

Inter- and intra-individual differences in fluid reasoning show distinct cortical responses

Rogier A. Kievit^{1,2*}

H. Steven Scholte²

Lourens J. Waldorp²

Denny Borsboom²

Conflict of Interest: The authors declare no competing financial interests.

* Corresponding, submitting and communicating author (Rogier.Kievit@mrc-cbu.cam.ac.uk)

¹Medical Research Council, Cognition and Brain Sciences Unit, 15 Chaucer Rd, Cambridge CB2 7EF,
Rogier.Kievit@mrc-cbu.cam.ac.uk

²Department of Psychology, University of Amsterdam, 1012 WX Amsterdam, Netherlands

H.S.Scholte@uva.nl, L.J.Waldorp@uva.nl, D.Borsboom@uva.nl

24 ***Abstract***

25 Fluid intelligence is a general cognitive ability associated with problem solving in the absence of
 26 task-specific knowledge. Neuroscientific studies of fluid intelligence have studied both fluid
 27 intelligence tasks of varying difficulty and individual differences in fluid intelligence ability, but have
 28 failed to appropriately distinguish the two dimensions. Here we use task-based fMRI (N=34) to show
 29 that *within* and *between* subject dimensions show both partial overlap and widespread differences.
 30 Individuals with higher ability showed widespread increased activity including bilateral frontoparietal
 31 systems, whereas more difficult items were associated with more focal activity increases in middle
 32 frontal gyri, frontal poles and superior frontal poles. Finally, we show that when difficulty is equated
 33 across individuals, those with higher ability tend to show more fronto-parietal activity, whereas low
 34 fluid intelligence individuals tend to show greater activity in higher visual areas. The fMRI and
 35 behavioural data for our paper are freely available in online repositories.

36

37

38

39

Introduction

Fluid intelligence is the ability to think logically and solve novel problems in the absence of task-specific knowledge (Horn and Cattell, 1966). It is a central component of psychometric theories of intelligence (Carpenter et al., 1990; Carroll, 1993; Engle et al., 1999) and closely related to core cognitive abilities including working memory (Engle et al., 1999), processing speed (Fry and Hale, 1996), attention (Engle, 2002), general intelligence (Blair, 2006) and executive functions (Salthouse et al., 2003). Individuals with higher fluid reasoning ability generally have better psychosocial outcomes (Strenze, 2007; Deary, 2012), lower instances of psychopathology (Gale et al., 2010) and lower morbidity and mortality (Deary et al., 2011). Moreover, fluid intelligence often declines rapidly in old age (Kievit et al., 2014; Aichele et al., 2015) with adverse consequences for the ability to live and function independently (Salthouse et al., 2003; Tucker-Drob, 2011).

Cognitive neuroscience has contributed a variety of insights into the neural processes and properties associated with fluid reasoning, including individual differences in (fluid) intelligence (Choi et al., 2008; Deary et al., 2010; Cole et al., 2012; Kievit et al., 2012, 2014; Ritchie et al., 2015), neural responses during fluid reasoning tasks of varying complexity (e.g. Duncan, 2000; Gray et al., 2003; Lee et al., 2006) and the adverse effects of lesions (Duncan et al., 1995; Roca et al., 2010). Together, these findings converge on a distributed parietal and frontal network associated with fluid reasoning (Kane and Engle, 2002; Jung and Haier, 2007; Fedorenko et al., 2013).

However, studies of fluid intelligence often implicitly conflate two sources of variation: Differences *between* subjects (i.e., differences in ability) and differences *within* subjects (i.e., differences in neural responses under varying task difficulty) (see also (Cronbach, 1957; Chabris, 2007). For instance, the Parieto-Frontal integration model (Jung and Haier, 2007) is a *process model* of reasoning behaviour (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. This is problematic, as it is well known that these dimensions can, and do, behave independently (Hamaker et al., 2005; Penke et al., 2011; Kievit et al., 2013). 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, affecting the findings to an unknown degree. Understanding this distinction in detail is crucial to our understanding of both the process of fluid reasoning and individual differences in fluid reasoning ability.

In the present paper, we use Item Response Theory (IRT, Embretson and Reise, 2013) to decompose neural responses during a fluid reasoning task into an inter-individual dimension and an intra-individual dimension. An IRT model combines item difficulty estimates (intra-individual parameters) with ability estimates for each subject (an inter-individual parameter). Using this model, we can separate neural systems that underlie individual differences 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. Crucially, by taking into account both dimensions we can compare individual differences in a novel manner: By selecting a differing subset of items for each individual tailored to their ability level, we can compare individual differences in terms of neural activity patterns whilst keeping intra-individual differences in subjective difficulty constant. Doing so sheds new light on the controversial notion of neural efficiency, and illustrates the power of simultaneously modelling inter- and intra-individual differences in a GLM framework.

Results

Behavioural results

To decompose the differential contributions of difficulty and ability in neural response, we fit a Rasch model to the response patterns of a set of Raven's Matrices (see Figure 1 for an example, see materials and methods for more detail).

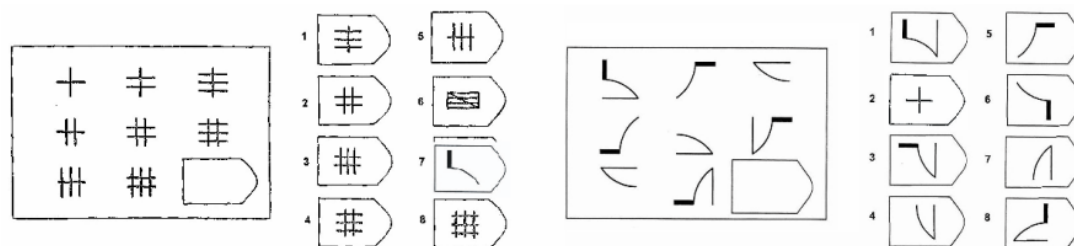


Figure 1. Stylized example an easy (left) and hard (right) Raven's Matrix fluid reasoning task.

A Rasch model is one from a family of Item Response Theory models (IRT; Hambleton et al., 1991; Embretson and Reise, 2013). Models such as the Rasch model 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 (e.g. Sluis et al., 2010). IRT models remain relatively rarely used in neuroimaging (but see Thomas et al., 2013). In the Rasch model, the difficulty of items is related to the ability of participants by means of a logistic function. The probability that person j with ability θ makes item i with difficulty β correctly can be described by equation (1).

$$1) P(X_{ji} = 1) = \frac{\varepsilon^{\theta_j - \beta_i}}{1 + \varepsilon^{\theta_j - \beta_i}}$$

Variants of Rasch models are widely used in fields such as educational testing (Bond and Fox, 2006) and more specific skills such as chess ability (van der Maas and 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 β , and each person has an ability parameter θ . We fit a Rasch model in

R (Team, 2014) using packages ltm (Rizopoulos, 2006) and eRm (Mair and 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, $SD=8.8$). The mean reaction time across individuals was 15.90s, $SD= 2.39$ s, with an item level RT ranging from 1.2 s to a maximum of 29.99 s. To best estimate the ability parameter (θ) of each participant, we fixed the difficulty parameters (β) 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. The Andersen Likelihood-Ratio test (Andersen, 1973) indicated that the response pattern fit the Rasch model adequately: $\chi^2(46, N=34)= 38.739, p=.767$. Figure 2 (top) shows the Item Response Curves of the 72 items and a histogram of the distribution of the ability scores (θ) of the participants (bottom).

To ensure that our sample of participants performed the test accurately, we next fit the model without the item-level constraints, to examine whether the difficulties estimated in our sample matched the difficulties based on the standardization sample. Despite a relatively small sample size, the betas showed a high degree of convergence with published standards ($r(70) = .85, p<.0001$). Further analyses showed that more difficult items (with higher betas) were associated with slower response times (Spearman's r of reaction times with correct response: $r(2446) = .55, p<.0001$), were less likely to be made correctly (Spearman's r difficulty with correct response= $r(2446) = -.59, p<.0001$) and were more likely to be null-responses (Spearman's r difficulty with null response = $r(2446) = .27, p<.0001$).

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 (β , based on the standardization sample) and ability (θ) to study the neural systems underlying differences in item difficulty and ability.

