

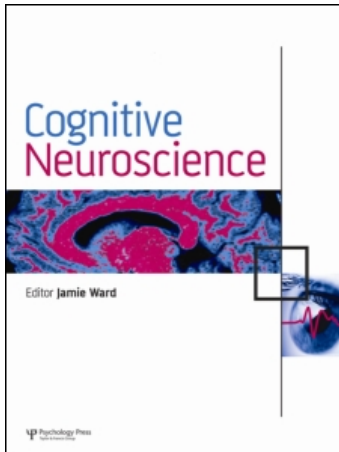
This article was downloaded by: [Brown University]

On: 15 July 2010

Access details: Access Details: [subscription number 907439020]

Publisher Psychology Press

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



## Cognitive Neuroscience

Publication details, including instructions for authors and subscription information:

<http://www.informaworld.com/smpp/title~content=t909559412>

### Is prefrontal cortex necessary for the storage and acquisition of relational concepts?

David Badre<sup>a</sup>

<sup>a</sup> Box 1978, Brown University, Providence, RI, USA

Online publication date: 07 June 2010

**To cite this Article** Badre, David(2010) 'Is prefrontal cortex necessary for the storage and acquisition of relational concepts?', Cognitive Neuroscience, 1: 2, 140 – 141

**To link to this Article:** DOI: 10.1080/17588921003782589

**URL:** <http://dx.doi.org/10.1080/17588921003782589>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.informaworld.com/terms-and-conditions-of-access.pdf>

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

## Discussion Paper

# Abstract relational categories, graded persistence, and prefrontal cortical representation

Ann Speed

*Sandia National Laboratories, Albuquerque, NM, USA*

Human behavior is characterized by creativity, flexibility, and adaptability. Psychologists have argued that this is a result of analogical reasoning processes. Neuroimaging studies point to PFC as a critical component of a larger network; however, it remains unclear how the brain accomplishes analogical reasoning. This paper presents a theory of prefrontal cortical function that attempts to explain the neural mechanisms of analogical processing in the context of the broader theoretical and empirical work on PFC. Specifically, the current paper proposes that neurons in PFC are particularly sensitive to relationship information, and that they develop response preferences for relationship information that increases in abstractness and complexity along the posterior–anterior axis. Further, this theory posits that representation formation in PFC is driven by frontostriatal circuits and that the persistence of these representations is determined by environmental consequences such that the longer the representation predicts reward or punishment, the longer the representation lasts. Finally, it is suggested that because analogy has been proposed as a core cognitive process, underlying many other interesting cognitive phenomena such as learning, creativity, and decision-making, it could serve as a useful tool for studying PFC function in general.

**Keywords:** Analogy; Prefrontal cortex.

## INTRODUCTION

Human behavior is characterized by creativity, flexibility, and adaptability. We rarely if ever experience precisely the same situation twice, and often find ourselves in new environments. Yet most of the time our behavior is sufficient, and we are able to learn about, make decisions about, solve problems in, and adapt to

new environments quickly. Psychologists argue that analogical reasoning is the cognitive process that enables this adaptability (Gick & Holyoak, 1980; Hofstadter, 2001; Holyoak & Thagard, 1997), yet the specific neural substrate for analogical reasoning remains unclear. Certainly a number of neuroimaging studies exist and all point to various areas in prefrontal cortex (PFC) as key components of a larger network

---

Correspondence should be addressed to: Ann Speed, Sandia National Laboratories, PO Box 5800 MS1011, Albuquerque, NM 87185-1011, USA. E-mail: aespeed@sandia.gov

Thanks very much to Jim Kroger, Keith Holyoak, Christy Warrender, and an anonymous reviewer for comments on an earlier draft of this paper, and to Jim for many, many helpful conversations.

Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy's National Nuclear Security Administration under Contract DE-AC04-94AL85000. This work was funded wholly by Sandia National Laboratories' Lab Directed Research and Development program, an internal funding source.

---

This manuscript has been authored by Sandia Corporation under Contract number DE-AC04-94AL85000 with the U.S. Department of Energy. The United States Government retains (and the publisher by accepting the article for publication acknowledges that the United States Government retains) a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of the manuscript, or allow others to do so, for United States Government purposes.  
www.psypress.com/cognitiveneuroscience DOI: 10.1080/17588921003660728

for accomplishing analogy (Bunge, Wendelken, Badre, & Wagner, 2005; Christoff et al., 2001b; Geake & Hansen, 2005; Green, Fugelsang, Kraemer, Shamosh, & Dunbar, 2006; Kroger et al., 2002; Luo et al., 2003; Mikkelsen, et al., 2010; Prabhakaran, Smith, Desond, Glover, & Gabrieli, 1997; Qiu, Li, Chen, & Zhang, 2008; Wharton et al., 2000). However, exactly how this network achieves analogical reasoning is still in question.

Since similar areas in PFC are implicated in other cognitive functions besides analogy including working memory, cognitive control, decision-making, and problem-solving (Badre & D'Esposito, 2007; Cabeza & Nyberg, 2000; Christoff & Gabrieli, 2000; Duncan & Owen, 2000; Koechlin, Ody, & Kouneiher, 2003; Miller & Cohen, 2001; O'Reilly & Frank, 2006; Rougier, Noelle, Braver, Cohen, & O'Reilly, 2006), it is necessary to provide an explanation of the physiological mechanisms underlying analogy in a way that is not limited to analogical reasoning. Therefore, this paper proposes a theory of PFC that provides an explanation of the nature and development of response preferences in PFC that not only attempts to explain how analogical reasoning might be accomplished, but also provides explanation for other empirical observations of PFC.

To this end, the discussion starts with an overview of analogical processing, including questions that must be addressed in order to understand how neural machinery might accomplish analogy-making. Then the proposed theory and empirical support for that theory are presented. The paper closes with a brief discussion on the possible utility of analogical reasoning as a tool for studying general PFC function.

## ANALOGY

Generally, an analogy is a mapping that is made between objects and relationships in a well-understood situation or problem retrieved from long-term memory (the *source*) and a situation or problem that is less well understood (the *target*). By mapping objects to objects and, more importantly, relationships to relationships, inferences can be made about the target based on what is known of the source. If a successful mapping is made (i.e., if the inferences drawn about the target are correct), then a general *schema*, or abstract representation of the problem class, is learned (Gentner, 1989; Gick & Holyoak, 1983).

Importantly, such a mapping is considered an analogy only if the source and target are not identical; however, they must share a common relational structure. That is, the *relationships between* objects, rather than the objects themselves, must match in order for

the analogy to be valid, and there is a preference for more cohesive sets of relationships over a set of disconnected ones (i.e., Gentner's *systematicity* principle; Gentner, 1983). Consequently, a distinction is made between "surface" and "structural" features of the source and target. Surface features are those attributes of problem exemplars that do not bear on the solution to the problem itself. Structural features are those relationships among elements of a situation that, when similar for two situations, are the relationships on which appropriate analogies are made (Gentner, 1983; Gentner & Markman, 1997). One important finding in the analogy literature is that domain experts pay attention preferentially to structural features whereas domain novices preferentially attend to surface features (Novick, 1988).

A standard example in the literature is an analogy between the solar system and Rutherford's model of the atom. In this example, the fact that the sun is a star is irrelevant to using the solar system to aid in understanding the relationship between electrons and the nucleus of an atom. The characteristic "star" would be a surface feature in this particular analogy. Structural features of this problem are the relationship between smaller objects and the larger object they orbit along with the "invisible" force that causes this interaction. In this case, the sun's gravity maps onto the electromagnetic forces present in the atomic nucleus. Importantly, for any given possible source problem, what constitutes a surface vs. a structural feature changes with the context in which the problem is being considered (Spellman & Holyoak, 1996).

One issue in understanding the neurophysiological mechanisms of analogy is: How is the brain able to identify the structural vs. the surface features of a given pair of problems in order to create an appropriate analogical mapping between them? Furthermore, how are situations we have encountered in the past represented in the brain in a way that enables flexible definition of surface vs. structural features? Since PFC is reliably recruited for analogy-making, as demonstrated by the numerous mentioned functional magnetic resonance imaging (fMRI), positron emission tomography, and electroencephalography (EEG) studies, and by at least one direct stimulation study (Borojerdi et al., 2001), clearly understanding how a PFC network might accomplish this is critical to answering these questions.

## PFC

Much work has been done on PFC from a variety of perspectives, and a number of theories have been posited on

its function and organization (e.g., Badre & D'Esposito, 2007; Barbas, 2000; Christoff, Ream, Geddes, & Gabrieli, 2003; Duncan, 2005; Fuster, 1997; Goldman-Rakic, 1987; Huey, Krueger, & Grafman, 2006; Koechlin et al., 2003; Miller & Cohen, 2001; Miller, Nieder, Freedman, & Wallis, 2003; Ramnani & Owen, 2004; Rougier et al., 2006). Many of these theories focus on PFC involvement in "executive" or cognitive control processes, and some hypothesize that different subareas execute different executive functions (e.g., Badre, Hoffman, Cooney, & D'Esposito, 2009; Bunge et al., 2005; Christoff & Gabrieli, 2000; Fuster, 2006; Koechlin et al., 2003; Miller & Cohen, 2001; Ramnani & Owen, 2004). Others propose that prefrontal neurons represent information (e.g., Badre & D'Esposito, 2007; Constantinidis, Franowicz, & Goldman-Rakic, 2001; Duncan, 2005; Freedman, Riesenhuber, Poggio, & Miller, 2002; Fuster, 2006; Krueger, Barbey, & Grafman, 2009; O'Reilly, Noelle, Braver, & Cohen, 2002; Wood & Grafman, 2003), and several of these theories focus on the relative importance of PFC for working memory function (e.g., O'Reilly & Frank, 2006).

However, none of the above theories offer enough information to answer the questions posed above regarding the physiological mechanisms enabling analogy. Are different areas of PFC executing the different stages of analogy-making that psychologists have proposed,<sup>1</sup> or do local networks in PFC selectively respond to different information and result in these mappings when activated (cf. Leech, Mareschal, & Cooper, 2008)? In either case, how do these response preferences or algorithms come to exist, and how are they implemented in the neural machinery? And more broadly, how are these response preferences or algorithms involved in the other cognitive processes for which PFC is selectively recruited?

## The relational category theory of PFC function

In an attempt to address these questions, the proposed theory offers thoughts on the nature and organization of response preferences in PFC along with thoughts on how

those response preferences are tuned. Following explication of the current theory is a comparison between what is being proposed and other theoretical perspectives.

### *Representation in PFC*

The current proposal is that PFC is selectively sensitive to relationship information: specifically relationships between environmental stimuli, concepts, characteristics, or actions. Building on the notion that cells in PFC represent information and that these representations are organized according to level of abstraction/complexity along the posterior–anterior areas axis (cf. Badre & D'Esposito, 2007; Christoff & Keramatian, 2007), the current construct is that in the most posterior areas of PFC (i.e., posterior areas of BA 9, BA44), cells preferentially respond to abstract, multimodal, yet simple "semantic" categories that are somewhat more abstract than those represented in anterior areas of unimodal sensory cortex (e.g., anterior inferotemporal cortex; Sigala, 2004; Sigala & Logothetis, 2002; Tamura & Tanaka, 2001). An example is the category *seed* where this category might represent any kind of plant seed, as well as analogous concepts such as *seed of an idea* or *seedling project*. These categories are defined by diagnostic characteristics: stimuli from the environment or less complex representations that have tended to be associated with that category.

In more anterior PFC areas (e.g., middle BA 9, anterior BA44, BA45, BA46), the categories represented begin to take on relationship information: such as *grows into* as in "seed *grows into* adult" or "seedling project *grows into* program." The diagnostic features to which local networks in this portion of PFC preferentially respond describe simple relationships between the abstracted categories represented in more posterior areas of PFC and/or areas of anterior temporal cortex (e.g., Badre & D'Esposito, 2007; Green et al., 2006) and thus tend to be more abstract.

In the most anterior areas of PFC such as frontal pole, local networks develop preferential responses to highly complex "relations of relations": The diagnostic features of these categories are the most abstract and complex. Examples include relationships between multiple objects that require simultaneous consideration of multiple independently varying relationships as in three-relation Raven's Progressive Matrices or systems of objects such as the solar system (Badre & D'Esposito, 2007; Christoff et al., 2001b; Geake & Hansen, 2005; Gentner, 1983; Green, et al., 2006; Kroger, et al., 2002, 2007; Prabhakaran et al., 1997; Wallis, Anderson, & Miller, 2001). To further extend the seed example, a response preference in this part of

<sup>1</sup> Specifically, cognitive psychologists have proposed that analogy consists of *retrieval* of a source analog from memory, *mapping* the source onto the current problem or situation that is poorly understood (the target), *adapting* the source solution to the details of the target, drawing *inferences* about the target based on what is known of the source, and finally *learning* an abstract representation about the class of problems exemplified by the source and the target (Gentner, 1989; Novick & Holyoak, 1991).

PFC might be for the concept of *lifecycle*, which by necessity requires understanding concepts of lesser abstraction such as *seed*, *grows into*, *death*, *maturity*, and the relationships between these concepts and their simpler yet systematic relationships with one another.

This framework is intentionally relatively abstract. It is the case that a number of studies have found somewhat different areas of PFC to be involved in what seems to be a differential (possibly hierarchical) organization, including areas of premotor cortex (Badre, 2008; Badre & D'Esposito, 2007; Badre et al., 2009), as well as dorsolateral, ventrolateral, and rostralateral PFC. The current argument is that these findings are not at odds with the current conception, but simply more detailed. It is expected that just as different areas of posterior cortex are specialized for different sensory modalities, and even different subsets of unimodal sensory processing (e.g., the fusiform face area), various areas of PFC should show analogous specialization. As such, individual relationships likely have different primary components, whereby those that involve significant motor components (e.g., remembering mappings between a simple stimulus and a correct response in the *response* condition of Badre & D'Esposito, 2007) would be closer to the premotor cortex while those with a strong emotional component might occur in more medial, rather than lateral, areas of PFC (e.g., Bechara & Damasio, 2005; Krueger et al., 2009). The current proposal simply focuses on the characteristics that determine organization of response preferences along the posterior–anterior gradient and how those representations might support analogy in addition to other typical prefrontal functions.

