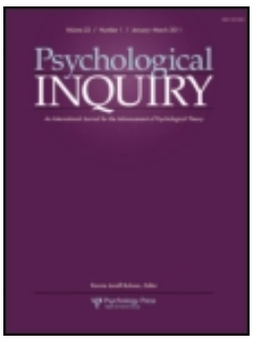
Getting Spearman off the skyhook

One more in a century (since Thomson, 1916) of attempts to vanquish g



Routledge
Taylor & Francis Group

Psychological Inquiry

An International Journal for the Advancement of Psychological Theory

ISSN: 1047-840X (Print) 1532-7965 (Online) Journal homepage: http://www.tandfonline.com/loi/hpli20

Getting Spearman off the Skyhook: One More in a Century (Since Thomson, 1916) of Attempts to Vanquish g

Ian J. Deary, Simon R. Cox & Stuart J. Ritchie

To cite this article: Ian J. Deary, Simon R. Cox & Stuart J. Ritchie (2016) Getting Spearman off the Skyhook: One More in a Century (Since Thomson, 1916) of Attempts to Vanquish g, Psychological Inquiry, 27:3, 192-199, DOI: 10.1080/1047840X.2016.1186525

To link to this article: http://dx.doi.org/10.1080/1047840X.2016.1186525

Published online: 02 Aug 2016.

PSYCHOLOGICAL INQUIRY
2016, VOL. 27, NO. 3, 192–199
http://dx.doi.org/10.1080/1047840X.2016.1186525

Routledge
Taylor & Francis Group

Getting Spearman off the Skyhook: One More in a Century (Since Thomson, 1916) of Attempts to Vanquish *g*

Ian J. Deary[a], Simon R. Cox[a,b], and Stuart J. Ritchie[a]

[a]Centre for Cognitive Ageing and Cognitive Epidemiology and Department of Psychology, University of Edinburgh, Edinburgh, United Kingdom; [b]Scottish Imaging Network, a Platform for Scientific Excellence (SINAPSE) Collaboration, Edinburgh, United Kingdom

Introduction

We provide seven responses to Kovacs and Conway's wide-ranging theory of intelligence differences. In the first six, we: reflect on the past theories that can be heard in this new one and how they have fared; discuss whether, in their present state, cognitive processes inferred from mental tests can be considered isomorphic with brain processes and can bear explanatory weight in theories of intelligence; and suggest that the positive manifold might be a formative biological latent trait while probably being a reflective psychological one. In the seventh, we attempt to test some hypotheses from Process Overlap Theory in our own Lothian Birth Cohort 1936 data.

Process Overlap Theory as a Palimpsest

Knowing the history of the many attempts to explain the positive correlations between mental tests almost hinders the assessment of this one. We kept seeing the ghosts of past theories in and between the lines of the writing. One of us has recently reexplained and reexamined (and recompared with others' theories) Godfrey Thomson's "bonds/sampling" theory of intelligence—to which Kristof Kovacs and Andrew Conway (this issue) concede they owe a debt—and we see strong similarities (Bartholomew, Allerhand, & Deary, 2013; Bartholomew, Deary, & Lawn, 2009). Well, yes and no: The account that Kovacs and Conway (this issue) give is more like Thomson's (1916) initial dice-throwing idea and is less informed by the later (e.g., Thomson, 1939) sampling/bonds theory, which did not propose a small number of separate processes, instead positing a huge number of neural/cognitive entities that are sampled by different tests: "Instead of showing that the mind has a definite structure, being composed of a few factors which work through innumerable specific machines, the low rank shows that the mind has hardly any structure" (Thomson, 1939, p. 270).

