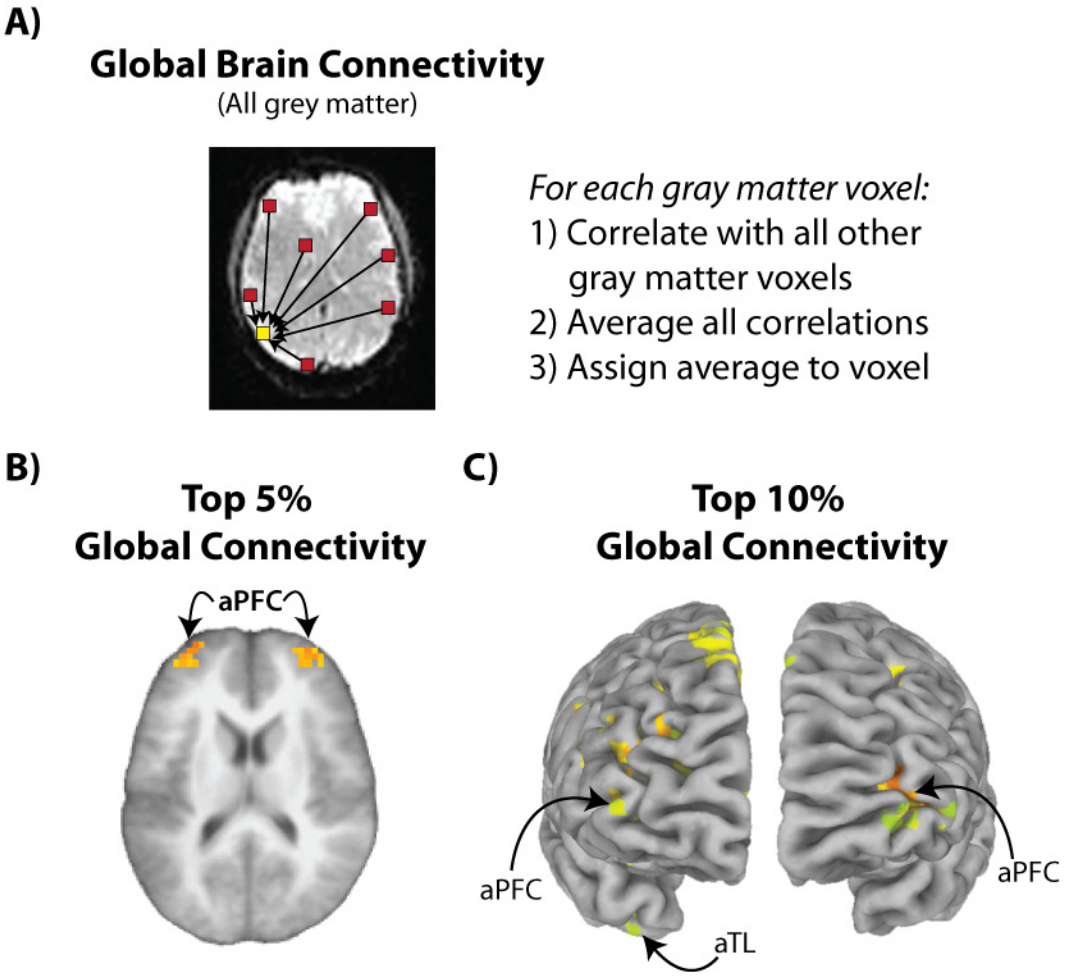

Figure 19 – aPFC and aTL Have Among the Highest Global Brain Connectivity
A novel method was developed to determine the regions with the highest global connectivity. The findings support the hypothesis that aPFC and aTL have among the highest global brain connectivity (GBC), possibly due to their role in integrating information from all over cortex.

4.5 DIFFERENCES BETWEEN MONKEYS AND HUMANS

> During hominid evolution, area 10 underwent a couple of additional changes: one involves a considerable increase in overall size, and the other involves a specific increase in connectivity, especially with other higher-order association areas. … Area 10 in the human brain appears to be specialized in size and organization, which suggests that functions associated with this part of the cortex have become particularly important during hominid evolution.
> – Semendeferi et al. (2001)

The prediction that brain regions with the largest differences between monkeys and humans would be especially involved in RITL (see section 1.2) was confirmed. Most prominently, two regions at the top of their respective hierarchies, aPFC (area 10) and aTL, were highly involved in RITL. Further, these two regions were confirmed to form a network (see Figure 11 and Figure 20A) and also to have high global connectivity (see Figure 19). Additionally, the cognitive control network (DLPFC, ACC/pSMA, IFJ, AIC, PMC, and PPC) was also involved in RITL and complex task performance (see Figure 14).

As mentioned previously, aPFC exploded in size between the common chimpanzee-human ancestor and modern humans (Figure 20). Indeed, the human aPFC is about 5X larger than bonobo chimpanzees, the species with the next-to-largest aPFC (Semendeferi et al., 2001). The present research's finding that aPFC is involved in RITL suggests that humans should be much better at RITL than bonobo chimpanzees. Indeed, the bonobo genius Kanzi was only able to get ~70% of simple instructions correct (the level of a 2½ year-old human, Savage-Rumbaugh et al., 1993), while young adult humans were able to get 92% of quite complex tasks correct on the first try (see section 2.2.1). However, future research investigating lesions of aPFC in both human and non-human primates is necessary to fully verify the functional necessity of aPFC for RITL. Other comparative studies of aPFC's function between monkeys and humans are also

necessary. Importantly, new methods have recently been developed to perform single-unit recording in this (very little explored) area in monkeys (Mitz, Tsujimoto et al., 2009).

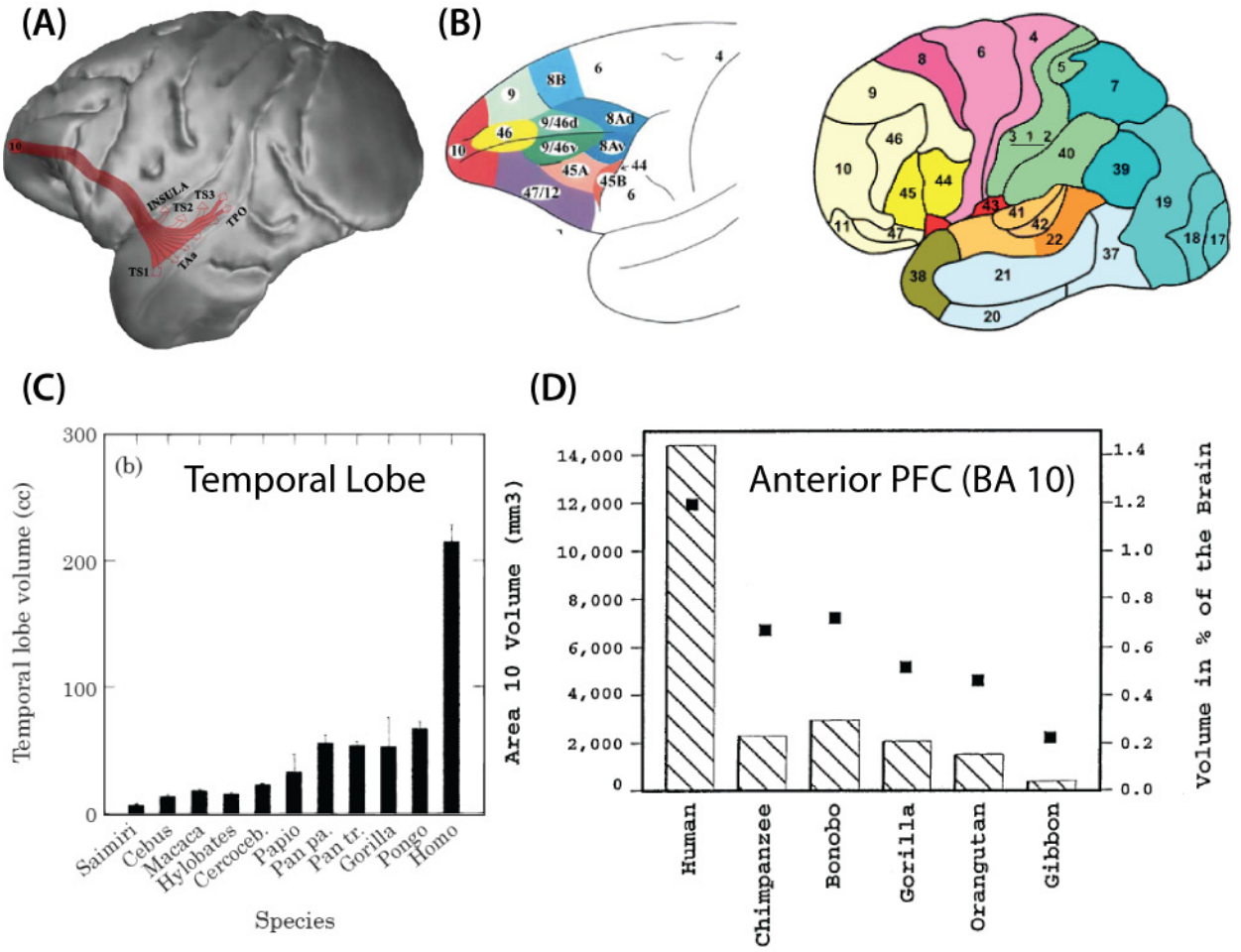

Figure 20 – Comparing Human and Monkey aPFC and aTL
A) The projection from aPFC to aTL in macaque monkeys is depicted. This anatomical projection from the frontal pole extends to the temporal pole. Adapted from (Petrides et al., 2007). **B**) The anatomical areas for macaque monkeys (left) and humans (right). These are not to scale. **C**) The volume of the temporal lobe across primate species is depicted. Adapted from (Rilling et al., 2002). **D**) The volume of aPFC across primate species is depicted. Adapted from (Semendeferi et al., 2001).