One empirical finding influential in the development of the current perspective is an fMRI study of analogy published by Green and colleagues (2006). This paper contrasted a typical **analogy** condition (two pairs of words with analogically similar relationships, e.g., *planet* is to *sun* as *electron* is to *nucleus*) with a **category** condition (two pairs of words with relationships between and within the two pairs, but without the analogical similarity; e.g., *cow* is to *milk*; *duck* is to *water* where *cow* and *duck* are both *animals*, *milk* and *water* are both *liquids* and therefore share semantic relationships across the pairs, but *cow* is not to *milk* as *duck* is to *water*), and a **semantic** condition (two pairs of words with within-pair relationships, but no between-pair relationships; e.g., *seed* is to *plant*; *car* is to *road*). The authors found a great deal of cortical activity that was shared by the analogy and category tasks, including several areas in posterior PFC (BAs 6 and 44), leading them to argue that categorization is a critical enabler of analogy-making. The analogy condition selectively recruited areas in

posterior PFC over the category condition (also in BAs 6 and 44), but also selectively recruited an area in left frontopolar cortex (BA9). Similar patterns have been reported by Bunge and her colleagues using different experimental manipulations (Bunge et al., 2005; Wendelken, Nakhabenko, Donohue, Carter, & Bunge, 2007), and recent EEG evidence indicates that very anterior areas of PFC are recruited for difficult (multirelational) vs. easy (single relation) analogies (Mikkelsen et al., 2010). Unfortunately in the Green et al. study, the results of contrasting the semantic condition to the analogy or category conditions were not reported. However, the current theory would predict that such contrasts would reveal recruitment selective to more posterior areas of PFC relative to the areas recruited for the category and analogy conditions.

Similarly, Geake and Hanson (2005) used letter string analogies from Hofstadter and Mitchell's Copycat computational model (Hofstadter, 1995; Mitchell, 1993), ranking the stimuli according to analogical depth (defined as the number of simultaneously manipulated relationships). While the authors did not explicitly ask where in cortex analogies of different difficulty seemed to be processed, they did find a linear increase in BOLD response as a function of analogical depth in two predetermined cortical areas of interest: BA9 and BA45/46 (Geake & Hansen, 2005).

If it is true that cells in PFC primarily become tuned to relational categories of increasing abstractness, then PFC accomplishes analogy via activation of those tuned categories upon presentation of exemplars of these categories. In this way, individual situations or problems can be freely used as source or target analogs based on a variety of interpretations, because the representation of individual situations is independent of the abstracted schemas that are represented in PFC. Therefore, one of the more difficult findings from the analogy literature—that the goals of the thinker can actually determine the analogical mapping that is made between two sets of stimuli—can be accounted for.

Calling this phenomenon the *pragmatic constraint* on analogy, Spellman and Holyoak (1996) demonstrated that when one relationship was emphasized above another in an ambiguous mapping situation, that emphasis was sufficient to cause subjects to make the suggested mapping significantly more often than the alternative, equally plausible mapping. Essentially, this work can be understood in terms of the top-down influence (i.e., control) that activation of a local network representing one set of relationships over the other has on the subsequent interpretation of the target analog (cf. Miller & Cohen, 2001). This perspective also provides a neural mechanism for the *systematicity* principle: the preference for complex

relations-of-relations instead of groups of unrelated relationships (Gentner, 1983; what *related* means in this context will be addressed in the next section).

This perspective also makes specific predictions about expert/novice differences. Basically, top-down controlled mapping can only happen if the “top” exists: If the subject has not acquired the schema of interest, recognizing new examples of the relationships involved and making appropriate mappings will be effortful at best, and a failure at worst. Indeed, many of the behavioral studies indicate that analogy is a difficult phenomenon to elicit in the laboratory, possibly because the domains used are novel to participants (Catrambone & Holyoak, 1989; Gick & Holyoak, 1983).

#### *Representation formation and graded representational persistence*

In order to explain how PFC accomplishes analogy-making and other cognitive functions, it is necessary to explain not only the kinds of categories/concepts to which PFC selectively responds, but also how those response preferences come to exist. The current proposal builds on ideas about the neural mechanisms of working memory and category learning in PFC (e.g., Ashby & Ennis, 2006; Ashby, Ennis, & Spiering, 2007; Frank, 2005; Miller & Cohen, 2001; O’Reilly & Frank, 2006) to propose that relational response preferences in PFC actually display a *graded* persistence that is determined by environmentally mediated basal ganglia function. That is, local circuits, and individual neurons, continue to preferentially respond to a particular set of relationships for as long as those relationships are behaviorally relevant<sup>2</sup> (e.g., it is useful to maintain knowledge of the set of characteristics that define horses and dogs, or those that define the concept of *orbits*). And, the tendency of these networks to respond to relevant categories becomes stronger over time (cf. Rainer & Miller, 2000). Those networks responding to categories that don’t continue to predict reinforcement or punishment consequences (e.g., stimuli typically used in working memory experiments such as the 1-AX, 2-BY task) maintain the short representational persistence characteristic of working memory phenomena, and are therefore easily

<sup>2</sup> It is important to note that use of terms such as “behavioral relevance,” “feedback” or “reinforcement and punishment” does not exclusively mean feedback from a teacher or some other observer, although that kind of feedback can be useful. In interacting with the environment, we often receive feedback directly as a result of performing an action incorrectly. We also receive subtle feedback signals as we interact with others in social situations: non-verbal behavior provides an enormous amount of information on other people’s responses to our actions and statements.

retuned (Freedman et al., 2002; Hazy, Frank, & O’Reilly, 2006). Relevance in this case is determined by the environmental contingencies that appear to be mediated by basal ganglia function (Chow & Cummings, 2007; Frank, 2005; O’Reilly & Frank, 2006). Thus, PFC develops relational response preferences over time through learning and experience. Some of these persist for very long periods of time; others do not. In this way, expertise in a particular domain can be developed, maintained, and modified; and locally relevant rule sets, such as those found in working memory tasks, can be learned quickly and forgotten just as quickly.

Relating back to one of the key questions from the analogy literature: How is the brain able to identify the structural vs. the surface features of a given pair of problems in order to create an appropriate analogical mapping between them? Structural features from this perspective are relationships or systems of relationships (i.e., systematicity; Gentner, 1983) that predict reinforcement or punishment (i.e., goal relevance; Spellman & Holyoak, 1996). Thus, only after an individual has had the opportunity to correlate a system of features with environmental contingencies can they then relate that system of relationships to other problems (i.e., create analogies).

A number of lines of evidence support this perspective. First is evidence from electrophysiological recordings that individual neurons in PFC can rapidly change the stimulus to which they respond. Specifically, in a number of studies, Earl Miller and his colleagues have demonstrated that individual neurons in PFC can come to respond selectively to a particular abstract rule or category (i.e., the same neurons will respond to different exemplars of “cat” vs. “dog”). Interestingly, however, when the rule for classifying the cat and dog stimuli changes, essentially reclassifying the same series of exemplars into new categories, the same neurons come to respond to these new categories (e.g., Freedman, Riesenhuber, Poggio, & Miller, 2001; Freedman, et al., 2002). The question remains: What drives this plasticity?

Building on evidence that basal ganglia is involved in the selection of appropriate motor responses, O’Reilly, Frank, and colleagues have proposed that similar dopaminergic striatal circuits actually provide the mechanism for this plasticity (Frank, 2005; O’Reilly & Frank, 2006). Specifically, when an organism receives an unexpected reward, there is a spike in dopamine release in the striatum which strengthens the association between the representation in working memory, the subsequent action, and the resulting reward. When a reward is expected and not received, the tonic dopamine levels drop, releasing

the clamp on prefrontal neurons, thereby enabling them to develop response preferences to stimuli other than the one that is no longer behaviorally relevant (e.g., Frank, 2005; O'Reilly & Frank, 2006). Rewards that are expected do not result in a change in tonic dopamine levels, and therefore do not result in a change in neuronal responses. Others have also provided empirical and theoretical support for this proposal (Ashby & Ennis, 2006). In their review, Ashby and Ennis (2006) argue that dopaminergic signals seem to act as a "teaching signal" for the brain, and provide evidence that damage to these circuits hinders category learning. Specifically, they cite the fact that Parkinson's and Huntington's patients show deficits on a variety of tasks (e.g., the weather prediction task, the Wisconsin Card Sorting task) that require development of categories/rule sets and switching between different categories (i.e., rule sets) for correct responding.

To provide a concrete illustration of the development of category representations in PFC, consider a young child learning the names of different animals. Presented with a husky, the child comes to associate the word "dog" with the appearance of the husky: its coloring, size, long fur, and multicolored eyes. If the husky is the only example of a dog the child encounters, he may initially have a difficult time rectifying the appearance of a pug with the word "dog" because of the surface differences between the two breeds: size, color, sound of their respective barks, even their behavior. However, in interacting with the pug and the husky, he comes to recognize similarities between them that make them both dogs: barking, personality characteristics, possession of four legs, chewing on objects, etc. He may also receive verbal feedback: "Yes, that's a dog too!" However, given only two exemplars, he has not fully learned the core set of features that are uniquely diagnostic to the category "dog" as opposed to other four-legged animals. Therefore, when confronted with a horse, he may initially behave as though it were a very large dog and offer it a bone. Feedback from the environment (i.e., the horse not responding to the bone the way the husky did) or from others (i.e., "That's not a dog: That's a horse!") enables the child to continue to distinguish the features of an animal that indicate it is a dog as opposed to a horse. These core features that are present across a wide number of exemplars are what analogy researchers would call structural features.

In terms of the importance of graded persistence, if this child lived in an environment where it was not useful to know the difference between a horse and a dog, the prediction is that this difference, and the separate multimodal representation of "horse" and "dog" in PFC would not survive: that they would devolve into a more

general representation of "animals with four legs" and the local networks that had taken on temporary representation of "horse" and "dog" would either participate in this more general category or would tend to respond selectively to some other category that was behaviorally relevant in this child's strange world. Likewise, if this child lived in a situation where the distinction between husky and pug was behaviorally relevant, local circuits of neurons would take on the task of selectively responding to these different breeds.

By way of contrast, as an example of a set of relationships that does not and should not attain longer-term persistence, consider the stimuli in a typical working memory task such as the 1-AX, 2-BY task. Because of the striatally-driven plasticity available in PFC, subjects are able to rapidly learn the relationships necessary to provide correct responses to the task. Local circuits of neurons in PFC take on those relational representations because of the innately reinforcing aspects of correctly performing the task, and subjects quickly master the necessary rules. Subjects are able to maintain these rules in working memory (i.e., cells in PFC continue to selectively represent these rules) for as long as is behaviorally necessary: in most cases for the duration of the experiment or block of trials. When the rules change, subjects can rapidly learn the new rule, and the responding neurons rapidly take on these new, behaviorally relevant response preferences (cf. Freedman, et al., 2001, 2002). If the 1-AX, 2BY rules proved to be behaviorally relevant after the end of the experiment and on into the subjects' lives, the current proposal states that the circuits that became tuned to these rules during the experiment would continue to respond to situations requiring application of these same rules: That is, these relational representations would become more persistent in time, and would remain so for as long as they predicted punishment or reinforcement.

Granted, these examples are much more concrete than those described earlier in the paper regarding the analogy between the solar system and Rutherford's model of the atom. However, the current proposal is that the mechanisms by which these more concrete features are learned and display persistence are fundamentally the same as those dictating learning more complex relationships and relations-of-relations. Both are striatally driven, both display a graded persistence depending on ongoing behavioral relevance, and both are multimodal representations occurring in PFC. They differ only with regard to their relative level of abstraction.

From an evolutionary perspective, this relevance-dependent plasticity is important because there is no way to determine *a priori* whether a particular set of relationships is important to remember or not. Therefore, the

ability for PFC to rapidly learn a set of relationships enables us to quickly bring our behavior in line with the operational contingencies present in the immediate environment. It is equally beneficial for us to be able to alter behavior that is no longer predictive of reward or punishment. However, there is a benefit to maintaining a particular relational categorical representation for a period of time after that relevance seems to have changed, as it may be that that category is only predictive of reward a certain percentage of the time (i.e., as in interval or ratio schedules of reinforcement). Therefore, this account predicts there should be some lag between the retuning of the response preferences of networks in PFC and the cessation of reward, and that lag should be correlated with the amount of time the relational category was predictive. Such a lag provides a mechanism for behavioral stability over changing or imperfectly correlated contingencies. Of course, this general process may be moderated for particularly strong consequences (e.g., extreme fear or pain) or in the event that the consequence-mediating circuits' functions are compromised (e.g., Frank, 2005). But the core point is that behavioral relevance determines response preference persistence in PFC.

### Relationship to other work

These ideas are not entirely novel. Many existing theories discuss hierarchical organization in PFC and cortex in general (e.g., Badre & D'Esposito, 2007; Christoff & Gabrieli, 2000; Fuster, 2006; Koechlin et al., 2003; Ramnani & Owen, 2004) and several others discuss tuning mechanisms for cells in PFC (Ashby & Ennis, 2006; Frank, 2005; Freedman, et al., 2001; Fuster, 2006; O'Reilly & Frank, 2006; Rainer & Miller, 2000). Others discuss the role of basal ganglia in development of transient representations or activity in PFC as a theory of observed plasticity in PFC-based working memory (O'Reilly & Frank, 2006; Rougier et al., 2006). Fuster, as well as Ashby and colleagues, has explored the general role of the limbic system and basal ganglia in the development of large networks of representations and categories (Ashby & Ennis, 2006; Fuster, 2006). And many theorists propose that PFC is fundamentally representational in nature (Ashby et al., 2007; Badre & D'Esposito, 2007; Constantinidis et al., 2001; Duncan, 2005; Freedman et al., 2002; Fuster, 2006; Goldman-Rakic, 1987; Krueger et al., 2009; Miller et al., 2003; O'Reilly et al., 2002; Rougier et al., 2006; Wood & Grafman, 2003). Finally, there is evidence that the basal ganglia are important in both Pavlovian conditioning and cognitive reinforcement learning and

that both error-related feedback and the integrity of the basal ganglia are essential for category learning (Ashby & Ennis, 2006). However, none of the above-cited work has explained the type of response preferences, the method of response tuning, and the varying length of time those preferences are maintained in a way that explains the role that PFC plays in the fundamental cognitive process of analogy.

The remainder of this section will compare the current proposal to several of these theories and is organized according to several key differences, namely, assumptions about explicit vs. implicit processing, the need to account for both long-term and short-term representations, and an alternative explanation for apparent cognitive control.

