Intelligence xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

Intelligence

A comment on "Fractionating Intelligence" and the peer review process

Richard J. Haier [a,*], Sherif Karama [b,c], Roberto Colom [d], Rex Jung [e], Wendy Johnson [f]

[a] School of Medicine (Emeritus), University of California, Irvine, CA, USA
[b] McConnell Brain Imaging Center, Montreal Neurological Institute, McGill University, Montreal, Canada
[c] Douglas Mental Health University Institute, Department of Psychiatry, McGill University, Montreal, Canada
[d] Universidad Autónoma de Madrid-Fundación CIEN/Fundación Reina Sofía, Spain
[e] Department of Neurosurgery, University of New Mexico, Albuquerque, NM, USA
[f] MRC Centre for Cognitive Ageing and Cognitive Epidemiology, Department of Psychology, University of Edinburgh, UK

ARTICLE INFO

Article history:
Received 6 February 2014
Received in revised form 18 February 2014
Accepted 20 February 2014
Available online xxxx

Keywords:
g-Factor
Intelligence
Factor analysis
Brain imaging
Cognitive testing

ABSTRACT

Hampshire and colleagues used factor analyses and simulations to conclude that the *g*-factor is not a valid construct for general intelligence because it could be accounted for by at least two independent components defined by distinct brain networks. In our view, their results depend on a number of assumptions and subjective decisions that, at best, allow for different interpretations. We also had a unique role in the review process of their paper prior to its publication when we were invited to write a Preview. We detail that role here and describe how non-transparent editorial decision-making rejected our Preview and allowed publication despite our major concerns. The main purpose of this report is to invite Hampshire and colleagues to respond to our specific scientific concerns that aim to clarify their work and contribute a constructive discussion about the meaning of their findings.

Hampshire and colleagues challenged fundamental concepts about the *g*-factor based on cognitive performance data from a battery of 12 tests taken by over 44,000 people on the internet and on fMRI data collected on 16 subjects performing the same cognitive test battery (Hampshire, Highfield, Parkin, & Owen, 2012). Their conclusions are derived from complex factor analyses and simulations that, in our view, are open to alternative interpretations because they depend on a number of arguable assumptions. We also have a unique perspective on the publication process for their paper that raises some troubling issues.

We detail these issues here for both their report and the process that led to its publication. The issues are intertwined, so we organize this paper according to the chronology that unfolded. We have an expectation that Dr. Hampshire and colleagues will respond to our substantive scientific points in a companion paper published simultaneously in this journal; he has been invited to do so both by us and by the editor. Such exchanges are common, constructive, and help advance the field. We have a hope but no expectation that the editors of Neuron will explain aspects of their peer review process that, in our view, created unnecessary confusion about the conclusions Hampshire et al. reached.

On July 13th, 2012 one of us (RH) received an invitation to write a Preview to highlight a paper by Hampshire et al. to be published in Neuron. The paper was due to be published soon and the deadline for the Preview was August 6th. RH agreed and received a copy of the accepted manuscript the same day. RH found many aspects of the paper quite difficult to understand and, more troubling, he worried that some main conclusions could be based on erroneous application and

* Corresponding author.
E-mail address: rich.haier@gmail.com (R.J. Haier).

http://dx.doi.org/10.1016/j.intell.2014.02.007
0160-2896

interpretation of factor analysis. Given the rapidly approaching deadline, and the potential attention that a paper on intelligence would receive in Neuron, RH sent copies of the accepted manuscript to four colleagues with considerable expertise in brain imaging and psychometrics, especially factor analysis (SK, RC, RJ, WJ) and asked for their comments. We exchanged emails and phone calls, which confirmed that we shared a number of major concerns about the manuscript. We jointly wrote a Preview that noted these concerns and gave a context for the reader to consider the conclusions presented by Hampshire et al. We submitted the Preview on August 6th, 2012 and in a cover letter we informed Neuron that our concerns were so serious, that had any of us been original reviewers, we would not have recommended acceptance without major clarifications. This cover letter and the Preview are in Appendix A. Throughout this paper, all the correspondences we reference are emails and all are preserved.

The next day, an editor at Neuron responded. They were clearly concerned and rightly wanted Hampshire et al. to respond to the issues we had raised. The editor invited us to send a more detailed critique that could be passed along to the original reviewers and to the authors. We did so on August 16th. That critique detailed 20 points; it is in Appendix B. We were not told how our 20 concerns of August 16th were communicated to the original reviewers and to the authors, or how they responded. On October 31st 2012 Neuron informed us that publication of the Hampshire et al. manuscript would go forward with some minor changes. We were also informed that, after considerable internal discussion, the editorial board had decided that our Preview would not be published; no reason was given. We objected and asked if we could submit a modified Preview based on the modified manuscript (which was not shared with us). Neuron declined. One editor asked to have a confidential phone call with RH and that call took place on December 2nd. RH respects that confidentiality and can only say that he found the editorial process and decision-making hard to understand.

The editorial decision-making became even more troubling when the Hampshire paper was published on December 20th, 2012. We saw the final version with the modifications for the first time two days earlier when a science writer sent RH an embargoed copy and asked for a comment on the importance of the paper. We were surprised to see that the final version did not address our concerns in any substantial way. For example, the key point we raised among the 20 concerns in Appendix B was whether using a factor analysis technique that forced rotated factors to be independent could objectively serve as the basis for a conclusion that there was no unitary *g*-factor and hence the conclusion about "fractionating" intelligence.

The authors issued a press release from their university (The University of Western Ontario in Canada) the day before the Neuron publication on December 20th. This press release is in Appendix C. The title is: "Western University-led research debunks the IQ myth." The press release received some attention mostly in non-science media outlets and hyped the study as demonstrating definitively that IQ was a meaningless concept. For example, the senior author, Adrian Owen, was quoted as saying: "When we looked at the data, the bottom line is the whole concept of IQ — or of you having a higher IQ than me — is a myth… There is no such thing as a single measure of IQ or a measure of general intelligence." (thestar.com, 12/19/12). Of course, most psychologists understand that this is a classic "straw man" argument since no one claims that an IQ score (which is a composite of a test battery) measures the whole of human intelligence. It is also widely understood that the *g*-factor is not synonymous with IQ.