However, we should also be fair to Kovacs and Conway, and state that Thomson's various writings can be perused to get a slightly different reading of the brain substrate for his theory, one that sounds similar to the Carroll–Horn–Cattell hierarchy (Carroll, 1993). At a 1939 symposium at which Spearman and Burt were also speaking, Thomson summed up as follows:

> I myself lean at the moment more towards Spearman's *g* and his later group factors than I do to Thurstone's, since they seem to me more in accord with the ideas of my own Sampling Theory. On that theory *g* is as it were the whole mind, and the tests are part of *g*, not *g* part of the tests. And were that mind entirely undifferentiated, structureless, *g* would be the only factor needed. As the complexity of the mind, and the complexity of the upper brain, is organized (partly by the maturing of hereditary bonds, mainly I fancy by education and life) and integrated into "pools", "clusters", call them what you will, so additional factors, additional descriptive coefficients, are needed. (Thomson, 1939/1940, p. 106)

We agree that, based on current evidence, one cannot now choose between Spearman's and Thomson's ideas either statistically or biologically (Bartholomew et al., 2009). However, two things about Thomson's ideas were not, but should be, recognized by Kovacs and Conway. First, with Cyril Burt, we agree that Thomson's mature theory might have been a different way of stating Spearman's theory:

> … (to put it crudely) a homogenous brain, consisting merely of a very large number of similar nerve cells, identical in nature and in strength, would obviously be a brain governed by a single general factor. In short, there is no mathematical difference between assuming only a single factor, varying continuously, and assuming an infinite (or indefinitely large) number of unit factors forming a single homogeneous "pool". A bushel of wheat is still a bushel, whether we call it corn or insist that it is composed of innumerable grains. (Burt, 1940, p. 160)

Second, Thomson made clear that his theory tried to make the explanatory construct an aspect(s) of the brain, whereas he saw that *g* was an aspect of the tests (which brings us on, later, to mental test-brain structure/function isomorphism):

> The difference in point of view between the sampling theory and the two-factor theory [of Spearman] is that the latter looks upon *g* as being part of the test, while the former looks upon the test as being part of *g*. The two-factor theory is therefore compelled to postulate specific factors to account for the remainder of the variance of the test, and has to go on to offer some suggestions as to what specific factors are—perhaps, neural engines [Cf Anderson, 1992]. The sampling theory simply says that the test requires only such and

CONTACT Ian J. Deary i.deary@ed.ac.uk Centre for Cognitive Ageing and Cognitive Epidemiology, Department of Psychology, 7 George Square, Edinburgh, EH8 9JZ.

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/hpli.

> such a fraction of the bonds of the whole mind—the same fraction which, on the two-factor theory, *g* forms of the variance of the test. (Thomson, 1939, pp. 281–282)

Given that Thomson is often recounted by many as Spearman's longest serving opponent of *g*, we should remind readers that he did not think that *g* was incorrect, only that it was, he judged, one among other possible explanations for the positive correlations among mental tests. He saw it as his duty to think of other possible accounts. However, he left it to the last paragraph of his long obituary of Spearman, that is, after Spearman could have heard him say it, to conclude, "Probably there is a general factor of intelligence" (Thomson, 1947, p. 382).

Can Cognitive Processes Bear Explanatory Weight?

One of us has already dealt with this issue at length in a book (Deary, 2000) that was devoted to assessing reductionist accounts of human intelligence differences. The levels of reduction considered were cognitive process accounts, accounts based on varieties of reaction time procedures, psychophysical accounts, and brain biology and genetics. The various cognitive process accounts were found to be unsatisfactory, because they did not reduce or explain:

> The nagging worry is that this area of research, frequently employing sophisticated modelling procedures, has done little more than neatly and attractively pull apart the layers of the psychometric layer cake. The slices can all be pushed together to reconstruct the cake, but they have not, in truth, revealed what we wanted: the ingredients and procedures of the recipes for different sponges and fillings. As Stauffer, Ree, and Carretta (1996, p. 193) commented, "Despite theoretical foundations and arguments, cognitive components tests appear to measure much the same thing as traditional paper-and-pencil tests." (Deary, 2000, p. 144)