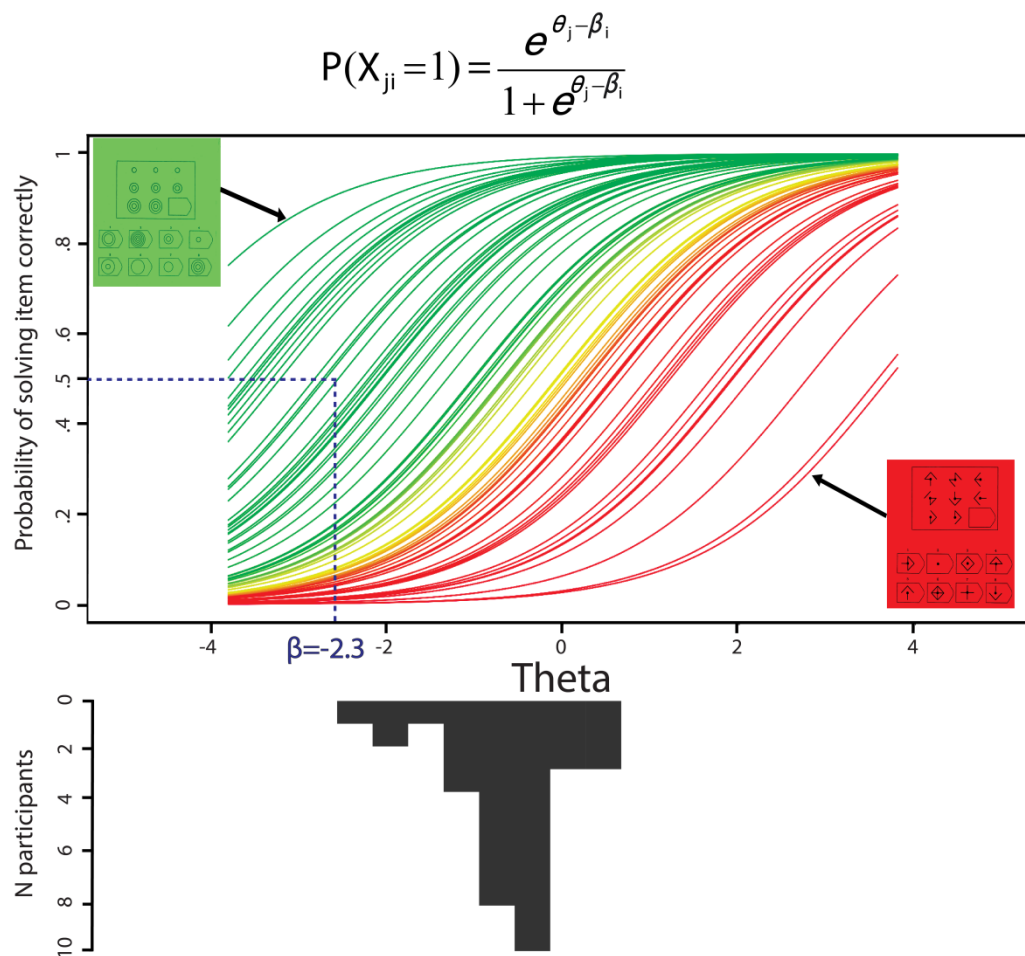
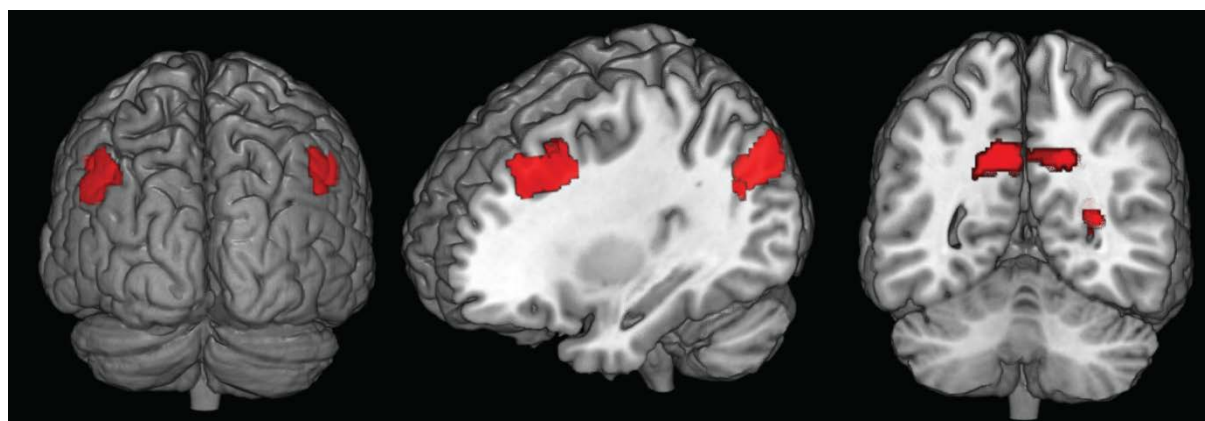


Figure 2: Item-characteristic curves for all 72 items (top). The 72 Raven's matrices items represented as ranging from easy (green/left) to hard (red/right). Ability is modelled such that person parameter theta corresponds to the probability of person j making item i correctly. The difficulty of an item (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).

Differences as a function of item difficulty

First, we examined which regions showed more activity for more difficult items when including all individuals and all difficulty levels. To do so, we take the beta estimates of the difficulty of the 72 items as shown in Figure 2, and used them to predict differential brain activity for each individual, controlling for individual differences in mean activity, with a FLAME random effects analysis. We include both correct and incorrect items, as 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. Figure 3 shows five clusters that show increased activity as a function of increasing difficulty. These clusters include the bilateral precunei, the bilateral superior parietal cortices, the superior frontal gyrus (L) and the precuneus bordering on the posterior cingulate. A large parietal cluster was formed by the right and left angular gyri in the parietal lobes (Brodmann areas 39), in line with work showing that activity in as varying parametrically in activity with an increase in complexity (Kroger et al., 2002). Together, this activity pattern is broadly in line with a broad, parieto-frontal network often associated with complex tasks (Jung and Haier, 2007; Duncan, 2010).



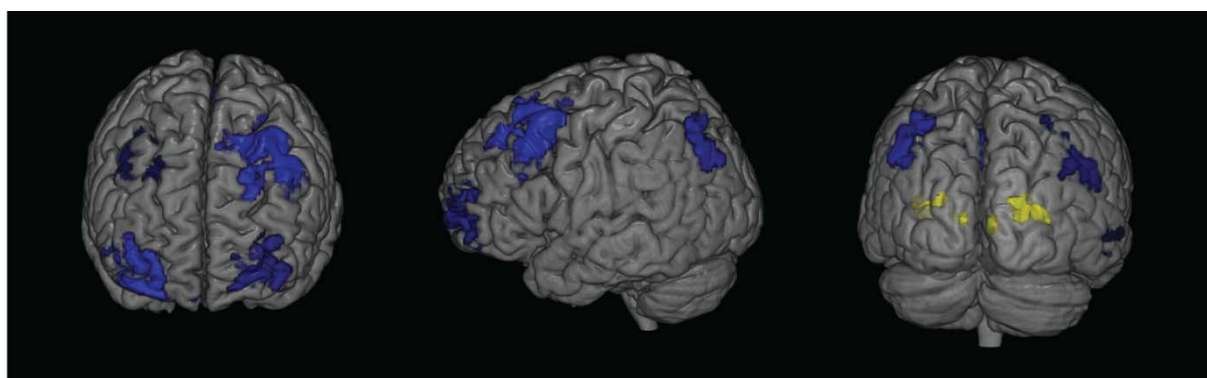
Difficulty Positive

Cluster Index	Voxels	P	Z-MAX X (mm)	Z-MAX Y (mm)	Z-MAX Z (mm)	Region
1	795	0.00171	26	-44	12	Precuneus, posterior cingulate
2	753	0.00251	-44	-68	34	Lateral occipital cortex, L
3	604	0.0103	-18	32	38	Superior frontal gyrus, L
4	561	0.0158	50	-68	46	Lateral occipital cortex, R
5	487	0.0336	-6	-56	34	Precuneus

Figure 3/Table 1: Spatial activity maps associated with an increase in item-difficulty (top), with peak activations of each of the five clusters shown below. No regions showed less activity for more difficult items.