#### *Explicit vs. implicit processing in PFC*

As mentioned above, the proposal that PFC is organized according to abstraction of representation or process along the posterior–anterior axis is certainly not unique (Badre & D'Esposito, 2007; Christoff & Gabrieli, 2000; Christoff & Keramatian, 2007; Fuster, 2006; Koechlin et al., 2003; O'Reilly et al., 2002; Ramnani & Owen, 2004). One account that is highly similar to the current one in terms of the types of stimuli to which PFC selectively responds, is the notion presented by Christoff et al. (2001b) that rostralateral PFC is involved in relational integration (combining two or more relationships in order to correctly answer the problem at hand), and other theoretical work by Christoff and colleagues (Christoff & Gabrieli, 2000; Christoff, Geddes, & Gabrieli, 2001a; Christoff & Keramatian, 2007; Christoff et al., 2003) which specifically posits that anterior regions of PFC (i.e., the intersection of the middle frontal gyrus and BA 10) are involved in processing information that is not directly tied to stimulus features. They posit that anterior PFC is involved in explicit consideration of internally generated information such as relationships, plans, goals, or information retrieved from long-term memory: That is, information that has been abstracted away from specific stimulus properties. They also state that the relative abstractness of this internally generated information will determine what portions of PFC are selectively recruited when such information is being considered.

While this theory can explain performance on tasks such as Raven's Progressive Matrices (i.e., Christoff et al., 2001b), one issue is that it specifically states that rostralateral PFC (RLPFC) is recruited only for tasks in which explicit attention is being paid to the internally generated information (Christoff et al., 2003). The strong version of this statement seems to exclude the possibility of differential involvement of RLPFC when a task has become fairly automatic. Interestingly,

a recent review of the practice effects literature indicated that practice with several higher-order cognitive tasks such as the Tower of London tends to display reduced prefrontal recruitment with increased practice (Kelly & Garavan, 2005), which seems to support Christoff's prediction. However, there is evidence from direct electrophysiological measurement that experience results in a smaller number of recruited PFC cells with narrower tuning curves (Rainer & Miller, 2000). Thus, it could be that automatic or implicit processing, which comes about through experience, recruits PFC networks as much as do explicit, effortful processes, but this activity is less detectable using fMRI because of the smaller cortical extent of these familiar representations. One way to test this might be to examine relative strength and extent of PFC recruitment in experts vs. novices.

#### *Long-term vs. short-term representations*

Another important difference between the current theory and those of several other researchers is the distinction between shorter-term and longer-term representations. Specifically, a core idea in the analogy literature is that a mapping is made between a current situation and something experienced at some indeterminate time in the past – often requiring retrieval of that information from a long-term store. However, the vastly greater part of the theoretical work on PFC is based on tasks that require only the formation of relatively short-term representations (Christoff & Keramian, 2007; Hazy et al., 2006; Reynolds & O'Reilly, 2009; Rougier et al., 2006; Wallis, Anderson, & Miller, 2001). This is true for any number of tasks used to study PFC including the 1-AX, 2-BY task employed by O'Reilly and colleagues (O'Reilly & Frank, 2006; Reynolds & O'Reilly, 2009), hierarchical tasks employed by Koechlin and colleagues (Koechlin, Corrado, Pietrini, & Grafman, 2000) and by Badre and colleagues (Badre & D'Esposito, 2007), and even many analogy tasks such as the letter string analogies employed by Geake and Hansen (2005) and arguably problems based on Raven's Progressive Matrices (e.g., Christoff et al., 2003; Kroger et al., 2002).

However, there exists neuroimaging evidence of analogy that directly taps into longer-term memory (e.g., Bunge et al., 2005; Green et al., 2006; Green, Kraemer, Fugelsang, Gray, & Dunbar, 2010). Recall that the stimuli used by Green et al. (2006) leveraged semantic relationships in four-term (A:B::C:D) word problems. These researchers found that as the nature of the relationship between and among these pairs of words increased in abstraction, recruitment of PFC moved to more anterior areas.

Bunge et al. (2005) reported similar effects using the same types of stimuli but different experimental manipulations. Specifically, subjects saw A:B::C:D word problems and were asked if the words were semantically or analogically related. In each condition, some were examples of analogy and some were examples of semantic relatedness (e.g., *bouquet:flower::chain:link* or *note:scale::rain:drought*). The researchers also manipulated the semantic strength between words on the semantic trials. They found that on analogy task trials frontopolar cortex (FPC, BA10) was selectively recruited, whereas semantic task trials recruited anterior left inferior PFC (aLIPC, BA 45, 47). They also found that semantic relatedness only modulated recruitment in aLIPC such that aLIPC was more heavily recruited for low-related than for high-related pairs. However, relatedness did not cause differential activity in FPC (although the trends they report indicate that FPC was recruited more heavily for low semantic relatedness than for high semantic relatedness).

Green et al. (2010) provide additional evidence for the notion that neurons in the frontal pole develop response preferences for relatively abstract, relational information and that these preferences are for information that has attained long-term persistence. Specifically, using four-term word analogies (A:B::C:D), they parametrically manipulated semantic distance between the first pair of words and the second pair of words. They also manipulated whether the two halves of the analogy were within-domain (i.e., both having to do with sensory stimuli as in *nose:scent::tongue:taste*) or occurred across domains (e.g., *nose:scent::antenna:signal*). They found simple effects for both semantic distance and domain such that left frontopolar cortex was increasingly active as a function of farther semantic distance and as a function of cross-domain vs. same-domain analogies. They also specifically tested whether this recruitment was due to the difficulty of the analogies (as measured by response time, correctness, and ratings of difficulty) and found no relationship. While these researchers interpret this finding to mean that the frontal pole is selectively recruited to perform analogical mapping (as opposed to other stages in the analogical process, see note 1), the fact that difficulty was not related to recruitment of frontal pole, but that abstractness of the relationship being considered in the analogy was, seems to indicate that frontal pole activity is specific to activation of abstract relationships, some of which may be long-term, persistent relationships. If frontal pole is selectively recruited for the mapping phase of analogical reasoning, it would seem that its level of recruitment would be directly related to difficulty.

The bottom line from this handful of studies is that they use stimuli that require subjects to rely on general knowledge stored in long-term memory in order to respond to the tasks, and these long-term memories involve, in some conditions, relatively abstract relationships. Interestingly, their processing recruits the same areas of PFC as are recruited for tasks that require shorter-term representations that are equally abstract (ventrolateral and dorsolateral for less abstract, frontal pole for more abstract). One could argue that this evidence points directly to the role that PFC neurons play in manipulating or controlling cognitive processes, as opposed to the nature of response preferences in different areas of PFC. However, this does not explain where these abstract, long-term representations are stored or how those longer-term representations come to be stored in a part of cortex other than the one in which shorter-term representations come to exist for the purposes of control (e.g., Koechlin & Hayfil, 2007; Reynolds & O'Reilly, 2009; Wallis et al., 2001). In other words, if representations of abstracted rules or relationships are created in PFC as tasks demand, only to be replaced by other relationships demanded by newer tasks (Badre & D'Esposito, 2007; Christoff & Keramatian, 2007; Koechlin et al., 2000; Miller, Erickson, & Desimone, 1996; Reynolds & O'Reilly, 2009; Wallis et al., 2001), where are these longer-term relationships such as “*revolves around*” or “*grows into*” that are called upon for semantic analogy problems such as those in Bunge et al. (2005) and Green et al. (2006, 2010), stored, how are they recruited during processing, and why are the same areas of PFC engaged when both short-term and long-term analogical relationships are being considered?

A much more parsimonious explanation for these results is that presented in the current theory: that individual neurons/local circuits in PFC develop graded response preferences for relationships that have demonstrated behavioral relevance (i.e., predict punishment or reinforcement) and that the length of time and number of situations across which those relationships hold their predictive power determines how persistent (i.e., long-term) those representations become. Thus, when a subject is learning a new relationship, whether that relationship is an inner-loop/outer-loop relationship, as in the 1AX – 2BY task, or is a fundamental concept in physics does not matter. What serves to distinguish between short-term and long-term is exactly that: how short-term or how long-term the relationship retains behavioral relevance. Thus, one would expect the same areas of PFC to be recruited for “short-term” relational tasks as well as for relational tasks that require tapping into long-term memory.

### *Cognitive control*

One set of hypotheses about PFC is that it engages in inhibition of inappropriate or competing responses (e.g., Badre et al., 2009; Koechlin & Hayfil, 2007; Koechlin et al., 2003; Miller & Cohen, 2001). Indeed, Holyoak and colleagues posit that this sort of inhibition is a critical aspect of analogy-making (Krawczyk et al., 2008; Richland, Morrison, & Holyoak, 2006; Viskontas, Morrison, Holyoak, Hummel, & Knowlton, 2004). A question that arises, however, is how this selective inhibition comes about. Much of the theorizing about PFC and cognitive control seems to attend mostly to the top-down biasing function of PFC. The current theory posits that this apparent inhibition is actually partially a result of the reinforcement-driven modulation by the striatum on prefrontal response tuning. Specifically the response preferences of a neuron or local circuit in PFC are tuned according to the diagnostic set of stimuli that predict reinforcement for a given set of stimuli or system of relationships in a given context. Over time, that diagnostic set is winnowed down to a minimum necessary set given the current environmental contingencies. When the category of interest is very new to the subject, he may respond in kind to many potential exemplars (those that share “surface” features), only a few of which provide the desired outcome. As he learns the critical cues (the “structural” features), those cues that define other situations to which he used to respond are now no longer linked to the forming PFC response preferences, thus his behavior may look like it is due to inhibition when it is really reinforcement-driven discrimination. When the predictive power of the relationships the subject has learned changes, there is some difficulty with regard to unclamping prefrontal neurons/circuits, allowing those circuits to be retuned to reflect the new environment. Thus, the subject sometimes reverts to previously learned behavior, only to re-experience the altered behavioral contingencies, thereby reactivating the release of that circuit to retuning. As the subject continues to encounter the new environmental contingencies, he comes to discover some subtle distinction between the prior associations and the current one (maybe some small detail previously ignored or possibly some temporal variable), and further refines his discrimination between the two contexts.

An empirical example of behavior that has been attributed to inhibition comes from a study of children’s analogical abilities as a function of age. Richland et al. (2006) had children of different ages respond to visual scene analogies (e.g., a picture of a woman chasing a cat chasing a rat as the source and a

picture of a man chasing a boy chasing a girl as the target). They found that when they included a featural distractor in the target problem (in this case a cat), younger children were less able to make the appropriate mapping in both one-relation and two-relation problems, even when the relation of interest was verbalized by the experimenter and children independently demonstrated that they understood, and could identify simple examples of that relation. The authors interpret this as an example of an inability to inhibit the choice of the distractor item. However, the fact that the distractor item was always similar to the item to be mapped (in this case, the experimenter would ask what item in the target problem was “like” the cat in the source problem) potentially calls into question whether the results are due solely to inhibition problems. There is significant evidence that people have a tendency to first appeal to surface features when they are unsure of the predictive structural features (Novick, 1988; Ross, 1987). Having had less experience with the world overall, it could be that the younger children had not fully identified the key features of the particular relationships used in the study and thus appealed to surface features when many items were present in the problem: They had not learned the diagnostic (structural) features of the relations in question. If this is the case, practice with multiple examples of the same relation, including feedback, should improve their performance in the face of cross-mapped information. If, on the other hand, inhibition is the primary culprit, such practice should have little to no effect on performance when such distractors are present.

Additional evidence that is difficult to account for in the inhibition interpretation of children’s lack of analogical sophistication (e.g., Bulloch & Opfer, 2009; Crone et al., 2009; Gentner & Toupin, 1986; Richland et al., 2006) is the different patterns of cortical activity seen between children and adults. There is evidence that this is the case for stimuli such as Raven’s Progressive Matrices and for visual stimuli that require a background level of verbal semantic knowledge (Crone et al., 2009; Wright, Matlen, Baym, Ferrer, & Bunge, 2008). Specifically, using drawings of common objects presented in an A:B::C:D format, Wright et al. found that aPFC is recruited later in analogical problem-solving for children than for adults, that frontal pole recruitment in both the analogy > fixation and semantic > fixation contrasts was positively correlated with children’s age, and that frontal pole recruitment was positively correlated with accuracy in adults. They also found that the specific cortical regions recruited in children are a subset of those recruited in adults, rather than finding a separate network for the two groups, and the analogy > semantic contrast for children reveals very

little activity unique to analogy in children: none of it in anterior areas of PFC.

Using stimuli derived from Raven’s Progressive Matrices, Crone et al. (2009) found that aPFC recruitment for the most difficult problems was greater, and sustained over a longer period of time for adults than for children. Specifically, the authors presented matrix problems with zero, one, and two relations to both adults (ages 18–25) and children (ages 8–12). In aPFC, they found that while both adults and children displayed different blood-oxygen-level-dependent (BOLD) responses to problems with one relation vs. two relations in early processing (4 to 8 s post-stimulus), adults demonstrated a stronger BOLD response for two-relation problems over one-relation problems for a much longer time (10 to 18 s post-stimulus) whereas children demonstrated increased processing for one-relation problems during this later time period. In dorsolateral PFC, children displayed a smaller BOLD response to one-relation problems than did adults early in processing; however, they sustained this response longer than did adults. Taking the behavioral data into account, the authors suggest that children initially attend to multiple relationships in the two-relation problems, but they don’t utilize both relations to solve these more difficult problems. Rather, it appears that children then revert to utilizing one of the two relations to provide an answer, a speculation borne out by the high error rates children displayed for two-relation problems.

To summarize, while the current proposal shares similarities with a number of other proposals, it is distinct in several critical ways. First, it can account directly for the role that PFC plays in analogy-making. Second, it accounts for the role of PFC in working memory as well as in tasks that require the use of longer-term general knowledge. Third, it explains specifically how these prefrontal response preferences come to exist and how they are modified.

## SPECIFIC EMPIRICAL PREDICTIONS

A number of specific predictions follow from the proposed theory. Discussed below are relational priming phenomena, and some predictions regarding lesions and disease processes in PFC.

### Relational priming effects

One of the physiological problems that must be solved in the analogy literature is: Exactly how are analogical mappings made between one situation and

another? In novices, this process seems to be supported by direct comparison of object information and is usually based on similarity of surface features (Novick, 1988). For example, if given two math problems about computers and one about cars, novices tend to map the two computer problems onto one another even if the proper match is actually between a computer problem and a car problem. In experts, analogical mapping is supported by direct comparison of relational information between objects that are usually not identical (or even similar).