As far as we are aware, the Hampshire paper was not covered as newsworthy in any major science publications. However, our attention was drawn to a blog written by the Neuroskeptic (http://blogs.discovermagazine.com/neuroskeptic) on December 24th, 2012. The Hampshire paper was summarized and the Neuroskeptic (anonymously written) made several critical observations. A series of reader comments followed over the next several weeks, most written anonymously. There were several comments that suggested knowledge of our unpublished Preview. We determined that a graduate student had overheard a relevant discussion and decided to comment on the blog anonymously without our knowledge. One commenter on the blog responded to some of the scientific critiques with a lengthy detailed technical argument (see Appendix D for the full comment). This detailed comment also concluded in part with these sentences: "Finally, a critical comment was submitted to Neuron however, there was no 'conspiracy'. It was decided, based on feedback from an independent reviewer, that the author of the comment was heavily biased and that the criticisms raised were lacking in substance. Also, the authors of the article demonstrated that they were both willing and able to address all of those criticisms point by point if the journal chose to publish them."

Obviously someone with inside knowledge of the review process wrote this comment. We sent this comment to Neuron and asked if it were true that our 20 detailed concerns were communicated only to one of the original reviewers who then determined our concerns did not have substance and were biased. We also requested that Neuron provide any written responses to our 20 points made by the original reviewers or the authors. Neuron replied that discussions were all by phone and there were no written responses. Neuron would not confirm that only one original reviewer determined that our concerns were biased or that they had not required a point-by-point response. Finally, we asked Neuron if we could submit comments on the Hampshire et al. paper under the category of "Viewpoint" or "Perspective" and allow the authors to respond. We felt that this would be constructive and educational. Neuron declined.

Over the last year, we have exchanged a series of emails with Dr. Hampshire. He clarified some points and sent us some key correlation matrices (that were not published) so we could better understand some of the analyses. He also noted that he had responded to some of our 20 points at the request of Neuron and that his responses had been sent for review and that the review agreed with all of them; publication followed. He added that he had offered to publish a point-by-point response but Neuron declined. We told Dr. Hampshire that we were writing this paper and he was positive about responding to the 20 points. We have common interests about the importance of combining neuroimaging with psychometrics. In our work, for example, we have used

Please cite this article as: Haier, R.J., et al., A comment on "Fractionating Intelligence" and the peer review process, *Intelligence* (2014), http://dx.doi.org/10.1016/j.intell.2014.02.007

imaging analyses to test and constrain the Parietal Frontal Integration Theory (PFIT) of intelligence (Colom et al., 2009; Haier et al., 2009; Jung & Haier, 2007).

One thing is clear: intelligence is a complex process and our measurement tools keep getting better. We are far from the point where any single study is likely to settle a long-standing research issue. The Hampshire paper reports an interesting but, in our view, flawed exercise that falls far short of being a new discovery. Their data can be interpreted in different ways and we look forward to their responses to our unpublished Preview (Appendix A) and to our 20 concerns (Appendix B). We also look forward to seeing more research efforts that combine neuroimaging and psychometrics, all of which contribute to understanding the *g*-factor and other aspects of intelligence in one way or another.

Appendix A. Cover letter and unpublished preview submitted to Neuron

(August 6th, 2012)

Dear [Editor],

I appreciate the opportunity to write a Preview for this paper. It has been a daunting challenge. It became clear to me after my first reading, that there were a number of confusing and possibly incorrect aspects of the paper so I asked four colleagues, all with considerable expertise, to take a look. They dug deep into the details and we exchanged many emails trying to understand exactly what the authors did. We were confused and no doubt many others also will be confused.

Frankly, had any of us reviewed the paper, we would have recommended major revisions for clarity and for interpretation, or outright rejection for incorrect interpretations and apparent misapplication of some factor analysis methods. The paper is quit complex and dense and presents a very narrow view of the extensive relevant literature. A detailed critique would involve technical aspects of factor analysis not readily amenable to the general reader. Our Preview aims to give the reader a broad perspective on current imaging/intelligence research and provide a skeptical context for evaluating the claims made. We hope we have struck the right balance.

PREVIEW

Submitted 8/6/12

Two "g" or not two "g": that is the question

Richard J. Haier, Sherif Karama, Roberto Colom, Rex Jung and Wendy Johnson

Abstract: Hampshire et al. factor analyzed fMRI data, identifying two brain networks related to intelligence factors. They concluded that existence of these networks argues against a unitary general intelligence factor. The results were driven by a series of methodological judgments open to other interpretations.

For over 100 years, research on the nature of human intelligence has analyzed the relations among various tests designed to quantify individual differences in cognitive abilities. One of the most robust findings in all of psychology is the observation that virtually all tests of mental abilities, irrespective of content and task demands, are positively correlated with each other, leading to the concept of a general factor, designated as "*g*". Sophisticated statistical techniques, especially factor analysis, have been used extensively to investigate the reliability and validity of the "*g*" construct and how *g* relates to other specific factors of intelligence (e.g. verbal ability, spatial ability, quantitative ability, etc.) and to basic cognitive abilities (e.g. short-term memory, information processing speed, executive updating, etc.). Past controversies about the factor structure of intelligence tests and whether *g* is an artifact of a particular method of factor analysis largely have been resolved (Carroll, 1993; Jensen, 1998).

Psychometrically, the *g* factor is found at the most basic level in the first un-rotated factor derived from a battery of cognitive tests — often accounting for about 50% of variance — and it is important to underscore that different methods (including hierarchical factor analysis identifying a higher-order *g*) yield remarkably similar results, provided each battery measures a broad range of cognitive abilities in a sample representing a broad range of population abilities (Jensen, 1998; Johnson, te Nijenhuis, & Bouchard, 2008). Considerable evidence shows that *g* predicts many real-world life outcomes (Gottfredson, 1997) and that genetic factors are important in explaining individual differences in *g* (Deary, Penke, & Johnson, 2010b). Contrary to the impression left by Hampshire et al. (Hampshire et al., 2012), however, intelligence researchers have never argued that a single construct like *g* can account for the "entire distribution of general intelligence" (Johnson & Bouchard, 2005).