We apologize for the extended MacArthur Park cake metaphor, but the point is that truly explanatory, reductionist science does not merely redescribe phenomena; rather, one needs lower level, validated concepts from which to build correspondences. We judge that psychology—including cognitive psychology and psychophysics—when not rooted in validated brain mechanisms has largely failed in this regard. Just as we criticize any psychometric intelligence researchers who imagine their hierarchical structural equation models that show patterns of cognitive covariance can actually tell them how the brain is fashioned, we still challenge cognitive psychologists to show a brain account (a mechanistic molecular account, not just correlations with, for example, brain imaging variables) of differences in a complex cognitive test that rises above naïve isomorphism, that is, claiming a distinct brain process can be seen in their atom-splitting of a mental test. We think we'll have to be patient in waiting for an account of why mental tests covary, because we understand too little as yet about which brain variables underlie cognitive differences.

So, to put together our first two points—that others have mostly been this way before, and that cognitive processes are rarely validated entities that can do explanatory work (they are "skyhooks" and not "cranes," according to Dennett's, 1995, typology)—and apply them to Kovacs and Conway's pithiest statement of their theory as follows (with our responses in brackets):

> The briefest possible summary of its central assumption is that any test item or cognitive task requires a number of domain-specific as well as domain-general cognitive processes [Spearman, Thomson, and Anderson, 1992, among others, said this sort of thing]. The domain-general processes that are central to performance on cognitive tests are primarily the ones that are identified as executive processes in cognitive psychology in general and the working memory literature in particular [So-called executive processes and working memory have been found empirically to be almost exact redescriptions of *g* and come from the sorts of mental tests that produce *g*; Diamond, 2013, said, in his Figure 4, that for two of the three higher level executive functions—reasoning and problem solving—"fluid intelligence is completely synonymous with these"; and Kyllonen and Christal, 1990, showed near-identity—between working memory and reasoning]. Such processes are recruited by a large number of test items, alongside domain-specific processes, which are tapped by items appearing in specific types of tests only. In turn, domain-general executive processes overlap with domain-specific processes more than the domain-specific processes overlap with one another. [These sound similar to Thomson's "pools"/"clusters" of bonds.]

Thus, the pattern of thinking here is a Spearman–Thomson–Anderson (1992) hybrid (pushing psychometric correlations around in an explanatory kaleidoscope), and we doubt the validity of the explanatory variable(s).

A Psychological "Grab Bag"

The Kovacs and Conway (this issue) article is to some extent a "grab bag" that includes both core content and some items that we think are of less quality and importance in the field of intelligence differences. It is our judgement call that the differentiation effect and the worst performance rule are, if they actually exist, relatively small-scale phenomena that are not particularly important for a general theory of intelligence to explain. One of us has also previously examined Duncan's goal neglect task and the kernel component that was supposed to account for *g* variance (Deary, 2000, pp. 136–140); the analysis found its construct validity wanting, although the correlations with the task were interesting. Another phenomenon brought to bear is the close correlation between fluid intelligence (Gf) and *g*; we do not agree that this is a cause for concern in the way that the authors do, and we have concerns about their psychometric argument to segregate them. This is based upon apparently different correlations between working memory and Gf ($r = .85$) and *g* ($r = .48$). However, the former statistic (Oberauer, Schulze, Wilhelm, & Süss, 2005) was based on a reanalysis of the same data as the latter (Ackerman, Beier, & Boyle, 2005). The difference was that Ackerman et al. (2005) opted to fix their manifest-to-latent loadings based on a previous model. The correlation between *g* (measured the same way in both publications) was substantially increased when these paths were not fixed (Oberauer et al., 2005). We deal with the further efforts to separate *g* and Gf in the empirical section of this commentary. In contrast, the crystallized–fluid intelligence division is a useful one, particularly for describing ageing effects. We see fluid intelligence being brain-as-knowledge-making-machine, using external or internal stimuli to operate on and crank out new stuff, and crystallized intelligence being brain-as-knowledge-warehouse, manifested when we bring already-stored items of knowledge to our or others' consciousness.

