Turtles all the way down? Psychometric approaches to the reduction problem
Kievit, R.A.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Chapter 7

**IRT modelling of neural activity during fluid reasoning: A dissociation between difficulty and ability throughout the cortex**

**Abstract**

Fluid intelligence is a general cognitive ability associated with problem solving in the absence of task-specific knowledge. It figures centrally in theories of general intelligence, and has been associated with a diverse range of psychosocial outcomes. Neuroscientific studies of fluid intelligence have focused both on fluid intelligence tasks of varying difficulty (a within-subject comparison) and individual differences in fluid intelligence ability (a between-subject comparison). These studies have yielded a variety of neural systems associated with fluid intelligence, including fronto-parietal systems, but have failed to consistently distinguish between these the intra-individual and the inter-individual domains. Here we apply psychometric (Rasch) modelling to model both the within and between subject dimensions simultaneously. We show that the two networks show partial overlap but also widespread differences. Individuals with higher fluid intelligence ability, as measured by the Cattell task, showed increased bilaterally symmetrical activity throughout the brain but most notably in frontoparietal systems. More difficult items were associated with smaller, more focal increases in activity in bilateral middle frontal gyri, bilateral frontal poles and bilateral superior frontal poles. We introduce the term 'neural ergodicity' to denote the subset of regions that show increased activity both on within- and between-subject dimensions, and discuss the central role of these regions. This study shows that psychometric techniques can successfully model person- and item parameters to answer substantive neuroscientific questions.

**This chapter is based on:**

Introduction

Fluid intelligence is the ability to think logically and solve novel problems in the absence of task-specific knowledge (Cattell, 1963; 1971; Horn & Cattell, 1966; Carroll, 1993). It is a central component of psychometric theories of intelligence (Carpenter, Just, & Schell, 1990; Carroll, 1993; Engle, Tuholski, Laughlin, & Conway, 1999) and closely related to core cognitive abilities including working memory (Kyllonen & Christal, 1990; Engle, Kane, & Tuholski, 1999), short term memory (Engle et al., 1999), processing speed (Fry & Hale, 1996), attention (Engle, 2002), general intelligence (Blair, 2006) and executive functions (Kane & Engle, 2002; Salthouse, Atkinson, & Berish, 2003).

Individual differences in fluid intelligence are related to range of socially relevant factors, such that people with higher fluid reasoning ability generally have better psychosocial outcomes (e.g. Huepe et al., 2012), improved socioeconomic factors including income and social mobility (Deary, 2012; Strenze, 2007), lower instances of psychopathology (Gale, Batty, Tynelius, Deary, & Rasmussen, 2010) and lower morbidity and mortality (Deary, Weiss, & Batty, 2010).

Cognitive neuroscience has contributed a variety of insights into the neural processes and properties associated with fluid reasoning. Previous research has examined neural differences associated with individual differences in (fluid) intelligence (Choi et al., 2008; Cole et al., 2012; Deary, Penke, & Johnson, 2010; Kievit et al., 2012), neural responses during fluid reasoning tasks of varying complexity (Gray, Chabris, & Braver, 2003; Duncan et al., 2000, Geake & Hansen, 2005; Kroger et al., 2002; Lee et al., 2006; Prabhakaran et al., 1997; Wendelken, O’Hare, Whitaker, Ferrer, & Bunge, 2011) and the effects of localized lesions on fluid reasoning performance (Duncan et al., 1995; Woolgar et al., 2010; Roca et al., 2010). Together, these findings have converged on a distributed parietal and frontal network associated with fluid reasoning (Jung & Haier, 2007; Duncan, 2010; Kane & Engle, 2002).