If, in fact, these relationships are learned and represented in PFC, as is proposed, then we should be able to prime those relationships and enhance mapping rates relative to non-primed relations in a manner similar to semantic priming effects. Research has provided some evidence for this (Leech et al., 2008; Spellman, Holyoak, & Morrison, 2001). However, this effect is reportedly difficult to achieve (James K. Kroger, personal communication, 2009; Spellman et al., 2001). Additional evidence on differences between domain experts, novices, and “journeymen” is needed. Specifically, the current proposal predicts that relational priming would be most effective for domain experts, that journeyman-level subjects would need additional support—maybe in the form of specific examples (see, e.g., Catrambone & Holyoak, 1989; Ross, 1987, 1989)—and that relational priming would be least helpful for domain novices because novices have not acquired a representation of that set of relationships that can be primed (cf. Ross, 1987; Spellman et al., 2001).

## Lesion and disease studies

If the current proposal is correct, it implies a balance that is struck between stability of representations and flexibility of representations: Damage to the circuits that control this balance (i.e., cortical–striatal circuits; Chow & Cummings, 2007) should lead to predictable behavioral and cognitive syndromes, depending on the type of damage.

Frontal patients do exhibit a wide variety of deficits from different types of memory difficulties, especially working memory; difficulties with planning and goal attainment, difficulties with task switching and changing behavior in a given task when the behavioral contingencies change, problems with apathy and inappropriate affect, and integrating information across time (Cummings & Miller, 2007; Fuster, 1997). Imbalances toward too much flexibility (which might manifest as various types of behavioral variability) might include attention problems (e.g., as in attention-deficit disorder), mood

instability, and stimulus-dependent behaviors such as utilization and imitation behaviors (Chow & Cummings, 2007). Conversely, disorders of too much stability (i.e., perseveration) might include an inability to switch tasks when reinforcement contingencies change (Bechara & Damasio, 2005), or apathy or amotivational disorders (i.e., stability in a state of doing nothing).

Specific predictions that follow from the current proposal with regard to prefrontal deficits include the following.

Patients should demonstrate difficulty in learning new classes of events that are defined by a specific set of relationships (e.g., a relational schema).

Patients should display differential impairment on analogy-making tasks as a function of analogical difficulty (measured by number of relationships modified simultaneously) and the site of the lesion such that the more anterior the lesion, the more specific is the difficulty with highly abstract or complex categories/analogy problems. Posterior prefrontal lesions may also present difficulty with more abstract categories because of the damage to areas of posterior PFC that selectively respond to components of those more abstract categories (Badre et al., 2009), but such damage should not interfere with the ability to learn new abstract categories. There is some evidence for this (e.g., Badre et al., 2009; Krawczyk et al., 2008; Waltz et al., 1999). Waltz and colleagues demonstrated that, relative to fronto-temporal patients with temporal degradation and relative to controls, fronto-temporal dementia patients with frontal degradation display markedly worse performance on Raven’s Progressive Matrices and this deficit is bigger for more complex problems. Badre and colleagues provided similar evidence from patients with focal damage to the inferior frontal sulcus and pre-dorsal premotor cortex (Badre et al., 2009).

Patients should demonstrate impairment in their ability to construct new analogical mappings in domains in which they have expertise because damage to prefrontal cortical areas should impair the strength of top-down influences on interpretation of novel incoming stimuli.

In addition, because of their specific reliance on category representations of different levels of abstraction and complexity, analogy tasks such as Raven’s Progressive Matrices or the Copycat letter string analogies (Hofstadter, 1995), once appropriately normed, might be useful for partially isolating the location of a lesion: at least along the posterior–anterior axis of lateral PFC. There is some evidence for the utility of analogy problems in diagnosis, in that both metaphor and proverb comprehension is sometimes used for assessment of PFC disorders (Fuster, 1997; Kramer &

Quitania, 2007). Metaphor is considered to be functionally similar to analogy by a number of researchers (Gentner, Bowdle, Wolff, & Boronat, 2001). This approach might be particularly useful, as location-sensitive, normed tests for prefrontal damage are still lacking (Stuss, 2007) and as these kinds of analogy problems can be made to be arbitrarily simple or complex. Also, there are experimental paradigms, such as that employed by Green and colleagues (2006) or Badre and D'Esposito (2007) that provide a series of control conditions that would allow for assessment of PFC damage or impairment of performance due to damage in more posterior areas of cortex (cf. Badre et al., 2009) once appropriately normed.

### CONCLUSION: BEHAVIORAL ADVANTAGE OF RELATIONAL CATEGORY REPRESENTATION AND ANALOGY-MAKING AS A METHOD TO STUDY PFC

As many psychologists have argued about the relative adaptive advantage of a reconstructive vs. a reproductive memory, the ability to develop knowledge about functionally similar situations without requiring the organism to remember all of the details of those situations clearly has a similar advantage. If our brains functioned in a reproductive manner, recording every detail of every experience, we would be virtually paralyzed when faced with apparently new situations that we

could easily deal with using analogical processes. Thus, because we never experience precisely the same situation twice, the ability to generalize from one situation to another situation (i.e., one that is functionally identical but on the surface dissimilar) is critical to our ability to function in everyday environments. So is our ability to acquire relational categories based on these mappings and the behavioral outcomes of these mappings.<sup>3</sup> In this way, an argument can be made that studying the general function of PFC from the perspective of analogy provides an additional framework for unifying the wide variety of cognitive behaviors for which PFC is selectively recruited. Analogy may not simply be an isolated task in which PFC is involved. Rather, as cognitive psychologists have argued, analogy is a core component of cognition, underlying many other cognitive phenomena for which PFC appears responsible including decision-making, problem-solving, and reasoning, to name a few (Hofstadter, 2001; Holyoak & Thagard, 1997). Thus, analogy paradigms and theory may be useful as a tool for unifying the currently complex literature on prefrontal function and organization.

Manuscript received 27 August 2009

Manuscript accepted 25 January 2010

First published online 18 March 2010

<sup>3</sup> Recall that that analogy-making entails, among other things, a *learning* phase in which a generalized representation, or *schema* of the source and target analogs, is formed (Gentner, 1989; Holyoak & Thagard, 1997; Keane, Ledgeway, & Duff, 1994).

## Commentaries

### Analogical reasoning: An incremental or insightful process? What cognitive and cortical evidence suggests

Alessandro Antonietti and Michela Balconi

Department of Psychology, Catholic University of the Sacred Heart, Largo Gemelli 1, 20123 Milano, Italy

E-mail: [michela.balconi@unicatt.it](mailto:michela.balconi@unicatt.it)

DOI: 10.1080/17588921003786606

**Abstract:** The step-by-step, incremental nature of analogical reasoning can be questioned, since analogy making appears to be an insight-like process. This alternative view of analogical thinking can be integrated in Speed's model, even though the alleged role played by dopaminergic subcortical circuits needs further supporting evidence.

In Speed's paper analogical reasoning is conceived, according to the prevalent view, as an incremental, step-by-step process that induces people, through the progressive discovery of new source-target similarities, to apply the set of relations embedded in the source to the target. However, the description of how individuals actually find the solution of a problem by

analogy suggests a different picture (Antonietti, 2001). In diary studies about the occurrence of analogical reasoning in everyday life, interviewees often reported that analogies did not come in mind through an incremental process. Furthermore, introspective reports showed that sometimes source–target correspondences are suddenly realized. These data are consistent with what was found in experiments aimed at testing how participants perceive the source–target correspondences across the task. Initially a source showing superficial similarities, but not a source showing structural similarities, was rated as being close to the target. However, when the first element of the target problem having an analogical correspondence with the source was provided, a sudden, significant increase of the similarity scores of the structural source was recorded. The awareness of the existence of only one relation between the source and the target, as proved by another series of experiments (Anolli, Antonietti, Cantoia, & Crisafulli, 2001), is enough to induce people to apply the source to solve the target. Finally, the investigation of individual differences highlighted that people possessing creative and insight skills, as well as intuitive–holistic cognitive styles, perform better in analogical problem-solving tasks than persons with analytical thinking skills and styles (Antonietti & Gioletta, 1995).

The overall picture that emerges from the above-mentioned studies suggests that analogical problem solving is an insight-like process that allows one to perceive, at a given moment, the whole source–target correspondence. This picture is partially consistent with the model proposed by Speed. In fact, recent evidence supports the notion that prefrontal cortex (PFC)—the core neurobiological mechanism involved, according to Speed, in analogy making—plays a role in the restructuring or insight phase of problem solving, namely, when individuals overcome the mental impasse that prevents them from finding the relevant answer and a new perspective emerges. Aziz-Zadeh, Kaplan, and Iacoboni (2009) found that the sudden discovery of an unexpected meaning in verbal problems is associated to the activation of a neural network including insula, anterior cingulate, and right PFC. A decreased alpha power in the right prefrontal region, indicating increased cortical activity, was reported by individuals who reached the solution of a problem by restructuring it (Sandkühler & Bhattacharya, 2008). Thus it appears that PFC is actually involved also in insight-like instances of analogical reasoning.

However, an aspect of Speed's model that needs further evidence concerns the link between analogical thinking, PFC, and dopaminergic subcortical circuits (such as basal ganglia). Although the cognitive and motor functions of dopaminergic subcortical structures are largely

demonstrated for many learning tasks, the neuromodulatory effect of dopamine on analogical transfer is not supported and actually demonstrated by clear empirical evidence (Shu-Chen & Sverker, 2002). In addition, the contribution of these circuits to PFC activity during analogical learning should be clarified. In other words, not only should the presence of an increased dopaminergic participation in the mediation of analogical learning be better evidenced, but the quantitative and qualitative decrease of the analogical transfer performance should be revealed in the case of pathologies that preclude a correct functioning of these cortical and subcortical structures. The direct reference to Parkinson's and Huntington's diseases, which were found to be related to deficits on a variety of analogical tasks, appears not relevant. The authors referred to studies in which specific cognitive tasks (such as the Card Sorting Task) were used, but such tasks more appropriately make evident the contribution of PFC and dopamine circuits to executive and metacognitive functions rather than to analogical thinking *per se*. This is the main reason why the basic question as to the contribution of different PFC areas in different phases of analogy—or, on the contrary, the local networks' responsiveness to different information—cannot receive an adequate response.

\* \* \*

## Development of reasoning: Behavioral evidence to support reinforcement over cognitive control accounts

**Julia R. Badger and Laura R. Shapiro**

*Psychology, Life and Health Sciences, Aston University, Aston Triangle, Birmingham B4 7ET, UK*  
*E-mail: badgerjr@aston.ac.uk*

DOI: 10.1080/17588921003786598

**Abstract:** Speed's theory makes two predictions for the development of analogical reasoning. Firstly, young children should not be able to reason analogically due to an undeveloped PFC neural network. Secondly, category knowledge enables the reinforcement of structural features over surface features, and thus the development of sophisticated, analogical, reasoning. We outline existing studies that support these predictions and highlight some critical remaining issues. Specifically, we argue that the development of inhibition must be directly compared alongside the development of reasoning strategies in order to support Speed's account.

Speed suggests that during development, neurons become organized along the posterior–anterior axis of the prefrontal cortex (PFC) such that anterior regions are tuned to increasingly abstract and complex situations. Speed’s theory proposes that these changes occur through “reinforcement-driven discrimination” (p. 134), such that the response preferences of neurons in PFC are tuned through reinforcement of particular relationships. We can therefore derive two key predictions for the development of analogical reasoning.

1. Young children should not be able to reason analogically due to an undeveloped PFC neural network (although Speed does not detail the timecourse of this development).
2. Analogical reasoning must be learned. Specifically, knowledge of the category of interest enables the reinforcement of “structural” features over “surface” features, and thus more sophisticated reasoning.

#### **SUPPORT FOR PREDICTION 1**

There is evidence that young children focus on perceptual over category information when making generalizations about the properties/behavior of new items. Sloutsky, Kloos, and Fisher (2007) found that four- to five-year-old children could successfully learn to group novel animals into categories, even when category information was pitted against obvious surface cues (shape/colour). However, in a subsequent induction task, children reverted to grouping objects in terms of obvious surface information, rather than basing induction decisions on category information. We have replicated this finding using biologically plausible novel animals (Badger & Shapiro, 2010), and additionally show that children shift from perceptual induction to category induction around age seven. These data suggest that young children’s natural default is to focus on surface features, and the ability to use structural features develops gradually during early childhood. According to Speed, this ability is constrained by the development of the PFC network.

#### **SUPPORT FOR PREDICTION 2**

There is evidence that children can be trained to focus on non-obvious or unobservable biological features within a particular domain. For example, Au et al. (2008) found that training that focused on the biological causal mechanism for cold and flu transmission was considerably more effective in impacting on children’s

reasoning about infectious diseases and preventive behavior than rule-based training programs. These data fit with Speed’s hypothesis that category knowledge reinforces PFC response preferences that correspond to structural over surface features of the domain.

### **OUTSTANDING ISSUES**

There is an alternative interpretation of these data. Specifically, Gelman (e.g., 2003; see also Bulloch & Opfer, 2009) argues that even young children are biased towards essentialist (internal, intrinsic) causes, and thus should show a natural default towards category induction. Thus, any bias towards surface/perceptual features in early reasoning must be interpreted as an inability to inhibit the “obvious” perceptual response over the less salient category-based default. Similarly, Au et al.’s findings must be interpreted as improving children’s ability to inhibit obvious/observable explanations, enabling their natural bias towards “essentialist” causes to be expressed.

Our research (Badger & Shapiro, 2010) directly tests perceptual-bias vs. inhibition interpretations of early perceptual induction. As in Sloutsky et al., we trained children to categorize novel insects according to a non-obvious category-membership rule (head shape). The salient perceptual cues (e.g., overall size, shape, colour) were not predictive of category membership. Children then performed a triad induction task in which they were asked to generalize a hidden property of the target to one of two test items (same-category choice or perceptual distractor). Unlike previous studies, we compared induction choices when the distractor items were at different levels of similarity to the target. As expected, children made significantly more perceptual (distractor) choices when this item was highly similar to the target. However, this effect did not interact with age. Thus, increased inhibition abilities cannot be driving children’s shift from perceptual to category induction. Instead, this shift is likely to be triggered by the development of more sophisticated reasoning abilities. These findings support Speed’s PFC account of analogical reasoning over competing accounts based on cognitive control, and highlight the need for developmental evidence to back up claims about the neural mechanisms of analogical processing.

### **ACKNOWLEDGMENT**

Supported by an ESRC studentship to the first author (EF/H016902/1).