Other Causes of *g* Variance

Kovacs and Conway (this issue) recognize that, across the human brain, the connecting white matter shows a latent factor whereby some people's connections tend generally to be healthier than others' and that this accounts for about 10% of the variance in the general factor extracted from multiple cognitive tests (Penke et al., 2012). So, they conclude, it might be that some general brain variance underlies most mental tasks, putting a limit on performance. We commend their pluralism here, in thinking that there might be some sources of general brain variance (they call it ζ) in addition to their favored cognitive processes in explaining the positive manifold. However, they never say precisely how much variance in *g* they predict to be explained by process overlap as compared to other sources of variance (like brain integrity or mutualism). The 10% figure comes from only one relevant brain measure; one of our recent publications ups the number to 20% with the addition of multiple other *g*-related measures of brain integrity, at least in older age (Ritchie et al., 2015). In the future, more advanced tools will probably increase the variance explained even further. If this proportion of explained variance rises markedly, will Kovacs and Conway still see room for cognitive processes as formative contributors to explaining the positive manifold?

We see no reason why the biological contributions to *g* should be reflective; rather, *g* could be characterized as the formative result of multiple (sometimes uncorrelated) aspects of biological makeup. This leaves open the possibility that *g* is a formative construct at the biological level and a reflective construct at the psychometric level. Vernon and Weese (1993) noted such a prospect with reference even to multiple uncorrelated (rotated) aspects of information processing contributing to *g* though, again, we would question the reductionist validity of these variables. We provide a small-scale empirical demonstration of this "formative biology of *g*" idea below.

We should state that we have doubts as to whether there is a level of explanatory constructs, at the cognitive level, that lies between *g* and specific test variance and "form" *g*, "results with tasks that are indeed elementary, and supposedly tap a small number of cognitive processes, show that *g* reflects a number of independent sources" (Kovacs & Conway, this issue, p. 162). As far as we see, performances on so-called elementary (they never are!) cognitive tests are reflections of, rather than formative of, *g* (see Luciano et al., 2005, and Plomin & Spinath, 2002, Figure 3, for discussion of this at the genetic level). Also, we judge that a set of biological (which of course includes environmental) formative variables that contribute to *g*—that is, a more or less efficient brain—is a more likely and tractable hypothesis than a set of psychological skyhooks, as Kovacs and Conway suggest when they argue that "tests indeed reflect specific abilities, which do have ontological reality [Really?!]" (p. 162).

We think it is likely that, at the biological level, there will be some contributors to domain-level and more specific cognitive performance, as well as to general cognitive ability. Insofar as Kovacs and Conway agree with this, it is a restatement of Anderson's (1992) theory of intelligence differences. He envisaged a "basic processing mechanism" on which all cognitive tests were implemented, which had individual differences, and which therefore contributed variance to differences in all cognitive tests. He also thought there were "specific processors" that dealt with types of mental problems (he mentions, e.g., spatial and verbal) and that showed individual differences that might be uncorrelated with each other and with the basic processing mechanism. In retro terms, Anderson's ideas might be translated into a cassette player (the basic processing mechanism on which all one's tapes are played and that is more or less hi- or lo-fidelity) and one's collection of cassette tapes (the specific processors that will have to bear the limitations of the cassette player in order to be heard, and that have their own quality variance, which has aspects not shared by other tapes). That set of ideas—of there being mostly general brain limitations, and some limitations that affect only specific types of test—accords quite well with models by Spearman (1904, 1927) and data collected from then onward.

So, when Kovacs and Conway (this issue) write, "Even if someone were, in theory, capable of successful performance on the domain-specific aspect of a mental test item, he or she might be unable to arrive at a correct answer because of failing to meet its executive attention demands" (p. 162), these are the limitations modelled by Anderson (1992; i.e., that a perfectly serviceable cassette tape cannot be heard on a damaged cassette player) and can sit on a "basic processing mechanism" that is a psychometric reflective *g* formed by partly uncorrelated biological influences (i.e. a generally more or less efficient brain). With reference to the item response theory equation, we think that the pattern of errors they strive to explain with a cognitive process model can be accounted for in part by biological influences on specific domains of cognitive functioning, influences that are additional to any effect they have on *g* (see below).