However, despite these advances, neuroscientific studies of fluid intelligence often inadvertently study two sources of variation simultaneously, namely differences between subjects (i.e., differences in ability) and differences within subjects (i.e., differences in the difficulty of a task under different experimental conditions). For instance, the Parieto-Frontal integration model proposed by Jung & Haier (2007) is a process model of reasoning behaviour (cf. Jung & Haier, 2007, p. 138). That is, it claims to describe the processes that happen within a subject during complex reasoning. However, it is largely based on neuroimaging studies concerning differences between individuals (e.g. p.140). Conversely, neuroimaging studies focusing on differential activity as a function of task complexity often neglect individual differences as a source of variation. For instance, participants in Prabhakaran and colleagues (1997) compared several different types of fluid reasoning tasks, but did so in 7 Stanford graduate students who solved ~88% of the items correctly. Similarly, Perfetti and colleagues (2009) compared a high and low IQ group, but the two groups did not differ in how many items they made correctly (p. 502). Other, better-controlled studies focused mostly on individual differences, with a binary variation in an attentional control task, but not on fluid intelligence (Gray, Chabris, & Braver, 2003). Studies that captured a parametric range of fluid intelligence task difficulty did not examine individual differences (Kroger et al., 2002).
This is problematic, as in psychology it is well known that these dimensions can, and do, behave independently. For instance, although response speed is positively correlated with accuracy across subjects (those who perform more quickly are, on average, more accurate, Jensen, 1998) there is an inverse relationship between speed and accuracy within subjects. This phenomenon, known as the speed-accuracy trade-off (Fitts, 1954; Mackay; 1982), illustrates the care we should take to decompose the underlying dimensions. Similarly, empirical work has shown that the personality dimensions that capture differences between people (e.g. the Big Five, Costa, & McCrae, 1992) do not capture variation within individuals over time (Hamaker, Dolan, & Molenaar, 2007). The inter- and intra-individual distinction has also been conflated in psychological research on general intelligence, where the between-individual factor of general intelligence (often called $g$) has mistakenly been described as a within-subject causal process variable (Kanazawa, 2010; but see Penke et al., 2011).

To summarize, studies so far have not yet successfully decomposed the two dimensions of difficulty and individual ability along two, well-defined, continuously varying dimensions. This leaves a fundamental ambiguity in what is meant, exactly, by the ‘neural substrate’ of fluid reasoning (e.g. Prabhakaran et al., 1997). Does this term refer to the question which neural systems are differentially recruited depending on the complexity of the task, or to which neural systems are differentially active between people of differing fluid reasoning ability? By not addressing the two dimensions of difficulty and ability simultaneously, studies that focus on either dimension implicitly treat the other dimension of variation as a source of noise. Failure to separate these sources of variance may partially explain the divergent results of many studies: as Norgate and Richardson (2007) critically note of a review on the neuroscience of intelligence; ‘What, for example, does consistently activated across 30% of the studies mean?’ (p.162).

In the present paper, we show how this problem can be solved. Applying an Item Response Theory (IRT) model, we decompose neural responses during a fluid reasoning task into an inter-individual dimension and an intra-individual dimension, and so separate neural systems that underlie individual differences in fluid intelligence from those that reflect differences in increasing task difficulty. We hypothesize that the neural networks that are differentially active within people with differing ability are not the same as neural networks that are differentially active within people across tasks of varying difficulty. These differences, if they exist, are crucial to our understanding of both the process of fluid reasoning and the individual differences construct of fluid intelligence. This provides novel insights into fluid intelligence, and offers a general methodological framework for separating task- and person-parameters in cognitive neuroscience.

**Materials and Methods**

*Participants:* 37 participants (19 female) with normal or corrected-to-normal vision participated and they received a financial reward for their participation. The participants were tested in accordance with the ethical guidelines of the American Psychological Association, and the study was approved by the University of Amsterdam Ethics Committee. Prior to analysis, three subjects were excluded because of excessive motion (N=1) or scanner malfunction.
(N=2), leaving 34 subjects (17 female). The final sample ranged in age from 18 to 30 (M=23.4, SD=2.8).

In the scanner, subjects performed a total of 72 Raven's matrix items, drawn from the Standard Progressive Matrices (36 items from sets C, D and E, Raven 1960) and Raven's Advanced Progressive Matrices items (Raven, Court, & Raven, 1996). Figure 7.1 shows an example of an easy item and a difficult item. The eight-option items were adapted for use in the scanner. Ravens matrices are considered good measurements of fluid reasoning ability (Carpenter, Just, & Schell, 1990) and figure centrally in psychometric analyses of general intelligence. The experiment was programmed using Presentation® software (Version 0.70, www.neurobs.com). Participants viewed the screen (61 cm x 36 cm) on which the stimuli were presented via a mirror mounted on the head coil. They had a four-button box in either hand to respond to the eight clearly marked answer options. Prior to the first scan subjects were able to practice pressing the buttons with visual feedback to ensure correct response mapping.