\* \* \*

## Is prefrontal cortex necessary for the storage and acquisition of relational concepts?

**David Badre**

*Box 1978, Brown University, Providence,  
RI 02912-1978, USA  
E-mail: David\_Badre@brown.edu*

DOI: 10.1080/17588921003782589

**Abstract:** The ability to make analogies requires building higher order relations and so keeping track of multiple independently varying dimensions of the concepts being compared. Frontal cortex may be well suited to support this type of function, as Speed's review makes clear. However, Speed goes further in arguing that PFC neurons necessarily support the storage and acquisition of relational concepts. This claim is evaluated in the context of broader perspectives on storage and acquisition of semantic knowledge.

It remains an open question how to characterize the functional organization of frontal cortex. For example, though there is growing evidence that progressively rostral frontal neurons support increasingly abstract control processing (Badre & D'Esposito, 2007; Christoff, Keramian, Gordon, Smith, & Madler, 2009; Koechlin, Ody, & Kouneiher, 2003), what makes those processes or neural representations more abstract remains controversial (Badre, 2008; Botvinick, 2008). Some have focused on timescale, noting that rostral portions of frontal cortex maintain information that provides the context for action over longer time intervals (Kouneiher, Charron, & Koechlin, 2009). My colleagues and I have emphasized policy abstraction or the degree to which a goal or rule that guides action generalizes over more specific goals or rules (Badre, Hoffman, Cooney, & D'Esposito, 2009). Speed articulates a third account, namely that prefrontal cortex (PFC) neurons are tuned for different degrees of relational complexity (also see Christoff et al., 2001b; Kroger et al., 2002).

These perspectives on rostro-caudal organization are not mutually exclusive. For example, abstract policy and complex relations both require keeping track of multiple independently varying dimensions. So, the ability to reason about higher order relations could rely on the same architecture that permits control over multileveled policy.

Importantly, however, Speed appears to go beyond assigning a role in relational reasoning to PFC, but makes the further claim that PFC is the site of storage of relational concepts and that striato-frontal mechanisms are necessary for their acquisition. For example,

she suggests that rostral PFC may respond preferentially to highly relational concepts, like that of a *life cycle*. Similarly, learning about Rutherford's model of an atom requires fronto-striatal learning circuits in order to tune frontal neurons for the relational representations that support this concept. In this claim, Speed appears to move somewhat beyond other perspectives on frontal organization. In particular, frontal neurons do not simply modulate processing of semantic representations stored elsewhere with respect to goal-relevant relations. Rather, processing relational semantics requires a necessary route through PFC because this is where this knowledge is stored.

A distinction is often drawn in cognitive neuroscience between stored semantic knowledge, like that of a life cycle, and control functions—including reasoning processes like analogy—that act on those stored representations to produce responses or to generatively discover new knowledge via inference. This distinction arises partly from the neuropsychological literature (Martin & Chao, 2001). Damage to posterior neocortical structures, particularly those in lateral temporal regions, is often associated with a loss of semantic knowledge (Damasio, Tranel, Grabowski, Adolphs, & Damasio, 2004; Martin & Chao, 2001). Conversely, PFC patients can have an intact semantic store but problems in strategically using this knowledge. For example, Sylvester & Shimamura (2002) studied a group of PFC patients who demonstrated deficits in tests of strategic episodic and semantic retrieval. The patients also performed triadic comparison and ordered similarity tasks that are highly dependent on stored relational semantics. Using multidimensional scaling, the “semantic space” of these patients was constructed from their responses on these tasks. In contrast to their deficits in strategic retrieval, their semantic space did not differ from the controls. Hence, stored relational semantics may not depend exclusively on PFC.

Similarly, not all relational semantics need be acquired via reinforcement-based striatal learning mechanisms. Consider that I could learn about the analogy between the solar system and Rutherford's model of the atom not just through a process of analogy but alternatively by reading Speed's review, which tells me the analogy between these concepts. In this latter case, my episodic memory system can immediately encode this novel relationship. Over time, this knowledge can be consolidated into semantic memory. This process requires the medial temporal lobe (MTL) memory system, rather than the striatum. Hence, amnesia arising from damage to MTL prevents acquisition of new semantic knowledge (Cohen, Eichenbaum, & Poldrack, 1997; Squire, 1994). Thus, fronto-striatal mechanisms may not be required for acquisition of relational semantics.

Recent perspectives on the functional organization of frontal cortex highlight its ability to keep track of multiple independent dimensions in order to internally guide thought and action. Such an architecture is important for analogy making. Also, PFC may form abstract representations of the types of relations that will be useful for generating future analogies. However, these abstract PFC representations may be distinct from abstract representations stored elsewhere that form our store of semantic knowledge.

### ACKNOWLEDGMENTS

Supported by the National Institutes of Health (NS065046).

\* \* \*

## Are all analogies created equal? Prefrontal cortical functioning may predict types of analogical reasoning

Evangelia G. Chrysikou  
and Sharon L. Thompson-Schill

University of Pennsylvania—Psychology, 3720  
Walnut Street, Philadelphia, PA 19104, USA  
E-mail: [evangelg@psych.upenn.edu](mailto:evangelg@psych.upenn.edu)

DOI: 10.1080/17588921003763977

**Abstract:** The proposed theory can account for analogies based on learned relationships between elements in the source and target domains. However, its explanatory power regarding the discovery of new relationships during analogical reasoning is limited. We offer an alternative perspective for the role of PFC in analogical thought that may better address different types of analogical mappings.

Analogical thought is often considered the cornerstone of abstract reasoning; it allows one to uncover relationships between a familiar situation in memory (the *source* domain) and a new situation (the *target* domain) that may not be well understood. Notably, analogy does not involve simple retrieval of information about the two domains, but a *mapping* between their surface elements based on shared abstract relationships.

Speed's framework attempts to illuminate the neural mechanisms underlying analogical thinking by

specifying how flexible encoding of relational information from experience in prefrontal brain circuits allows for the differentiation between surface and structural features during analogical mappings. Specifically, prefrontal cortex (PFC) neurons represent relationship information between different concepts, features, or actions; these neurons are attuned to relational categories of increasing abstractness following a hierarchical rostrocaudal organization. Upon establishment of successful mappings between a source and a target domain, PFC neurons are selectively activated to represent abstract longer-term relational categories; moreover, the response of PFC neurons persists depending on the behavioral relevance of the activated relationships.

An important product of analogical reasoning in the real world is the understanding of a target domain through the discovery of new relationships that were not previously known outside of the context of the analogy. Although Speed's account may provide a neural framework for understanding how abstract relationships are learned and stored progressively in PFC neurons from experience and how available relational information may be activated during A:B::C:D analogical reasoning problems that have been traditionally used to study analogy in the laboratory (e.g., *brain:thought :: stomach: ?*), this presents a question for the current theory: If the target domain is not well understood, how are these new or unlearned relationships between the domain's surface features explicitly represented in PFC neurons?

For example, when one is using an object in a novel way instead of a typical object, to achieve an ad hoc goal (e.g., using a *baseball bat* as *rolling pin*), how is the mapping of the abstract relationship between the objects represented in PFC? When Watson and Crick mapped analogically the known helical structure of the alpha-keratin molecule (source domain), to discover the unknown structure of the DNA molecule (target domain), how was relational information about the source activated without predicting in advance their potential for reward or punishment? How was relational information about the target domain represented, given that it was this very abstract relational information about the DNA structure that was poorly understood? Such analogies may require different representation in PFC neurons relative to analogical mappings between well-understood source and target domains (e.g., the analogy "planet is to sun as electron is to nucleus" refers to *already established* relationships about the structure of the solar system and atom).

We propose here that, instead of representing relational categories of increased abstractness, the PFC

functions as a domain-general, biasing mechanism that sculpts the representational response space (Frith, 2000), focusing attention on certain aspects or features of a representation during analogical reasoning, while ignoring others. Such a conceptualization of the PFC may allow for explicit predictions regarding the extent of the involvement of this region depending on the type of analogical reasoning. According to this approach, PFC might be involved in analogies that are based on strong preexisting knowledge of abstract structural relationships in the source and target domains. In such cases, biasing the response space would allow for focus only on the relevant aspects of these relationships for a successful analogical mapping between the source and target domains. In contrast, PFC regions may not be involved to the same extent for analogies that are not based on explicit preexisting knowledge and which—if successful—might lead to new discoveries. In such cases, biasing the response space may be counterproductive, given that one may not know in advance which relationships will become of optimal behavioral relevance (see Chrysikou & Thompson-Schill, in press; Thompson-Schill, Ramscar, & Chrysikou, 2009).

We argue that such an approach to PFC offers a neural framework for analogical reasoning that is able to account for both types of analogy, which may further our understanding of analogical transfer (or its failure) in real-life circumstances.

#### ACKNOWLEDGMENTS

This work was supported by NIMH grant R21MH083029 and NIDCD grant R01DC009209 to Sharon L. Thompson-Schill.

\* \* \*

## Does the PFC model of analogy account for decision making, problem solving, reasoning, flexibility, adaptability, and even creativity?

Joaquín Barutta<sup>1</sup>, Raphael Guex<sup>2</sup>,  
and Agustín Ibáñez<sup>3</sup>

<sup>1</sup>*Italian Hospital University, Buenos Aires, Argentina*

<sup>2</sup>*University of Geneva, Geneva, Switzerland*

<sup>3</sup>*Institute of Cognitive Neurology, (INECO), Favaloro University, National Scientific and Technical Research Council (CONICET), Buenos Aires, Argentina, and Universidad Diego Portales, Santiago, Chile*

*E-mail: aibanez@neurologiacognitiva.org*

DOI: 10.1080/17588921003786580

**Abstract:** From everyday cognition to scientific discovery, analogical processes play an important role: bringing connection, integration, and interrelation of information. Recently, a PFC model of analogy has been proposed to explain many cognitive processes and integrate general functional properties of PFC. We argue here that analogical processes do not suffice to explain the cognitive processes and functions of PFC. Moreover the model does not satisfactorily integrate specific explanatory mechanisms required for the different processes involved. Its relevance would be improved if fewer cognitive phenomena were considered and more specific predictions and explanations about those processes were stated.

Speed proposes a novel PFC model of analogical processing. This model explains analogical processes as a progressive integration from posterior to more anterior areas of PFC, during which the information processing increases in abstractness and complexity. The frontostriatal circuits would bring the basis for analogy formation and persistence, sustained by learning and prediction of reward/punishment. The model is discussed in relation to other approaches to PFC and also to several processes involved, such as explicit and implicit processing, long vs. short-term representations, and cognitive control. More importantly, this model is presented as a useful tool for integrating the multiple functions of PFC in order to understand complex behaviors, such as decision making, problem solving, reasoning, flexibility, adaptability, and even creativity.

In spite of the main merit of this work, which lies in an effort to integrate the different roles of PFC and the analogical processes in order to understand complex behaviors, there are several caveats that raise doubts about the model's usefulness.

Although analogy would be a very important factor in wide-ranging cognitive processes, it is hard to imagine how a general cognitive skill such as analogy could be enough to explain as many cognitive processes as proposed by Speed. Would the same analogical model explain decision making, reasoning, creativity, and other very disparate processes? How is it possible for such a model to achieve this goal? Is there an identical neuronal substrate for all these cognitive processes? No precise description or insight on these main issues can be found in the paper. In the same vein, those complex

cognitive skills usually involve a conjugation of several processes (e.g., decision making can engage reversal learning and inhibition, risk-taking, emotion, executive function, and working memory, and some of those skills are known to be processed in other areas than the PFC; Dunn, Dalgleish, & Lawrence, 2006). No clear pathways that explain the sufficiency of analogy to account for such disparate cognitive processes are addressed in the PFC analogy model.

Moreover, the model doesn't sufficiently specify the kinds of analogies involved in such cognitive processes. Even though analogical explanation, analogical description, and analogical reasoning share a common characteristic (all make use of analogies), the kinds of information they provide are substantially different (Copi, 1994; Gamut, 1991). In fact, in spite of the well-known relevance of analogical reasoning in cognitive processes such as decision making or problem solving, the paper seems to specifically focus on analogical explanation and description. This is also apparent in the examples provided by the author.

Today it is widely accepted that complex cognition recruits large and relatively specific networks, including but also going beyond the PFC, and with very detailed cognitive properties. This is especially relevant when considering decision making (Frith & Singer, 2008), reasoning (Reijneveld, Ponten, Berendse, & Stam, 2007), creativity (Yeats & Yeats, 2007), or problem solving (Unterrainer & Owen, 2006). The proposed model does not fit as an explanatory mechanism of the neurocognitive functions required to address such different cognitive and neurophysiological processes. If no specific behavioral or neurophysiological predictions can be stated for each cognitive phenomenon addressed by this model, the extreme extension of the phenomena considered by the model becomes an enormous difficulty instead of being advantageous over alternative PFC explanations.

In brief, although Speed's proposal is novel and interesting, it sounds too ambitious and at the same time lacks the wide range of model predictions and explanations expected to account for such a variety of phenomena. Possibly, a model improvement would consist in a less ambitious range of cognitive phenomena and, simultaneously, the development of a more specific set of predictions and explanations.

#### ACKNOWLEDGMENTS

Supported by the CONICET career grant to Agustín Ibáñez.

\* \* \*

## What role for the anterior cingulate in analogical reasoning?

Michael W. O'Boyle

*Department of Human Development and Family Studies, Texas Tech University, Lubbock, TX 7940, USA*

*E-mail: michael.oboyle@ttu.edu*

DOI: 10.1080/17588921003802064

**Abstract:** While prefrontal and frontal cortex of the brain are well documented to mediate many executive functions, including creativity, flexibility, and adaptability, the anterior cingulate cortex (ACC) is known to be involved in error detection and conflict resolution, and is crucial to reward-based learning. A case is made for the notion that any neural model of analogical reasoning must incorporate the critical (and specialized) contributions of the ACC.

In her target article, Ann Speed does an admirable job of outlining a model designed to capture the neural circuitry underlying analogical reasoning in the brain. In particular, she suggests that different neurons along the anterior-posterior axis of the prefrontal cortex (PFC) are differentially sensitive to the abstractness and relatedness of the informational components comprising analogies, and that the persistence of the representations used for analogy solution is mediated by fronto-parietal neural circuits that are sensitive to environmental consequences (i.e., their potential for success/reward or failure/punishment).

One aspect of the model that seems to have been overlooked, however, is the engagement of the anterior cingulate cortex (ACC) during higher-order cognition. Note that considerable research is being done on this brain region (Brodmann area 32 and others), which has revealed a crucial role both anatomically and behaviorally for the ACC in the performance of a variety of higher-order cognitive tasks: contributions that would presumably extend to the analogical reasoning process.

