The neural bases of intelligence: a perspective based on functional neuroimaging

Sharlene D. Newman & Marcel Adam Just
Center for Cognitive Brain Imaging, Carnegie Mellon University, Department of Psychology, Pittsburgh, PA 15213

Address correspondence to:
Marcel Adam Just
D.O. Hebb Professor of Psychology
Director, Center for Cognitive Brain Imaging
Department of Psychology
Carnegie Mellon University
Baker Hall 327
5000 Forbes Ave.
(412) 268 –2791(ph)
(412) 268 – 2804 (fax)
Abstract

We present a theory of neural processing that is derived from the use of functional magnetic resonance imaging (fMRI). According to the neural principles described here, fluid intelligence arises from a highly adaptive, flexible neural system. The theory is composed of a set of operating principles for cortical computation. Together these principles suggest that intelligence arise from a dynamically configured set of brain areas that collaborate adaptively to meet a cognitive challenge.
Introduction

The study of intelligence has provided two major and enduring contributions to the understanding of human thought: a comprehensive characterization of human intelligence and a method to measure the variation in intelligence among individuals. These contributions have been based almost exclusively on behavioral measures of intelligence, using primarily paper and pencil tests. The development of brain imaging technology at the end of the twentieth century provided the ability to measure brain activity in individuals during the performance of tasks like those that compose intelligence tests. These brain imaging measures have the potential of providing a new and possibly more comprehensive view of intelligence and the basis of individual differences. In this chapter, we sketch the very beginnings of this new approach to intelligence that may provide a new comprehensive characterization of intelligence enriched by insights from recent brain imaging findings. This new approach may also provide suggestions of new methods to measure individual differences.

Intelligence is very difficult to define, and in fact, there is no consensus among scientific researchers as to what is meant by intelligence (Jensen, 1998). A general definition provided by Sternberg and Salter (1982) that we will use is “goal-directed adaptive behavior.” Intelligent behavior is adaptive in that it changes to confront and meet successfully with challenges. Because it is not enough for intelligent behavior to simply be adaptive, it is also thought to be goal-directed, or purposeful. However, it is the adaptive nature of intelligence that will be the primary focus of this chapter.

Spearman situated g at the apex of a hierarchy of abilities. g represents an individual’s general problem-solving skill, accounts for a person performing well on a variety of cognitive tasks, and is sometimes referred to as fluid intelligence. According to Spearman, one of the
factors that determine g is “mental energy”. Mental energy “enters into the measurement of ability of all kinds, and which is throughout constant for any individual, although varying greatly for different individuals” (1927, p. 411; Jensen, 1998). Because very little was known about brain function in the 1920’s, Spearman was unable to elaborate further as to what corresponded to “mental energy.” However, our proposal below implicitly includes an energy facet.

The conventional psychometric study of behavioral performance has been accompanied by attempts to correlate individual differences in intelligence with biological measures. In some sense these attempts have been in search of a definition of “mental energy”. For example, for over a hundred years researchers have been examining the correlation between head circumference (a proxy for brain size) and intelligence measures, generally suggesting that the larger the brain, the more intelligent the individual. Although many studies have found a modest correlation, these studies have been quite controversial [for a review see Van Valen (1974); Jensen & Sinha (1992)] and have not provided insights into either the nature of intelligence nor the measurement of individual differences.

In the 1980’s, Jensen hypothesized that it was not necessarily the size of the brain but the speed of processing that was central to intelligence, showing a relationship between reaction time and intelligence (Brody, p. 56). This relationship suggested that the characteristics of the nervous system determine reaction time, and that individuals whose nervous systems function more effectively and rapidly develop more complex intellectual skills. Electrophysiological recordings (ERP) of electrical activity measured on the scalp have also shown a relationship between neural processing characteristics and intelligence. Studies using ERP have revealed consistent correlations with intelligence (Jensen, 1998; Jensen & Sinha, 1992) and have been used to measure individual differences both in the normal population (McGarry-Roberts et al.,
1992; King & Kutas, 1995; Vos & Friederici, 2003) and those with psychiatric and neurological dysfunctions (John, 1994). The electrophysiological approach attempts to relate the electrical activity of the brain to the ongoing cognitive information processing. For example, this approach has found that individuals who are extreme in their ability (e.g., good vs. poor comprehenders; King & Kutas, 1995) have distinguishable electrical signatures visual sentence during a comprehension task. This approach has been successful in showing that there are electrophysiological differences that correlate with individual differences in performance, but the electrical measures are indirect and are not related to a comprehensive theory of intelligence.