Figure 7.1. Example of an easy (left) and a hard (right) item. Edited from original for illustrative purposes.

Scanning procedure
Prior to the scanning session, subjects read instructions and performed 12 practice trials (not used in the study) to ensure they understood the task. After ensuring the instructions were clear, participants were placed in the scanner. Each block consisted of 12 Raven's matrices, interspersed by a 16 second inter-trial interval. The blocks were pseudo-randomized such that each of the six blocks contained 12 fixed items spanning the complete range of difficulty (from easy to difficult), but were randomized within each block. This ensured that subjects did not 'give up' because trials became increasingly complex across blocks.

Image acquisition and preprocessing
Imaging data were obtained at the University of Amsterdam Spinoza Centre for Functional Magnetic Resonance Imaging using a 3-T Philips scanner using a 8-channel head coil. During the presentation of the Raven tasks we recorded BOLD-MRI (GE-EPI, TR=2346 ms, TE=30 ms, FA=90°, transversal recording, FOV=200×2 mm, matrix size=80, 39 slices, slice thickness=3, slice gap=0.3, ascending acquisition). We also acquired a high-resolution anatomical recording (3DT1, TR=8.1 ms, TE=3.74, FA=8°, FOV=240×220×188 mm, voxel size=1 mm³) for normalization purposes. Foldable foam pads were used to minimize head motion. Data were analysed using FSL (FMrib’s Software Library, www.fmrib.ox.ac.uk/fs (Smith et al., 2004), MATLAB (Version 7.10.0, The Mathworks, Inc., Natick, MA, USA), and R (Team, 2013). Functional data
were analysed using FEAT (FMRI Expert Analysis Tool Version 5.98), in which we performed motion correction, slice time correction, spatial smoothing (5mm) and low pass filtering (100 s). We generated explanatory variables for each individual presented item of the Raven’s progressive matrices using the double gamma model of the hemodynamic response function. This yielded 12 explanatory variables (EV’s) per run. These EVs were subsequently combined using a model in which we specified both the mean activation level and the item difficulty for each item. This yielded an estimate, per subject, of the extent to which activity of voxels differed across items varying in item difficulty. At the between-subject level we specified a model in which the average activity of the covariate fit from the fixed effects pooling stage was entered and in which the ability of the individual subjects was included as a predictor. In this way we were able to estimate item difficulty independently of subject ability. Higher-level analysis was carried out using FLAME (FMRIB's Local Analysis of Mixed Effects) stage 1 and stage 2 with automatic outlier detection (Beckmann, 2003, Woolrich et al., 2004; Woolrich, 2008). Statistics where thresholded using cluster-based correction at $z=2.3$ and a corrected cluster significance threshold of 0.05 (Worsley, 2001).

**Results**

**Behavioural results**

Previous studies that examine the impact of changes in task demands have often divided stimuli into ‘easy’ and ‘hard’ (e.g. Kalbfleish, Van de Meter, & Zeffiro, 2007; Perfetti et al., 2009). Here, we model difficulty continuously, so as to better capture the complete parametric space of difficulty offered by the stimuli. To decompose the differential contributions of difficulty and ability in neural response, we fit a Rasch model to the response patterns. A Rasch model is one from a family of Item Response Theory models (IRT; Hambleton, Swaminathan, & Rogers, 1991). In the Rasch model, the difficulty of items is related to the ability of participants using a logistic function to predict the likelihood of making an item correctly. Variants of Rasch models are widely used in both fields of general abilities (e.g. educational testing; Bond & Fox, 2013) and specific skills (e.g. modelling chess ability; Van der Maas & Wagenmakers, 2005). In the Rasch model we model $M$ dichotomously scored items ($1$=correct, $0$=incorrect) for $N$ persons. Each item has a difficulty parameter $\beta$, and each person has an ability parameter $\theta$. The probability that person $j$ with ability $\theta$ makes item $i$ with difficulty $\beta$ correctly can be described by the logistic function shown in Figure 7.2.
Turtles all the way down?