The aTL is another region verified to be involved in RITL and is also much larger in humans than monkeys (Figure 20C). Evidence suggests this region is at the apex of the ventral ('what') stream of perceptual processing (Ungerleider et al., 1994), with representations of

complex integrated semantics (Pobric et al., 2007; Taylor et al., 2009). The temporal lobes are much larger (12X) in humans than monkeys, and is larger for humans than predicted for a monkey with a human-sized brain (Rilling et al., 2002). These features suggest that humans likely have a much richer ability to represent complex integrated semantics. This is consistent with aTL's likely role in facilitating RITL in humans, a behavior in which integrating existing rule semantics into novel task sets is essential.

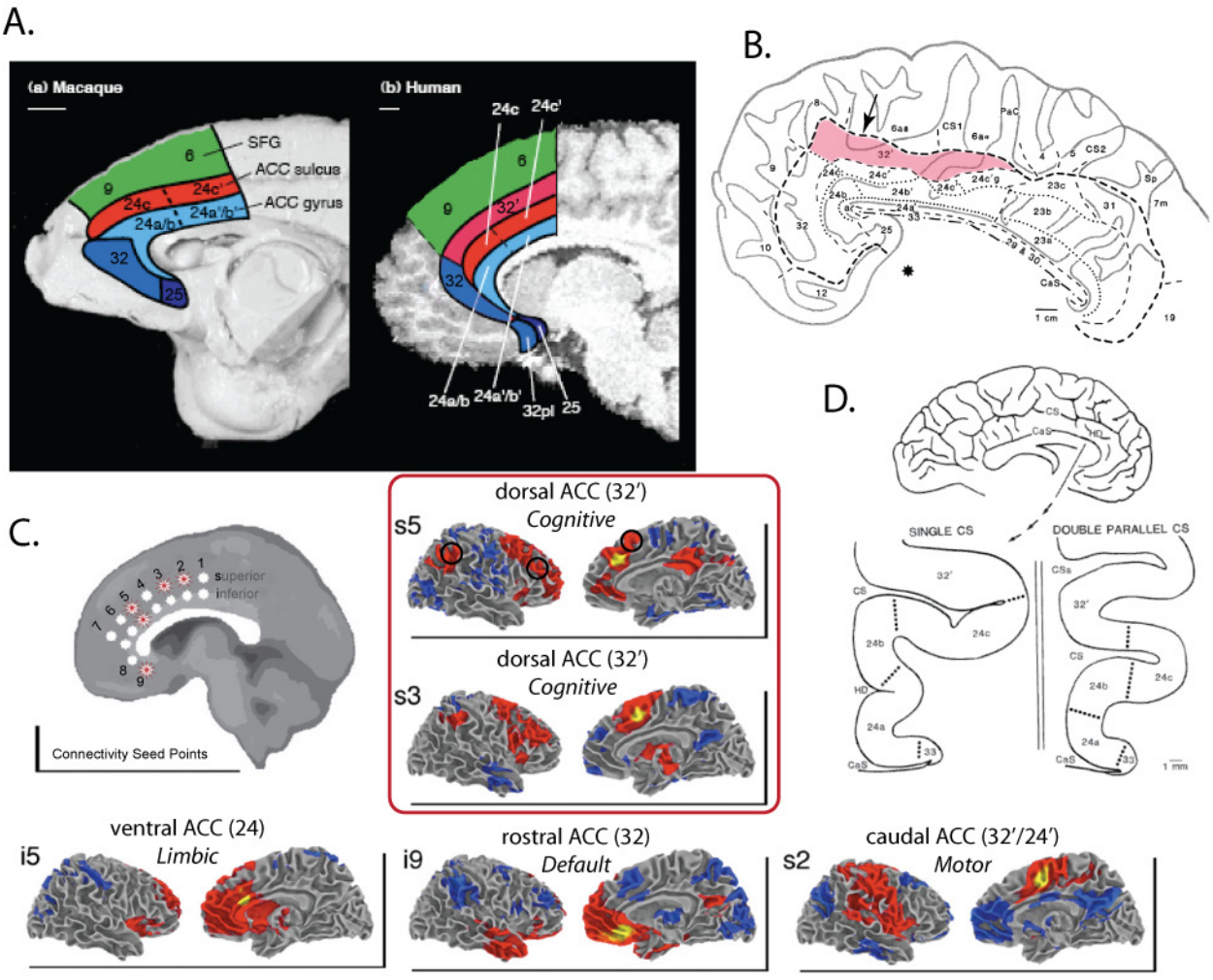

Figure 21 – Anatomy of Anterior Cingulate Cortex
A) Macaque (left) and human (right) primate brains are illustrated with corresponding anatomical areas illustrated. Importantly, a dorsal-caudal extension of area 32, area 32', is present only in the human brain. Figure adapted from (Rushworth, Walton et al., 2004). **B**) A flatmap illustration of human anterior cingulate, illustrating the location and extent of area 32'. The arrow indicates that

area 32' often extends onto the dorsal medial wall surface. Figure adapted from (Vogt, Nimchinsky et al., 1995). **C**) Resting-state functional connectivity MRI maps of seed points along human ACC. Area 32' extends from superior point 3 (s3) to s6, and shows connectivity with cognitive control regions posterior parietal cortex, DLPFC, and possibly nearby medial frontal regions. This region is separated connectively from caudal ACC in posterior area 32'/24' (which appears to be a motor area), ventral ACC in area 24 (which appears to be a limbic area), and rostral ACC in area 32 (which is part of the "default state" network; (Raichle et al., 2001)). Figure adapted from (Margulies, Kelly et al., 2007). **D**) The anatomy of ACC is extremely variable between individuals. 30-50% of humans have a double cingulate in at least one hemisphere (Paus, Tomaiuolo et al., 1996; Paus, 2001). Figure from (Vogt et al., 1995).

So far, the species differences discussed follow Darwin's claim that "there is no fundamental difference between man and the higher animals in their mental faculties" (Darwin, 1871). Darwin claimed that the inter-species differences were a matter of degree, not of kind. Recent evidence suggests, however, that a new anatomical region exists in dorsal-caudal anterior cingulate cortex (ACC), a region active during the PRO paradigm task performance (see Figure 14). This new region, the caudal extension of area 32 (Figure 21), may be important for cognitive control of non-motor semantic decision processes (van Veen and Carter, 2005). This interpretation is consistent both with the area's involvement in complex semantic judgment task performance (see Figure 14) and the lack of abstract conceptualization in non-human primates (Preuss, 2006). Further research is necessary to characterize this new anatomical region, as well as the exact functional properties it provides.

4.6 HEMISPHERIC DIFFERENCES DURING TASK PREPARATION

Classic neuropsychological research showed that the human cortical hemispheres are functionally sub-specialized (Gazzaniga, Bogen et al., 1962; Gazzaniga, 2005). Their studies conclude that the left hemisphere is more adept at language, logic, procedural coordination, and

developing causal explanations, that the right hemisphere is more adept at visuospatial processing and semantic integration, and that humans' hemispheres are more sub-specialized than monkeys' (for review see Gazzaniga, 2005). These findings may shed light on differences in activity across the hemispheric found here, as explained below.

The fMRI and MEG experiments both revealed interhemispheric differences in activity across conditions. Based on the timing of activity across conditions, it appears left aPFC and right aTL are involved in task set integration. Findings from split-brain patients indicate that the left hemisphere is involved with procedural organization, suggesting that left aPFC coordinates the meta-task of task set integration during RITL. In contrast, split-brain and hemispheric priming research indicates that the right hemisphere is specifically involved in semantic integration and comprehension (Chiarello, 1998; St George, Kutas et al., 1999), possibly even for maintaining an integrated sense of reality itself (Sacks, 1985), suggesting that right aTL integrates the task component semantics during task set integration. Together these findings suggest that left aPFC likely controls right aTL, coordinating semantic integration during task set formation. This is consistent with known anatomical projections *from* aPFC *to* aTL (see Figure 20A) However, there is also evidence for bidirectional connectivity between these regions (see Figure 17A and (Rempel-Clower and Barbas, 2000)).