Individual differences in neural networks during fluid reasoning

Second, we examined individual differences in neural responses during the fluid reasoning task. The ability estimates theta, one for each participant (shown in Figure 2, bottom), were entered into a FLAME random effects analysis to account for individual differences in ability. The results in Figure 4 show that individuals with higher estimated ability showed greater activity in a widespread, bilateral set of regions. 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. Lee et al., 2006; Choi et al., 2008). The regions of greater activity in individuals with higher ability show overlap with what is known as the Multiple Demand System (Duncan, 2010; Fedorenko et al., 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 et al., 2013) both in humans and in single cell recordings in non-human primates (Kusunoki et al., 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., 2010). Conversely, a set of three regions (shown in Table 2 and Figure 4, yellow) showed more activity in people with lower ability, most notably in bilateral lateral occipital cortex.



Ability Positive

Cluster Index	Voxels	P	Z-MAX X (mm)	Z-MAX Y (mm)	Z-MAX Z (mm)	Region
1	2349	1.98E-08	-46	26	46	Middle Frontal gyrus, L
2	2168	5.96E-08	28	60	-10	Frontal pole, R
3	1288	2.92E-05	12	-60	42	Precuneus, R
4	1106	0.000122	-36	-72	36	Lateral occipital cortex, L
5	1031	0.000225	-18	70	-4	Frontal pole, L
6	938	0.00049	40	-62	34	Lateral occipital cortex, R
7	864	0.000929	38	22	38	Middle Frontal gyrus, R
8	674	0.00523	56	-10	-24	Middle Temporal gyrus, R

Ability Negative

Cluster Index	Voxels	P	Z-MAX X (mm)	Z-MAX Y (mm)	Z-MAX Z (mm)	Region
1	1091	0.000138	-26	-74	8	Occipital Pole, L
2	454	0.0475	-64	4	4	Precentral Gyrus, L
3	451	0.049	-36	4	10	Central Opercular/Insular cortex L

Figure 4/ Table 2: Spatial activity maps showing clusters of greater activity in individuals of higher fluid reasoning ability (blue) and clusters with significantly less activity in individuals with higher ability (yellow). Peak activations of each of the clusters are shown in the table.

Conjunction and disjunction of intra- and inter- individual dimensions:

Next, we formally examine the (dis)similarities between the inter- and intra-individual dimensions. This is of interest for both methodological and intrinsic reasons. First, the extent to which these dimensions *differ* will illustrate the methodological perils of conflating the inter- and intra-individual dimensions. To examine to what extent these two dimensions differed, we examine where in the brain the activity associated with differences in item difficulty was greater than the activity

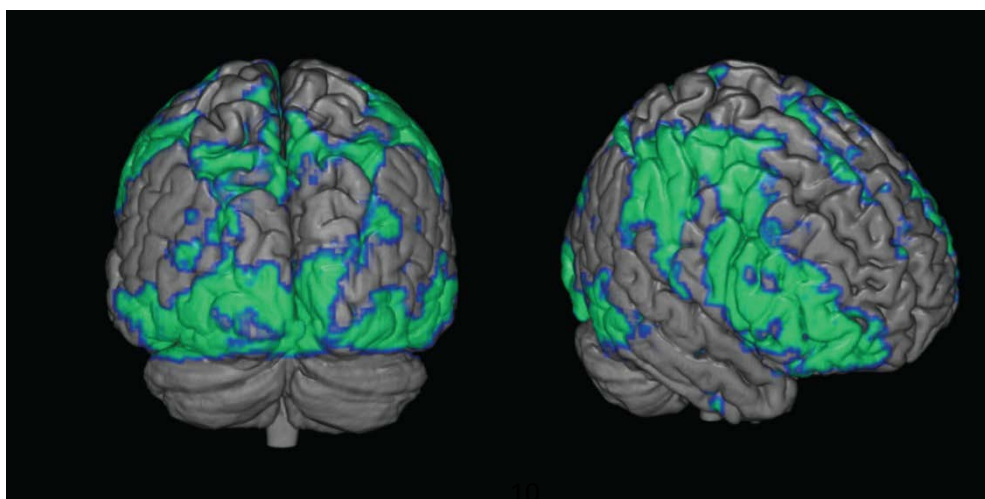


Figure 5: Regions showing greater activity as a function of individual differences in ability than as a function of difficulty. This contrast clearly illustrates the necessity of separating the two sources of variation.

associated with differences in ability, and vice versa. The former contrast yielded no significant regions of interest, suggesting there are no regions of the brain that are more active for more difficult trials that aren't also more active in people with higher ability. However, the converse contrast (where in the brain are individual differences in ability greater than differences due to item difficulty) showed widespread bilateral regions where the differences as a function of ability were greater than the differences associated with difficulty (Figure 5). These analyses suggest that neural patterns as a function of individual differences seem to be more spatially distributed than intra-individual differences as a function of task difficulty. This is the case despite the fact that the items ranged from extremely easy (made correctly by every person) to extremely hard (not being made correctly by any participant).

The second question we can ask is not where these two dimensions differ, but where they overlap. It has long been known, although often ignored, that only in highly specific circumstances can we infer intra-individual processes from inter-individual differences (or vice versa). This inference is only valid when a process is *ergodic* (Molenaar, 2004). Ergodicity implies that the statistical characterization of within-subject variation is (asymptotically) identical to the variation at the level of the group (Molenaar, 2004; Molenaar and Campbell, 2009), which is very unlikely for most psychological constructs (Kievit et al., 2013). Although ergodicity is usually framed within the context of (natural) variation over time, it can be equally useful to describe intra-individual differences in task complexity (in which case it is closely related to *inter-individual measurement invariance*, e.g. see Adolf et al., 2014). Although some neuroimaging techniques have been developed that can test for ergodicity in neuroimaging for specific designs, e.g. in event-related connectivity (Gates et al., 2011), it is still a relatively neglected topic. Generally, these techniques examine '*global*' ergodicity, that is, are the observed patterns as a whole identical for inter and intra-individual comparisons. We here propose a more lenient, but conceptually useful, form of ergodicity for neuroimaging: Where in the brain does intra-individual manipulation yield *the same differential activity* as that which characterizes

inter-individual differences on that task (e.g. Sliwinski et al., 2010; Raz and Lindenberger, 2011; Voelkle et al., 2014)? We refer to this pattern as *local ergodicity*.

Such an analysis is useful for a variety of reasons. First, it forces us to make explicit the distinction between intra- and inter-individual differences, an issue often neglected in cognitive neuroscience. Second, and more importantly, regions that are relevant in explaining both dimensions are more likely to be fundamental to the cognitive phenomenon of interest. As the convergence of inter- and intra-individual this patterns is a non-trivial requirement, cases where it does apply they are likely to tell us something about the mechanics underlying the phenomenon of interest. Conversely, the lack of overlap between the two domains may be informative in terms of the cognitive insight gained into the phenomenon. If the neural systems underlying inter- and intra-individual differences are clearly separable, then explanatory accounts of the two dimensions should differ accordingly. Much like the relationship between speed and accuracy can be explained by qualitatively different mechanisms within people (a negative relationship, due to a speed accuracy trade-off) and between people (a positive relationship, due to the positive manifold), violations of ergodicity in neuroimaging may lead to more refined proposed neural mechanisms that capture each dimension separately.