We fit a Rasch model in R (Team, 2013) using the package ltm (Rizopoulos, 2006) and eRm (Mair & Hatzinger, 2007). We considered both null-responses (no response within the 30 second time limit) and incorrect responses as incorrect, giving each participant a potential range of 0 to 72 correct. The 34 participants made an average of 39.6 items correct (range: min=19, max=53, $SD=8.8$), and took an average of 16143 ms to respond to items (range: min=1199 ms, max=29990 ms, $SD=4240$ ms). To best estimate the ability parameter ($\theta$) of each participant given the sample size, we fixed the difficulty parameters ($\beta$) of the 72 items based on the Ravens standardization sample (Raven, Court, & Raven, 1996). The difficulty parameters of the items ranged from -3.59 to 4.8, capturing a wide range of difficulties.

We then fit the Rasch model with the prior constraints on the difficulty estimates of the items. The Andersen Likelihood-Ratio test (Andersen, 1973) indicated that the response pattern fit the Rasch model adequately: $\chi^2(48, N=34)=41.071, p=.75$. Figure 7.2 shows the Item Response Curves of the 72 items and a histogram of the distribution of the ability scores ($\theta$) of the participants.

**Figure 7.2** The 72 Raven’s matrices items represented as ranging from easy (green/left) to hard (red/right). Ability is modelled such that person parameter theta corresponds to the probability of person $j$ making item $i$ correctly. The difficulty of an item (beta) can be read off by looking up the position on the X-axis that corresponds to a probability of .5 of making that item correctly (example shown in blue).
To ensure that our sample of participants performed the test accurately, we correlated the betas estimated on the basis of our restricted sample with betas based on the standardization sample. Despite a relatively small sample size, the betas showed a high degree of convergence with published standards ($r(70)=.90, p<.0001$). Further analyses showed that more difficult items (with higher betas) were associated with slower response times (correlation of z-scored rt's within subjects: $r(2242)=.60, p<.001$), were less likely to be made correctly (point-biserial correlation: $r_{pb}(2242)=.59, p<.001$) and were more likely to be null-responses (point-biserial correlation: $r_{pb}(2242)=.27, p<.001$).

Together these behavioural analyses suggest that the behavioural manipulation of fluid intelligence was successful: Participants took longer to respond to more difficult items, were less likely to respond correctly and were more likely to fail to respond within the time limit. Moreover, the pattern of responses was well-described by a Rasch model. For all further neuroimaging analyses we use the estimates of difficulty ($\beta$, based on the standardization sample) and ability ($\theta$) to study the neural systems underlying differences in item difficulty and ability.

**Individual differences in neural recruitment during fluid reasoning**

First, we examined individual differences in neural responses during the fluid reasoning task. The ability estimates ($\theta$), one for every person, were entered as predictors in a FLAME random effects analysis. The results in Figure 7.3 show widespread, bilateral increased activity as a function of theta such that individuals with higher estimated ability showed greater activity. The regions of greater activity include bilateral inferior and superior parietal cortices, a large portion of the paracingulate gyrus, bilateral (but greater on the left) Brodmann areas 10, and bilateral middle frontal gyri. These regions have been associated with a wide variety of executive functioning tasks. The large parietal network, consisting of Brodmann areas 7, 39 and 40 (inferior and superior parts of the parietal cortex) have been reported in a variety of imaging studies associated with individual differences in intelligence (e.g. Choi et al., 2008; Lee et al., 2006). The regions of greater activity in individuals with higher ability show considerable overlap with what is known as the Multiple Demand System (MDS, Duncan, 2010; Duncan & Owen, 2000; Fedorenko, Duncan, & Kanwisher, 2013). The multiple demand system is a distributed set of regions throughout the cortex known to be differentially active in a wide range of complex tasks such as working memory, interference monitoring, and mathematical problem solving (Fedorenko, Duncan, & Kanwisher, 2013), both in humans and in single cell recordings in non-human primates (Kusunoki, Sigala, Gaffan, & Duncan, 2009). Lesions to regions within the MDS lead to disproportionate problems in tasks of executive functioning compared to other cognitive abilities (Duncan et al.; 1995; Woolgar et al., 2011).
We illustrate the partial overlap of the inter-individual differences in ability with the MDS in Figure 7.4. Three key bilateral regions show overlap between inter-individual differences in fluid intelligence and the MDS: bilateral inferior parietal sulci, bilateral middle frontal gyri and the paracingulate gyri. However, the Multiple Demand System is usually associated with increased activity as a function of difficulty, a contrast we will explore in the next section. The findings here also show partial overlap with the region-based descriptions of the Parieto-Frontal Integration Theory (P-FIT, Jung & Haier, 2007, page 150). More importantly, we will later argue that the reason the P-FIT only partially overlaps with the current contrast is because of the lack of separation between the two dimensions we here attempt to disentangle.