At the anatomical level it is well documented that bilateral premotor and dorsolateral PFC are highly interconnected and that each of these regions projects directly to the ACC, which in turn is highly interconnected with virtually all other frontal areas of the brain (Petrides & Pandya, 1999). Moreover at the behavioral level, the ACC is known to mediate and facilitate the online monitoring of performance primarily through error checking and conflict resolution

(Carter et al., 1998), and is thus particularly sensitive to the success/failure (i.e., consequences) of the current cognitive resource allocation strategy being utilized to solve an ongoing task. In light of its extensive anatomical interconnections with the PFC and its important role in the fine tuning of successful task performance, particularly in reward-based learning circumstances (Bush et al., 2002) similar to those required to solve analogies, one wonders how a discussion of the ACC and its specialized functions could be omitted from the target article as the ACC is certain to be an integral component of the brain circuitry underlying analogical reasoning.

The importance of the ACC to high-level cognition, specifically in the domain of expert/novice performance, can be seen in several papers that colleagues and I have published indicating that one of the key brain characteristics that differentiates average math (AM) ability children from those who are math gifted (MG) (i.e., those who are in the top one-half of 1% in mathematical reasoning ability as measured by their performance on the SAT-Math) is the intense (BOLD) engagement of the ACC (see O'Boyle, et al., 2005). For example, when performing 3-D mental rotation, the MG and AM children activate the same prefrontal and parietal regions (although the pattern is predominantly bilateral in the MG and more right hemispheric in the AM). But in the MG, the ACC is significantly more active than in the AM group. This finding suggests that the ACC serves an important executive function in terms of online and real-time monitoring of how well the MG individuals are performing the task at hand (in this case 3-D rotation of objects). And it appears to play a crucial role in the switching (and fine tuning) of the allocation of cognitive resources by the brain (i.e., either increasing or decreasing the original activation level) to optimize task performance. In light of its specialized functions, it would seem essential to incorporate the contributions of the ACC in any attempt at modeling the neural underpinnings of analogical reasoning.

\* \* \*

## The neurophysiology of magnitude: One example of extraction analogies

Carmelo Mario Vicario<sup>1,2,3</sup>  
and Davide Martino<sup>4</sup>

<sup>1</sup>University of Rome, and IRCCS Fondazione Santa Lucia, Rome, Italy

<sup>2</sup>Dipartimento di Psicologia, Università di Roma "La Sapienza", Via dei Marsi 78, 00185, Rome, Italy

<sup>3</sup>S. I. S. S. A – I. S. A. S International School for Advanced Studies, via Beirut 2-4, Trieste, Italy

<sup>4</sup>University of Bari, Bari, Italy

E-mail: [carmelo.vicario@uniroma1.it](mailto:carmelo.vicario@uniroma1.it)

DOI: 10.1080/17588921003763969

**Abstract:** Brain mapping for analogical information seems based on extraction of similarity on various levels, including perceptual similarity, abstract conceptual dimensions, and goals. Given the utility of analogical inferences on quantities that would extract covariance of time, space, and numbers, we discuss here the processing of time, space, and quantity as an example of the process of extraction of analogical information. This view is supported by evidence documenting common activation of the prefrontal cortex (PFC) in the processing of all of these magnitudes.

The brain regions that mediate analogical reasoning are not well explored. There is increasing interest in understanding the neural systems that mediate analogical processing, which is essential for learning and fluid intelligence (Wartenburger, Heekeren, Preusse, Kramer, & van der Meer, 2009) as well as to generate predictions of future relevant experiences (Bar, 2007).

Given the importance of analogical processing in human cognition, it is striking that there is such a lack of knowledge regarding the cerebral correlates underlying analogical processing. Some types of analogical thinking have been found to activate the lateral and medial PFC (Bunge, Wendelken, Badre, & Wagner, 2005; Waltz et al., 2000). Other studies report also the involvement of parietal brain regions, particularly during the processing of geometric analogies (Wartenburger et al., 2009; Wharton et al., 2000).

In her article Speed addresses this issue, proposing that neurons in PFC are particularly sensitive to relationship information that increases in abstractness and complexity along the posterior–anterior axis. In approaching this issue, however, no previous works investigating common cortical metrics of time, space, and quantity were taken into account. Given the utility of relational inferences on quantities that would extract covariance of time, space and numbers, this review offers the intriguing possibility to hypothesize that principles proposed in the theory of magnitude (Walsh, 2003) represent an exemplar case of analogical processing. Indeed, the extraction process of analogies can be based on similarity at various levels, including not only perceptual similarity but also

abstract conceptual dimensions and goals (Hummel & Holyoak, 2003). The theory of magnitude is based on the view that the linking function of the multiple capacities of the frontal and parietal cortex is the need to encode information about magnitudes in the external world needed during motor action (Walsh, 2003). Walsh (2003) posits the existence of a generalized magnitude system in our cognitive system, in which physical distances, temporal durations, and numerical distances are subserved by the same “pool” of neurons documented in the primate and human brains. Under this view one could suppose some cognitive mechanism extracting magnitude analogies along a common code, during the processing of space, time, and quantity information. This common code seems to assume a vectorial texture, so that the spatial concepts “few” and “many” map onto “short” and “long,” respectively (Vicario et al., 2008). Two nonhuman primate studies have addressed the cortical processing of duration (Leon

and Shadlen, (2003; Onoe et al., 2001). The two areas activated were dorsolateral prefrontal cortex (DLPFC) and the inferior parietal lobe: areas that have shown both number and spatial properties in single unit studies. The view for an involvement of frontal regions in the extraction of magnitude analogies agrees with the discovery of “numerons” (Nieder, 2003) localized in the prefrontal cortex (PFC) region, which seems selective for both spatial and temporal information (Harrington, Haaland, & Knight, 1998; Rao, Mayer, & Harrington, 2001).

Therefore, the transversal activation of the PFC regions following the exposure to temporal, numerical, and spatial information might cover up some general mechanism involved in the extracting of magnitude analogies. This process, in turn, might give rise to priming and/or interference phenomena between magnitudes such as reported in the literature.

\* \* \*

---

## Reply to Commentaries

# Architecture of PFC supports analogy, but PFC is not an analogy machine

Ann Speed

*Sandia National Laboratories, Albuquerque, NM, USA*

In the preceding discussion paper, I proposed a theory of prefrontal cortical organization that was fundamentally intended to address the question: How does prefrontal cortex (PFC) support the various functions for which it seems to be selectively recruited? In so doing, I chose to focus on a particular function, analogy, that seems to have been largely ignored in the theoretical treatments of PFC, but that does underlie many other cognitive functions (Hofstadter, 2001; Holyoak & Thagard, 1997). At its core, this paper was intended to use analogy as a foundation for exploring one possibility for prefrontal function in general, although it is easy to see how the analogy-specific interpretation arises (as in the comment by Ibáñez). In an attempt to address this more foundational question, this response will step away from analogy as a focus, and will address first the various comments from the perspective of the initial motivation for developing this theory, and then specific issues raised by the commentators.

The original motivation for this theory began with an effort to build a computational model of analogy-making that was more closely tied to neurophysiological mechanisms (Speed, Verzi, Benz, Dixon, & Warrender, in preparation). In analyzing existing models of analogy (e.g., Eliasmith & Thagard, 2001; Falkenhainer, Forbus, & Gentner, 1989; Hummel & Holyoak, 2003; Kokinov & Petrov, 2001; Leech, Mareschal, & Cooper, 2008; Mitchell, 1993), we failed to find an answer to the question: Where does the information come from on which PFC performs analogy (Speed, 2008)? All such models involve representation of sparse, and fairly abstract, symbolic concepts. The most physiologically realistic representations appear in two different models: LISA and DRAMA (Eliasmith & Thagard, 2001; Hummel & Holyoak, 1997, 2003). LISA is a hybrid symbolic-connectionist model with inputs to nodes representing concepts such as “male” or “emotion-object.” DRAMA uses holographic reduced representations (512-dimension vectors) to represent complex concepts in a distributed manner. However, these concepts comprise the nouns, adjectives, and verbs in verbal analogies such as the mapping between Cupid’s arrow and the pang felt in the heart of the lovesick. But the hard problem remains: How do these models “know” anything about what it means to be male or

lovesick in any way that enables them to make analogies the way the brain might?

Thus, in building the model, we attempted to address this question in a way that did not limit its explanatory power to analogy alone. That is, this same knowledge representation question must be addressed for whatever PFC function one is considering, be that working memory, inhibition, decision making, etc. Therefore, questions that had to be addressed in order to implement computational code included: Are neurons in PFC representational or do they perform some kind of transformation (an issue raised in Badre’s comment; see also Wood & Grafman, 2003)? If PFC neurons represent information, how do those representations come to be and what is their nature? Since it is unlikely they are hard-coded (i.e., present from birth), they must be learned—but how? On the other hand, if PFC neurons perform operations (e.g., biasing), are those operations hard-coded or are they learned? If PFC representations/operations are learned, is this process supervised or is it unsupervised? That is, is there a teaching signal or are random associations made by simply encountering huge amounts of data (as in unsupervised machine learning techniques)? Given the widespread argument that PFC exercises some kind of control over other areas

of cortex—whether biasing, inhibition, or otherwise, the question remained—how? If PFC is the ever-elusive “executive,” how does it “know” what to inhibit? Or to bias? Or to otherwise control?

Most of theoretical work in PFC did not yield answers to these questions detailed enough to implement code. The work that began to provide some answers came from O’Reilly’s computational modeling and theoretical work in strially based working memory, and from Miller’s empirical demonstrations that individual cells in primate PFC respond to abstract rules (e.g., Freedman, Riesenhuber, Poggio, & Miller, 2003; O’Reilly & Frank, 2006; Wallis, Anderson, & Miller, 2001). Thus, the current proposal attempted to address these questions. It describes what PFC neurons might represent/do and how that functionality is tuned.

Some of the predictions made in the target article have already been supported by new data (e.g., Badger & Shapiro’s comment, which provides evidence against a strict inhibition interpretation of PFC function), and there may be other such examples in the literature. Ultimately, differences between the current proposal and other accounts are empirical questions. Several suggestions along these lines are made below.

### PROPOSALS FOR FUTURE EMPIRICAL WORK

Regarding the involvement of anterior cingulate cortex (ACC) in development of behaviorally relevant response preferences in PFC raised by O’Boyle, one difficulty with a strict basal ganglia explanation for response tuning is that there is no clear mechanism for the effects of punishment—only for removal of expected rewards (e.g., Frank, 2005). Thus, if ACC does provide a mechanism for learning via presentation of aversive consequences (i.e., positive punishment), in addition to augmenting the striatal sensitivity to reward removal (i.e., negative punishment), one would expect to see increased ACC activity due to experimental manipulations that involve aversive presentation (e.g., Nishijo et al., 1997). Similarly, damage to ACC or to ACC–PFC pathways should cause difficulty learning new abstract categories that predict aversive consequences.

Related to this is the concern raised by Antonetti and Balconi in which they question the role of striato-frontal circuits in analogical mapping. This question reflects a possible point of confusion regarding the primary goal of the target article. This circuit was not intended as an explanation for the process of analogical mapping. Rather, it was intended as the method by which relational representations come to exist in

PFC. It is these representations, then, that are assumed to enable analogical mapping via activation of those abstract concepts from environmental stimuli and top-down biasing of posterior areas selectively responsive to the particular objects present in the current source and target pair (whether this process happens explicitly, or implicitly, as in insight problems the authors cite). A link between basal ganglia recruitment and category learning has been presented by Ashby and colleagues (e.g., Ashby & Ennis, 2006).

Regarding the question raised by Chrysikou and Thompson-Schill, “How are these new or unlearned relationships between the domain’s surface features explicitly represented in PFC neurons?” (p. 141), one must consider that analogy is typically between a target that is poorly understood and a source that is better understood. Thus, in order to actually make a valid analogy between two problems, the structure of one (the source) already enjoys response tuning in PFC. However, it is also important to realize that often novices make inappropriate analogies (e.g., Novick, 1988; Ross, 1989). In the education and training literature, inappropriate analogies are known as negative transfer—situations in which performance on the target is actually worse for having experienced the source because the two are not a valid analogical pair (Cormier, 1987). Relatedly, I would argue that the reason analogy is difficult to achieve in the laboratory (e.g., Catrambone & Holyoak, 1989; Gick & Holyoak, 1983) is that when representations of the source and target are based on single examples, they are insufficient for making valid analogies. As was specified in the target article, and as supported by behavioral evidence (e.g., Catrambone & Holyoak, 1989; Gick & Holyoak, 1983; Novick, 1988), only after experience with multiple examples of a class of problems can an individual reliably identify the structural features of that class. Reward and punishment are simply teaching signals for indicating the validity of the analogy, which then influences *future* behavior. There is no a priori way to predict when an analogy will be valid, and often they are not (i.e., a Westerner failing to remove his shoes at the door of a Japanese home). It is only by making mistakes (invalid analogies) that we come to learn the structural features of problems (cf. Ghodsian, Bjork, & Benjamin, 1997).

The proposal by Vicario and Martino, that processes underlying understanding magnitude may be an example of analogical processing, is intriguing and fits well with studies indicating a frontoparietal network involved in solving Raven’s-like matrix problems (Kroger et al., 2002; Prabhakaran, Smith, Desond, Glover, & Gabrieli, 1997). Subsequent work that directly compares activity in parietal vs. prefrontal

cortices as a function of responses to subordinate (e.g., individual representations of the number 1), ordinate (e.g., the concept of the number 1), and superordinate (e.g., the concept of prime numbers, real numbers, integers, etc.) stimuli in progressively more complex problems could test this hypothesis.

Regarding the location of semantic knowledge representation question raised by Badre, an area of research that seems lacking in the PFC literature is in the development of knowledge: from childhood to adult (which exists in the analogy literature to an extent, e.g., Crone et al., 2009), from novice to expert, and from immediately post-injury to years out. Such analysis could be very instructive, and should complement the many snapshots of functioning we have now. For example, Sylvester & Shimamura (2002) examine the semantic categorization abilities of several frontal patients who average 11 years post-injury. They found that patients group common animals in the same way that an uninjured age-controlled group does. However, if the current theory is correct, 11 years is plenty of time for patients to have reacquired these categorical representations in undamaged areas of PFC. Thus, testing to see whether this categorization changes over time post-injury could be instructive.