In the last 25 years, research on the nature of intelligence has combined psychometric methods with brain-imaging techniques to discover structural and functional correlates of *g* and other intelligence factors. Many cognitive researchers have now joined psychometricians in such efforts, often with a renewed emphasis on individual differences. A key goal of the imaging studies in this area has been to determine whether specific, unique brain systems underlie *g*, or whether *g* depends on a combination of systems that support specific factors or elemental cognitive processes. In 2007, a review of relevant imaging literature proposed a parieto-frontal network model of general intelligence and noted that not all parts of the network may be relevant to *g* in any one individual (Jung & Haier, 2007). For instance, males and females, as well as young and older adults, may differ in the brain areas underlying differences in intelligence (Haier, Jung, Yeo, Head, & Alkire, 2005). If so, *g* may not be manifested in all brains in the same way, further complicating the concept of "neuro *g*".

A special issue of Intelligence devoted to brain imaging studies included the question, "Is there a neuro-g?" (Haier et al., 2009). None of the imaging evidence to date supports a single network (Colom, Karama, Jung, & Haier, 2010), but there is no logical reason to conclude that *g* is not unitary because two or more brain networks may be involved. It would be quite interesting, however, if some brain networks were related to *g* in some individuals, but other networks were related to *g* in other individuals.

The Hampshire et al. report is a complex paper (Hampshire et al., 2012) but the main focus is the application of factor analysis to explore the relation between: (a) functional brain networks determined by fMRI activations common to 12 cognitive tasks performed by 16 individuals, and (b) intelligence factors derived from another sample of 44,000+ who completed an online battery of the same 12 tasks. They identified two independent brain networks (one for memory and one for reasoning) involved in task performance, concluding that their data did not support a unitary general factor of intelligence. In our view, there are several issues that challenge this interpretation.

The key feature of their report is the use of factor analysis on brain image voxels to find clusters interpreted as brain networks. This is a worthy approach, used by others. Penke et al. (2012), for example, used factor analysis on Diffusion Tensor Imaging data to identify a common factor of white matter integrity that predicted intelligence mediated by information processing speed. As factor analysis was applied in the Hampshire et al. report, however, several points require clarification. For readers unfamiliar with the technical aspects of factor analysis, we focus on conceptual issues related to the core finding of the putative identification of separate brain networks for working memory and reasoning that may relate to intelligence factors. Since we have limited space we will not comment on their supporting analyses.

Factor analysis involves several methodological steps and analytical judgments are required at each one. Its purpose is to identify the fundamental dimensions or 'factors' underlying multivariate data and their relations to the measured data. But the factor definitions are arbitrary, as the factors can be "rotated" in many ways. Doing this is a judgmental decision that dictates the perspective from which the data will be viewed, in much the same way that standing in front of a building gives a different perspective on it than standing on the roof, though both may be considered "real" viewpoints. Typically, *g* is defined as the first un-rotated factor because this minimizes such judgment calls.

Hampshire et al. identified a strong un-rotated general factor and several other factors in their imaging data (supplemental Table 1) but then went on to rotate these factors to impose their independence. These independent factors were the basis of their interpretations and conclusions about two independent networks. As is the case in test score data, the un-rotated factor solution that shows the strong general factor reflects brain organization to the same degree as the statistically independent factors, and very likely could be tied back to the test score data exactly as Hampshire et al. demonstrated for the independent factors. The un-rotated factor, for example, could well be a "neuro-g" and it is important to consider the alternative interpretation. The authors should publish the basic correlation matrices as supplemental data so other researchers can explore other possible interpretations.

Basic information about the 16 subjects in the imaging study is also lacking. We do not know ages, proportion of males and females, or the range of intelligence scores. Reporting such information is generally considered standard practice and central to interpreting the analyses and for potential replication. Some may believe that any 16 subjects would produce the same results as any other 16 subjects, but we disagree. If, for example, most of the subjects were college students, there may be restricted ranges on some measures that could influence the basic correlation matrices on which the factor analyses are based. Importantly, fMRI studies have found brain activation differences during cognitive testing when subjects selected for high or average intelligence were compared (Graham et al., 2010). There was evidence of such differences among the 16 subjects here as well. The authors mention having analyzed subjects individually and report that 3 out of the 16 did not follow the two-factor model. It would have been useful to have more details to assess to what extent the individual factor solutions of the other 13 subjects conformed to the original whole-group two-factor model.

The study's design appears to have suffered from a conceptual confusion that sometimes appears in studies using factor analysis. Hampshire et al. used averaged voxel activations (in a limited set of brain areas) across all 12 tests and all 16 subjects as the basis for identifying the two brain networks but then related these *average* across-subject voxel activation differences to *individual differences* in task performance in the much larger internet sample. The interpretation of these associations as indicating that neurological factors underlie variation in task performance is questionable.

Overall, the interpretation of results is discussed by relating the findings to a limited number of previous reports. A broader view may well change the interpretation. For example, studies of human lesion patients with large samples have shown detailed brain mapping of intelligence factors (Barbey et al., 2012; Glascher et al., 2010), n = 241 and n = 182, respectively. The authors mention one smaller lesion study but integrating their results with these studies may well modify their conclusion about *g*.

Finally, it also must be remembered that fMRI data show aspects of blood flow indirectly related to neuron activity. Habituation to repeated stimuli during testing depresses activations and subtraction of activation from a resting baseline may not be optimal for finding a common general brain factor involving intelligent performance. A general factor of intelligence may also rely on white matter characteristics (Penke et al., 2012) or other brain parameters invisible to the fMRI technique used here. Other imaging techniques have also been used to identify intelligence correlates including magnetoencephalography (with better temporal resolution) and/or proton magnetic resonance spectroscopy (with better chemical resolution). The fact that the brain is a complex electro-chemical organ, with both temporal and spatial characteristics that function in "small world" networks, likely limits the ability of fMRI (or any single neuroimaging modality for that matter) to detect any general characteristic underlying cognition that might be defined as *g*.

This study is based on an interesting data set and we applaud the effort. For the reasons stated, however, we think the definitive tone of the interpretations and conclusions is not justified. In our view, the report does not clarify the nature of *g*, but it does illustrate the richness of the questions and the technical challenges confronting intelligence researchers today. We are still a long way from identifying the brain characteristics of general intelligence and why some people reason better, remember more, and learn faster than other people. Yet there has been much progress from Spearman's original 1904 formulation of *g*, and he undoubtedly would enjoy discussing and dissecting the vast quantities of additional data available to us today.