To test for neural ergodicity, we performed a conjunction analysis (Nichols et al., 2005) based on

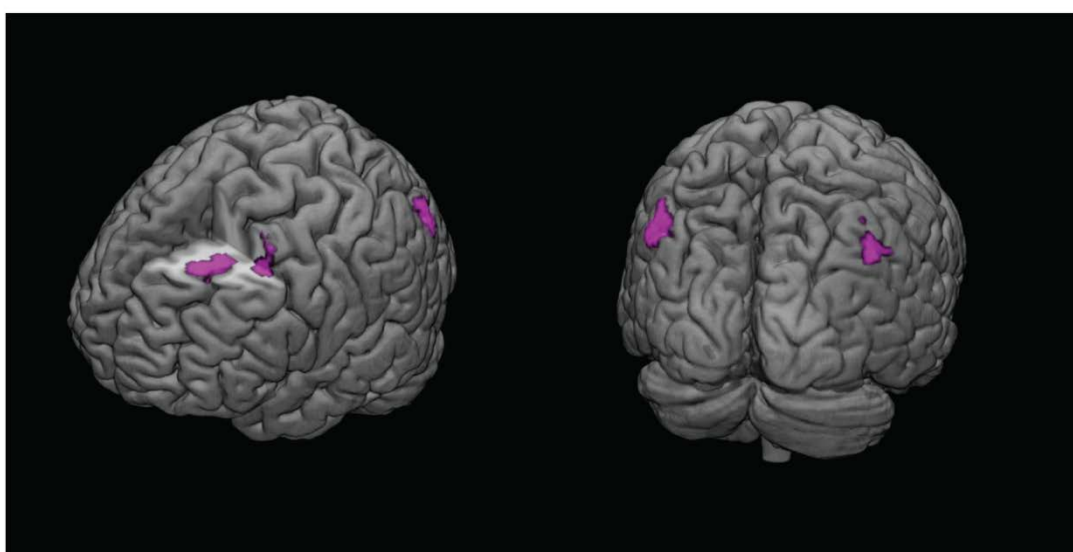


Figure 6: Locations that show evidence of neural ergodicity. These three clusters (bilateral angular gyri, bilateral precuneus and left superior frontal gyrus) showed increased activity both for more difficult items and for individuals of greater ability.

the statistical maps of the first two analyses (intra-individual differences in difficulty shown in Figure 3 and inter-individual differences in ability shown in Figure 4). Figure 6 shows the parietofrontal regions that are more active both as a function of increasing item difficulty and increased ability, and can be said to display *local neural ergodicity*. These regions can be described broadly as three clusters: The bilateral angular gyri in the superior parietal cortices, the bilateral precunei and the right middle and superior frontal gyri. Most notably the left lateral frontal region shows a striking convergence with a recent finding that greater connectivity of the LPFC predicts better fluid reasoning performance (Cole et al., 2012). In our final analysis we use IRT analysis to reexamine the ongoing debate on neural efficiency.

Correcting for individual differences: Neural (in)efficiency

Finally, we examine how we may use the parameters estimated in our IRT model to get a better grasp of possible mechanisms differentiating individuals. Appropriately taking into account both intra- and inter-individual variation is important for a particular, hotly contested, hypothesis, namely that of *neural efficiency* (Haier et al., 1988; Neubauer et al., 2002; Neubauer and Fink, 2009; Poldrack, 2014). In its most common form, neural efficiency is the claim that individuals of higher ability show less activity during cognitive tasks because they 'display a more focused cortical activation during cognitive performance resulting in lower total brain activation than in less intelligent individual' (Neubauer et al., 2002, p. 515). However, this concept has been challenged recently for being little more than a tautological redescription of the data (cf. Poldrack, 2014, p. 2) such that 'those of higher ability' finding the same task 'less hard work'. A more relevant question, we argue, is to compare individuals of different abilities when they are forced to 'work equally hard' (We note that Neubauer and Fink, 2009, do mention task complexity as a possible moderator of neural efficiency, e.g. pp. 1013). In other words, to meaningfully study differences in the processes that occur when individuals are being challenged cognitively (high difficulty), we must control for baseline differences in ability. Doing so, we can study the more relevant question of whether differences between individuals in fluid reasoning ability are associated with different cognitive patterns that would be suggestive of

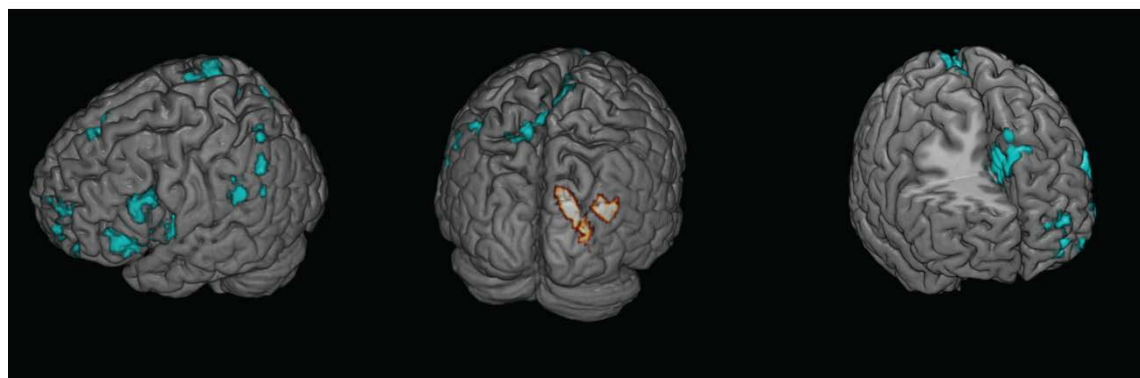
different cognitive strategies when performing items of equal difficulty. Item response theory allows for an easy way to separate difficulty and ability in precisely this manner.

For this analysis, we must ensure we select a subset of items that are equally difficult for all individuals. To do so we can simply subtract individual ability scores (θ) from the difficulty of each item (β), to get a *corrected difficulty score* for each item, for each person. Next, we select a subset of all 72 items for each individual, such that the range of items is equal in difficulty across all individuals. If after controlling for individual differences we find different patterns of activation, this would suggest that people who score more highly on Raven's matrices don't simply make more items correctly: They recruit different neural networks than those with lower ability, even when making items of equal (subjective) difficulty.

We selected a subset of 30 items for every individual such that the *mean corrected difficulty* (defined as the difficulty of the items minus the ability of the participants) of those items was equal for every individual. We compared these subsets across every pair of individuals and found no significant differences in corrected item difficulties (all p 's > 0.069). However, significance tests are poorly equipped to quantify the absence of effects (Wagenmakers, 2007), so we reran this comparison using default Bayesian t-tests (Morey and Rouder, 2013). This analysis showed no evidence for significant differences and considerable evidence for an absence of such differences (mean BF_{01} = 2.95, max BF_{01} = .92). Next, we repeated the analysis shown in Figure 3 by first; calculating per subject, a dummy contrast for the trials that were equated for difficulty (e.g. equally difficult given the capacity of the subject) and using these in a FLAME analysis in which we entered ability estimates (θ) to answer the follow question: Which regions (if any) are more active in individuals with higher ability, and which regions (if any) are more active in individuals of low ability, given that we have equated the subjective difficulty of the items included in the analysis? Figure 7 shows both results: In cyan, we can see five main clusters that are more active in individuals who have

greater fluid reasoning ability¹. These clusters include the Anterior Cingulate Cortex, a small posterior cluster bordering on the right supramarginal gyrus and a most pronounced frontal set of clusters in the orbitofrontal cortex and the dorsolateral prefrontal cortex, extending to the left frontal pole. Moreover, it shows that, contrary theories of neural efficiency, in our sample individuals with greater ability generally show *greater* frontal and prefrontal activity when performing items of the same subjective difficulty. Second, we examine whether there are any regions that are more active in individuals with lower fluid reasoning ability. This analysis yielded a single, left-lateralized cluster, shown in yellow/red, bordering the extrastriate cortex in the superior occipital lobe, on the intersection between Brodmann area 18 and 19.

¹ Note that due to a smaller subset of available items we did not model the intercept for the dummy variables at subject levels, so the t-values at the individual level are higher than the t-values in the analysis presented in Figure 2.



Ability positive Matched

Cluster Index	Voxels	P	Z-MAX X (mm)	Z-MAX Y (mm)	Z-MAX Z (mm)	Region
1	1866	1.19E-08	-30	50	2	Frontal pole, L
2	1735	5.96E-08	-54	-40	30	Supramarginal Gyrus, L
3	1108	9.89E-06	-10	22	38	Paracingulate Gyrus, L
4	377	0.0386	34	-38	24	White matter bordering Angular Gyrus, R
5	360	0.0488	10	-40	78	Postcentral Gyrus, R

Ability negative matched

Cluster Index	Voxels	P	Z-MAX X (mm)	Z-MAX Y (mm)	Z-MAX Z (mm)	Region
1	0.0104	476	40	-82	34	Lateral Occipital Complex, superior division, R

Figure 7/Table 3: Greater activity in people with higher fluid intelligence (cyan) and lower fluid intelligence (yellow/orange) once item difficulty has been equated across individuals. This pattern suggests that, beyond the unidimensionality of the behavioural variables, there exist individual differences in neural responses during fluid reasoning items. Using Neurosynth suggests these differences are compatible with a more memory- and rule maintenance based strategy for individuals with high fluid intelligence compared to a more visual, object-oriented strategy by individuals with lower ability.