Badre also raises the issue that striatal circuits may not be necessary for the development of relational

knowledge. However, one finding from the education literature is that the way knowledge is learned in the classroom (i.e., by being told) can produce “inert” knowledge. That is, while the student may be able to restate the concept, he is unable to transfer it to a novel situation or problem (Bransford, Sherwood, Vye, & Rieser, 1986; Smith, Ford, & Kozlowski, 1997). Imaging studies comparing semantic knowledge that can only be restated vs. semantic knowledge that can be transferred analogically will elucidate whether the type of knowledge addressed by the current proposal is different from that discussed by Badre.

There are certainly many empirical data to collect in order to determine the viability of the current proposal. However, this account does raise some important questions for current theoretical perspectives regarding the physical mechanisms underlying those proposals. As I have argued elsewhere (Speed, 2008), there is a need to pay increasing attention to the actual physical mechanisms that underlie theoretical accounts of PFC function (see also, Hazy, Frank, & O’Reilly, 2006; O’Reilly & Frank, 2006). Irrespective of the ultimate fate of the current proposal, I hope that empirical tests pitting it against other perspectives, and additional physiologically based computational modeling efforts, will result in a more complete understanding of the physical mechanisms underlying PFC function.

---

## References from the Discussion Paper, the Commentaries, and the Reply

- Anolli, L., Antonietti, A., Cantoia, M., & Crisafulli, L. (2001). Accessing source information in analogical problem solving. *Quarterly Journal of Experimental Psychology*, *54A*, 237–261.
- Antonietti, A. (2001). *Analogical discoveries: Identifying similarities to solve problems*. Rome: Carocci.
- Antonietti, A., & Gioietta, M. A. (1995). Individual differences in analogical problem solving. *Personality and Individual Differences*, *18*, 611–619.
- Ashby, F. G., & Ennis, J. M. (2006). The role of the basal ganglia in category learning. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 46, pp.1–36). New York: Elsevier.
- Ashby, F. G., Ennis, J. M., & Spiering, B. J. (2007). A neurobiological theory of automaticity in perceptual categorization. *Psychological Review*, *114*(3), 632–656.
- Au, T. K., Chan, C. K. K., Chan, T.-K., Cheung, M. W. L., Ho, J. Y. S., & Ip, G. W. M. (2008). Folkbiology meets microbiology: A study of conceptual and behavioral change. *Cognitive Psychology*, *57*, 1–19.
- Aziz-Zadeh, L., Kaplan, J. T., & Iacoboni, M. (2009). Aha! The neural correlates of verbal insight solutions. *Human Brain Mapping*, *30*, 908–916.
- Badger, J. R., & Shapiro, L. R. (2010). The shift from perceptual to category induction is independent of featural distraction. *Manuscript submitted for publication*.
- Badre, D. (2008). Cognitive control, hierarchy, and the rostro-caudal organization of the frontal lobes. *Trends in Cognitive Science*, *12*(5), 193–200.
- Badre, D., & D’Esposito, M. (2007). Functional magnetic resonance imaging evidence for a hierarchical organization of the prefrontal cortex. *Journal of Cognitive Neuroscience*, *19*(12), 2082–2099.
- Badre, D., Hoffman, J., Cooney, J. W., & D’Esposito, M. D. (2009). Hierarchical cognitive control deficits following damage to the human frontal lobe. *Nature Neuroscience*, *12*(4), 515–522.
- Bar, M. (2007). The proactive brain: Using analogies and associations to generate predictions *Trends in Cognitive Sciences*, *11*(7), 280–289.

- Barbas, H. (2000). Connections underlying the synthesis of cognition, memory, and emotion in primate prefrontal cortices. *Brain Research Bulletin*, *52*, 319–330.
- Bechara, A., & Damasio, A. R. (2005). The somatic marker hypothesis: A neural theory of economic decision. *Games and Economic Behavior*, *52*, 336–372.
- Borojerdi, B., Phipps, M., Kopylev, L., Wharton, C. M., Cohen, L. G., & Grafman, J. (2001). Enhancing analogic reasoning with rTMS over the left prefrontal cortex. *Neurology*, *56*, 526–528.
- Botvinick, M. M. (2008). Hierarchical models of behavior and prefrontal function. *Trends in Cognitive Sciences*, *12*(5), 201–208.
- Bransford, J. D., Sherwood, R., Vye, N. J., & Rieser, J. (1986). Teaching thinking and problem solving. *American Psychologist*, *41*, 1078–1089.
- Bulloch, M. J., & Opfer, J. E. (2009). What makes relational reasoning smart? Revisiting the perceptual-to-relational shift in the development of generalization. *Developmental Science*, *12*(1), 114–122.
- Bunge, S. A., Wendelken, C., Badre, D., & Wagner, A. D. (2005). Analogical reasoning and prefrontal cortex: Evidence for separable retrieval and integration mechanisms. *Cerebral Cortex*, *15*, 239–249.
- Bush, G., Vogt, B. A., Holmes, J., Dale, A. M., Greve, D., Jenike, M. A., et al. (2002). Dorsal anterior cingulate cortex: A role in reward-based decision making. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(1), 523–528.
- Cabeza, R., & Nyberg, L. (2000). Imaging cognition II: An empirical review of 275 PET and fMRI studies. *Journal of Cognitive Neuroscience*, *12*(1), 1–47.
- Carter, C. S., Braver, T. S., Barch, D. M., Botvinick, M. N., Noll, D., & Cohen, J. D. (1998). Anterior cingulate cortex, error detection and the on-line monitoring of performance. *Science*, *280*(5364), 747–749.
- Catrambone, R., & Holyoak, K. J. (1989). Overcoming contextual limitations on problem-solving transfer. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 1147–1156.
- Chow, T. W., & Cummings, J. L. (2007). Frontal-subcortical circuits. In B. L. Miller & J. L. Cummings (Eds.), *The human frontal lobes: Functions and disorders* (2nd ed., pp. 25–43). New York: Guilford Press.
- Christoff, K., & Gabrieli, J. D. E. (2000). The frontopolar cortex and human cognition: Evidence for a rostrocaudal hierarchical organization within the human prefrontal cortex. *Psychology*, *28*, 168–186.
- Christoff, K., Geddes, L. M. T., & Gabrieli, J. D. E. (2001a). Rostrolateral prefrontal cortex involvement in evaluating self-generated information. *NeuroImage*, *13*, S649.
- Christoff, K., & Keramatian, K. (2007). Abstraction of mental representations: Theoretical considerations and neuroscientific evidence. In S. A. Bunge & J. D. Wilson (Eds.), *Perspectives on rule-guided behavior* (pp. 107–126). Oxford, UK: Oxford University Press.
- Christoff, K., Keramatian, K., Gordon, A. M., Smith, R., & Madler, B. (2009). Prefrontal organization of cognitive control according to levels of abstraction. *Brain Research*, *1286*, 94–105.
- Christoff, K., Prabhakaran, V., Dorfman, J., Zhao, Z., Kroger, J. K., Holyoak, K. J., et al. (2001b). Rostrolateral prefrontal cortex involvement in relational integration during reasoning. *NeuroImage*, *14*(5), 1136–1149.
- Christoff, K., Ream, J. M., Geddes, L. M. T., & Gabrieli, J. D. E. (2003). Evaluating self-generated information: Anterior prefrontal contributions to human cognition. *Behavioral Neuroscience*, *117*(6), 1161–1168.
- Chrysikou, E. G., & Thompson-Schill, S. L. (in press). Dissociable brain states linked to common and creative object use. *Human Brain Mapping*.
- Cohen, N. J., Eichenbaum, H., & Poldrack, R. A. (1997). Memory for items and memory for relations in the procedural/declarative memory framework. *Memory*, *5*, 131–178.
- Constantinidis, C., Franowicz, M. N., & Goldman-Rakic, P. S. (2001). The sensory nature of mnemonic representation in the primate prefrontal cortex. *Nature Neuroscience*, *4*(3), 311–316.
- Copi, I. M. (1994). *Introduction to logic* (9th ed.). New York: Macmillan.
- Cormier, S. M. (1987). The structural processes underlying transfer of training. In S. M. Cormier & J. D. Hagman (Eds.), *Transfer of learning: Contemporary research and applications* (pp. 152–182). San Diego, CA: Academic Press.
- Crone, E. A., Wendelken, C., van Leijenhorst, L., Honomichl, R. D., Christoff, K., & Bunge, S. A. (2009). Neurocognitive development of relational reasoning. *Developmental Science*, *12*(1), 55–66.
- Cummings, J. L., & Miller, B. L. (2007). Conceptual and clinical aspects of the frontal lobes. In B. L. Miller & J. L. Cummings (Eds.), *The human frontal lobes: Functions and disorders* (2nd ed., pp. 12–21). New York: Guilford.
- Damasio, H., Tranel, D., Grabowski, T., Adolphs, R., & Damasio, A. (2004). Neural systems behind word and concept retrieval. *Cognition*, *92*(1–2), 179–229.
- Duncan, J. (2005). Task models in prefrontal cortex. In U. Mayr, E. Awh, S. W. Keele, & M. I. Posner (Eds.), *Developing individuality in the human brain: A tribute to Michael I. Posner* (pp. 87–108). Washington, DC: American Psychological Association.
- Duncan, J., & Owen, A. M. (2000). Common regions of the human frontal lobe recruited by diverse cognitive demands. *Trends in Neuroscience*, *23*, 475–483.
- Dunn, B. D., Dalgleish, T., Lawrence, A. D. (2006). The somatic marker hypothesis: A critical evaluation. *Neuroscience & Biobehavioral Reviews*, *30*(2), 239–271.
- Eliasmith, C., & Thagard, P. (2001). Integrating structure and meaning: A distributed model of analogical mapping. *Cognitive Science*, *25*, 245–286.
- Falkenhainer, B., Forbus, K. D., & Gentner, D. (1989). The structure mapping engine: Algorithm and examples. *Artificial Intelligence*, *41*, 1–63.
- Frank, M. J. (2005). Dynamic dopamine modulation in the basal ganglia: A neurocomputational account of cognitive deficits in medicated and nonmedicated Parkinsonism. *Journal of Cognitive Neuroscience*, *17*(1), 51–72.
- Freedman, D. J., Riesenhuber, M., Poggio, T., & Miller, E. K. (2001). Categorical representation of visual stimuli in the primate prefrontal cortex. *Science*, *291*, 312–316.
- Freedman, D. J., Riesenhuber, M., Poggio, T., & Miller, E. K. (2002). Visual categorization and the primate prefrontal cortex: Neurophysiology and behavior. *Journal of Neurophysiology*, *88*, 929–941.
- Freedman, D. J., Riesenhuber, M., Poggio, T., & Miller, E. K. (2003). A comparison of primate prefrontal and inferior temporal cortices during visual categorization. *Journal of Neuroscience*, *23*(12), 5235–5246.

- Frith, C. (2000). The role of dorsolateral prefrontal cortex in the selection of action as revealed by functional imaging. In S. Monsell & J. Driver (Eds.), *Control of cognitive processes* (pp. 549–565). Cambridge, MA: MIT Press.
- Frith, C. D., Singer, T. (2008). The role of social cognition in decision-making. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363, 3875–3886.
- Fuster, J. M. (1997). *The prefrontal cortex: Anatomy, physiology, and neuropsychology of the frontal lobe* (3rd ed.). Philadelphia: Lippincott-Raven.
- Fuster, J. M. (2006). The cognit: A network model of cortical representation. *International Journal of Psychophysiology*, 60, 125–132.
- Gamut, L. T. F. (1991). *Logic, language and meaning*. Chicago: University of Chicago Press.
- Geake, J. G., & Hansen, P. C. (2005). Neural correlates of intelligence as revealed by fMRI of fluid analogies. *NeuroImage*, 26(2), 555–564.
- Gelman, S. A. (2003). *The essential child: Origins of essentialism in everyday thought*. New York: Oxford University Press.
- Gentner, D. (1983). Structure mapping: A theoretical framework for analogy. *Cognitive Science*, 7, 155–170.
- Gentner, D. (1989). The mechanisms of analogical mapping. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 199–241). New York: Cambridge University Press.
- Gentner, D., Bowdle, B., Wolff, P., & Boronat, C. (2001). Metaphor is like analogy. In D. Gentner, K. J. Holyoak, & B. N. Kokinov (Eds.), *The analogical mind: Perspectives from cognitive science*. Cambridge, MA: MIT Press.
- Gentner, D., & Markman, A. B. (1997). Structure mapping in analogy and similarity. *American Psychologist*, 52, 45–56.
- Gentner, D., & Toupin, C. (1986). Systematicity and surface similarity in the development of analogy. *Cognitive Science*, 10, 277–300.
- Ghodsian, D., Bjork, R. A., & Benjamin, A. S. (1997). Evaluating training during training: Obstacles and opportunities. In M. A. Quinones & A. Ehrenstein (Eds.), *Training for a rapidly changing workplace: Applications of psychological research* (pp. 63–88). Washington, DC: American Psychological Association.
- Gick, M. L., & Holyoak, K. J. (1980). Analogical problem solving. *Cognitive Psychology*, 12, 306–355.
- Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology*, 15, 1–38.
- Goldman-Rakic, P. S. (1987). Circuitry of primate prefrontal cortex and regulation of behavior by representational memory. *Handbook of Physiology*. Bethesda, MD: American Psychological Society.
- Green, A. E., Fugelsang, J. A., Kraemer, D. J. M., Shamos, N. A., & Dunbar, K. N. (2006). Frontopolar cortex mediates abstract integration in analogy. *Brain Research*, 1096, 125–137.
- Green, A. E., Kraemer, D. J. M., Fugelsang, J. A., Gray, J. R., & Dunbar, K. N. (2010). Connecting long distance: Semantic distance in analogical reasoning modulates frontopolar cortex activity. *Cerebral Cortex*, 20, 70–76.
- Harrington, D. L., Haaland, K., & Knight, R. (1998). Cortical networks underlying mechanism of time perception. *Journal of Neuroscience*, 18, 1085–1095.
- Hazy, T. E., Frank, M. J., & O'Reilly, R. C. (2006). Banishing the homunculus: Making working memory work. *Neuroscience*, 139, 105–118.
- Hofstadter, D. R. (1995). *Fluid concepts and creative analogies*. New York: Basic Books.
- Hofstadter, D. R. (2001). Epilogue: Analogy as the core of cognition. In D. Gentner, K. J. Holyoak, & B. N. Kokinov (Eds.), *The analogical mind: Perspectives from cognitive science* (pp. 499–538). Cambridge, MA: MIT Press.
- Holyoak, K. J., & Thagard, P. (1997). The analogical mind. *American Psychologist*, 52, 35–44.
- Huey, E. D., Krueger, F., & Grafman, J. (2006). Representations in the human prefrontal cortex. *Current Directions in Psychological Science*, 15, 167–171.
- Hummel, J. E., & Holyoak, K. J. (1997). Distributed representations of structure: A theory of analogical access and mapping. *Psychological Review*, 104, 427–466.
- Hummel, J. E., & Holyoak, K. J. (2003). A symbolic-connectionist theory of relational inference and generalization. *Psychological Review*, 110, 220–264.
- Keane, M., Ledgeway, T., & Duff, S. (1994). Constraints on analogical mapping: A comparison of three models. *Cognitive Science*, 18, 387–438.
- Kelly, A. M. C., & Garavan, H. (2005). Human functional neuroimaging of brain changes associated with practice. *Cerebral Cortex*, 15, 1089–1102.
- Koechlin, E., Corrado, G., Pietrini, P., & Grafman, J. (2000). Dissociating the role of the medial and lateral anterior prefrontal cortex in human planning. *Proceedings of the National Academy of Sciences of the United States of America*, 97(13), 7651–7656.
- Koechlin, E., & Hayfil, A. (2007). Anterior prefrontal function and the limits of human decision-making. *Science*, 318, 594–598.
- Koechlin, E., Ody, C., & Kouneiher, F. (2003). The architecture of cognitive control in human prefrontal cortex. *Science*, 302, 1181–1185.
- Kokinov, B. N., & Petrov, A. A. (2001). Integrating memory and reasoning in analogy-making: The AMBR model. In D. Gentner, K. J. Holyoak, & B. N. Kokinov (Eds.), *The analogical mind: Perspectives from cognitive science* (pp. 59–124). Cambridge, MA: MIT Press.
- Kouneiher, F., Charron, S., & Koechlin, E. (2009). Motivation and cognitive control in the human prefrontal cortex. *Nature Neuroscience*, 12(7), 939–945.
- Kramer, J. H., & Quinlan, L. (2007). Bedside frontal lobe testing. In B. L. Miller & J. L. Cummings (Eds.), *The human frontal lobes: Functions and disorders* (2nd ed., pp. 279–291). New York: Guilford.
- Krawczyk, D. C., Morrison, R. G., Viskontas, I., Holyoak, K. J., Chow, T. W., Mendez, M. F., et al. (2008). Distraction during relational reasoning: The role of prefrontal cortex in interference control. *Neuropsychologia*, 46, 2020–2032.
- Kroger, J. K., Saab, F. W., Fales, C. I., Bookheimer, S. Y., Cohen, M. S., & Holyoak, K. J. (2002). Recruitment of anterior dorsolateral prefrontal cortex in human reasoning: A parametric study of relational complexity. *Cerebral Cortex*, 12, 477–485.
- Kroger, J. K., Speed, A., Anderson, J. P., Mikkelsen, E. J., Spring, D. K., & Polsky, A. L. (2007, October). *An ERP study of analogical reasoning*. Paper presented at the 47th Annual Meeting of the Society for Psychophysiological Research, Savannah, GA.