Other right hemisphere activity included right DLPFC during semantic WM loading, and right aPFC during complex logic processing (see Figure 12). Both of these regions' activity patterns are consistent with an emphasis on semantic processing. Right DLPFC appears to be involved in loading and maintaining the semantics of the task instructions, while right aPFC appears to be involved in the semantics of logical processing. Further research is necessary to verify the consistency of these inter-hemispheric differences.

4.7 THE ROLE OF HIERARCHY IN NEUROCOGNITIVE PROCESSING

In cognitive neuroscience hierarchies are typically thought to provide abstraction (Figure 22A). In PFC, this abstraction is one of process both in time (e.g., WM and sequential behaviors) (Fuster, 2001; Botvinick, 2008) and in the nature of processing (e.g., maintaining vs. integrating information) (Braver et al., 2002). In the temporal lobe, this abstraction is one of representation both of objects (Ungerleider et al., 1994; Taylor et al., 2009) and general semantics (Rogalsky et al., 2008). In light of *both* aPFC and aTL being active (and functionally connected) during RITL, I would like to posit a more parsimonious view. It is possible that both these hierarchies function to reduce interference between similar representations using integrated higher-order representations. Such functionality may be necessary to allow representation of large sets of finely distinguished ideas and behaviors.

Evidence for this novel view comes from several sources, including both research on PFC and the temporal lobe. Badre & D'Esposito (2007) used different levels of processing abstraction to reveal a functional hierarchy within PFC (Figure 22B). Critically, they used *interference* at each level of the hierarchy to induce regulation by the level above it. This suggests that the role of the hierarchy in PFC is to resolve interference at the level below. The means by which this interference is resolved has yet to be completely understood, but it appears likely that integrated/abstract representations at a higher level of the hierarchy may resolve 'disputes' at lower levels of the hierarchy.

According to this hypothesis, resolving interference via abstract representations should also be fundamental to the temporal lobe hierarchy. Rogers et al. (2006) found that aTL is more active for concrete semantic categories like 'robin' than more abstract semantic categories like 'animal' for the same images of objects (Figure 22C). This finding is certainly problematic for

theories of abstraction in brain hierarchies, and yet a new hypothesis can account for it. Instead of increasing abstraction in the classic sense (i.e., increasing generality with each level; e.g., bird → animal), the temporal lobe hierarchy likely represents *increasing integration*, which can either lead to abstract concepts of relationships among words in a sentence (Rogalsky et al., 2008) or concrete multi-modal object representations that help reduce interference at a lower level of representation (Rogers et al., 2006; Taylor et al., 2009).

In the case of Rogers et al., the concrete concept 'robin' likely interferes a great deal with the concepts 'blue jay' and 'chicken', requiring a complex integration (e.g., wings + can fly + yellow beak + orange chest) to identify. In contrast, 'animal' is a relatively distinct concept that requires relatively few features to identify; in the case of an image of a robin possibly only integration of the legs and eyes. Of course, an exemplar farther from the 'animal' prototype, such as a sea sponge, would likely require integration of a more complex set of semantics to properly categorize, and is therefore more likely to involve a region at the top of the hierarchy (i.e., aTL).

The present research involved the highest levels of both the PFC and temporal lobe hierarchies (see section 2.2.5), suggesting some shared function between them. The novel hypothesis posited here suggests that this shared function may be to represent integrated semantics (both in terms of processing and representation) for the purpose of reducing interference at lower levels of the hierarchies. This suggests that RITL does not *necessarily* involve aPFC and aTL. Rather, these regions may be only necessary when there are similar interfering alternative tasks, as is the case here (each task shared either one or two features with several other tasks; see Figure 2).

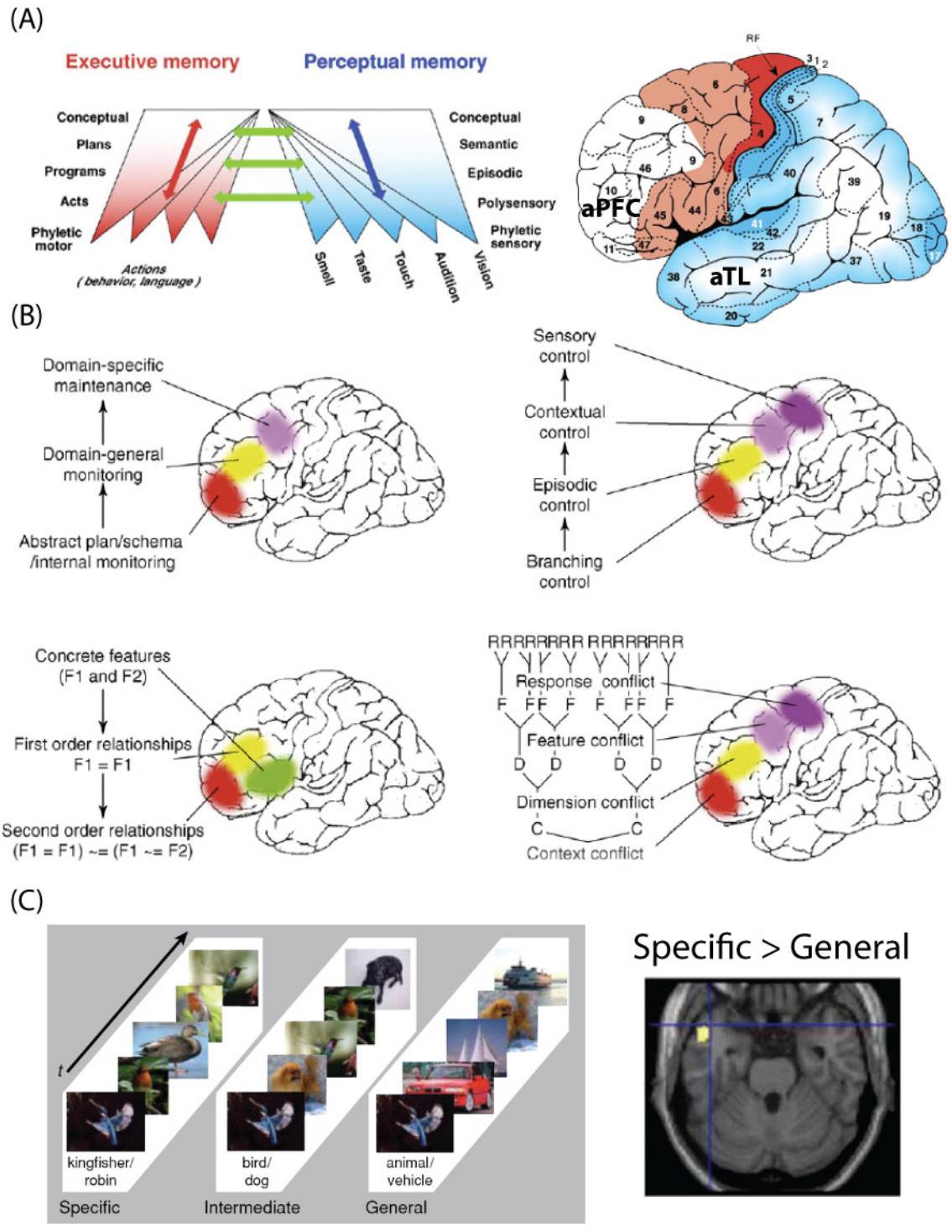

Figure 22 – Executive and perceptual hierarchies in neocortex
A) The neocortex is separated into two hierarchies, with the frontal hierarchy processing action and the posterior hierarchy processing perception. Figure adapted from (Fuster, 2006). **B**) Four theories of hierarchy in prefrontal cortex. Badre's, which is most compatible with the theory proposed here, is in the lower right corner. Figure adapted from (Badre, 2008). **C**) Three levels of semantic abstraction were tested by Rogers et al. (2006). The 'specific' level activated aTL more than the 'general' level. Figure adapted from (Rogers et al., 2006).

Further research is necessary to determine if aPFC and aTL are necessary for RITL. It seems that learning simple tasks that are clearly distinguished from other tasks is unlikely to require regions at the top of the hierarchies. However, in adulthood the need to learn such tasks is rare. Either novel tasks are complex (composed of already-learned sub-tasks), or they are simple sequences (e.g., driving directions) that are very similar to other sequences. Further, if a task is truly unique from all others then RITL may not even be possible since RITL relies on integrating (from instruction) previously learned tasks (Schneider et al., 1991). Thus, while aPFC and aTL may not be required in all cases of RITL, they are almost certainly necessary for the vast majority of task learning scenarios in adults.