Although we find overlap with both these broad models, there are also several distinctions. For instance, we find no differences between individuals in cerebellar, early visual, or insular activity. Similarly, we find focal increases in activity in the left inferior temporal gyrus. Activity in this region is commonly associated with greater object and object feature processing (Wei et al., 2012) and conceptual object processing (Martin, 2007). This may reflect the
idiosyncrasies of the task, as Raven’s Matrices requires the decomposition of the objects into relevant constituent relevant visual features and inferring likely consecutive patterns (Carpenter et al., 1990). We will return to the overlap and distinctions between these various models in more detail later. First, we examine whether neural systems associated with an increase in difficulty are similar or distinct from those associated with individual differences in ability.

**Figure 7.4.** Three main regions of overlap between the Multiple Demand System and regions that show greater activity in people with higher fluid intelligence scores. Red show activity associated with higher fluid intelligence, green is a Multiple Demand mask based on Fedorenko et al., yellow shows regions of overlap: A) Bilateral Inferior Parietal lobules B) Bilateral middle/inferior frontal gyri C) Bilateral paracingulate gyri.

**Neural network differences in difficulty**

Next, we examined which neural systems showed increased activity as a function of difficulty. To do so, we took the beta estimates of the difficulty of the 72 items as represented in Figure 7.2, and used them to predict differential brain activity for each individual, controlling for individual differences in mean activity as described above. We included all items in the analysis, as the analyses of the reaction times and considerations from item response theory suggest that the cognitive processes that ultimately lead to incorrect answers are as much
part of fluid reasoning as the cognitive processes that lead to correct responses. Modelling only correct responses would neglect the neural systems recruited when individuals reach the upper limits of complexity. Figure 7.5 shows three clusters of differential activity as a function of increasing difficulty. These include the left frontal pole, Brodmann’s area 10 bordering on 11, converging with findings in a study that varied relational complexity and found that ‘a greater degree of anterior left prefrontal involvement was also apparent in the analysis of areas selectively activated by the two highest levels of complexity’ (Kroger et al., 2002, p. 482). The specific role of the left BA10 during tasks of high relational complexity was replicated by Bunge, Helskog and Wendelken (2009), who found that ‘left RLPFC is engaged only when participants must consider the higher-order relationship between two individual relations’ (p. 338). This region of the prefrontal cortex has been suggested to mature relatively slowly, with the selectivity activation of BA 10 as a function of task complexity appearing relatively late during development, around 15 to 18 years, with the strongest effect also in the left BA 10 (Wendelken et al., 2011). We will return to this finding in detail in the conjunction analysis below.

A large occipito-parietal cluster was formed by the right Lateral Occipital Complex (Brodmann area 18 bordering on 19) and the superior parietal sulcus bordering on precuneus (BA 7). These latter two regions, but not the frontal pole, show overlap (yellow) with parts of the MDS. The bilateral parietal lobes have previously been reported as varying parametrically in activity with an increase in complexity (Kroger et al., 2002). Notably, increased activity as a function of difficulty is not as broad and far less widespread than the differences between people, despite the items ranging from being made correctly by every person to not being made correctly by any participant, a pattern previously also found in Duncan and colleagues (2000). This suggests that the differences between people correspond to greater neural differences than differences in neural responses to stimuli of varying difficulty within people, despite the fact that in this sample, the range of difficulties is greater
than the range of abilities, both in terms of IRT parameters and in terms of floor or ceiling effects. In the next section we formally examine the (dis)similarities between the inter- and intraindividual dimensions.