Generally, we think there is some naïve cognitive process–brain structure/function isomorphism in the target article. For example, Kovacs and Conway (this issue) state that "test developers devote a lot of time and effort to constructing unidimensional measures, tests that purportedly tap a single ability only" (p. 165). But do they—and do the test developers—really think we know the abilities, in terms of processes in the brain, that are tapped by these tests? We can describe test similarities, but we are wise to be agnostic about what stimulus-mincing and computing goes on in the head to solve them. Some of the material in the piece that appears to suggest that one can divine the brain's functional lineaments from what we can rationally think about a mental test's contents recapitulates the dry Casaubonian scholarship of, for example, Carpenter, Just, and Shell (1990) on Raven's matrices (see the critique by Deary, 2000). We think one must understand the processing structure and limitations of the brain and then join that to mental test performance; mental test performance will tell us only so much—perhaps not much—about what the brain does and how. Translated to the kidney, in the study of cognitive differences we are still admiring and classifying the variety of colors in our urine while we await the discovery of the nephron.

The Mysterious Figure 8

We stared at Kovacs and Conway's (this issue) key Figure 8—their core astrological chart purporting to explain why some people are cleverer than others—for ages, trying to work out what it stated explicitly and how to test that. If, we thought, we

could crack the code of this mandala, we might find a make-or-break hypothesis in the article. Our plight was not helped by the fact that the relevant section of the article— titled "Process Overlap Theory"—stops short of clearly elucidating the overlap of the cognitive processes at the domain level of their hierarchy in Figure 8. Instead, executive functions are shown as a constellation of indistinguishable black dots. The degree to which one dot equates to another across domains remains opaque. There are tantalizing hints in the text (such as the idea that cognitive inhibition is required across number series items, verbal analogies and matrix reasoning), yet the missed opportunity to render this, and other such specifics, more clearly diminishes opportunities to create a testable, falsifiable theory. We confess we feel as if we might not fully have unpicked and understood Figure 8 and its accompanying text, and we should like to have grilled the authors on it; we do not rule out that we could have missed some key ideas.

Big Theory, Small Data

Intelligence research, as one of us has previously argued, has a plethora of flashy and eye-catching "big theories" that, ultimately, have not been productive:

> Like trying to decorate a house while a hyperactive toddler runs around messing things up and forcing one to do trivial tidying instead of long-term renovation, a theory can keep one busy refuting or operationalising its aspects instead of focussing on less immediately compelling, but fundamentally more important, sensible empirical advances. … Big theories divert people from the available empirical evidence and get them arguing instead about the evidence can be forced into their scheme. (Deary, 2000, pp. 108–109)

We data-gathering wallflowers can therefore appear grumpy and jealous, as we follow our hair-shirt credos that, first, gathering relevant and preferably large amounts of data from both brain and behavior and creatively understanding their associations is likely to be helpful and, second, recognizing and admitting that the tools and concepts are probably not in place yet to truly understand intelligence differences. More evidence-based intelligence research is required. We admit that this, though perhaps correct, is rather boring:

> At the risk of appearing unutterably dull, and to compound the felony of being against fanciful theory, one has to urge more replicated studies, more inter-laboratory agreements on the operationalisation of constructs and parameters to be measured, and generally larger masses of data on the same topic so that one may hypothesise from solid ground. To listen to discussions within the intelligence community is sometimes like watching an archaeologist who has dug a trench one foot square and is speculating from that rather than widening the trench. (Deary, 2000, p. 110)

To be clear, the problem is not with the constructing of a theory per se, it is the distance between the theory and the relevant data. To understand cognitive differences and how variance in them is parsed in the brain, one needs enough good cognitive and brain data, and sufficient isomorphism between them. We have types of mental tests—for which some are "desperately seeking a mental cytology" (Deary, 2000, p. 88)—and a good idea about how they covary, and models that arrange and display that covariance. We don't have the mechanistic brain constructs to which we can map these packets of covariance beyond relatively gross measures (such as those of brain macrostructure, blood oxygenation, and neuroanatomy, which provide only indirect—though valuable—intimations of the true neurobiological nature of cognitive processes; e.g. Zald, 2007). Identifying the existence of a cognitive process using psychometric properties alone does not necessarily correspond to the way in which the human brain gives rise to the behavioral phenomenon being measured.