4.8 GENERAL CONCLUSIONS

Two neuroimaging methods were used here in conjunction with a new cognitive paradigm to better understand RITL. The findings suggest RITL is made possible in humans via a combination of language, enhanced WM manipulation (aPFC), and enhanced integrated semantic representation (aTL). aPFC likely has a larger role in cognition than its involvement in WM manipulation, however, as aPFC has the connectivity necessary to support integration and coordination of much of the brain (Burgess, Gilbert et al., 2007) (also see Figure 19). In the context of RITL, it is likely that aPFC 'scaffolds' novel task processes until other regions can learn to perform those processes independently.

As a primary means for behavioral transmission, RITL is relevant to much of human behavior. For instance, future RITL research could help improve education (e.g., by speeding

learning by optimizing RITL), artificial intelligence (e.g., by building machines that learn like humans), or helping those with cognitive disabilities (e.g., by aiding older adults learning new technologies). There are also a number of applications within cognitive neuroscience itself, given that nearly every cognitive paradigm involves an initial practice session utilizing RITL. In order to maximally impact the world, future RITL research should recursively accelerate research by making optimal use of RITL to share findings and new methodologies with others, as I hope was accomplished here.

APPENDIX A

CORTICAL SURFACE RECONSTRUCTION AND SEGMENTATION METHODS

The following text describing the Freesurfer methods was obtained from the creators of Freesurfer (http://surfer.nmr.mgh.harvard.edu/), and is included for completeness:

Cortical reconstruction and volumetric segmentation was performed with the Freesurfer image analysis suite, which is documented and freely available for download online (http://surfer.nmr.mgh.harvard.edu/). The technical details of these procedures are described in prior publications (Dale and Sereno, 1993; Dale, Fischl et al., 1999; Fischl, Sereno et al., 1999a; Fischl et al., 1999b; Fischl and Dale, 2000; Fischl, Liu et al., 2001; Fischl, Salat et al., 2002; Fischl, Salat et al., 2004a; Fischl, van der Kouwe et al., 2004b; Segonne, Dale et al., 2004; Han, Jovicich et al., 2006; Jovicich, Czanner et al., 2006)). Briefly, this processing includes motion correction and averaging of multiple volumetric T1 weighted images (when more than one is available), removal of non-brain tissue using a hybrid watershed/surface deformation procedure (Segonne et al., 2004), automated Talairach transformation, segmentation of the subcortical white matter and deep gray matter volumetric structures (including hippocampus, amygdala, caudate, putamen, ventricles) (Fischl et al., 2002; Fischl et al., 2004a) intensity normalization (Sled, Zijdenbos et al., 1998), tessellation of the gray matter white matter boundary, automated

topology correction (Fischl et al., 2001; Segonne, Pacheco et al., 2007), and surface deformation following intensity gradients to optimally place the gray/white and gray/cerebrospinal fluid borders at the location where the greatest shift in intensity defines the transition to the other tissue class (Dale et al., 1993; Dale et al., 1999; Fischl et al., 2000). Once the cortical models are complete, a number of deformable procedures can be performed for in further data processing and analysis including surface inflation (Fischl et al., 1999a), registration to a spherical atlas which utilized individual cortical folding patterns to match cortical geometry across subjects (Fischl et al., 1999b), parcellation of the cerebral cortex into units based on gyral and sulcal structure (Fischl et al., 2004b; Desikan et al., 2006), and creation of a variety of surface based data including maps of curvature and sulcal depth. This method uses both intensity and continuity information from the entire three dimensional MR volume in segmentation and deformation procedures to produce representations of cortical thickness, calculated as the closest distance from the gray/white boundary to the gray/CSF boundary at each vertex on the tessellated surface (Fischl et al., 2000). The maps are created using spatial intensity gradients across tissue classes and are therefore not simply reliant on absolute signal intensity. The maps produced are not restricted to the voxel resolution of the original data thus are capable of detecting submillimeter differences between groups. Procedures for the measurement of cortical thickness have been validated against histological analysis (Rosas, Liu et al., 2002) and manual measurements (Kuperberg, Broome et al., 2003; Salat, Buckner et al., 2004). Freesurfer morphometric procedures have been demonstrated to show good test-retest reliability across scanner manufacturers and across field strengths (Han et al., 2006).

APPENDIX B

GLOBAL BRAIN CONNECTIVITY METHODS

Participants

We included 14 right-handed subjects (7 male, 7 female), aged 19 to 29 (mean age 22) in the study. These subjects were recruited from the University of Pittsburgh and surrounding area. Subjects were excluded if they had any medical, neurological, or psychiatric illness, any contraindications for MRI scans, or were left-handed. All subjects gave informed consent.

MRI data collection

Image acquisition was carried out on a 3T Siemens Trio MRI scanner. Thirty-eight transaxial slices were acquired every 2000 ms **(**FOV: 210 mm, TE: 30 ms, Flip angle: 90°, voxel dimensions: 3.2 X 3.2 X 3.2 mm), with a total of 300 echo-planar imaging (EPI) volumes collected per run. Three-dimensional anatomical MP-RAGE images and T2 structural in-plane images were collected for each subject. One ten-minute resting fMRI run was collected just after collecting anatomical images and before subjects performed any experimental tasks. A static white screen with a black central fixation cross was projected onto a screen visible to each subject via a mirror mounted inside the MRI scanner. Subjects were instructed to keep their eyes

open (to avoid known eyes-closed oscillations in visual cortex and to help the subjects avoid falling asleep) with instructions to stay relaxed with eyes still.

Preprocessing

Freesurfer (Desikan et al., 2006) was used with each subject's MP-RAGE to segment gray matter, white matter, and ventricle voxels. Following this step, fMRI preprocessing methods as implemented in AFNI (Cox, 1996) were used. Standard slice timing correction and motion correction procedures were used. As is standard with rs-fcMRI analyses, time series were then extracted from white matter and ventricle voxels (with some distance from the gray matter in order to ensure no overlap), and linear regression was used to remove any correlations between the gray matter voxels and these nuisance covariates (as well as the motion parameters). The global signal was not regressed out from the gray matter voxels in order to avoid artifactual anti-correlations (Murphy et al., 2008), and also because this was a signal of interest. As is typical of rs-fcMRI, temporal filtering was applied with $0.009 < f(Hz) < 0.08$ in order to remove any remaining cardiac and respiratory effects not removed by the regression. Finally, all voxels outside the gray matter were masked out. Note that, unlike most fMRI analyses, spatial smoothing was not performed during preprocessing. This was to avoid any contamination of the signals by non-gray matter sources prior to GBC analysis.

Global brain connectivity (GBC) analysis

We developed the GBC analysis method initially using MATLAB, switching to AFNI for the findings reported here. GBC analysis involves seed-based correlation of each gray matter voxel with all other gray matter voxels. These values are then averaged together and the resulting

value (i.e., that voxel's GBC) is assigned to that voxel in a new brain map. Fisher's z transformation is applied to each (Pearson's r) correlation prior to averaging, and then converted back to r-values afterward (according to standard statistical practice). This process repeats for all gray matter voxels, resulting in a single brain map reflecting GBC.

In order to perform a group analysis, each subject's GBC map was fit to a Talairach (Talairach, 1988) template (AFNI's version of 'colin27') using a 12 degrees-of-freedom affine transformation. Spatial smoothing (6mm FWHM) was then applied to the GBC map in order to help account for intersubject variability in the location and size of functional regions. Potential contamination across hemispheres due to spatial smoothing was avoided by applying smoothing to each hemisphere separately. Note that for the first statistical test described below, spatial smoothing was only applied after differencing GBC and grand mean GBC values. The grand mean GBC value for each subject was determined by averaging all the GBC values across all gray matter voxels (reflecting the average GBC across the brain).

GBC statistical test: GBC vs. zero GBC t-test

The 'highest percent' threshold was determined by raising the t-statistic (relative to '0') threshold until exactly ten percent of the total gray matter voxels remained. Nearest-neighbor clusters (among the voxels surviving the t-statistic threshold) smaller than 15 voxels were removed prior to calculating the percentage. The total number of gray matter voxels for the group (42314±1765) was determined by averaging across the total number of gray matter voxels of each subject (prior to smoothing, post-Talairaching).

Top 5% GBC amplitude map

Since the top 10% GBC t-test map reflects statistical confidence from zero GBC rather than GBC amplitude alone, we sought to find the highest globally connected regions by GBC amplitude (i.e., the highest r-values). A GBC amplitude threshold leaving the top 5% voxels was used in order to ensure the included regions were truly among the top globally connected regions. Further, in order to rule out any potential sampling artifacts or biases created by spatial smoothing, no spatial smoothing was used and any region near a sampling artifact interface was removed. Finally, since the highest GBC amplitude regions should also be consistent across subjects and analyses, we only included regions that were also present in the GBC vs. grand mean GBC map (reported elsewhere) as well as the top 10% GBC t-test map. The map resulting from this conservative approach was designed to identify the brain regions that we can confidently say, with minimal doubt regarding potential confounds, are among the highest globally connected regions in the brain.