**Focal conjunction of difficulty and ability: Bilateral frontal poles, parietal cortices and middle frontal gyri**

To examine where differential activity along both dimensions was similar we performed a conjunction analysis (Nichols, Brett, Andersson, Wager, & Poline, 2005) based on the statistical maps of the first two analyses. Figure 7.6 shows regions that were differentially active both as a function of difficulty and ability. These regions can be described broadly as three bilateral systems: The inferior and superior parietal cortices, the middle and superior frontal gyri and bilateral frontal poles, or BA 10. The middle and superior frontal gyri and the parietal cortices largely overlap with the multiple demand system and described in some detail above. An interesting point of overlap is bilateral activity profiles in the frontal poles, or Brodmann's area 10. Activity in BA 10 is far from specific to fluid reasoning ability – it has been shown to be differentially active in a wide range of cognitive and non-cognitive abilities ((e.g. emotion regulation, Gilbert et al. 2006). It has been described as one of the least well-understood areas of the cortex (Gilbert et al., 2006), and undergoes pronounced functional changes during development. Lesions in this region have been associated with specific fluid reasoning deficits (Barbey, Colom, Paul, & Grafman, 2013) and a wide range of strategic planning problems in both the presence (Duncan, Burgess, & Emslie 1999; Roca et al., 2010) and absence (Shallice & Burgess 1991) of measurable differences in intelligence. Based on a meta-analysis of 104 studies, Gilbert and colleagues (2006) show that subdivisions of BA 10 are active during a variety of higher cognitive functions including working memory, planning, reasoning, goal maintenance, prospective memory and episodic memory. A variety of functional accounts have been used to describe the processes dependent on prefrontal cortex, including the maintenance of goals and subgoals (Koechlin et al., 1999), integrating multiple cognitive operations in pursuit of a higher behavioural goals (Ramnani & Owen, 2004), and hierarchical information processing (Koechlin & Summerfield, 2007). The so-called *gateway hypothesis* of the rostral prefrontal cortex, which suggests that BA10 is essential to ‘attend to externally or internally generated complex representations’ (Burgess, Dumontheil, & Gilbert, 2007; Christoff et al. 2001). This aligns well with successful solving of Raven’s matrices, which requires the online generation, maintenance and hypothesis-driven inference of rules during prolonged (>20 seconds) periods of time. The increased complexity of high-difficulty items would place larger demands on the ability to simultaneously entertain the multiple self-generated rules seen in the stimulus. Similarly, people who perform better on the task may be either more able to represent multiple subgoals simultaneously, or more able at deducing the relevant subcomponents in the first place.
Conjunction of intra- and inter- individual dimensions: Evidence for neural ergodicity?

The results above suggest that, at least within the parametric range studied here, the three bilateral regions shown in Figure 7.6 display similar inter-individual differences in activity as intra-individual differences. This is important for substantive reasons, but also ties closely to a fundamental question in psychology, namely the relationship between inter- and intra-individual phenomena. These have been called the ‘two disciplines of scientific psychology’ (Cronbach, 1957), and despite calls for better integration and more focus on intra-individual processes (e.g. Molenaar & Campbell, 2009) the two domains are rarely integrated. This is potentially problematic because it is only in highly specific circumstances where we can infer intra-individual processes from inter-individual differences, namely when the process is ergodic (Molenaar, 2004). Ergodicity implies that the statistical characterization of within-subject variation (e.g. the mean and variance) is (asymptotically) identical to those at the level of the group. This is very unlikely for most psychological constructs. Although ergodicity is often framed within the context of (natural) variation over time, it can be equally useful to describe intra-individual differences in task demands.

The importance of the question of ergodicity for neuroscience is increasingly being realized. Recent work suggests that even within a relatively well-controlled network analysis, ergodicity is violated (i.e. differences in network connectivity within and between subjects do not converge, Medaglia, Ramanathan, Venkatesan, & Hillary, 2013). Similarly, a recent study examined the convergence between functional network connectivity (functional connections within individuals, between brain regions), structural network connectivity (covariation of structural brain properties between subjects) and maturational change (structural covariation between regions within subjects across developmental time) (Alexander-Bloch, Raznahan, Bullmore, & Giedd, 2013). They found that although the three dimensions (inter, intra and long-term intra-individual comparisons) showed partial overlap (correlations between network characteristics of ~.3), there were also considerable qualitative differences in network properties such as modularity, and quantitative differences between the three dimensions. In the context of neurocognitive ageing this question is essential, as the age range that can be tested cross-
sectionally is far greater than the range that can be tested longitudinally. Crucially then, mechanistic inferences from cross-sectional ageing studies rely on what is called the convergence assumption, namely the assumption that the cross-sectional age-related differences between people of varying ages accurately reflect the process of ageing within each individual, an assumption homologous to ergodicity. This assumption is crucial, as the most valuable inferences that can be drawn for studies are intra-individual: How can an individual affect his or her trajectory through life, or how may we therapeutically intervene to cause benefits within an individual?