Metaphorizing again, the effort to understand the psychobiology of intelligence has a resemblance with digging the tunnel between England and France: We hope, with workers on both sides having a good sense of direction, that we can meet and marry brain biology and cognitive differences. To date, though many have used them to begin the biology-side-digging, we have to admit that variables like brain size and white matter "integrity," though they have produced interesting and replicable correlations with intelligence, are not close to the sort of mechanistic understanding a true reductionist desires. However, it is (using Dennett's, 1995, concepts again) at least some progress using "cranes" rather than psychological process "skyhooks."

Some Empirical Tests

Consistent with our role as biology-side tunnelers, our task to provide commentary would be incomplete without putting our backs into some empirical testing of several points arising from the target article. We address two specific predictions gleaned from the Kovacs and Conway article, followed by a more general point: (a) the strength of the positive manifold varies as a function of frontal lobe atrophy; (b) *g* cannot be localized, whereas Gf can; and (c) the formative biology of *g*. We test each of these using cognitive, genetic, and brain-imaging data from the second wave of the Lothian Birth Cohort 1936 (for which details can be obtained from Deary et al., 2007; Deary, Gow, Pattie, & Starr, 2012; Wardlaw et al., 2011). Although we are still unclear as to whether the following are genuinely unique predictions of process overlap theory, one of the benefits of "big theory" is that it raises several points that one can empirically test.

Domain-Generality of the Positive Manifold and Frontal Lobe Atrophy

In their final paragraph, Kovacs and Conway (this issue) describe a number of predictions made by process overlap theory. One is that

> process overlap theory predicts that age patterns of the maturation as well as aging of the prefrontal cortex and thus of executive processes should determine the domain-generality of the positive manifold. However, this prediction might be difficult to test, because different executive processes show very different developmental and ageing patterns, and there is a large individual variation the maturation and aging process itself. (p. 172)

We take this to mean that the positive manifold of intelligence should become stronger as a function of greater prefrontal atrophy (the structural integrity of which is central to executive processes). An adequate test of this must also address the additional two caveats provided by Kovacs and Conway (this issue). First, executive processes show different ageing patterns. One plausible reason for

reports of heterochronicity in the ageing of executive functions may be because not all executive processes are equally supported by the frontal cortex (Andrés, Guerrini, Phillips, & Perfect, 2008), nor do all such functions necessarily receive equal support from precisely the same frontal subregions (Kievit et al., 2014; MacPherson, Della Sala, Cox, Girardi, & Iveson, 2015). Comparative differences in executive test reliabilities and/or the psychometric treatment of memory and fluid variables may also partly drive their observed differential age effects (Johnson, Logie, & Brockmole, 2010; Kievit et al., 2014). Kovacs and Conway are consistent in their attribution of executive processes to the frontal lobes in general, and particularly with respect to Gf and Gv (their Figure 8). Thus, one could infer that a measure of prefrontal atrophy would more strongly index the age effects on those executive processes more heavily supported by this region. In their second caveat, they rightly acknowledge that the link between chronological age and biological aging varies from person to person. Fortuitously, the sample in which we test the prediction, the Lothian Birth Cohort 1936, has an extremely narrow age range (all were born in 1936), minimizing this concern.

In this sample of 681 participants with useable MRI data at a mean age of 72.64 years (SD = 0.72), we used Freesurfer v.5.3 (http://surfer.nmr.mgh.harvard.edu/) and the Desikan-Killiany atlas (Desikan et al., 2006) to derive a measure of each participant's frontal lobe volume (summing the volumes of the following regions: superior frontal, middle frontal, rostral middle frontal, middle orbitofrontal, lateral orbitofrontal, frontal pole, rostral and caudal anterior cingulate and the inferior frontal pars opercularis, pars triangularis and pars orbitalis). We corrected the measure for intracranial volume (maximum healthy brain size in younger adulthood) to produce a proxy measure of frontal lobe atrophy.