This analysis shows there are differences in neural activity between individuals of higher and lower fluid reasoning ability even when they perform equally challenging tasks. Next we may ask what these differences reveal about the cognitive processes underlying these differences. Inferring mental states and styles from neuroimaging patterns is notoriously complex (Henson, 2005; Poldrack, 2006). However, we can use Neurosynth (<http://neurosynth.org/>, Yarkoni et al., 2011), an automated meta-analysis tool using automated analysis of neuroimaging data and keyword frequency across more

than 11,000 studies. This database makes it possible to provide a descriptive heuristic of cognitive states possibly associated with activity patterns, based on what has been reported in previous studies. Specifically, we can use the peak activations as shown above for both difficulty and ability to compute the posterior probability of a certain keyword being mentioned with high frequency (>1 in 1000) in any of the articles included in the database that also report activity in that cluster. Performing this analysis for the peak cluster above shows that the peak activation of the cluster that is more active in people with lower gf (X= 40mm, Y=-82mm, Z=34mm) has a posterior probability of .79 with the keyword 'objects'. This is in line, tentatively, with the hypothesis that individuals with lower fluid intelligence, on average, rely on more purely visual strategies when performing fluid reasoning tasks. In contrast, the peak cluster for individuals with higher fluid intelligence (X= -30mm, Y=52mm, Z=2mm) is instead associated with keywords such as 'memory' (posterior probability=.7), 'retrieval' (posterior probability=.74) and 'maintaining' (posterior probability=.83). This is compatible with the hypothesis that individuals with higher fluid intelligence rely more on more frontal and prefrontal regions associated with memory and rule based strategies to solve fluid reasoning items. However, specific paradigms (e.g. selective disruption of via TMS of frontal versus occipital regions) would be necessary to support this hypothesis, as Neurosynth in isolation can only provide part of the picture (cf. Yarkoni, 2015a, 2015b). If our inference above is correct, higher gf individuals would be more adversely affected by lateral frontal stimulation whereas lower gf individuals would be more affected by high visual disruption.

Together, these results suggest that when individuals who vary in ability have been matched to perform tasks that are equally difficult, there are noticeable differences in neural patterns, such that people of higher ability show relatively more left lateralized prefrontal activity, whereas individuals with lower ability show more right lateralized higher visual activity. These findings do not support the general neural efficiency hypothesis, instead suggesting more complex *qualitative* differences between high and low ability individuals. Our analysis shows how we can refine this question using psychometric techniques such that neuroimaging can reveal individual differences beyond a well-

fitting unidimensional of purely behavioural data. We agree with Poldrack (2014) that neural efficiency as it is often operationalized will rarely be the question we are interested in, and suggest that psychometric techniques are more commonly used to refine the question at hand.

Discussion

In this study we decompose two distinct but equally important dimensions of fluid intelligence: Intra-individual differences in neural responses to items of varying difficulty and inter-individual differences in neural activity for individuals of differing fluid reasoning ability. 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 including the angular gyri and the precunei. In addition to these differences, we find a subset of three cortical systems, namely bilateral parietal, bilateral middle frontal and bilateral prefrontal gyri, which show 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 *neural ergodicity*. The importance of the question of ergodicity for neuroscience is increasingly being realized, as 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 et al., 2011). Moreover, this subset of regions may reflect key processes emphasised by cognitive theories that link inter- and intra-individual processes in intelligence (e.g. Van Der Maas et al., 2006; Kovacs and Conway, 2016).

Crucially, we show that IRT can be used to equate item difficulty across a sample of individuals that differ in ability, thus making the neural comparisons more psychologically meaningful and comparable. Doing so suggests that high ability individuals show more frontal activity whereas individuals of lower ability show more occipital activity. These findings are compatible with, although certainly not conclusive of, strategy differences in reasoning style for individuals of higher and lower ability. Crucially, our approach can be extended to related cognitive domains that study both ability

and difficulty levels, such as working memory (e.g. Pessoa et al., 2002 versus (Osaka et al., 2003) cognitive control (e.g. Botvinick et al., 2001 and Hester et al., 2004) and many more.

Applying IRT in this context provides not only a cautionary note in conflating the two dimensions, but illustrates how neuroimaging data may allow us to better understand the mechanisms of individual differences: Although the fit of the Rasch model to the behavioural responses was good (suggesting unidimensionality of the behavioural responses), a more detailed investigation of the neural responses showed residual differences between individuals of high and low ability informative of the underlying cognitive processes. Practically, the dissociation of intra- and inter-individual dimensions suggest that many study designs may not be tailored to answer the question of interest in the most efficient, or even most accurate, way – Findings of individual differences will depend, in part, on the range of item difficulties presented, and findings of parametric difficulty will depend on the mean ability and range of the population being studied.

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. Future research may extend these findings by the better integration of the temporal dynamics of the cognitive processes underlying fluid reasoning. Recently developed psychometric models have combined intra-individual processes as an information-accumulation process with inter-individual differences in ability (van der Maas et al., 2011). Such models could be fruitfully combined with similar integrative developments in neuroimaging methods that combine the high spatial resolution of fMRI with more time-sensitive methods, leading to (combined) M/EEG or fMRI/EEG.

The central role of fluid reasoning in human cognition means that the potential payoffs of a better (mechanistic) understanding are considerable. 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 (Duncan et al., 1995; Woolgar et al., 2010; Barbey et al., 2014) and the promise of clinical assessment of cognitive faculties by means of standardized neuroimaging tests (Allen and Fong, 2008). However, before we can get close to a mechanistic level of understanding require to achieve these ambitious goals, careful study of two fundamentally different dimensions of psychological variation is crucial: The inter-individual domain and the intra-individual domain. Over 50 years ago, Cronbach (1957) referred to these domains as ‘the two disciplines of scientific psychology’ and it has been questioned to which extent these two domains have been brought closer together since (Borsboom et al., 2009; Voelkle et al., 2014). By using formal models to best capture inter- and intra-individual phenomena in neuroimaging studies, the two sub-disciplines of cognitive neuroscience may more quickly, and more fruitfully, converge.

Materials and Methods

Participants: 37 participants (19 female) with normal or corrected-to-normal vision participated. 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, and received a financial reward for their participation. 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 1 shows a stylized 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 and figure centrally in psychometric analyses of general intelligence (Carpenter et al., 1990). The experiment was programmed using Presentation® (Neurobs, 2011). Participants viewed the screen (61 cm x 36 cm) on which the stimuli were presented via a mirror mounted on the head coil. Participants had a four-button box in each 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.

All raw behavioural data and an analysis script written in R (Team, 2014) are freely available on Figshare (both behavioural data and basic demographics https://figshare.com/articles/Raw_datafiles/2077552 as well as the R code https://figshare.com/articles/Analysis_code_written_in_R/2077555). The neuroimaging first level data and key contrasts are freely available on Neurovault (Gorgolewski et al., 2015) at <http://neurovault.org/collections/1160/>.

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, with a maximum 30 second response window for each item. 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 within or across blocks.