Although some neuroimaging techniques have been developed that can test for ergodicity in neuroimaging for specific designs, e.g. in event-related connectivity (Gates, Molenaar, Hillary, & Slobounov, 2011), it is still a relatively neglected topic. Moreover, these techniques models compare global ergodicity, that is, is the model as a whole identical for inter and intra-individual comparisons. This is a very strict requirement, which will likely rarely hold. We here propose a more lenient, but conceptually useful, form of ergodicity for neuroimaging: Where in the cortex does intra-individual manipulation yield the same differential activity as that which characterizes inter-individual differences on that task (e.g. Lindenberger & Von Oertzen, 2006; Sliwinski, Hoffman, & Hofer, 2010)? We propose to call this local ergodicity: By formally testing where (if anywhere) the differential brain activity as a function of inter-individual differences is the same as the regions that show intra-individual differences (for the same construct), we can see if regions display such neural ergodicity. In our analysis, the parietal, frontal and prefrontal systems shown in Figure 7.6 display a localized form of neural ergodicity. In this restricted set of brain regions, parametric increases of activity associated with more difficult items are similar to the differences between people of varying ability. Of course, this is a relatively weak form of neural ergodicity, as it only captures a linear function of two parameters, and holds for certain, but not all, regions.

Nonetheless, neural ergodicity is a useful concept for a variety of reasons. For example, a violation of neural ergodicity may be informative in terms of the cognitive insight gained into the phenomenon. If the neural systems underlying both dimension are different, then this suggest that the causal and mechanistic processes underlying the two dimensions are likely to differ. Much like the relationship between speed and accuracy can be explained by qualitatively different mechanisms within and between people, violations of ergodicity in neuroimaging may lead to more refined neural mechanisms of the phenomena of interest.

Secondly, neural ergodicity may simply be a useful heuristic: As ergodicity is a priori unlikely to hold (cf. Molenaar, 2004), it forces us to shift the default assumption to the (inter- and intra) dimensions being different, unless shown otherwise. For this reason, neuroimaging results drawn from inter-individual studies should not be used to draw process-based inferences at the level of the individual, as the data will be agnostic about this dimension. But most importantly, as the convergence of inter- and intra-individual patterns is a non-trivial requirement, cases where it does apply (even in the weak form proposed above) they are likely to tell us something about the mechanics underlying the phenomenon of interest. For example, in neuroimaging of bipolar disorders, disrupted prefrontal/amygdalar connectivity relating to depressive episodes has been found in a variety of cross-sectional studies (Strakowski et al., 2004). A longitudinal analysis focused on the intra-individual
Turtles all the way down?

variation, showing that, in a small cohort of bipolar patients, the within-subject variation of depressive symptoms was associated with similar prefrontal-amygdalar disruptions (Cerullo et al., 2012). Crucially, a recent study used focal TMS to disrupt brain activity in the prefrontal cortex in a sample of treatment-resistant depressive patients. This intervention led a considerable decrease (50%) in depressive symptoms in the majority of patients. Crucially, the patients that responded to the treatment displayed increases in prefrontal/amygdalar connections in a direction more similar to controls, whereas in non-responders, the disrupted pattern in connectivity pattern was not improved (Martinot et al., 2011). Although these are preliminary findings, such results suggest the potential fruitfulness of studying the mechanistic role of regions or systems for which inter- and intra-individual variations converge (and therefore show local neural ergodicity).

Discussion

In this study, we use a parametric IRT model to show that greater ability in fluid intelligence is associated with broad, bilateral increases in activation of fronto-parietal regions, whereas increases in activity within individuals as a function of difficulty are associated with a more focal set of regions constrained to parietal and prefrontal regions. These patterns partially replicate previous models that were based on both dimensions simultaneously, illustrating the benefits of decomposing the two dimensions. In addition to these differences, a subset of three cortical systems, namely bilateral parietal, bilateral middle frontal and bilateral prefrontal gyri, showed increases both as a function of difficulty and as a function of increased ability. We propose to describe this convergence of intra- and inter individual neural responses as localized neural ergodicity.