References

Barbey, A.K., Colom, R., Solomon, J., Krueger, F., Forbes, C., and Grafman, J. (2012). An integrative architecture for

general intelligence and executive function revealed by lesion mapping. Brain 135, 1154–1164.

Carroll, J.B. (1993). Human cognitive abilities: a survey of factor-analytic studies (Cambridge; New York: Cambridge University Press).

Colom, R., Karama, S., Jung, R.E., and Haier, R.J. (2010). Human intelligence and brain networks. Dialogues Clin Neurosci 12, 489–501.

Deary, I.J., Penke, L., and Johnson, W. (2010). The neuroscience of human intelligence differences. Nature Reviews Neuroscience 11, 201–211.

Duncan, J. and Owen, A. M., Common regions of the human frontal lobe recruited by diverse cognitive demands, Trends Neurosci 23, 2000, 475–83.

Glascher, J., Rudrauf, D., Colom, R., Paul, L.K., Tranel, D., Damasio, H., and Adolphs, R. (2010). Distributed neural system for general intelligence revealed by lesion mapping. Proceedings of the National Academy of Sciences of the United States of America 107, 4705–4709.

Gottfredson, L.S. (1997). Why g matters: The complexity of everyday life. Intelligence 24, 79–132.

Graham, S., Jiang, J.Y., Manning, V., Nejad, A.B., Zhisheng, K., Salleh, S.R., Golay, X., Berne, Y.I., and McKenna, P.J. (2010). IQ-Related fMRI Differences during Cognitive Set Shifting. Cerebral Cortex 20, 641–649.

Haier, R.J., Colom, R., Schroeder, D.H., Condon, C.A., Tang, C., Eaves, E., and Head, K. (2009). Gray matter and intelligence factors: Is there a neuro-g? Intelligence 37, 136–144.

Haier, R.J., Jung, R.E., Yeo, R.A., Head, K., and Alkire, M.T. (2005). The neuroanatomy of general intelligence: sex matters. Neuroimage 25, 320–327.

Jensen, A.R. (1998). The g factor: The Science of Mental Ability (Westport: Praeger).

Johnson, W., and Bouchard, T.J. (2005). The structure of human intelligence: It is verbal, perceptual, and image rotation (VPR), not fluid and crystallized. Intelligence 33, 393–416.

Johnson, W., te Nijenhuis, J., and Bouchard, T.J. (2008). Still just 1 g: Consistent results from five test batteries. Intelligence 36, 81–95.

Jung, R.E., and Haier, R.J. (2007). The Parieto-Frontal Integration Theory (P-FIT) of intelligence: Converging neuroimaging evidence. Behavioral and Brain Sciences 30, 135–154.

Penke, L., Maniega, S.M., Bastin, M.E., Valdes Hernandez, M.C., Murray, C., Royle, N.A., Starr, J.M., Wardlaw, J.M., and Deary, I.J. (2012). Brain white matter tract integrity as a neural foundation for general intelligence. Mol Psychiatry.

Appendix B. 20 point critique submitted to Neuron

Dear [Editor]

Here are the detailed concerns that were the basis for our Preview. As you know, our Preview tried to alert the reader that the conclusions in the report should be meet with skepticism. Our main concern with the *in press* report was that few readers would go beyond the abstract and get the wrong take-away message that the report was a new discovery that challenged the idea of a general factor of intelligence. As you can see by the details below, we think the paper has series flaws that cannot be corrected by clearer writing alone. We think that this paper, should it be published in its current form, will just confuse the literature with incorrect conclusions, especially for researchers who specialize in intelligence (these authors clearly do not — as evidenced by the poor scholarship in both the Introduction and the Discussion, as well as the methodological flaws noted below). The general reader certainly would get incorrect and misleading information about intelligence research. As reviewers, our recommendation would be REJECTION.

The details below are written as if we had been reviewers, so please feel free to cut and paste them to the authors, or to the original reviewers (but, if you do please keep our names confidential at this point). We understand your difficult position and we have made a determined effort to be constructive.

Best wishes,

Rich

****

This paper is about psychometrics and whether brain-imaging data from one sample adds new insights about the general factor of intelligence (*g*). The paper depends on a series of subjective choices about how to use factor analysis techniques. Many details require clarification and alternative interpretations need to be acknowledged. Overall, there appear to be fundamental flaws (**noted below in BOLD type**) in the presentation that severely limit any conclusions about the nature of *g*.

1. The introduction is misleading about whether there is a general factor of intelligence or whether it may be a statistical artifact. The authors have cited mostly older papers that raise issues now mostly resolved. The works of Carroll (Carroll, 1993) or Jensen (Jensen, 1998) are central to these issues and need to be cited, along with newer papers that are quite relevant for both the Introduction and the Discussion (Barbey et al., 2012; Colom et al., 2010; Deary, Penke, & Johnson, 2010a; Glascher et al., 2010; Haier et al., 2009; Johnson et al., 2008; Karama et al., 2011; Penke et al., 2012).
2. The Introduction notes that the imaging analysis is based on brain areas identified in Duncan & Owen (2000); those areas are mostly in frontal lobe, whereas the cited Jung & Haier model identifies areas distributed across the brain. The two models are not really the same, contrary to the authors' suggestion, and they have provided no rationale for using the one and not the other as a basis for the analysis. They did analyze the entire brain in addition to the Duncan & Owen areas, and found three factors instead of two; this whole brain analysis is potentially quite important and deserves more detailed discussion and a figure.
3. The introduction is also misleading about factor analysis. It is true that it is indeterminate (the same data can generate literally an infinite number of 'solutions') and how any of them generated by behavioral data relate to any of them generated by biological data is a wide-open question. What isn't true, however, is that there is one best factor solution waiting to be discovered. Rather all solutions can be appropriate in some circumstances and

not in others, and evaluating any solution is a matter of judgment, just like deciding whether the most useful view of a building in which you need to do some kind of work depends on whether you're going to be sitting at a desk in an office or fixing the roof. **Because this is true of factor analysis in both brain and behavioral data, the fact of finding associations between one kind of solution in brain and behavioral data doesn't necessarily say anything about whether or not there may be similar associations between another kind of solution in the same brain and behavioral data.** In fact, because you're working with the same two covariance matrices either way, such similarity of associations is effectively inevitable, as biometric decompositions of factor solutions in twin data have repeatedly made clear. **This gets at the very heart of the design of this study: it really cannot say much of anything definitive about *g* or no *g* or its 'location' in the brain beyond a demonstration that brain and behavioral data can be modeled in a similar way.**