BIBLIOGRAPHY

Akaike, H. (1976). "An information criterion (AIC)." Math Sci **14**(5).

Anderson, J. R. (1996). The Architecture of Cognition, Lawrence Erlbaum Associates.

Anderson, J. R. (2000). Cognitive psychology and its implications, WH Freeman.

Asaad, W. F., G. Rainer and E. K. Miller (2000). "Task-Specific Neural Activity in the Primate Prefrontal Cortex." Journal of Neurophysiology.

Avants, B., P. Schoenemann and J. Gee (2006). "Lagrangian frame diffeomorphic image registration: Morphometric comparison of human and chimpanzee cortex." Medical Image Analysis **10**(3): 397-412.

B. Klein, S. (2008). Learning: Principles and Applications, SAGE.

Badre, D. (2008). "Cognitive control, hierarchy, and the rostro–caudal organization of the frontal lobes." Trends in Cognitive Sciences **12**(5): 193-200.

Badre, D. and M. D'Esposito (2007). "Functional magnetic resonance imaging evidence for a hierarchical organization of the prefrontal cortex." Journal of Cognitive Neuroscience **19**(12): 2082-99.

Beggs, J. M. and D. Plenz (2003). "Neuronal avalanches in neocortical circuits." J Neurosci **23**(35): 11167-77.

Bennett, M. R. and P. Hacker (2008). "History of Cognitive Neuroscience." 312.

Biederman, I. and M. Shiffrar (1987). "Sexing day-old chicks: A case study and expert systems analysis of a difficult perceptual-learning task." Journal of Experimental Psychology: Learning, Memory, and Cognition **13**(4): 640-645.

Botvinick, M. M. (2008). "Hierarchical models of behavior and prefrontal function." Trends in Cognitive Sciences **12**(5): 201-8.

Bovair, S. and D. E. Kieras (1991). "Toward a model of acquiring procedures from text." Handbook of reading research **2**: 206–229.

Brakke, K. E. and E. S. Savage-Rumbaugh (1995). "The development of language skills in bonobo and chimpanzee—I. Comprehension." Language and Communication **15**(2): 121-148.

Brass, M. and D. Y. von Cramon (2004). "Decomposing components of task preparation with functional magnetic resonance imaging." J Cogn Neurosci **16**(4): 609-20.

Braver, T. S. and D. M. Barch (2006). "Extracting core components of cognitive control." Trends in Cognitive Sciences **10**(12): 529-32.

Braver, T. S. and S. R. Bongiolatti (2002). "The role of frontopolar cortex in subgoal processing during working memory." Neuroimage **15**(3): 523-36.

Brown, J. S. and T. S. Braver (2005). "Learned predictions of error likelihood in the anterior cingulate cortex." Science **307**(5712): 1118-1121.

Bullmore, E. and O. Sporns (2009). "Complex brain networks: graph theoretical analysis of structural and functional systems." Nat. Rev. Neurosci. **10**(3): 186-198.

Bunge, S. A., I. Kahn, J. D. Wallis, E. K. Miller and A. D. Wagner (2003). "Neural circuits subserving the retrieval and maintenance of abstract rules." J Neurophysiol **90**(5): 3419-28.

Bunge, S. A., C. Wendelken, D. Badre and A. D. Wagner (2005). "Analogical Reasoning and Prefrontal Cortex: Evidence for Separable Retrieval and Integration …." Cerebral Cortex.

Burgess, P. W. (1997). "Theory and methodology in executive function research." Methodology of frontal and executive function.

Burgess, P. W., I. Dumontheil and S. J. Gilbert (2007). "The gateway hypothesis of rostral prefrontal cortex (area 10) function." Trends in Cognitive Sciences **11**(7): 290-8.

Burgess, P. W., S. J. Gilbert and I. Dumontheil (2007). "Function and localization within rostral prefrontal cortex (area 10)." Philos Trans R Soc Lond, B, Biol Sci **362**(1481): 887-99.

Castellanos, N. and V. Makarov (2006). "Recovering EEG brain signals: Artifact suppression with wavelet enhanced independent component analysis." Journal of Neuroscience Methods **158**(2): 300-312.

Chein, J. M. and W. Schneider (2005). "Neuroimaging studies of practice-related change: fMRI and meta-analytic evidence of a domain-general control network for learning." Brain Res Cogn Brain Res **25**(3): 607-23.

Chiarello, C. (1998). "On codes of meaning and the meaning of codes: Semantic access and retrieval within and between hemispheres." Right hemisphere language comprehension: Perspectives from cognitive neuroscience: 141-160.

Cohen, L., S. Lehéricy, F. Chochon, C. Lemer, S. Rivaud and S. Dehaene (2002). "Language-specific tuning of visual cortex? Functional properties of the Visual Word Form Area." Brain **125**(Pt 5): 1054-69.

Cole, M. W. and W. Schneider (2007). "The cognitive control network: Integrated cortical regions with dissociable functions." Neuroimage **37**(1): 343-60.

Constantinidis, C. and X. J. Wang (2004). "A neural circuit basis for spatial working memory." Neuroscientist **10**(6): 553-65.

Cordes, D., V. M. Haughton, K. Arfanakis, J. D. Carew, P. A. Turski, C. H. Moritz, M. A. Quigley and M. E. Meyerand (2001). "Frequencies contributing to functional connectivity in the cerebral cortex in "resting-state" data." AJNR Am J Neuroradiol **22**(7): 1326-33.

Cordes, D., V. M. Haughton, K. Arfanakis, G. J. Wendt, P. A. Turski, C. H. Moritz, M. A. Quigley and M. E. Meyerand (2000). "Mapping functionally related regions of brain with functional connectivity MR imaging." AJNR Am J Neuroradiol **21**(9): 1636-44.

Cover, T. M. and J. A. Thomas (2006). "Elements of Information Theory." John Wiley and Sons.

Cox, R. W. (1996). "AFNI: software for analysis and visualization of functional magnetic resonance neuroimages." Comput Biomed Res **29**(3): 162-73.

Croft, R. J. and R. J. Barry (2000). "Removal of ocular artifact from the EEG: a review." Neurophysiologie clinique = Clinical neurophysiology **30**(1): 5-19.

Curtis, C. E., M. W. Cole, V. Y. Rao and M. D'Esposito (2005). "Canceling planned action: an FMRI study of countermanding saccades." Cereb Cortex **15**(9): 1281-9.

Curtis, C. E. and M. D'Esposito (2003). "Persistent activity in the prefrontal cortex during working memory." Trends Cogn Sci **7**(9): 415-423.

D'Esposito, M., B. R. Postle, D. Ballard and J. Lease (1999). "Maintenance versus manipulation of information held in working memory: an event-related fMRI study." Brain Cogn **41**(1): 66-86.

Dale, A. M., B. Fischl and M. I. Sereno (1999). "Cortical surface-based analysis. I. Segmentation and surface reconstruction." Neuroimage **9**(2): 179-94.

Dale, A. M. and E. Halgren (2001). "Spatiotemporal mapping of brain activity by integration of multiple imaging modalities." Curr Opin Neurobiol **11**(2): 202-8.

Dale, A. M., A. K. Liu, B. R. Fischl and R. L. Buckner (2000). "Dynamic Statistical Parametric Mapping Combining fMRI and MEG for High-Resolution Imaging of …." Neuron.

Dale, A. M. and M. I. Sereno (1993). "Improved localization of cortical activity by combining EEG and MEG with MRI cortical surface reconstruction: a linear approach." J Cogn Neurosci **5**: 162-176.

Darwin, C. (1871). "The Descent of man and selection in relation to sex." Published, e.g., by American Home Library in 1902: 688.

De Pisapia, N., J. Slomski and T. Braver (2007). "Functional Specializations in Lateral Prefrontal Cortex Associated with the Integration and Segregation of Information in Working Memory." Cerebral Cortex **17**(5): 993-1006.

Desikan, R. S., F. Ségonne, B. Fischl, B. T. Quinn, B. C. Dickerson, D. Blacker, R. L. Buckner, A. M. Dale, R. P. Maguire, B. T. Hyman, M. S. Albert and R. J. Killiany (2006). "An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest." Neuroimage **31**(3): 968-80.

di Pellegrino, G., L. Fadiga, L. Fogassi, V. Gallese and G. Rizzolatti (1992). "Understanding motor events: a neurophysiological study." Exp Brain Res **91**(1): 176-80.

Dosenbach, N. U., K. M. Visscher, E. D. Palmer, F. M. Miezin, K. K. Wenger, H. C. Kang, E. D. Burgund, A. L. Grimes, B. L. Schlaggar and S. E. Petersen (2006). "A core system for the implementation of task sets." Neuron **50**(5): 799-812.

Ehrlich, A. H. (2008). "The Dominant Animal: Human Evolution and the Environment." 428.