We then used a moderated confirmatory factor analysis model (Tucker-Drob, 2009) to calculate the extent of (de)differentiation of cognitive abilities—indexed by a varied battery of thirteen tests, organized into four domains as previously described by Tucker-Drob, Briley, Starr, and Deary (2014), and corrected for age and sex—according to the extent of frontal atrophy. We found a result that was, to an extent, in line with the prediction of process overlap theory: The estimated factor communality (the % of the total variance across the cognitive tests explained by the factor) was 23.6% higher in individuals with the greatest rates of atrophy than in those with the least atrophy (52.7% vs. 29.1%). However, the wide confidence interval on the estimate, as shown in Figure 1, means that this communality difference was not statistically significant.

As previously noted, we are not certain whether this prediction is specific to process overlap theory. We would expect individuals with more atrophy, and thus smaller frontal lobes, to have lower intelligence. Thus, the prediction can be seen as simply a restatement of the idea of ability differentiation. If this is so, it is certainly not a new prediction. Nevertheless, we provide the result here for further discussion.

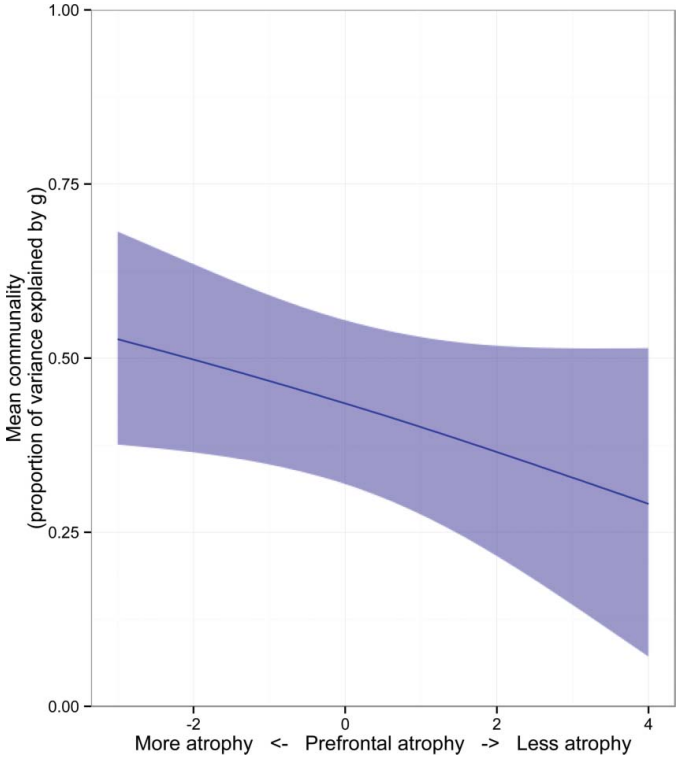


Figure 1. Test of prefrontal differentiation. *Note.* Shaded area around the mean communality line is the 95% confidence interval. The *x*-axis is in standard deviation units.

Localization of g and Gf in the Brain

Kovacs and Conway (this issue) make the following statement in their section "Overlapping Networks in the Brain": "…even though [Gf] is statistically identical to *g*, imaging studies demonstrate their dissociability; whereas *g* cannot be localized, Gf is linked to the prefrontal (primarily dorsolateral) and partly to the (primarily posterior) parietal cortex with remarkable consistency" (p. 167). The strong claim that *g* cannot be localized, whereas Gf can, in spite of their statistical near-unity, is to ignore the raft of potential cross-study differences, low sample sizes, and imaging modality limitations, as well as some studies that do identify neural correlates of *g* in the very areas Kovacs and Conway assert are the exclusive preserve of Gf (reviewed in, e.g., Colom & Thompson, 2011). Moreover, the claim that studies "demonstrate their dissociability" would require at least one study to have directly compared the neural correlates of *g* and Gf within the same sample, finding the former to be absent and/or nonoverlapping with the latter. Because we are not aware of any such study, we attempted one here.