- Krueger, F., Barbey, A. K., & Grafman, J. (2009). The medial prefrontal cortex mediates social event knowledge. *Trends in Cognitive Sciences*, *13*, 103–109.
- Leech, R., Mareschal, D., & Cooper, R. P. (2008). Analogy as relational priming: A developmental and computational perspective on the origins of a complex cognitive skill. *Behavioral and Brain Sciences*, *31*, 357–378.
- Leon, M. I., & Shadlen, M. N. (2003). Representation of time by neurons in the posterior parietal cortex of the macaque. *Neuron*, *38*, 317–327.
- Luo, Q., Perry, C., Peng, D., Jin, Z., Xu, D., Ding, G., et al. (2003). The neural substrate of analogical reasoning: An fMRI study. *Cognitive Brain Research*, *17*, 527–534.
- Martin, A., & Chao, L. L. (2001). Semantic memory and the brain: Structure and processes. *Current Opinion in Neurobiology*, *11*(2), 194–201.
- Mikkelsen, E. J., Speed, A., Anderson, J. P., Spring, D. K., Polsky, A. L., & Kroger, J. K. (2010 *Manuscript submitted for publication*). Dissociated processing of representational structure and surface features in an analogical reasoning task: An ERP study.
- Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Reviews of Neuroscience*, *24*, 167–202.
- Miller, E. K., Erickson, C. A., & Desimone, R. (1996). Neural mechanisms of visual working memory in prefrontal cortex of the macaque. *Journal of Neuroscience*, *16*, 5154–5167.
- Miller, E. K., Nieder, A., Freedman, D. J., & Wallis, J. D. (2003). Neural correlates of categories and concepts. *Current Opinion in Neurobiology*, *13*, 198–203.
- Mitchell, M. M. (1993). *Analogy-making as perception*. Cambridge, MA: MIT Press.
- Nieder, A., & Miller, E. K. (2004). A parieto-frontal network for visual numerical information in the monkey. *Proceedings of the National Academy of Sciences of the United States of America*, *101*, 7457–7462.
- Nishijo, H., Yamamoto, Y., Ono, T., Uwano, T., Yamashita, J., & Yamashita, T. (1997). Single neuron responses in the monkey anterior cingulate cortex during visual discrimination. *Neuroscience Letters*, *227*, 79–82.
- Novick, L. R. (1988). Analogical transfer, problem similarity, and expertise. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 510–520.
- Novick, L. R., & Holyoak, K. J. (1991). Mathematical problem solving by analogy. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*, 398–415.
- O'Boyle, M. W., Cunnington, R., Silk, T., Vaughan, D., Jackson, G., Syngeniotis, A., et al. (2005). Mathematically gifted male adolescents activate a unique brain network during mental rotation. *Cognitive Brain Research*, *25*, 583–587.
- Onoe, H., Komori, M., Onoe, K., Takechi, H., Tsukada, H., & Watanabe, Y. (2001). Networks recruited for time perception: A monkey positron emission tomography (PET) study. *NeuroImage*, *13*, 37–45.
- O'Reilly, R. C., & Frank, M. J. (2006). Making working memory work: A computational model of learning in the prefrontal cortex and basal ganglia. *Neural Computation*, *18*, 283–328.
- O'Reilly, R. C., Noelle, D. C., Braver, T. S., & Cohen, J. D. (2002). Prefrontal cortex and dynamic categorization tasks: Representational organization and neuromodulatory control. *Cerebral Cortex*, *12*, 246–257.
- Petrides, M., & Pandya, D. N. (1999). Dorsolateral prefrontal cortex: Comparative cytoarchitectonic analysis in the human and the macaque brain and corticocortical connection patterns. *European Journal of Neuroscience*, *11*(3), 1011–1036.
- Prabhakaran, V., Smith, J. A. L., Desond, J. E., Glover, G. H., & Gabrieli, J. D. E. (1997). Neural substrates of fluid reasoning: An fMRI study of neocortical activation during performance of the Raven's Progressive Matrices Test. *Cognitive Psychology*, *33*, 43–63.
- Qiu, J., Li, H., Chen, A., & Zhang, Q. (2008). The neural basis of analogical reasoning: An event-related potential study. *Neuropsychologia*, *46*, 3306–3013.
- Rainer, G., & Miller, E. K. (2000). Effects of visual experience on the representation of objects in the prefrontal cortex. *Neuron*, *27*, 179–189.
- Ramnani, N., & Owen, A. M. (2004). Anterior prefrontal cortex: Insights into function from anatomy and neuroimaging. *Nature Reviews Neuroscience*, *5*, 184–194.
- Rao, S. M., Mayer, A. R., & Harrington, D. L. (2001). The evolution of brain activation during temporal processing. *Nature Neuroscience*, *4*, 317–323.
- Reijneveld, J. C., Ponten, S. C., Berendse, H. W., & Stam, C. J. (2007). The application of graph theoretical analysis to complex networks in the brain. *Clinical Neurophysiology*, *118*(11), 2317–2331.
- Reynolds, J. R., & O'Reilly, R. C. (2009). Developing PFC representations using reinforcement learning. *Cognition*, *113*(3), 281–292.
- Richland, L. E., Morrison, R. G., & Holyoak, K. J. (2006). Children's development of analogical reasoning: Insights from scene analogy problems. *Journal of Experimental Child Psychology*, *94*, 249–273.
- Ross, B. H. (1987). This is like that: The use of earlier problems and the separation of similarity effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*, 629–639.
- Ross, B. H. (1989). Distinguishing types of superficial similarities: Different effects on the access and use of earlier problems. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 456–468.
- Rougier, N. P., Noelle, D. C., Braver, T. S., Cohen, J. D., & O'Reilly, R. C. (2006). Prefrontal cortex and flexible cognitive control: Rules without symbols. *Proceedings of the National Academy of Sciences of the United States of America*, *102*, 7338–7343.
- Sandkühler, S., & Bhattacharya, J. (2008). Deconstructing insight: EEG correlates of insightful problem solving. *PLoS ONE*, *3*, e1459.
- Shu-Chen, L., & Sverker, S. (2002). Integrative neurocomputational perspectives on cognitive aging, neuromodulation, and representation. *Neuroscience & Biobehavioral Reviews*, *26*(2), 795–808.
- Sigala, N. (2004). Visual categorization and the inferior temporal cortex. *Behavioural Brain Research*, *149*, 1–7.
- Sigala, N., & Logothetis, N. K. (2002). Visual categorization shapes feature selectivity in the primate temporal cortex. *Nature*, *415*, 318–320.
- Sloutsky, V. M., Kloos, H., & Fisher, A. V. (2007). When looks are everything: Appearance similarity versus kind information in early induction. *Psychological Science*, *18*(2), 179–185.
- Smith, E. M., Ford, J. K., & Kozlowski, S. W. J. (1997). Building adaptive expertise: Implications for training design

- strategies. In M. A. Quinones & A. Ehrenstein (Eds.), *Training for a rapidly changing workplace* (pp. 89–118). Washington, DC: American Psychological Association.
- Speed, A. (2008). Computational modeling of analogy: Destined ever to only be metaphor? *Behavioral and Brain Sciences*, *31*(3), 397–398.
- Speed, A., Verzi, S., Benz, Z., Dixon, K. R., & Warrender, C. (in preparation). A computational model of analogy making linking primary sensory to prefrontal cortex.
- Spellman, B. A., & Holyoak, K. J. (1996). Pragmatics in analogical mapping. *Cognitive Psychology*, *31*, 307–346.
- Spellman, B. A., Holyoak, K. J., & Morrison, R. G. (2001). Analogical priming via semantic relations. *Memory and Cognition*, *29*(3), 383–393.
- Squire, L. R. (1994). Declarative and nondeclarative memory: Multiple brain systems supporting learning and memory. In D. L. Schacter & E. Tulving (Eds.), *Memory systems* (pp. 203–231). Cambridge, MA: MIT Press.
- Stuss, D. T. (2007). New approaches to prefrontal lobe testing. In B. L. Miller & J. L. Cummings (Eds.), *The human frontal lobes: Functions and disorders* (2nd ed., pp.292–305). New York: Guilford.
- Sylvester, C. Y., & Shimamura, A. P. (2002). Evidence for intact semantic representations in patients with frontal lobe lesions. *Neuropsychology*, *16*(2), 197–207.
- Tamura, H., & Tanaka, K. (2001). Visual response properties of cells in the ventral and dorsal parts of the macaque inferotemporal cortex. *Cerebral Cortex*, *11*, 384–399.
- Thompson-Schill, S. L., Ramscar, M., & Chrysikou, E. G. (2009). Cognition without control: When a little frontal lobe goes a long way. *Current Directions in Psychological Science*, *18*, 259–263.
- Unterrainer, J. M., Owen, A. M. (2006). Planning and problem solving: From neuropsychology to functional neuroimaging. *Journal of Physiology – Paris*, *99*(4–6), 308–317.
- Vicario, C. M., Pecoraro, P., Turriziani, P., Kock, G., Caltagirone, C., & Oliveri, M. (2008). Relativistic compression and extension of experiential time in the left and right space. *PLoS ONE*, *5*(3), e1716.
- Viskontas, I., Morrison, R. G., Holyoak, K. J., Hummel, J. E., & Knowlton, B. J. (2004). Relational integration, inhibition and analogical reasoning in older adults. *Psychology and Aging*, *19*, 581–591.
- Wallis, J. D., Anderson, K. C., & Miller, E. K. (2001). Single neurons in prefrontal cortex encode abstract rules. *Nature*, *411*, 953–956.
- Walsh, V. (2003). A theory of magnitude: Common cortical metrics of time, space and quantity. *Trends in Cognitive Sciences*, *7*, 483–488.
- Waltz, J. A., Knowlton, B. J., Holyoak, K. J., Boone, K. B., Mishkin, F. S., de Menezes Santos, M., et al. (1999). A system for relational reasoning in human prefrontal cortex. *Psychological Science*, *10*(2), 119–125.
- Wartenburger, I., Heekeren, H. R., Preusse, F., Kramer, J., & van der Meer, E. (2009). Cerebral correlates of analogical processing and their modulation by training. *NeuroImage*, *48*, 291–302.
- Wendelken, C., Nakhabenko, D., Donohue, S. E., Carter, C. S., & Bunge, S. A. (2007). “Brain is to thought as stomach is to ?”: Investigating the role of rostralateral prefrontal cortex in relational reasoning. *Journal of Cognitive Neuroscience*, *20*(4), 682–693.
- Wharton, C. M., Grafman, J., Flitman, S. S., Hansen, E. K., Bauner, J., Marks, A., et al. (2000). Toward neuroanatomical models of analogy: A positron emission tomography study of analogical mapping. *Cognitive Psychology*, *40*, 173–197.
- Wood, J. N., & Grafman, J. (2003). Human prefrontal cortex: Processing and representational perspectives. *Nature Reviews Neuroscience*, *4*, 139–147.
- Wright, S. B., Matlen, B. J., Baym, C. L., Ferrer, E., & Bunge, S. A. (2008). Neural correlates of fluid reasoning in children and adults. *Frontiers in Human Neuroscience*, *1*, 1–8.
- Yeats, R. M., & Yeats, M. F. (2007). Business change process, creativity and the brain: A practitioner’s reflective account with suggestions for future research. *Annals of the New York Academy of Sciences*, *1118*, 109–121.