Fischl, B. and A. M. Dale (2000). "Measuring the thickness of the human cerebral cortex from magnetic resonance images." Proc Natl Acad Sci U S A **97**(20): 11050-5.

Fischl, B., A. Liu and A. M. Dale (2001). "Automated manifold surgery: constructing geometrically accurate and topologically correct models of the human cerebral cortex." IEEE Trans Med Imaging **20**(1): 70-80.

Fischl, B., D. H. Salat, E. Busa, M. Albert, M. Dieterich, C. Haselgrove, A. van der Kouwe, R. Killiany, D. Kennedy, S. Klaveness, A. Montillo, N. Makris, B. Rosen and A. M. Dale (2002). "Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain." Neuron **33**(3): 341-55.

Fischl, B., D. H. Salat, A. J. van der Kouwe, N. Makris, F. Segonne, B. T. Quinn and A. M. Dale (2004a). "Sequence-independent segmentation of magnetic resonance images." Neuroimage **23 Suppl 1**: S69-84.

Fischl, B., M. I. Sereno and A. M. Dale (1999a). "Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system." Neuroimage **9**(2): 195-207.

Fischl, B., A. van der Kouwe, C. Destrieux, E. Halgren, F. Segonne, D. H. Salat, E. Busa, L. J. Seidman, J. Goldstein, D. Kennedy, V. Caviness, N. Makris, B. Rosen and A. M. Dale (2004b). "Automatically parcellating the human cerebral cortex." Cereb Cortex **14**(1): 11-22.

Fischl, B. R., M. I. Sereno, R. B. Tootell and A. M. Dale (1999b). "High-resolution intersubject averaging and a coordinate system for the cortical surface." Hum. Brain Mapp. **8**(4): 272-84.

Forman, S. D., J. D. Cohen, M. Fitzgerald, W. F. Eddy, M. A. Mintun and D. C. Noll (1995). "Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): use of a cluster-size threshold." Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine **33**(5): 636-47.

Fuster, J. (2006). "The cognit: A network model of cortical representation." International Journal of Psychophysiology **60**(2): 125-132.

Fuster, J. M. (2001). "The prefrontal cortex--an update: time is of the essence." Neuron **30**(2): 319-33.

Fuster, J. M. (2004). "Upper processing stages of the perception-action cycle." Trends in Cognitive Sciences **8**(4): 143-5.

Fuster, J. M. and J. P. Jervey (1982). "Neuronal firing in the inferotemporal cortex of the monkey in a visual memory task." Journal of Neuroscience **2**(3): 361-375.

Gazzaniga, M. (2008). "Human: The Science Behind What Makes Us Unique." 352.

Gazzaniga, M., J. Bogen and R. Sperry (1962). "Some functional effects of sectioning the cerebral commissures in man." Proceedings of the National Academy of Sciences **48**(10): 1765-1769.

Gazzaniga, M. S. (2005). "Forty-five years of split-brain research and still going strong." Nat. Rev. Neurosci. **6**(8): 653-9.

Genovese, C. R., N. A. Lazar and T. Nichols (2002). "Thresholding of statistical maps in functional neuroimaging using the false discovery rate." Neuroimage **15**(4): 870-8.

Geweke, J. (1982). "Measurement of Linear Dependence and Feedback Between Multiple Time Series." Journal of the American Statistical Association.

Goebel, R., A. Roebroeck, D. S. Kim and E. Formisano (2003). "Investigating directed cortical interactions in time-resolved fMRI data using vector autoregressive modeling and Granger causality mapping." Magn Reson Imaging **21**(10): 1251-61.

Golanov, E. V., S. Yamamoto and D. J. Reis (1994). "Spontaneous waves of cerebral blood flow associated with a pattern of electrocortical activity." Am J Physiol **266**(1 Pt 2): R204-14.

Goldberg, R. F., C. A. Perfetti and W. Schneider (2006a). "Distinct and common cortical activations for multimodal semantic categories." Cognitive, Affective & Behavioral Neuroscience **6**(3): 214-22.

Goldberg, R. F., C. A. Perfetti and W. Schneider (2006b). "Perceptual knowledge retrieval activates sensory brain regions." J Neurosci **26**(18): 4917-21.

Goldenberg, G. and H. O. Karnath (2006). "The neural basis of imitation is body part specific." J Neurosci **26**(23): 6282-7.

Grabowski, T. J., H. Damasio, D. Tranel, L. L. Ponto, R. D. Hichwa and A. R. Damasio (2001). "A role for left temporal pole in the retrieval of words for unique entities." Hum. Brain Mapp. **13**(4): 199-212.

Granger, C. W. J. (1969). "Investigating Causal Relations by Econometric Models and Cross-spectral Methods." Econometrica **37**(3): 424-438.

Grinband, J., J. Hirsch and V. Ferrera (2006). "A Neural Representation of Categorization Uncertainty in the Human Brain." Neuron **49**(5): 757-763.

Hämäläinen, M. and R. Hari (2004). Magnetoencephalographic Characterization of Dynamic Brain Activation: Basic Principles and Methods of Data Collection and Source Analysis. Brain mapping : the methods. A. W. Toga and J. C. Mazziotta. Amsterdam, Boston: Academic Press.

Hämäläinen, M. S. and R. Ilmoniemi (1994). "Interpreting magnetic fields of the brain: minimum norm estimates." Medical and Biological Engineering and Computing **32**(1): 35-42.

Han, X., J. Jovicich, D. Salat, A. van der Kouwe, B. Quinn, S. Czanner, E. Busa, J. Pacheco, M. Albert, R. Killiany, P. Maguire, D. Rosas, N. Makris, A. Dale, B. Dickerson and B. Fischl (2006). "Reliability of MRI-derived measurements of human cerebral cortical thickness: the effects of field strength, scanner upgrade and manufacturer." Neuroimage **32**(1): 180-94.

He, P., G. Wilson and C. Russell (2004). "Removal of ocular artifacts from electro-encephalogram by adaptive filtering." Medical and Biological Engineering and Computing.

Horsey, R. (2002). "The art of chicken sexing." UCL Working Papers in Linguistics **14**.

Horwitz, B. (2003). "The elusive concept of brain connectivity." NeuroImage **19**(2): 466-470.

Hurford, J. (2001). "Language beyond our grasp: what mirror neurons can, and cannot, do for language evolution." New Scientist **169**(2275): 22.

Inuso, G., F. La Foresta, N. Mammone and F. C. Morabito (2007). "Wavelet-ICA methodology for efficient artifact removal from Electroencephalographic recordings." Neural Networks.

Jackson, P. L., A. N. Meltzoff and J. Decety (2006). "Neural circuits involved in imitation and perspective-taking." Neuroimage **31**(1): 429-39.

Jacobs, B., L. Driscoll and M. Schall (1997). "Life-span dendritic and spine changes in areas 10 and 18 of human cortex: a quantitative Golgi study." The Journal of Comparative Neurology **386**(4): 661-80.

Jacobs, B., M. Schall, M. Prather, E. Kapler, L. Driscoll, S. Baca, J. Jacobs, K. Ford, M. Wainwright and M. Treml (2001). "Regional dendritic and spine variation in human cerebral cortex: a quantitative golgi study." Cereb Cortex **11**(6): 558-71.

Jovicich, J., S. Czanner, D. Greve, E. Haley, A. van der Kouwe, R. Gollub, D. Kennedy, F. Schmitt, G. Brown, J. Macfall, B. Fischl and A. Dale (2006). "Reliability in multi-site structural MRI studies: effects of gradient non-linearity correction on phantom and human data." Neuroimage **30**(2): 436-43.

Kieras, D. E. and S. Bovair (1986). "The acquisition of procedures from text: A production-system analysis of transfer of training." Journal of Memory and Language **25**(5): 507-524.

Koechlin, E., G. Basso, P. Pietrini, S. Panzer and J. Grafman (1999). "The role of the anterior prefrontal cortex in human cognition." Nature **399**(6732): 148-51.

Koechlin, E., C. Ody and F. Kouneiher (2003). "The Architecture of Cognitive Control in the Human Prefrontal Cortex." Science.

Kohler, W. (1927). The Mentality of Apes, Routledge and Kegan Paul.

Kumar, P., R. Arumuganathan, K. Sivakumar and C. Vimal (2008). "Removal of Ocular Artifacts in the EEG through Wavelet Transform without using an EOG Reference Channel." Int. J. Open Problems Compt. Math **1**(2).

Kuperberg, G. R., M. R. Broome, P. K. McGuire, A. S. David, M. Eddy, F. Ozawa, D. Goff, W. C. West, S. C. Williams, A. J. van der Kouwe, D. H. Salat, A. M. Dale and B. Fischl (2003). "Regionally localized thinning of the cerebral cortex in schizophrenia." Arch Gen Psychiatry **60**(9): 878-88.