4. Complete information on participant recruitment and demographics for the imaging participants needs to be provided. This is a standard reporting requirement in the field. Educational levels, sex and ages are particularly relevant here because restriction of the range of test scores can influence the basic correlation matrices on which factor analysis of both brain and behavioral data depend (there are also sex differences on some speeded tests). fMRI studies of cognitive tasks also can show different results depending on the IQ of subjects (Graham et al., 2010).
5. More information should also be provided on how the scanning and testing of these participants were carried out and how the brain mapping of the images was carried out. This should state formally whether or not the fMRI sequences varied in length between participants and tasks, as well as whether task order was counter-balanced and whether and how rest breaks were allowed. From the description that is provided variation in length should be the case as some of the tasks were obviously not of equal length for each participant given that some ended only after 3 errors while others had fixed time lengths. This said, variation in fMRI sequence length as a function of subject performance is surprising since the protocols (and associated length) usually need to be programmed BEFORE the fMRI session starts. Details about the baseline condition also need to be added.
6. PCA relies on the assumption that the subjects of analysis are independent and identically distributed. In the primary PCA on which this paper is based, average voxel activations are the 'subjects'. Yet, activations in the various voxels are clearly not independent due, in part, to the imposed smoothing and issues inherent to the fMRI signal. Further, whether activation patterns are identically distributed across tasks is not a given. It's not clear what all this does to the results, but some of the simulation skills of these authors should be applied to figuring this out.
7. **The decision to rotate the components using varimax rotation was completely subjective on the part of the authors and dictated that they would get independent factors that would not represent general intelligence in any recognizable way. Supplemental Table 1 makes very clear that the voxel data they had, before rotation, showed a very clear general factor, every bit as strong as they generally appear in behavioral data. A focus on this might well change the conclusions dramatically.** This whole study, with some alterations as noted below, should be repeated using this general factor and the second one from that solution (for direct comparability in number of factors). This is very important because *g* is USUALLY defined in psychometric studies as this first UNROTATED principal component (or factor) specifically to remove the subjectivity involved in every rotation method. But even if some rotation methods were to be considered, it would make a lot more sense to use an oblique rotation method that will estimate the correlations between the factors, so that the data can tell you to what degree a 2-factor solution should be considered to have generated independent factors. **The authors should publish the basic correlation matrices as supplemental data so other researchers can explore other possible interpretations.**
8. At the top of page 6, they're checking whether their PCA solution is robust to variations in their specific choices of brain areas. However, they kept the same regions but calculated the center of gravity of these regions. They did rerun the same analysis on the same regions but this time reduced the size of the a priori region to only say, 10 mm in radius. They did the same thing again, increasing by increments of 5 mm until they reached 25 mm. In other words, they looked at the same a priori regions but simply varied their size by using ROIs based on the centers of gravity of the original ROIs. They called these different ROIs but, obviously, they were essentially the same ROIs but of different size. Further, it's really not clear why, only for this analysis of differently sized ROIs, did they eliminate negative activations (i.e. they don't appear to have eliminated the negative activations from the first analysis on the full ROIs). The apparent different standard for data analysis needs to be justified. Also, what was the rationale for eliminating deactivations that might indicate inhibitory networks?
9. Much more information is needed on the PCAs run on individual participant voxel data. What were the solutions like for the 3 participants who didn't generate the same number of components? What about the ones who did generate 2 components? How did their components correlate with the overall group components? **This is important, because if 3/16 participants didn't generate the same number of factors as the overall group, it suggests that the 2-component solutions of the other 13 might not be very similar, and thus there may be substantial individual differences in brain usage that have been averaged out of the overall 2-component solution on which the rest of the analyses, and the paper's inferences and conclusions, were based.**
10. What is reported as having been done immediately following mention of the 3 individual-participant solutions with more than 2 components on page 6 needs to be made much clearer. We couldn't follow it at all.

Please cite this article as: Haier, R.J., et al., A comment on "Fractionating Intelligence" and the peer review process, *Intelligence* (2014), http://dx.doi.org/10.1016/j.intell.2014.02.007