We show that psychometric models in general and IRT models specifically can simultaneously model two dimensions commonly of interest to cognitive neuroscientists. Models such as the Rasch model applied here can decompose two dimensions simultaneously, better deal with measurement error, explicitly test the assumption that only one ability is being tested, and (in certain cases) a properly specified measurement model can increase statistical power (Van der Sluis et al., 2010). Although certain psychometric models such as structural equation models have been widely applied in neuroimaging (e.g. Büchel & Friston, 1997; McIntosh & Gonzalez-Lima, 1994; Kievit et al., 2012; Penke et al., 2010), IRT models remain relatively rarely used (but see Thomas et al., 2013). The application of IRT models fits with a recent trend towards using formal models to decompose cognitive processes in cognitive neuroscience (Forstmann et al.; 2011; Kievit et al., 2012; Kriegeskorte & Kievit, 2013).

Limitations

Although the current approach represents a step forward in modelling the con- and divergence of two psychologically relevant dimensions, we are aware that we implicitly assume homogeneity in several other potentially dimensions. For instance, our study focuses on an age range (18-30) within which fluid intelligence is relatively stable. This means that in a sample with a larger age range, there is possibility that the neural systems underlying individual differences in fluid intelligence will be distinct from individual differences (of the same magnitude) seen in our sample. Similarly, although we examine a gender-balanced sample of considerable ability differences, as in most cognitive neuroscientific studies, our subjects are predominantly WEIRD (Western, Educated, Industrialized, Rich and democratic, Henrich, Heine, & Norenzayan,
For these reasons, the generalization of our findings to other cultures and age ranges remains an open empirical question. Moreover, as for all studies, a larger number of subjects and/or trials is likely to increase the set of regions showing differential activity (Gonzalez-Castillo et al., 2012; Button et al., 2013). Nonetheless, the clear, bilateral patterns we found that replicate focal subsets of previous models suggest that our design was adequate to model the broad differences between the two dimensions.

Future research may extend these findings in various directions. One of those directions is the better integration of temporal cognitive dynamics. Recently developed psychometric models have combined intra-individual processes as an information-accumulation process with inter-individual differences in ability (Van der Maas, Molenaar, Maris, Kievit, & Borsboom, 2011). Such models could be fruitfully combined with similar integrative developments in neuroimaging methods that combine localizationist approaches with more time-sensitive methods, leading to (combined) M/EEG or fMRI/EEG.

Conclusion

Fluid intelligence is perhaps the most domain-general ability studied by psychologists and neuroscientists alike. Fluid intelligence can be studied as a general ability that certain people display, consistently, to a greater degree than others, and it can be seen as an abstract description of the process that occurs, to a greater or lesser degree, when problem of varying difficulty have to be solved in the absence of domain-specific knowledge. For a full understanding of fluid intelligence, careful study of both dimensions is necessary. The promise of mechanistic, neuroscientific understanding can only be fulfilled if both dimensions of variation are properly understood. The central role of fluid reasoning in human cognition means that the potential payoffs of a better (mechanistic) understanding are great. A better mechanistic understanding of fluid reasoning is essential for the promise of targeted behavioural or neurological training delaying deterioration of fluid reasoning during old age (e.g. Salthouse, 2009) or recovery of fluid reasoning associated difficulties after strokes or lesions (Barbey et al., 2013; Duncan et al., 1995, Woolgar et al., 2011; Roca et al., 2010) and may improve the accuracy of clinical assessment of cognitive faculties by means of standardized neuroimaging tests (Allen & Fong, 2008).

However, before we can reach a mechanistic level of understanding, we first need to separate two fundamentally different dimensions of psychological variation: The inter-individual domain and the intra-individual domain. Over 50 years ago, Cronbach referred to these domains as ‘the two disciplines of scientific psychology’ (Cronbach, 1957), and it has been questioned to which extent these two domains have been brought closer together since (Borsboom, Kievit, Cervone, & Hood, 2009). By using formal models to best capture inter- and intra-individual phenomena in neuroimaging studies, the two subdisciplines of cognitive neuroscience may more quickly, and more fruitfully, converge.