Lah, S., T. Lee, S. Grayson and L. Miller (2008). "Changes in retrograde memory following temporal lobectomy." Epilepsy & behavior : E&B **13**(2): 391-6.

Li, R. and J. C. Principe (2006). "Blinking artifact removal in cognitive EEG data using ICA." Conference proceedings : Annual International Conference of the IEEE Engineering in Medicine and Biology Society IEEE Engineering in Medicine and Biology Society Conference **1**: 5273-6.

Lin, F., T. Witzel, S. Ahlfors, S. Stufflebeam, J. W. Belliveau and M. Hamalainen (2006). "Assessing and improving the spatial accuracy in MEG source localization by depth-weighted minimum-norm estimates." Neuroimage **31**(1): 160-171.

Lin, F. H., J. W. Belliveau, A. M. Dale and M. S. Hämäläinen (2005). "Distributed current estimates using cortical orientation constraints." Hum. Brain Mapp. **27**(1): 1-13.

Lloyd, E. A. (2004). "Kanzi, evolution, and language." Biology and Philosophy **19**(4): 577-588.

Margulies, D. S., A. M. Kelly, L. Q. Uddin, B. B. Biswal, F. X. Castellanos and M. P. Milham (2007). "Mapping the functional connectivity of anterior cingulate cortex." Neuroimage **37**(2): 579-88.

Maunsell, J. H. and D. C. van Essen (1983). "The connections of the middle temporal visual area (MT) and their relationship to a cortical hierarchy in the macaque monkey." J Neurosci **3**(12): 2563-86.

Miller, E. K. and J. D. Cohen (2001). "An integrative theory of prefrontal cortex function." Annu Rev Neurosci **24**: 167-202.

Miller, E. K., D. J. Freedman and J. D. Wallis (2002). "The Prefrontal Cortex: Categories, Concepts and Cognition." Philosophical Transactions: Biological Sciences **357**(1424): 1123-1136.

Milner, B. (1963). "Effects of different brain lesions on card sorting." Archives of Neurology **9**: 90–100.

Mitz, A. R., S. Tsujimoto, A. J. Maclarty and S. P. Wise (2009). "A method for recording single-cell activity in the frontal-pole cortex of macaque monkeys." Journal of Neuroscience Methods **177**(1): 60-6.

Miyashita, Y., M. Kameyama, I. Hasegawa and T. Fukushima (1998). "Consolidation of visual associative long-term memory in the temporal cortex of primates." Neurobiology of Learning and Memory **70**(1-2): 197-211.

Monsell, S. (1996). Control of mental processes. Unsolved mysteries of the mind: Tutorial essays in cognition**:** 93–148.

Monsell, S. (2003). "Task switching." Trends in Cognitive Sciences.

Murphy, K., R. Birn, D. A. Handwerker, T. Jones and P. A. Bandettini (2008). "The impact of global signal regression on resting state correlations: Are anti-correlated networks introduced?" Neuroimage.

Newell, A. and H. A. Simon (1972). Human problem solving, Prentice-Hall.

Noelle, D. (1997). "A connectionist model of instructed learning." Doctoral dissertation, UC San Diego.

Noelle, D. and G. Cottrell (1996). "Modeling interference effects in instructed category learning." Proceedings of the 18th Annual Conference of the Cognitive Science Society: 475–480.

Norman, D. A. and T. Shallice (2000). "Attention to action: Willed and automatic control of behavior." Cognitive neuroscience: a reader.

O'Reilly, R. C. and Y. Munakata (2000). "Computational Explorations in Cognitive Neuroscience."

Ogawa, S., D. W. Tank, R. Menon, J. M. Ellermann, S. G. Kim, H. Merkle and K. Ugurbil (1992). "Intrinsic signal changes accompanying sensory stimulation: functional brain mapping with magnetic resonance imaging." Proc Natl Acad Sci U S A **89**(13): 5951-5.

Okada, Y., J. Jung and T. Kobayashi (2007). "An automatic identification and removal method for eye-blink artifacts in event-related magnetoencephalographic measurements." Physiol. Meas. **28**(12): 1523-1532.

Orban, G., K. Claeys, K. Nelissen, R. Smans, S. Sunaert, J. Todd, C. Wardak, J. Durand and W. Vanduffel (2006). "Mapping the parietal cortex of human and non-human primates." Neuropsychologia **44**(13): 2647-2667.

Pandya, D. and E. Yeterian (1985). "Architecture and connections of cortical association areas." Cerebral Cortex **4**: 3-61.

Passingham, R. (2009). "How good is the macaque monkey model of the human brain?" Curr Opin Neurobiol.

Paus, T. (2001). "Primate anterior cingulate cortex: where motor control, drive and cognition interface." Nat. Rev. Neurosci. **2**(6): 417-24.

Paus, T., F. Tomaiuolo, N. Otaky, D. MacDonald, M. Petrides, J. Atlas, R. Morris and A. C. Evans (1996). "Human cingulate and paracingulate sulci: pattern, variability, asymmetry, and probabilistic map." Cereb Cortex **6**(2): 207-14.

Petrides, M. and D. N. Pandya (2007). "Efferent association pathways from the rostral prefrontal cortex in the macaque monkey." J Neurosci **27**(43): 11573-86.

Pobric, G., E. Jefferies and M. A. Ralph (2007). "Anterior temporal lobes mediate semantic representation: mimicking semantic dementia by using rTMS in normal participants." Proc Natl Acad Sci USA **104**(50): 20137-41.

Preuss, T. M. (2006). "Who's afraid of Homo sapiens?" Journal of biomedical discovery and collaboration **1**: 17.

Quintana, J. and J. M. Fuster (1992). "Mnemonic and predictive functions of cortical neurons in a memory task." Neuroreport **3**(8): 721-4.

Quintana, J. and J. M. Fuster (1999). "From perception to action: temporal integrative functions of prefrontal and parietal neurons." Cereb Cortex **9**(3): 213-21.

Rabbitt, P. (1997). "Methodology of Frontal and Executive Function." books.google.com.

Raichle, M. E., A. M. MacLeod, A. Z. Snyder, W. J. Powers, D. A. Gusnard and G. L. Shulman (2001). "A default mode of brain function." Proc Natl Acad Sci USA **98**(2): 676-82.

Ramnani, N. and A. Owen (2004). "Anterior prefrontal cortex: insights into function from anatomy and neuroimaging." Nat. Rev. Neurosci. **5**(3): 184-94.

Reed, S. K. and C. A. Bolstad (1991). "Use of examples and procedures in problem solving." Journal of Experimental Psychology: Learning, Memory, and Cognition **17**(4): 753-766.

Rempel-Clower, N. L. and H. Barbas (2000). "The laminar pattern of connections between prefrontal and anterior temporal cortices in the Rhesus monkey is related to cortical structure and function." Cereb Cortex **10**(9): 851-65.

Reynolds, J. R., K. B. McDermott and T. S. Braver (2006). "A Direct Comparison of Anterior Prefrontal Cortex Involvement in Episodic Retrieval and Integration." Cerebral Cortex.

Rilling, J. K. and R. A. Seligman (2002). "A quantitative morphometric comparative analysis of the primate temporal lobe." Journal of Human Evolution **42**(5): 505-33.

Roddy, J. (1952). Harvard-trained dog. LOOK**:** 17-20.

Roebroeck, A., E. Formisano and R. Goebel (2005). "Mapping directed influence over the brain using Granger causality and fMRI." Neuroimage **25**(1): 230-42.

Rogalsky, C. and G. Hickok (2008). "Selective Attention to Semantic and Syntactic Features Modulates Sentence Processing Networks in Anterior Temporal Cortex." Cerebral Cortex **19**(4): 786-796.

Rogers, T. T., J. Hocking, U. Noppeney, A. Mechelli, M. L. Gorno-Tempini, K. Patterson and C. J. Price (2006). "Anterior temporal cortex and semantic memory: reconciling findings from neuropsychology and functional imaging." Cognitive, Affective & Behavioral Neuroscience **6**(3): 201-13.

Rosas, H. D., A. K. Liu, S. Hersch, M. Glessner, R. J. Ferrante, D. H. Salat, A. van der Kouwe, B. G. Jenkins, A. M. Dale and B. Fischl (2002). "Regional and progressive thinning of the cortical ribbon in Huntington's disease." Neurology **58**(5): 695-701.

Rougier, N. P., D. C. Noelle, T. S. Braver, J. D. Cohen and R. C. O'Reilly (2005). "Prefrontal cortex and flexible cognitive control: rules without symbols." Proc Natl Acad Sci U S A **102**(20): 7338-43.

Rushworth, M. F., M. E. Walton, S. W. Kennerley and D. M. Bannerman (2004). "Action sets and decisions in the medial frontal cortex." Trends Cogn Sci **8**(9): 410-7.

Sacks, O. (1985). "The Man Who Mistook His Wife for a Hat." Duckworth.