11. What is involved in the ICA is also not at all clear though its point is. The basic idea of checking what you have using another method is very solid, but **using another method that also imposes independence on the factors, which they say ICA does, is not at all helpful when it's exactly the arbitrary imposition of independence on the factors that is the problem with the original PCA.** So this section, mostly on page 7, basically says, if you use another technique that makes the same arbitrary assumption as the first method in the same data and try some random data simulations again relying on the same assumption, you get very similar results. But what if you don't make this arbitrary assumption? Or make some rival arbitrary assumption such as that there is ONE dominant component? Our guess is that these procedures would replicate that too.
12. More information should be provided on how participants were excluded from the Internet survey data. More than half were excluded. See also #20 below.
13. In the second paragraph of their "-predictive power" section, the authors write, "The loadings of the tasks on the MDwm and MDr networks from the ICA were formed into two vectors. These were regressed onto each individual's set of 12 standardized task scores with no constant term. When each individual's MDwm and MDr 'component scores' were varied in this manner…". The 'component scores' were varied in what matter? A plot/graph/figure would be helpful.
14. The bottom of page 8 is not at all clear. But the bottom line is: the fact that the brain component data accounted for 34.3% of the variance and the first two components of the behavioral data accounted for 36.6% of the variance makes it all too possible that there was no overlap at all between the two. The checks performed using random vectors do make complete lack of overlap unlikely, but nothing is reported about how much better the intended vectors predicted the performance data than the random ones did, and nothing provides any evidence that the overlap was particularly high. **This makes the conclusion that was drawn (i.e. that the intended vectors accounted for greater than 50% of the systematic variance in the performance data) simply unwarranted, as it relies on the assumption of 100% overlap.**
15. **There is another problem with this and later parts of the analysis. This is that the brain components, whether from PCA or ICA, were based on AVERAGE activations across all 16 imaging participants. Thus the participants' individual differences in activations were completely lost (effectively considered to be noise in the data). How these average activations even SHOULD be related to individual differences in test scores, even if everything works in the brain exactly according to the two identified components, is far from clear. It's analogous to using patterns of average test scores just from counties in Massachusetts to understand individual differences in test scores within China. Put this way, most people would be able to see this is not a particularly good idea.**
16. **This means that the correlations reported at the bottom of page 9 are between average activations and individual differences in performance. They thus appear to be between what is noise according to the activation data but signal in the performance data. Within the two identified behavioral components, the loadings of the neural components in Table 2 track task reliabilities almost exactly except for some Digit Span, suggesting that mostly this whole analysis is just correlating absence of noise with absence of noise, with noise meaning individual differences in activation in the neural data and test unreliability in the behavioral data. How to interpret that in any meaningful way is not clear.**
17. Bottom of page 11 and page 12: They say they generated two simulated data sets, one based on the 2-component behavioral data solution and one based on the 3-component solution. They then, apparently, describe just what they did for the 3-component simulation set. We assume they did the analogous analysis for the 2-component simulation set, but they should state that more clearly. What they generated though was essentially 20 sets of task scores systematized specifically and only by the ICA-generated behavioral components. That is, the only aspects of these scores that were not noise were imposed by the common imposition of the ICA loadings. It's not at all clear what they then did, but they appear to have done PCAs on these sets of scores and correlated the (average??) task factor loadings generated to those from the Internet data itself. **This is confusing but it sounds like they showed that, when you organize random noise according only to the components that the behavioral data had, you get results that correlates pretty well with the only organization system it has. This isn't at all amazing and doesn't say ANYTHING about the robustness of the components of the organizational system they imposed. Of course, we could be wrong about what they did, but then they need to write more clearly about this complex sequence.**
18. **Their more important conclusion on page 12 is similarly meaningless, assuming we guessed right about what they did.**
19. **There's no surprise in their section on page 13 – all this has been demonstrated many times and it doesn't help to substantiate their claim that the components they identify are KEY to intelligence in the brain, nor does it refute anything about g.**
20. There are problems with the ways they administered and scored several of the tasks in the Internet sample. More information is needed to figure out what the impact might be, but it likely reduced the size of the first principal component in these data, contributed to the large number of participants they had to throw away, and reduced the reliability of the test scores considerably. **The problem is that they used some but not all of the aspects of computerized adaptive testing in a way that wrecked the ability to consider the scores to be on interval scales**. In computerized adaptive testing, the idea is to shorten the time to administer a test to an accurate score by giving a harder next item when the participant gets an item right and an easier item when the participant gets an item wrong. As long as a participant answers systematically, the test can stop as soon as item performance converges on a sufficient

Please cite this article as: Haier, R.J., et al., A comment on "Fractionating Intelligence" and the peer review process, *Intelligence* (2014), http://dx.doi.org/10.1016/j.intell.2014.02.007

degree of accuracy of measurement, which is generally shorter than the full test. These authors, however, administered tasks this way but for a fixed duration of time, so that faster responders got more items and some participants were measured much more accurately than others. For other tasks there was a cut-off of three errors, and the score was the difficulty level on which the failures accrued, which depends heavily (more so than just counting items right) on the difficulty levels being interval-scaled (not necessarily matched to, say, number of words to be remembered) and errors made being systematically associated with ability. **Even more seriously, for some tasks, the score increment received for getting an item right was greater for harder items so the whole scoring system was not interval-based.**

References

Barbey, A. K., Colom, R., Solomon, J., Krueger, F., Forbes, C., & Grafman, J. (2012). An integrative architecture for general intelligence and executive function revealed by lesion mapping. *Brain, 135*, 1154–1164. doi: Doi 10.1093/Brain/Aws021

Carroll, J. B. (1993). *Human cognitive abilities: a survey of factor-analytic studies*. Cambridge; New York: Cambridge University Press.

Colom, R., Haier, R. J., Head, K., Alvarez-Linera, J., Quiroga, M. A., Shih, P. C., & Jung, R. E. (2009). Gray matter correlates of fluid, crystallized, and spatial intelligence: Testing the P-FIT model. *Intelligence, 37*(2), 124–135.

Colom, R., Karama, S., Jung, R. E., & Haier, R. J. (2010). Human intelligence and brain networks. *Dialogues Clin Neurosci, 12*(4), 489–501.

Deary, I. J., Penke, L., & Johnson, W. (2010a). The neuroscience of human intelligence differences. *Nat Rev Neurosci, 11*(3), 201–211. doi: nrn2793 [pii]10.1038/nrn2793

Deary, I. J., Penke, L., & Johnson, W. (2010b). The neuroscience of human intelligence differences. *Nature Reviews Neuroscience, 11*(3), 201–211. doi: Doi 10.1038/Nrn2793

Duncan, J. and Owen, A. M., Common regions of the human frontal lobe recruited by diverse cognitive demands, Trends Neurosci 23, 2000, 475–83.

Glascher, J., Rudrauf, D., Colom, R., Paul, L. K., Tranel, D., Damasio, H., & Adolphs, R. (2010). Distributed neural system for general intelligence revealed by lesion mapping. *Proceedings of the National Academy of Sciences of the United States of America, 107*(10), 4705–4709. doi: Doi 10.1073/Pnas.0910397107

Gottfredson, L. S. (1997). Why g matters: The complexity of everyday life. *Intelligence, 24*(1), 79–132.

Graham, S., Jiang, J. Y., Manning, V., Nejad, A. B., Zhisheng, K., Salleh, S. R., . . . McKenna, P. J. (2010). IQ-Related fMRI Differences during Cognitive Set Shifting. *Cerebral Cortex, 20*(3), 641–649. doi: Doi 10.1093/Cercor/Bhp130

Haier, R. J., Colom, R., Schroeder, D. H., Condon, C. A., Tang, C., Eaves, E., & Head, K. (2009). Gray matter and intelligence factors: Is there a neuro-g? *Intelligence, 37*(2), 136–144.

Haier, R. J., Jung, R. E., Yeo, R. A., Head, K., & Alkire, M. T. (2005). The neuroanatomy of general intelligence: sex matters. *Neuroimage, 25*(1), 320–327.