In this chapter we present a theory of neural processing that is derived from the use of functional neuroimaging, particularly functional MRI (fMRI). Magnetic resonance imaging, primarily fMRI and possibly diffusion tensor imaging (DTI) in the future, has the potential to provide a clearer characterization of the neural bases of intelligence. A key contribution of fMRI is its ability to provide information about several important properties of the large-scale neural networks that underlie cognition. These properties include the specification of the set of brain regions that are involved in a given task; the temporal profile of the activation, or a reflection of the neural processing time course; and the degree of synchronization between pairs of activated regions which reflects the functional connectivity between regions.

The theory presented in this chapter is provides an initial account for g, or fluid intelligence. Intelligence is born out of networks of cortical areas and therefore, the investigation of the behavior of these large-scale cortical networks that may lead to an explanation of individual differences in ability. The major proposal of this chapter is that how well the neural system can adapt to changes in the environment will affect the quality and efficiency of its processing, thereby constituting a major source of individual differences. The theory is
composed of a set of operating principles for cortical computation put forth by Just and Varma (2003). These principles include:

1. Energy is consumed during the performance of cognitive tasks and each cortical area has a limited resource capacity. This principle has direct implications for individual differences in intelligence. First it suggests that the amount of resources available, or the resource capacity within the neural system may vary across individuals. Secondly, it may be that the amount of resources required to perform a task may differ across individuals due to variations in efficiency.

2. The topology (cortical composition) of neurocognitive networks associated with a given task changes dynamically, adapting itself to the demands of a given task. Therefore, the efficiency with which this topological change occurs may contribute to individual differences in task performance.

3. Cortical regions function collaboratively to perform tasks. Variation in the degree of synchronization or efficiency of the communication between regions may contribute to individual differences in task performance.

4. The quality of the white matter tracts connecting cortical areas may also affect processing speed. The variation in the degree or quality of the anatomical connections between processing regions may contribute to individual differences in task performance.

The principles outlined above suggest possible sources of individual differences in intelligence. The remainder of this chapter further explores these properties and provides citations of supporting experimental data.
Processing capacity

Thinking is biological work that requires resources, and is thus constrained by their availability. In any biological system, there is an upper limit on resource availability. Certainly there are upper bounds on thinking, such that one can do only so much thinking per unit time. It turns out to be helpful to consider such limitations as resource availability. Tasks that attempt to impose a load greater than the maximum that the resources permit will produce performance that is errorful, slow, or does not meet some task requirement, deteriorations consistent with decreased resources. This phenomenon is evident in the differences in cognition observed as a function of individual differences in working memory capacity (Just & Carpenter, 1992).

Recent neuroimaging research has provided extensive support for the resource consumption perspective. The amount of cortical activation within a given region increases with the computational demands that are placed on the region, as demonstrated in several types of cognitive tasks, including sentence comprehension (Just et al., 1996; Keller et al., 2001; Röder et al., 2002), working memory (Braver et al., 1997; Rypma et al., 1999), and mental rotation tasks (Carpenter et al., 1999; Just et al., 2001). For example, in language comprehension, the volume of fMRI-measured cortical activation in both Broca’s area and Wernicke’s area has been shown to increase with linguistic complexity of the sentence that is being comprehended (Just et al., 1996). These findings indicate that as a task places additional computational demands on a cortical region, it consumes more resources, eliciting greater fMRI-measured activation.

One of the implications of the resource consumption approach is that individuals may differ in resource availability and/or their efficiency. In other words, those with above-average performance may have either a greater computational capacity, use the available resources more efficiently, or both. There is evidence that lends support to the efficiency hypothesis: several
PET studies have reported negative correlations between psychometrically-measured abilities and the volume of cortical activation produced by tasks that draw upon these abilities (Just et al., 2003; Haier et al., 1988; Parks et al., 1988, 1989). Reichle, Carpenter and Just (2000) conducted a fMRI study that tested this hypothesis. The study examined the relation between individual differences in cognitive ability (verbal or spatial ability) and the amount of cortical activation engendered by two strategies (linguistic vs. visual-spatial) in a sentence-picture verification task. The study showed that the fMRI-measured activation was correlated with behaviorally assessed cognitive abilities in the two processing domains. The direction of the correlation is consistent with the idea that higher ability individuals use their resources more efficiently: higher ability individuals showed less fMRI-measured activation than did less proficient individuals. Specifically, individuals with better verbal proficiency (as measured by the reading span test) had less activation in Broca’s area when they used the verbal strategy, while individuals with better visual-spatial proficiency (as measured by the Vandenberg, 1971, mental rotation test) had less activation in the left parietal cortex when they used the visual-spatial strategy (see Figure 2).