Image acquisition and pre-processing

Imaging data were obtained at the University of Amsterdam Spinoza Centre for Functional Magnetic Resonance Imaging using a 3-T Philips Achieva TX scanner using an 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=200² mm, matrix size=80², 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/fsl, MATLAB (Version 7.10.0, The Mathworks, Inc., Natick, MA, USA), and R (Team, 2014). 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

443 voxels differed across items varying in item difficulty. At the between-subject level we specified a
 444 model in which the average activity of the covariate fit from the fixed effects pooling stage was
 445 entered and in which the ability of the individual subjects was included as a predictor, so we could the
 446 estimate the effects of item difficulty independently of subject ability. Higher-level analysis was
 447 carried out using FLAME (FMRIB's Local Analysis of Mixed Effects) stage 1 and stage 2 with automatic
 448 outlier detection (Woolrich et al., 2004; Woolrich, 2008). Statistics were thresholded using cluster-
 449 based correction at $z=2.3$ and a corrected cluster significance threshold of 0.05 (Worsley, 2001)
 450
 451

References

- Adolf J, Borkenau P, Borsboom D, Dolan CV (2014) Quantitative Psychology and Measurement Measurement invariance within and between subjects : A distinct problem in testing the equivalence of intra- and inter-individual model structures . Measurement invariance within and between individuals : A distinct.
- Aichele S, Rabbitt P, Ghisletta P (2015) Life Span Decrements in Fluid Intelligence and Processing Speed Predict Mortality Risk. Psychol Aging aop Available at: <http://psycnet.apa.orgpsycarticles/2015-27315-001> [Accessed July 9, 2015].
- Allen MD, Fong AK (2008) Clinical application of standardized cognitive assessment using fMRI. I. Matrix reasoning. Behav Neurol 20:127–140 Available at: <http://www.ncbi.nlm.nih.gov/pubmed/19641250> [Accessed September 3, 2015].
- Andersen EB (1973) A goodness of fit test for the rasch model. Psychometrika 38:123–140 Available at: <http://link.springer.com/10.1007/BF02291180> [Accessed June 15, 2014].
- Barbey AK, Colom R, Paul EJ, Grafman J (2014) Architecture of fluid intelligence and working memory revealed by lesion mapping. Brain Struct Funct 219:485–494 Available at: <http://www.ncbi.nlm.nih.gov/pubmed/23392844> [Accessed June 15, 2014].
- Blair C (2006) How similar are fluid cognition and general intelligence? A developmental neuroscience perspective on fluid cognition as an aspect of human cognitive ability. Behav Brain Sci 29:109–125; discussion 125–160 Available at: http://journals.cambridge.org/abstract_S0140525X06009034 [Accessed June 15, 2014].
- Bond TG, Fox CM (2006) Applying the Rasch Model: Fundamental Measurement in the Human Sciences, Second Edition. Available at: http://books.google.com/books?hl=nl&lr=&id=MRr_AQAAQBAJ&pgis=1 [Accessed June 15, 2014].
- Borsboom D, Kievit RA, Cervone D, Hood SB (2009) The Two Disciplines of Scientific Psychology, or: The Disunity of Psychology as a Working Hypothesis. In: Dynamic Process Methodology in the Social and Developmental Sciences, pp 67–97. New York: Springer-Verlag.
- Botvinick MM, Braver TS, Barch DM, Carter CS, Cohen JD (2001) Conflict monitoring and cognitive control. Psychol Rev 108:624–652 Available at: <http://psycnet.apa.orgjournals/rev/108/3/624> [Accessed February 10, 2016].
- Carpenter PA, Just MA, Shell P (1990) What one intelligence test measures: A theoretical account of the processing in the Raven Progressive Matrices Test. Psychol Rev 97:404–431.
- Carroll JB (1993) Human cognitive abilities: A survey of factor-analytic studies.
- Chabris CF (2007) Cognitive and neurobiological mechanisms of the Law of General Intelligence. In: Integrating the Mind: Domain General Versus Domain Specific Processes in Higher Cognition, pp 520. Psychology Press. Available at: <https://books.google.com/books?hl=en&lr=&id=4LI5AgAAQBAJ&pgis=1> [Accessed February 10, 2016].
- Choi YY, Shamosh NA, Cho SH, DeYoung CG, Lee MJ, Lee J-M, Kim SI, Cho Z-H, Kim K, Gray JR, Lee KH (2008) Multiple bases of human intelligence revealed by cortical thickness and neural activation. J Neurosci 28:10323–10329 Available at: <http://www.jneurosci.org/content/28/41/10323.short> [Accessed June 11, 2014].
- Cole MW, Yarkoni T, Repovs G, Anticevic A, Braver TS (2012) Global connectivity of prefrontal cortex

495 predicts cognitive control and intelligence. *J Neurosci* 32:8988–8999 Available at:
496 <http://www.jneurosci.org/content/32/26/8988.short> [Accessed May 23, 2014].

497 Cronbach LJ (1957) The two disciplines of scientific psychology. *Am Psychol* 12:671–684 Available at:
498 <http://psycnet.apa.orgjournals/amp/12/11/671> [Accessed June 18, 2014].

499 Deary IJ (2012) Intelligence. *Annu Rev Psychol* 63:453–482 Available at:
500 <http://www.annualreviews.org/doi/abs/10.1146/annurev-psych-120710-100353> [Accessed June
501 3, 2014].

502 Deary IJ, Penke L, Johnson W (2010) The neuroscience of human intelligence differences. *Nat Rev*
503 *Neurosci* 11:201–211 Available at: <http://dx.doi.org/10.1038/nrn2793> [Accessed May 30, 2014].

504 Deary IJ, Weiss A, Batty GD (2011) Intelligence and Personality as Predictors of Illness and Death: How
505 Researchers in Differential Psychology and Chronic Disease Epidemiology Are Collaborating to
506 Understand and Address Health Inequalities. *Psychol Sci Public Interes* 11:53–79 Available at:
507 <http://psi.sagepub.com/content/11/2/53.extract> [Accessed June 6, 2014].

508 Duncan J (2000) A Neural Basis for General Intelligence. *Science* (80-) 289:457–460 Available at:
509 <http://www.sciencemag.org/content/289/5478/457.short> [Accessed June 5, 2014].

510 Duncan J (2010) The multiple-demand (MD) system of the primate brain: mental programs for
511 intelligent behaviour. *Trends Cogn Sci* 14:172–179 Available at:
512 <http://www.sciencedirect.com/science/article/pii/S1364661310000057> [Accessed January 4,
513 2014].

514 Duncan J, Burgess PW, Emslie H (1995) Fluid intelligence after frontal lobe lesions. *Neuropsychologia*
515 33:261–268 Available at: <http://www.sciencedirect.com/science/article/pii/0028393294001248>
516 [Accessed January 13, 2014].

517 Embretson SE, Reise SP (2013) Item Response Theory. Psychology Press. Available at:
518 <https://books.google.com/books?hl=en&lr=&id=9Xm0AAAAQBAJ&pgis=1> [Accessed October 29,
519 2015].

520 Engle RW (2002) Working Memory Capacity as Executive Attention. *Curr Dir Psychol Sci* 11:19–23
521 Available at: <http://cdp.sagepub.com/content/11/1/19.short> [Accessed May 24, 2014].

522 Engle RW, Tuholski SW, Laughlin JE, Conway ARA (1999) Working memory, short-term memory, and
523 general fluid intelligence: A latent-variable approach. *J Exp Psychol Gen* 128:309–331 Available
524 at: <http://psycnet.apa.orgjournals/xge/128/3/309> [Accessed May 17, 2014].

525 Fedorenko E, Duncan J, Kanwisher N (2013) Broad domain generality in focal regions of frontal and
526 parietal cortex. *Proc Natl Acad Sci U S A* 110:16616–16621 Available at:
527 <http://www.pnas.org/content/early/2013/09/20/1315235110.abstract> [Accessed December 17,
528 2013].

529 Fry AF, Hale S (1996) Processing Speed, Working Memory and Fluid Intelligence: Evidence for a
530 Developmental Cascade. *Psychol Sci* 7:237–241 Available at:
531 <http://pss.sagepub.com/content/7/4/237.short> [Accessed May 14, 2014].

532 Gale CR, Batty GD, Tynelius P, Deary IJ, Rasmussen F (2010) Intelligence in early adulthood and
533 subsequent hospitalization for mental disorders. *Epidemiology* 21:70–77 Available at:
534 <http://www.ncbi.nlm.nih.gov/pubmed/19907333> [Accessed June 12, 2014].

535 Gates KM, Molenaar PCM, Hillary FG, Slobounov S (2011) Extended unified SEM approach for
536 modeling event-related fMRI data. *Neuroimage* 54:1151–1158 Available at:
537 <http://www.sciencedirect.com/science/article/pii/S105381191001133X> [Accessed June 15,
538 2014].