Hampshire, A., Highfield, R. R., Parkin, B. L., & Owen, A. M. (2012). Fractionating Human Intelligence. *Neuron, 76*(6), 1225–1237. doi: Doi 10.1016/J.Neuron.2012.06.022

Jensen, A. R. (1998). *The g factor: The Science of Mental Ability*. Westport: Praeger.

Johnson, W., & Bouchard, T. J. (2005). The structure of human intelligence: It is verbal, perceptual, and image rotation (VPR), not fluid and crystallized. *Intelligence, 33*(4), 393–416. doi: Doi 10.1016/J.Intell.2004.12.002

Johnson, W., te Nijenhuis, J., & Bouchard, T. J. (2008). Still just 1 g: Consistent results from five test batteries. *Intelligence, 36*(1), 81–95. doi: Doi 10.1016/J.Intell.2007.06.001

Jung, R. E., & Haier, R. J. (2007). The Parieto-Frontal Integration Theory (P-FIT) of intelligence: Converging neuroimaging evidence. *Behavioral and Brain Sciences, 30*(02), 135–154.

Karama, S., Colom, R., Johnson, W., Deary, I. J., Haier, R., Waber, D. P., . . . Grp, B. D. C. (2011). Cortical thickness correlates of specific cognitive performance accounted for by the general factor of intelligence in healthy children aged 6 to 18. *Neuroimage, 55*(4), 1443–1453. doi: Doi 10.1016/J.Neuroimage.2011.01.016

Penke, L., Maniega, S. M., Bastin, M. E., Valdes Hernandez, M. C., Murray, C., Royle, N. A., . . . Deary, I. J. (2012). Brain white matter tract integrity as a neural foundation for general intelligence. *Mol Psychiatry*. doi: mp201266 [pii]10.1038/mp.2012.66

Appendix C

Press Release

Western University-led research debunks the IQ myth 12/19/12 9:20 AM

http://www.eurekalert.org/pub_releases/2012-12/uowo-wur121712.php Page 1 of 2

Public release date: 19-Dec-2012

Contact: Jeff Renaud

jrenaud9@uwo.ca

519-661-2111 x85165

University of Western Ontario

Western University-led research debunks the IQ myth

After conducting the largest online intelligence study on record, a Western University-led research team has concluded that the notion of measuring one's intelligence quotient or IQ by a singular, standardized test is highly misleading.

The findings from the landmark study, which included more than 100,000 participants, were published today in the journal *Neuron*. The article, "Fractionating human intelligence," was written by Adrian M. Owen and Adam Hampshire from Western's Brain and Mind Institute (London, Canada) and Roger Highfield, Director of External Affairs, Science Museum Group (London, U.K.).

Utilizing an online study open to anyone, anywhere in the world, the researchers asked respondents to complete 12 cognitive tests tapping memory, reasoning, attention and planning abilities, as well as a survey about their background and lifestyle habits.

"The uptake was astonishing," says Owen, the Canada Excellence Research Chair in Cognitive Neuroscience and Imaging and senior investigator on the project. "We expected a few hundred responses, but thousands and thousands of people took part, including people of all ages, cultures and creeds from every corner of the world."

The results showed that when a wide range of cognitive abilities are explored, the observed variations in performance can only be explained with at least three distinct components: short-term memory, reasoning and a verbal component.

No one component, or IQ, explained everything. Furthermore, the scientists used a brain scanning technique known as functional magnetic resonance imaging (fMRI), to show that these differences in cognitive ability map onto distinct circuits in the brain.

With so many respondents, the results also provided a wealth of new information about how factors such as age, gender and the tendency to play computer games influence our brain function. "Regular brain training didn't help people's cognitive performance at all yet aging had a profound negative effect on both memory and reasoning abilities," says Owen.

Hampshire adds, "Intriguingly, people who regularly played computer games did perform significantly better in terms of both reasoning and short-term memory. And smokers performed poorly on the short-term memory and the verbal factors, while people who frequently suffer from anxiety performed badly on the short-term memory factor in particular".

To continue the groundbreaking research, the team has launched a new version of the tests at http://www.cambridgebrainsciences.com/theIQchallenge

"To ensure the results aren't biased, we can't say much about the agenda other than that there are many more fascinating questions about variations in cognitive ability that we want to answer," explains Hampshire.

###

For interviews with Adrian M. Owen or Adam Hampshire from Western's Brain and Mind Institute, please contact Jeff Renaud, Senior Media Relations Officer, Western University at 519-661-2111, ext.

Western University-led research debunks the IQ myth 12/19/12 9:20 AM

http://www.eurekalert.org/pub_releases/2012-12/uowo-wur121712.php Page 2 of 2

85165 (office), 519-520-7281 (mobile), jrenaud9@uwo.ca

For full article, images and video, please visit http://www.uwo.ca/its/brain/news/iqmyth.html

AAAS and EurekAlert! are not responsible for the accuracy of news releases posted to EurekAlert! by contributing institutions or for the use of any information through the EurekAlert! system.

HOME DISCLAIMER PRIVACY POLICY TERMS & CONDITIONS CONTACT US TOP

Copyright ©2012 by AAAS, the science society.

Appendix D

One comment on the Neuroskeptic blog in response to earlier critical comments about the Hampshire et al. paper (italics added)

Anonymous said...

Neuroskeptic,

This seems like a mostly accurate assessment of the article however your closing example is a bit misleading. The article claims, that different types of intelligence relate to different brain networks. It also claims, that while one can generate a higher order 'g' factor from cross-component correlations, the neural basis of that factor is ambiguous. The article also suggests that the brain imaging data may be used to determine what the likely neural basis of that 'g' factor is. The cross-component correlations that may be used to generate a higher order 'g' factor are reported in one of the main figures in the article. However, what is evident is that those correlations are accurately predicted by the fact that some of the tasks have substantial loadings on multiple brain networks.

You write that 'although there was a 'g' factor statistically, it was explained by the fact that tasks required both the memory and the logic networks' and that consequently, 'it doesn't matter. If all tasks require both memory and reasoning, 'then the sum of someone's memory and reasoning ability is in effect a g score'.