Salat, D. H., R. L. Buckner, A. Z. Snyder, D. N. Greve, R. S. Desikan, E. Busa, J. C. Morris, A. M. Dale and B. Fischl (2004). "Thinning of the cerebral cortex in aging." Cereb Cortex **14**(7): 721-30.

Savage-Rumbaugh, E. S., J. Murphy, R. A. Sevcik, K. E. Brakke, S. L. Williams, D. M. Rumbaugh and E. Bates (1993). "Language Comprehension in Ape and Child." Monographs of the Society for Research in Child Development **58**(3/4).

Schmolesky, M. T., Y. Wang, D. P. Hanes, K. G. Thompson, S. Leutgeb, J. D. Schall and A. G. Leventhal (1998). "Signal timing across the macaque visual system." J Neurophysiol **79**(6): 3272-8.

Schneider, W. and J. M. Chein (2003). "Controlled & automatic processing: behavior, theory, and biological mechanisms." Cognitive Science **27**(3): 525-559.

Schneider, W., A. Eschman and A. Zuccolotto (2002). "E-Prime: User's Guide.
." Psychology Software Inc.

Schneider, W. and W. Oliver (1991). "An Instructable Connectionist/Control Architecture: Using Rule-Based Instructions to Accomplish Connectionist Learning in a Human Time Scale." Architectures for intelligence: the twenty-second Carnegie Mellon Symposium on Cognition: 113.

Schneider, W. and R. Shiffrin (1977). "Controlled and automatic human information processing: I. Detection, search, and attention." Psychological Review(84): 1-66.

Seeley, W. W., V. Menon, A. F. Schatzberg, J. Keller, G. H. Glover, H. Kenna, A. L. Reiss and M. D. Greicius (2007). "Dissociable Intrinsic Connectivity Networks for Salience Processing and Executive Control." Journal of Neuroscience **27**(9): 2349.

Segonne, F., A. M. Dale, E. Busa, M. Glessner, D. Salat, H. K. Hahn and B. Fischl (2004). "A hybrid approach to the skull stripping problem in MRI." Neuroimage **22**(3): 1060-75.

Segonne, F., J. Pacheco and B. Fischl (2007). "Geometrically accurate topology-correction of cortical surfaces using nonseparating loops." IEEE Trans Med Imaging **26**(4): 518-29.

Semendeferi, K., E. Armstrong, A. Schleicher, K. Zilles and G. W. Van Hoesen (2001). "Prefrontal cortex in humans and apes: A comparative study of area 10." American Journal of Physical Anthropology **114**(3): 224-241.

Serences, J. T. (2004). "A comparison of methods for characterizing the event-related BOLD timeseries in rapid fMRI." Neuroimage **21**(4): 1690-700.

Seth, A. K. (2005). "Causal connectivity of evolved neural networks during behavior." Network **16**(1): 35-54.

Shannon, C. E. (1948). "A mathematical theory of communication." Bell System Tech. J. **27**: 379-423, 623-656.

Shehzad, Z., A. Kelly, P. Reiss, D. Gee, K. Gotimer, L. Uddin, S. Lee, D. Margulies, A. Roy, B. Biswal, E. Petkova, F. Castellanos and M. Milham (2009). "The Resting Brain: Unconstrained yet Reliable." Cerebral Cortex: 21.

Shiffrin, R. and W. Schneider (1977). "Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory." Psychological Review **84**(2): 127-190.

Skinner, B. F. (1951). "How to teach animals." Scientific American(185): 26-29.

Sled, J. G., A. P. Zijdenbos and A. C. Evans (1998). "A nonparametric method for automatic correction of intensity nonuniformity in MRI data." IEEE Trans Med Imaging **17**(1): 87-97.

St George, M., M. Kutas, A. Martinez and M. I. Sereno (1999). "Semantic integration in reading: engagement of the right hemisphere during discourse processing." Brain **122 (Pt 7)**: 1317-25.

Stephan, K. M. (2004). "On the role of general system theory for functional neuroimaging." J Anat **205**(6): 443-70.

Sticht, T. G. (1977). Comprehending reading at work. Cognitive Processes in Comprehension. M. A. J. P. A. Carpenter. Hillsdale, NJ, Erlbaum**:** 221-246.

Talairach, J. (1988). Co-planar stereotaxic atlas of the human brain.

Taulu, S. and R. Hari (2008). "Removal of magnetoencephalographic artifacts with temporal signal-space separation: Demonstration with single-trial auditory-evoked responses." Human brain mapping.

Taulu, S., J. Simola and M. Kajola (2004). "MEG recordings of DC fields using the signal space separation method (SSS)." Neurology & clinical neurophysiology : NCN **2004**: 35.

Taulu, S., J. Simola and M. Kajola (2005). "Applications of the Signal Space Separation Method." Signal Processing.

Taylor, K., E. Stamatakis and L. Tyler (2009). "Crossmodal integration of object features: Voxel-based correlations in brain-damaged patients." Brain **132**(3): 671-683.

Team, R. D. C. (2009). R: A Language and Environment for Statistical Computing. Vienna, Austria, R Foundation for Statistical Computing.

Toro, R., P. Fox and T. Paus (2008). "Functional Coactivation Map of the Human Brain." Cerebral Cortex **18**(11): 2553-2559.

Turing, A. M. (1950). "Computing machinery and intelligence." Mind.

Uddin, L., A. Clare Kelly, B. Biswal, F. Xavier Castellanos and M. Milham (2009). "Functional connectivity of default mode network components: Correlation, anticorrelation, and causality." Hum. Brain Mapp. **30**(2): 625-637.

Ungerleider, L. G. and J. V. Haxby (1994). "'What' and 'where' in the human brain." Curr Opin Neurobiol **4**(2): 157-65.

van Veen, V. and C. S. Carter (2005). "Separating semantic conflict and response conflict in the Stroop task: A functional MRI study." Neuroimage **27**: 497–504.

Vendrell, P., C. Junque, J. Pujol, M. A. Jurado, J. Molet and J. Grafman (1995). "The role of prefrontal regions in the Stroop task." Neuropsychologia **33**(3): 341-52.

Vincent, J. L., G. H. Patel, M. D. Fox, A. Snyder, J. T. Baker, D. C. Van Essen, J. M. Zempel, L. H. Snyder, M. Corbetta and M. E. Raichle (2007). "Intrinsic functional architecture in the anaesthetized monkey brain." Nature **447**(7140): 83-6.

Vogt, B. A., E. A. Nimchinsky, L. J. Vogt and P. R. Hof (1995). "Human cingulate cortex: surface features, flat maps, and cytoarchitecture." The Journal of Comparative Neurology **359**(3): 490-506.

Wager, T. D., J. Jonides and S. Reading (2004). "Neuroimaging studies of shifting attention: a meta-analysis." Neuroimage **22**(4): 1679-93.

Wallis, J. D. and E. K. Miller (2003). "From rule to response: neuronal processes in the premotor and prefrontal cortex." J Neurophysiol **90**(3): 1790-806.

Wallstrom, G. L., R. E. Kass, A. Miller, J. F. Cohn and N. A. Fox (2004). "Automatic correction of ocular artifacts in the EEG: a comparison of regression-based and component-based methods." International journal of psychophysiology : official journal of the International Organization of Psychophysiology **53**(2): 105-19.

Wendelken, C., D. Nakhabenko, S. E. Donohue, C. S. Carter and S. A. Bunge (2008). ""Brain is to thought as stomach is to ??": investigating the role of rostrolateral prefrontal cortex in relational reasoning." Journal of Cognitive Neuroscience **20**(4): 682-93.

Wheeler, M. E., S. E. Petersen and R. L. Buckner (2000). "Memory's echo: vivid remembering reactivates sensory-specific cortex." Proc Natl Acad Sci U S A **97**(20): 11125-9.

Williams, S. L., K. E. Brakke and E. S. Savage-Rumbaugh (1997). "Comprehension skills of language-competent and nonlanguage-competent apes." Language and Communication **17**(4): 301-317.

Wise, S. P., D. Boussaoud, P. B. Johnson and R. Caminiti (1997). "Premotor and parietal cortex: corticocortical connectivity and combinatorial computations." Annu Rev Neurosci **20**: 25-42.

Wood, J. and J. Grafman (2003). "Human prefrontal cortex: processing and representational perspectives." Nat. Rev. Neurosci. **4**(2): 139-147.

Xiong, J., L. M. Parsons, J. H. Gao and P. T. Fox (1999). "Interregional connectivity to primary motor cortex revealed using MRI resting state images." Hum Brain Mapp **8**(2-3): 151-6.

Yarkoni, T., D. M. Barch, J. Gray, T. Conturo, T. S. Braver and B. Baune (2009). "BOLD Correlates of Trial-by-Trial Reaction Time Variability in Gray and White Matter: A Multi-Study fMRI Analysis." PLoS ONE **4**(1): e4257.