539 Gorgolewski KJ, Varoquaux G, Rivera G, Schwarz Y, Ghosh SS, Maumet C, Sochat V V, Nichols TE,
540 Poldrack RA, Poline J-B, Yarkoni T, Margulies DS (2015) NeuroVault.org: a web-based repository
541 for collecting and sharing unthresholded statistical maps of the human brain. *Front Neuroinform*
542 9:8 Available at:
543 [http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4392315&tool=pmcentrez&rendert](http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4392315&tool=pmcentrez&rendertype=abstract)
544 [ype=abstract](http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4392315&tool=pmcentrez&rendertype=abstract) [Accessed February 10, 2016].

545 Gray JR, Chabris CF, Braver TS (2003) Neural mechanisms of general fluid intelligence. *Nat Neurosci*
546 6:316–322 Available at: <http://dx.doi.org/10.1038/nn1014> [Accessed May 27, 2014].

547 Haier RJ, Siegel B V., Nuechterlein KH, Hazlett E, Wu JC, Paek J, Browning HL, Buchsbaum MS (1988)
548 Cortical glucose metabolic rate correlates of abstract reasoning and attention studied with
549 positron emission tomography. *Intelligence* 12:199–217 Available at:
550 <http://www.sciencedirect.com/science/article/pii/0160289688900165> [Accessed September 24,
551 2014].

552 Hamaker EL, Dolan C V., Molenaar PCM (2005) Statistical Modeling of the Individual: Rationale and
553 Application of Multivariate Stationary Time Series Analysis. *Multivariate Behav Res* 40:207–233
554 Available at: http://dx.doi.org/10.1207/s15327906mbr4002_3 [Accessed June 15, 2014].

555 Hambleton RK, Swaminathan H, Rogers HJ (1991) Fundamentals of item response theory. London:
556 SAGE.

557 Henson R (2005) What can functional neuroimaging tell the experimental psychologist? *Q J Exp*
558 *Psychol A* 58:193–233 Available at:
559 <http://www.tandfonline.com/doi/abs/10.1080/02724980443000502#.VeWoVPI3lpg> [Accessed
560 April 26, 2015].

561 Hester R, Fassbender C, Garavan H (2004) Individual differences in error processing: a review and
562 reanalysis of three event-related fMRI studies using the GO/NOGO task. *Cereb Cortex* 14:986–
563 994 Available at: <http://cercor.oxfordjournals.org/content/14/9/986.short> [Accessed January 17,
564 2016].

565 Horn JL, Cattell RB (1966) Refinement and test of the theory of fluid and crystallized general
566 intelligences. *J Educ Psychol* 57:253–270.

567 Jung RE, Haier RJ (2007) The Parieto-Frontal Integration Theory (P-FIT) of intelligence: converging
568 neuroimaging evidence. *Behav Brain Sci* 30:135–154; discussion 154–187 Available at:
569 http://journals.cambridge.org/abstract_S0140525X07001185 [Accessed January 21, 2014].

570 Kane MJ, Engle RW (2002) The role of prefrontal cortex in working-memory capacity, executive
571 attention, and general fluid intelligence: An individual-differences perspective. *Psychon Bull Rev*
572 9:637–671 Available at: <http://www.springerlink.com/index/10.3758/BF03196323> [Accessed
573 May 18, 2014].

574 Kievit RA, Davis SW, Mitchell D, Taylor JR, Duncan J, Cam-CAN, Henson RN (2014) Distinct aspects of
575 frontal lobe structure mediate age-related differences in fluid intelligence and multitasking. *Nat*
576 *Commun*.

577 Kievit RA, Frankenhuis WE, Waldorp LJ, Borsboom D (2013) Simpson’s paradox in psychological
578 science: a practical guide. *Front Psychol* 4:513 Available at:
579 <http://www.frontiersin.org/Journal/10.3389/fpsyg.2013.00513/abstract> [Accessed February 20,
580 2014].

581 Kievit RA, van Rooijen H, Wicherts JM, Waldorp LJ, Kan K-J, Scholte HS, Borsboom D (2012)
582 Intelligence and the brain: A model-based approach. *Cogn Neurosci* 3:89–97 Available at:
583 <http://dx.doi.org/10.1080/17588928.2011.628383> [Accessed January 4, 2014].

584 Kovacs K, Conway AR. (2016) Process Overlap Theory: A Unified Account of the General Factor of
585 Intelligence. *Psychol Inq* In press.

586 Kroger JK, Sabb FW, Fales CL, Bookheimer SY, Cohen MS, Holyoak KJ (2002) Recruitment of anterior
587 dorsolateral prefrontal cortex in human reasoning: A parametric study of relational complexity.
588 *Cereb Cortex* 12:477–485.

589 Kusunoki M, Sigala N, Gaffan D, Duncan J (2009) Detection of fixed and variable targets in the monkey
590 prefrontal cortex. *Cereb Cortex* 19:2522–2534 Available at:
591 <http://cercor.oxfordjournals.org/content/19/11/2522.short> [Accessed May 29, 2014].

592 Lee KH, Choi YY, Gray JR, Cho SH, Chae J-H, Lee S, Kim K (2006) Neural correlates of superior
593 intelligence: stronger recruitment of posterior parietal cortex. *Neuroimage* 29:578–586
594 Available at: <http://www.sciencedirect.com/science/article/pii/S1053811905005550> [Accessed
595 June 4, 2014].

596 Mair P, Hatzinger R (2007) Extended Rasch modeling: The eRm package for the application of IRT
597 models in R. *J Stat Softw* 20:1–20.

598 Medaglia JD, Ramanathan DM, Venkatesan UM, Hillary FG (2011) The challenge of non-ergodicity in
599 network neuroscience. Available at:
600 <http://informahealthcare.com/doi/abs/10.3109/09638237.2011.639604> [Accessed June 15,
601 2014].

602 Molenaar PCM (2004) A Manifesto on Psychology as Idiographic Science: Bringing the Person Back
603 Into Scientific Psychology, This Time Forever. *Meas Interdiscip Res Perspect* 2:201–218 Available
604 at: http://dx.doi.org/10.1207/s15366359mea0204_1 [Accessed June 15, 2014].

605 Molenaar PCM, Campbell CG (2009) The New Person-Specific Paradigm in Psychology. *Curr Dir*
606 *Psychol Sci* 18:112–117 Available at: <http://cdp.sagepub.com/content/18/2/112.short> [Accessed
607 June 2, 2014].

608 Morey RD, Rouder JN (2013) Package “BayesFactor.”

609 Neubauer AC, Fink A (2009) Intelligence and neural efficiency: Measures of brain activation versus
610 measures of functional connectivity in the brain. *Intelligence* 37:223–229 Available at:
611 <http://www.sciencedirect.com/science/article/pii/S0160289608001566> [Accessed May 30,
612 2014].

613 Neubauer AC, Fink A, Schrausser DG (2002) Intelligence and neural efficiency. *Intelligence* 30:515–536
614 Available at: <http://www.sciencedirect.com/science/article/pii/S0160289602000910> [Accessed
615 June 15, 2014].

616 Neurobs (2011) Presentation® software. Available at: www.neurobs.com.

617 Nichols T, Brett M, Andersson J, Wager T, Poline J-B (2005) Valid conjunction inference with the
618 minimum statistic. *Neuroimage* 25:653–660 Available at:
619 <http://www.sciencedirect.com/science/article/pii/S1053811904007505> [Accessed May 25,
620 2014].

621 Osaka M, Osaka N, Kondo H, Morishita M, Fukuyama H, Aso T, Shibasaki H (2003) The neural basis of
622 individual differences in working memory capacity: an fMRI study. *Neuroimage* 18:789–797
623 Available at: <http://www.sciencedirect.com/science/article/pii/S1053811902000320> [Accessed
624 February 10, 2016].

625 Penke L, Borsboom D, Johnson W, Kievit RA, Ploeger A, Wicherts JM (2011) Evolutionary psychology
626 and intelligence research cannot be integrated the way Kanazawa (2010) suggested. *Am Psychol*
627 66:916–917 Available at: <http://psycnet.apa.org/journals/amp/66/9/916> [Accessed June 18,

2014].

Pessoa L, Gutierrez E, Bandettini PA, Ungerleider LG (2002) Neural Correlates of Visual Working Memory. *Neuron* 35:975–987 Available at: <http://www.sciencedirect.com/science/article/pii/S0896627302008176> [Accessed February 5, 2016].