In one sense this is the case, the tendency for tasks to load on multiple system in the brain is likely to be a large part of the basis of the 'g' factor. Indeed, this is the conclusion drawn in the article. However, the problem is that not all tasks did require both networks, or at least, not to a significant extent. Specifically, in some task contexts, the networks were very strongly dissociated when measured relative to rest. That is, some tasks had very little in the way of loading on one functional brain network alongside a very heavy loading on another – this is also reported in the article. This observation from the brain imaging analysis is paralleled by the very weak bivariate correlations between the self-same tasks in the behavioural analysis. For example, the short-term memory task – basically a variant on Corsi block tapping – correlated at about $r = 0.05$ with the deductive reasoning task. Clearly, these depend upon quite separate abilities, as both have good communalities with the battery of tasks as a whole but have a miniscule correlation with each other. One can design all sorts of tasks that load heavily on multiple processes; undoubtedly complex tasks will always load on many different systems in the brain and multiple abilities. However, the study provided little evidence for the influence of a monolithic intelligence factor over those abilities when the brain imaging data were taken into account. Thus, they should be considered independent from one another.

As for whether a composite score, generated from all factors is a better predictor of demographic variables. This issue is also addressed directly in the article. There are instances, in which such a score would show differences in two distinct population measures, when the underlying basis of those differences was quite distinct. Thus, a multifactor model is more informative.

Similarly, some correlations were greater when first level components were examined separately. Thus, a multifactor model may be more sensitive to population differences as well.

Finally, a critical comment was submitted to Neuron however, there was no 'conspiracy'. It was decided, based on feedback from an independent reviewer, that the author of the comment was heavily biased and that the criticisms raised were lacking in substance. Also, the authors of the article demonstrated that they were both willing and able to address all of those criticisms point by point if the journal chose to publish them. This is a highly controversial topic.

No doubt many other researchers will wish to comment on the results and will hold different views. Only those that raise sensible questions should be published. As for comment

13:28, anyone who takes that type of tone in a scientific debate, is self evidently a troll!

25 December 2012 17:23

References

Barbey, A. K., Colom, R., Solomon, J., Krueger, F., Forbes, C., & Grafman, J. (2012). An integrative architecture for general intelligence and executive function revealed by lesion mapping. *Brain*, *135*, 1154–1164. http://dx.doi.org/10.1093/Brain/Aws021.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies.* Cambridge; New York: Cambridge University Press.

Colom, R., Haier, R. J., Head, K., Alvarez-Linera, J., Quiroga, M. A., Shih, P. C., et al. (2009). Gray matter correlates of fluid, crystallized, and spatial intelligence: Testing the P-FIT model. *Intelligence*, *37*(2), 124–135.

Colom, R., Karama, S., Jung, R. E., & Haier, R. J. (2010). Human intelligence and brain networks. *Dialogues in Clinical Neuroscience*, *12*(4), 489–501.

Deary, I. J., Penke, L., & Johnson, W. (2010). The neuroscience of human intelligence differences. *Nature Reviews Neuroscience*, *11*(3), 201–211. http://dx.doi.org/10.1038/nrn2793 (nrn2793 [pii]).

Deary, I. J., Penke, L., & Johnson, W. (2010). The neuroscience of human intelligence differences. *Nature Reviews Neuroscience*, *11*(3), 201–211. http://dx.doi.org/10.1038/Nrn2793.

Duncan, J., & Owen, A. M. (2000). Common regions of the human frontal lobe recruited by diverse cognitive demands. *Trends Neurosci*, *23*, 475–483.

Glascher, J., Rudrauf, D., Colom, R., Paul, L. K., Tranel, D., Damasio, H., et al. (2010). Distributed neural system for general intelligence revealed by lesion mapping. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(10), 4705–4709. http://dx.doi.org/10.1073/pnas.0910397107 (0910397107 [pii]).

Gottfredson, L. S. (1997). Why g matters: The complexity of everyday life. *Intelligence*, *24*(1), 79–132.

Graham, S., Jiang, J. Y., Manning, V., Nejad, A. B., Zhisheng, K., Salleh, S. R., et al. (2010). IQ-related fMRI differences during cognitive set shifting. *Cerebral Cortex*, *20*(3), 641–649. http://dx.doi.org/10.1093/Cercor/Bhp130.

Haier, R. J., Colom, R., Schroeder, D. H., Condon, C. A., Tang, C., Eaves, E., et al. (2009). Gray matter and intelligence factors: Is there a neuro-g? *Intelligence*, *37*(2), 136–144.

Haier, R. J., Jung, R. E., Yeo, R. A., Head, K., & Alkire, M. T. (2005). The neuroanatomy of general intelligence: Sex matters. *NeuroImage*, *25*(1), 320–327.

Hampshire, A., Highfield, R. R., Parkin, B. L., & Owen, A. M. (2012). Fractionating human intelligence. *Neuron*, *76*(6), 1225–1237. http://dx.doi.org/10.1016/J.Neuron.2012.06.022.

Jensen, A. R. (1998). *The g factor: The science of mental ability.* Westport: Praeger.

Johnson, W., & Bouchard, T. J. (2005). The structure of human intelligence: It is verbal, perceptual, and image rotation (VPR), not fluid and crystallized. *Intelligence*, *33*(4), 393–416. http://dx.doi.org/10.1016/J.Intell.2004.12.002.

Johnson, W., te Nijenhuis, J., & Bouchard, T. J. (2008). Still just 1 g: Consistent results from five test batteries. *Intelligence*, *36*(1), 81–95. http://dx.doi.org/10.1016/J.Intell.2007.06.001.

Jung, R. E., & Haier, R. J. (2007). The parieto-frontal integration theory (P-FIT) of intelligence: Converging neuroimaging evidence. *Behavioral and Brain Sciences*, *30*(02), 135–154.

Karama, S., Colom, R., Johnson, W., Deary, I. J., Haier, R., Waber, D. P., et al. (2011). Cortical thickness correlates of specific cognitive performance accounted for by the general factor of intelligence in healthy children aged 6 to 18. *NeuroImage*, *55*(4), 1443–1453. http://dx.doi.org/10.1016/J.Neuroimage.2011.01.016.

Penke, L., Maniega, S. M., Bastin, M. E., Valdes Hernandez, M. C., Murray, C., Royle, N. A., et al. (2012). Brain white matter tract integrity as a neural foundation for general intelligence. *Molecular Psychiatry*. http://dx.doi.org/10.1038/mp.2012.66 (mp201266 [pii]).