To summarize, the two studies above show the adaptation of the brain to the magnitude of the computational load. The first shows that the amount of resources consumed, as measured by fMRI, increases as a function of task demand. The second suggests that highly proficient individuals use their resources more efficiently than do less proficient individuals. Together, these two results show the adaptiveness of resource consumption and that the consumption rate is influenced by individual differences in ability.

**Malleability of processing networks**

Intelligent responding at the cortical level must include the ability to arbitrarily map inputs and outputs (Garlick, 2002). At the large-scale cortical network level, this suggests that
the network of cortical areas activated in a given task – its composition and topological pattern of collaboration – is neither structurally fixed nor static. Rather, it varies dynamically during task performance. The previous conception of the neural basis of intelligence was that some fixed volume of brain tissue in a fixed set of brain areas (i.e. a fixed hardware infrastructure) is used to perform a particular task, like mental rotation or reasoning. According to the dynamic view we advocate, the “underlying hardware” is a moving target, changing not only from one type of stimulus item to another, but also changing from moment to moment during the processing of a given item.

There are at least two circumstances that may necessitate a dynamic change in the neural underpinnings of a cognitive task: 1) changes in the availability of cortical resources and 2) fluctuations in the computational demands of a task. As the resource pool of an area with a given set of specializations is exhausted, some overflow of its functions migrates from a more specialized area to less specialized areas. Although there is a typical set of areas activated in a given type of task, additional areas can become activated if the task is made significantly more demanding. For example, when a sentence comprehension task is made progressively more difficult by increasing the structural complexity of the sentences, activation in the right hemisphere homolog of Wernicke’s area (left posterior superior temporal gyrus) systematically increases from a negligible level to a substantial level (Just et al., 1996). One of the sources of individual differences in cognition may be flexibility with which additional regions are recruited.

The second situation that may necessitate dynamic self-assembly of a large-scale cortical network is a fluctuation in the computational demands of a given task. The dynamic assembly of neurocognitive networks is incremental or continuous, not all-or-none. This provides for just-in-time, as-needed, neural support for cognitive processing. This principle is demonstrated in a
study of verbal reasoning conducted by Newman et al. (2002). There, two conditions were
presented that varied the location of the maximal reasoning load within a sentence. In the first
(early/low load) condition, the reasoning load occurred early in the sentence; in the second
(late/high load) condition, the maximal reasoning load occurred late in the sentence (see Table 1). The time of occurrence of the maximal activation of prefrontal cortex varied as a function of
the location of the maximal reasoning load in the expected direction (see Figure 3). This
difference in the time course of activation supports the idea that cortical regions are recruited as
needed. The ability to dynamically recruit additional resources may very well be a source of
individual differences.

Dynamic self-assembly may be the physiological manifestation of the adaptive nature of
thought. When a task becomes too difficult for the current strategy, a new one is “devised”. The
ability to switch strategies and dynamically change the cortical landscape related to a given task
may contribute to individual differences. In fact, Garlick (2002) showed that an artificial neural
network which was better able to adapt its connections to the environment learned to read faster,
accommodated information from the environment better, and scored higher on fluid intelligence
tests. Each of these properties are characteristic of people with higher g.

**Functional connectivity**

A number of cortical regions are involved in performing any cognitive task. These
regions must be coordinated, possibly by passing information back and forth. Evidence of such
intercommunication pathways between cortical areas in humans performing a cognitive task
comes from two sources. The first is the existence of anatomical pathways between areas
(discussed in the next section). The corpus callosum is a prime example of an anatomical
pathway between potentially collaborating cortical areas. In addition, many other cortico-cortico
pathways are known from primate neuroanatomical studies (see Mesulam, 2000) as well as from more recent diffusion tensor imaging studies of white matter tracts in humans that are related to cognitive function (Klingberg et al., 2000). Furthermore, there are many additional anatomical links between cortical areas via subcortical regions, such as the thalamus.

The second source of evidence for coordination among the activated areas during cognitive activity is found in functional neuroimaging. The activation in a set of cortical areas is highly synchronized, indicating collaboration among areas. An increasingly used technique measures the correlation of the activation levels in two activated areas over some time period, and generally shows systematic synchronization between areas, modulated by a number of variables. The synchronization is taken as evidence of functional connectivity [or effective connectivity (Friston, 1994; Horwitz et al., 1998)]. Functional connectivity in the context of brain imaging refers to indirect evidence of communication or collaboration between various brain areas. The general assumption is that the functioning of voxels whose activation levels rise and fall together is coordinated.