Poldrack RA (2006) Can cognitive processes be inferred from neuroimaging data? *Trends Cogn Sci* 10:59–63 Available at: <http://www.sciencedirect.com/science/article/pii/S1364661305003360> [Accessed July 10, 2014].

Poldrack RA (2014) Is “efficiency” a useful concept in cognitive neuroscience? *Dev Cogn Neurosci* Available at: <http://www.sciencedirect.com/science/article/pii/S1878929314000413> [Accessed June 14, 2014].

Prabhakaran V, Smith JA, Desmond JE, Glover GH, Gabrieli JD (1997) Neural substrates of fluid reasoning: an fMRI study of neocortical activation during performance of the Raven’s Progressive Matrices Test. *Cogn Psychol* 33:43–63 Available at: <http://www.sciencedirect.com/science/article/pii/S0010028597906599> [Accessed June 18, 2014].

Raz N, Lindenberger U (2011) Only Time Will Tell : Cross-Sectional Studies Offer No Solution to the Age – Brain – Cognition Triangle : Comment on Salthouse (2011). 137:790–795.

Ritchie SJ, Bastin ME, Tucker-Drob EM, Maniega SM, Engelhardt LE, Cox SR, Royle NA, Gow AJ, Corley J, Pattie A, Taylor AM, Valdes Hernandez M d. C, Starr JM, Wardlaw JM, Deary IJ (2015) Coupled Changes in Brain White Matter Microstructure and Fluid Intelligence in Later Life. *J Neurosci* 35:8672–8682 Available at: <http://www.jneurosci.org/content/35/22/8672.short> [Accessed June 4, 2015].

Rizopoulos D (2006) ltm: An R package for latent variable modeling and item response theory analyses. *J Stat Softw* 17:1–25 Available at: <https://lirias.kuleuven.be/handle/123456789/38163> [Accessed June 18, 2014].

Roca M, Manes F, Gleichgerricht E, Watson P, Ibáñez A, Thompson R, Torralva T, Duncan J (2013) Intelligence and executive functions in frontotemporal dementia. *Neuropsychologia* 51:725–730 Available at: <http://www.sciencedirect.com/science/article/pii/S0028393213000146> [Accessed December 20, 2013].

Roca M, Parr A, Thompson R, Woolgar A, Torralva T, Antoun N, Manes F, Duncan J (2010) Executive function and fluid intelligence after frontal lobe lesions. *Brain* 133:234–247 Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2801324&tool=pmcentrez&rendertype=abstract> [Accessed January 4, 2014].

Salthouse TA, Atkinson TM, Berish DE (2003) Executive functioning as a potential mediator of age-related cognitive decline in normal adults. *J Exp Psychol Gen* 132:566–594 Available at: <http://www.ncbi.nlm.nih.gov/pubmed/14640849> [Accessed November 11, 2013].

Sliwinski M, Hoffman L, Hofer SM (2010) Evaluating Convergence of Within-Person Change and Between-Person Age Differences in Age-Heterogeneous Longitudinal Studies. *Res Hum Dev* 7:45–60 Available at: <http://dx.doi.org/10.1080/15427600903578169> [Accessed June 18, 2014].

Sluis S van der, Verhage M, Posthuma D, Dolan CV (2010) Phenotypic complexity, measurement bias, and poor phenotypic resolution contribute to the missing heritability problem in genetic association studies. *PLoS One* 5 Available at: <http://dare.uva.nl/en/record/416495> [Accessed June 18, 2014].

Strenze T (2007) Intelligence and socioeconomic success: A meta-analytic review of longitudinal

673 research. *Intelligence* 35:401–426 Available at:
674 <http://www.sciencedirect.com/science/article/pii/S0160289606001127> [Accessed May 31,
675 2014].

676 Team (2014) R: a language and environment for statistical computing.

677 Thomas ML, Brown GG, Thompson WK, Voyvodic J, Greve DN, Turner JA, Mathalon DH, Ford J, Wible
678 CG, Potkin SG (2013) An application of item response theory to fMRI data: prospects and pitfalls.
679 *Psychiatry Res* 212:167–174 Available at:
680 <http://www.sciencedirect.com/science/article/pii/S0925492713000279> [Accessed June 16,
681 2014].

682 Tucker-Drob EM (2011) Neurocognitive functions and everyday functions change together in old age.
683 *Neuropsychology* 25:368–377 Available at: <http://psycnet.apa.orgjournals/neu/25/3/368>
684 [Accessed September 3, 2015].

685 Van Der Maas HLJ, Dolan C V., Grasman RPPP, Wicherts JM, Huizenga HM, Raijmakers MEJ (2006) A
686 dynamical model of general intelligence: The positive manifold of intelligence by mutualism.
687 *Psychol Rev* 113:842–861.

688 van der Maas HLJ, Molenaar D, Maris G, Kievit RA, Borsboom D (2011) Cognitive psychology meets
689 psychometric theory: On the relation between process models for decision making and latent
690 variable models for individual differences. *Psychol Rev* 118:339–356 Available at:
691 <http://psycnet.apa.orgjournals/rev/118/2/339> [Accessed December 10, 2014].

692 van der Maas HLJ, Wagenmakers E-J (2005) A psychometric analysis of chess expertise. *Am J Psychol*
693 118:29–60 Available at: <http://cat.inist.fr/?aModele=afficheN&cpsidt=16647448> [Accessed June
694 18, 2014].

695 Voelkle MC, Brose A, Schmiedek F, Lindenberger U (2014) Toward a Unified Framework for the Study
696 of Between-Person and Within-Person Structures: Building a Bridge Between Two Research
697 Paradigms. *Multivariate Behav Res* 49:193–213 Available at:
698 <http://dx.doi.org/10.1080/00273171.2014.889593> [Accessed June 18, 2014].

699 Wagenmakers E-J (2007) A practical solution to the pervasive problems of p values. *Psychon Bull Rev*
700 14:779–804 Available at: <http://www.springerlink.com/index/10.3758/BF03194105> [Accessed
701 June 16, 2015].

702 Woolgar A, Parr A, Cusack R, Thompson R, Nimmo-Smith I, Torralva T, Roca M, Antoun N, Manes F,
703 Duncan J (2010) Fluid intelligence loss linked to restricted regions of damage within frontal and
704 parietal cortex. *Proc Natl Acad Sci U S A* 107:14899–14902 Available at:
705 <http://www.pnas.org/content/107/33/14899.short> [Accessed January 13, 2014].

706 Woolrich M (2008) Robust group analysis using outlier inference. *Neuroimage* 41:286–301 Available
707 at: <http://www.sciencedirect.com/science/article/pii/S1053811908001778> [Accessed May 27,
708 2014].

709 Woolrich MW, Behrens TEJ, Beckmann CF, Jenkinson M, Smith SM (2004) Multilevel linear modelling
710 for FMRI group analysis using Bayesian inference. *Neuroimage* 21:1732–1747 Available at:
711 <http://www.sciencedirect.com/science/article/pii/S1053811903007894> [Accessed May 23,
712 2014].

713 Worsley K (2001) Statistical analysis of activation images. In: *Functional MRI: An Introduction to*
714 *Methods* (Jezzard P, Matthews PM, Smith SM, eds), pp 251–270. Oxford: Oxford University
715 Press.

716 Yarkoni T (2015a) Still not selective: comment on comment on comment on Lieberman & Eisenberger
717 (2015). Still not Sel comment comment comment Lieberman Eisenberger Available at:

<http://www.talyarkoni.org/blog/2015/12/14/still-no>.

Yarkoni T (2015b) No, the dorsal anterior cingulate is not selective for pain: comment on Lieberman and Eisenberger (2015). Available at: <http://www.talyarkoni.org/blog/2015/12/05/no-the-dorsal-anterior-cingulate-is-not-selective-for-pain-comment-on-lieberman-and-eisenberger-2015/>.

Yarkoni T, Poldrack RA, Nichols TE, Van Essen DC, Wager TD (2011) Large-scale automated synthesis of human functional neuroimaging data. Nat Methods 8:665–670 Available at: <http://dx.doi.org/10.1038/nmeth.1635> [Accessed March 24, 2015].

Acknowledgements

RAK, HSS, LW and DB conceived of the experiment. RAK programmed the experiment, performed the behavioural analysis, main authored the article and created the figures. HSS and RAK performed the neuroimaging analysis. All authors contributed to the final manuscript.