A consistent finding is that more demanding conditions tend to produce higher functional connectivity than qualitatively similar but less demanding conditions (Diwadkar et al., 2000; Hampson et al., 2002). For example, in the domain of language there is a demonstrable functional connectivity between Broca’s and Wernicke’s areas both when participants are listening to texts and when they are at rest, the connectivity is substantially higher when they are listening to texts (Hampson et al., 2002). Another example, of this increased functional connectivity with increased demand was observed when an object recognition task is made more demanding by deleting more of the object contour (Diwadkar et al., 2003). In this case, the
degree of synchronization between the inferior temporal (ventral) area and the parietal (dorsal) area increases with difficulty, as shown in Figure 4.

There is also evidence that functional connectivity increases with learning (Buchel et al., 1999). In that study fMRI was used to examine the neural basis of associative learning of visual objects and their locations. The study found an increase in the functional connectivity between cortical regions associated with spatial and object processing with learning in the task. In addition, it was shown that the time course of the changes in functional connectivity was closely correlated with the time course of the changes in behavioral performance. The functional connectivity became higher at those times when performance improved.

Both of these adaptations, that functional connectivity increases with task difficulty and that it increases with learning, provide support for the idea that the brain’s ability to adapt to the environment may be a key characteristic of intelligence. An unexplored hypothesis states that high-ability individuals have higher functional connectivity than lower ability individuals. However, like any correlation, functional connectivity does not indicate the causality underlying the measure. Nevertheless, this technique allows for the exploration of the level of coordination between cortical regions across individuals, which may provide further insights to the biological underpinnings of individual differences in task performance.

Anatomical connectivity

Recently, a novel magnetic resonance imaging technique (diffusion tensor imaging or DTI) has been developed that can potentially provide information regarding the microstructure of white matter in vivo (Basser et al., 1994. DTI has been used to examine anatomical connectivity, or the physical neuronal connections between regions. The anatomical connections between cortical regions are essential to inter-region communication. In fact, the quality of these
connections has been suggested to directly affect processing speed. For example, recent developmental research has shown that the neural changes that take place during the first two years of life include a dramatic increase in the number of synaptic connections and an increase in the thickness of the myelin sheath that envelopes nerve cell axons (Siegler, 1998; Anderson, 2000). These two changes are important because they both affect conduction speed which is thought to, in turn, affect processing speed. Combined with fMRI, information about white-matter tracts has the potential to reveal important information about neurocognitive networks, which may help to elucidate the neural basis of individual differences.

Given that DTI is such a new technique, there have been very few studies that have used it. One of the first studies, that of Klingberg and colleagues (2000), compared the white matter tracts within the temporo-parietal region of poor and normal readers. There, Klingberg et al. found significant group differences in the myelination of the white matter in both the left and right hemispheres. In addition, Klingberg et al. found a high positive correlation between the DTI measure of the left hemisphere and reading ability, as measured by the Word ID test (Woodcock, 1987). Their results show not only the importance of the temporo-parietal region in language processing, but also that differences in the white matter tracts contribute significantly to individual differences observed in reading. It will be interesting to learn from future DTI studies whether the properties of white matter tracts are related to individual differences in cognitive abilities or to conventional measures of intelligence. As this technique is further developed, it promises to shed further light onto the neurological basis of intelligence.

**Is intelligence localized in the brain?**

Both g and the frontal lobe have often been linked to executive functions such as control processing, strategy formulation, planning and monitoring the contents of working memory
Support for this idea has come from both behavioral studies of normal and patient populations (Duncan et al., 1995, 1996) and a recent neuroimaging study (Duncan et al., 2000). For example, in a recent neuroimaging study conducted by Duncan and colleagues (2000), an attempt was made to determine the cortical area that underpins g. In that study, two variables were manipulated, the g loading (low or high) and test type (verbal or spatial), (an example problem is shown in Figure 4). Duncan et al. found that in both the verbal and spatial conditions, the frontal cortex revealed greater activation for the high-g condition compared to the low-g condition, supporting the idea that g reflects functions of the frontal lobe. Further support for the importance of the frontal lobe in intelligence was found in a recent review of the neuroimaging literature. Frontal activation similar to that observed during the high-g condition was also elicited by such processing demands as novelty, response competition, working memory load, and perceptual difficulty (Duncan and Owen, 2000).

We do not dispute that the frontal lobes play an important role in problem-solving and intelligence, but suggest instead that the biological basis of intelligence extends beyond the frontal lobe. In fact, intact frontal functions are somewhat unrelated to intelligence, as measured by psychometric tests (Teuber, 1972). IQ scores are rarely affected by damage to that region. We have argued here that intelligence does not lie in any particular brain region, but is instead a function of a more distributed, dynamically configured set of areas. According to this theory, the commonality or generality of processing that g represents refers to the ability of the neural system to adapt and be flexible. More specifically, g may represent the neural system’s ability to adapt to dynamic changes in the quantity and quality of changing computation demands. A study conducted by Duncan et al. (1996) found that the frontal process most central to g was goal
neglect and goal activation. This finding is in agreement with our dynamic processing account, because in order to adapt to changes in strategy there must be efficient goal switching. Therefore, the theory presented in this chapter suggests that intelligence cannot be localized to any particular brain region. It arises, instead from the coordination and collaboration of several neural components.

**Summary**

This chapter has attempted to describe a possible biological basis for fluid intelligence, or g. According to the principles described here, fluid intelligence may be the product of an adaptive, flexible neural system. More specifically, fluid intelligence may represent the neural system’s ability to adapt to dynamic changes in a complex cognitive process.

The principles outlined here are not thought to be exhaustive but are meant to be a springboard from which exciting new studies and theories of individual differences can emerge. We now have the technological capability to explore the human brain in its active state with the use of fMRI and soon will be able to investigate the integrity of its white matter tracts *in vivo* with DTI. With the combination of these two imaging techniques as well as computational modeling like that conducted by Garlick (2002) we are able to ask and possibly answer several innovative questions regarding the neural basis of intelligence.
References


Duncan, H., Emslie, H., & Williams, P. (1996). Intelligence and the frontal lobe: the


Haier, R. J., Siegel, B. V., Neuchterlein, K. H., Hazlett, E., Wu, J. C., Paek, J., Browning, H. L.,
and attention studied with Positron Emission Tomography. *Intelligence, 12*, 199-217.

Hampson, M., Peterson, B. S., Skudlarski, P., Gatenby, J. C., & Gore, J. C. (2002). Detection of

gyrus in normal reading and dyslexia. *Proceedings of the National Academy of Sciences USA,*


<table>
<thead>
<tr>
<th>Early/Low load</th>
<th>Late/High load</th>
</tr>
</thead>
<tbody>
<tr>
<td>The first month after April</td>
<td>The day before my favorite day</td>
</tr>
<tr>
<td>is the month before my favorite month.</td>
<td>Is the first day after Monday.</td>
</tr>
<tr>
<td>June, July, Other</td>
<td>Thursday, Friday, Other</td>
</tr>
</tbody>
</table>
Figure Captions

Figure 1: (adapted from Reichle et al., 2000) The relation between visual-spatial skill (as measured by the Vandenberg, 1971, mental rotation task) and the volume of cortical activation generated in the left (Panel A) and right (Panel B) parietal regions of interest (ROIs), as a function of gender. The best-fitting regression lines indicate that visual-spatial skill was negatively correlated with activation volume in both the left ($r = -.74$) and right ($r = -.61$) hemispheres. Panels C and D show the relation between individual differences in verbal skill (as measured by the Daneman and Carpenter, 1980, reading span task) and the volume of cortical activation generated in the left (Panel A) and right (Panel B) inferior frontal regions of interest (ROIs), as a function of gender. As the best-fitting regression lines indicate, verbal skill was negatively correlated with activation volume in the left hemisphere ($r = -.49$), but not the right ($r = .16$).

Figure 2: (adapted from Newman et al., 2002) The blue curve depicts the time course observed in the left dorsolateral prefrontal cortex during the Early/Low load condition and the pink curve the late/high load condition. Box 2 encompasses images related to the first phrase of the problem (e.g., the first month after April), box 3 encompasses images related to the second phrase (e.g., is the month before my favorite month), and box 4 encompasses images related to the response phase. As shown, the early/low load condition engenders more activation during the early phase of the problem compared to the late/high load condition, while the late/high load condition induces more activation later in the problem. The delay in peak activation for the late/high load condition corresponds to the increased recruitment of DLPFC processing later in this problem type.
Figure 3: (from Diwadkar et al., 2003) Increase in functional connectivity with workload in an object recognition task, where workload is increased by deleting more of the object contour.

Figure 4: (from Duncan et al., 2000) Materials from the high-g and low-g spatial task. Display elements were four panels, each containing one or more shapes, symbols or drawings. One panel differed in some respect from the others. Compared to the low-g problems, extensive problem solving was necessary to identify the “different” panel in the high-g problems.
Figure 1

(A.)

(B.)

(C.)

(D.)

Number of Activated Voxels in Verbal Strategy

Number of Activated Voxels in Visual-Imagery Strategy

Male
Female

Reading Span

Vandenberg Test Scores

Male
Female
Figure 3

Easier

Harder

Parietal/Dorsal Activation
Inferior Temporal/Ventral Activation
Figure 4