Kovacs and Conway argue that current brain research reports neural correlates of *g* are so diverse that consistent localization is prohibited, in contrast to the correlates of Gf, which include mainly dorsolateral prefrontal and parietal cortices. A direct test of the contention that *g* and Gf are neuroanatomically dissociable requires an adequately powered study in which these two factor scores could be created in the same population using appropriate, but nonoverlapping, cognitive tests, and on whom brain MRI data are available. To this end, we (again using data from the Lothian Birth Cohort 1936) examined the subregional volume and surface area correlates of *g* and Gf across the frontal and parietal lobes. To construct *g*, we used Wechsler Adult Intelligence Scale–III Digit-Symbol Substitution, a test of Choice Reaction Time, Wechsler Memory Scale–III Verbal Paired Associates, the National Adult Reading Test, and Verbal Fluency (see Deary et al., 2007, for all references and descriptions). To construct Gf, we used Matrix

Table 1. Associations between frontal and parietal cortical surface area, *g*, and Gf.

Lobe	Region	Hemisphere	Association with *g*	*sig*	Association with Gf	*sig*
Frontal						
	Dorsolateral	L	.126	**	.123	**
		R	.134	**	.127	**
	Inferior frontal	L	.089	*	.073	
		R	.053		.042	
	Lateral orbital	L	.095	*	.084	*
		R	.151	***	.130	**
	Medial orbital	L	.091	*	.082	*
		R	.163	***	.142	***
	Caudal ACC	L	.058		.050	
		R	.066		.066	
	Rostral ACC	L	.157	***	.141	***
		R	.127	**	.125	**
	Caudal middle	L	.058		.050	
		R	.066		.066	
Parietal						
	Superior	L	.107	*	.102	*
		R	.097	*	.084	*
	Inferior	L	.087	*	.078	
		R	.101	*	.099	*

Note. Variables corrected for age at scan or testing, respectively, and sex, prior to inclusion in model. Pearson's *r* reported. Association between *g* and fluid intelligence (Gf): $r = .983$, $p < .001$. L = left; R = right; ACC = Anterior Cingulate Cortex.
$^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

Reasoning, Block Design, Letter-Number Sequencing (from the Wechsler Adult Intelligence Scale–III) and Spatial Span (from the Wechsler Memory Scale–III). We estimated factor scores from a confirmatory factor model of both latent variables. We corrected both the MRI and cognitive measures for age and sex. The results are displayed in Tables 1 and 2. As mentioned by Kovacs and Conway, *g* and Gf were near-perfectly correlated ($r = .98$), but their cerebral correlates did not behave as the theory would predict. Not only were some regions associated with *g* (consistently left dorsolateral, left rostral cingulate, and bilateral parietal), but the magnitude of associations for all subregions for *g* and Gf were near-identical (vector correlation for surface area, $r = .98$, and for volume, $r = .99$). These data provide clear evidence that *g* and Gf are virtually identical in terms of bivariate associations, and with respect to their cortical correlates.

Table 2. Associations between frontal and parietal cortical volume, *g*, and Gf.

Lobe	Region	Hemisphere	Association With *g*	*sig*	Association With Gf	*sig*
Frontal						
	Dorsolateral	L	0.083	*	0.081	*
		R	0.070		0.067	
	Inferior frontal	L	0.075		0.067	
		R	0.028		0.020	
	Lateral orbital	L	0.037		0.031	
		R	0.045		0.029	
	Medial orbital	L	−0.003		0.001	
		R	−0.002		−0.013	
	Caudal ACC	L	0.007		−0.002	
		R	0.018		0.015	
	Rostral ACC	L	0.101	*	0.097	*
		R	0.059		0.059	
	Caudal middle	L	0.007		−0.002	
		R	0.018		0.015	
Parietal						
	Superior	L	0.128	**	0.121	**
		R	0.115	**	0.104	*
	Inferior	L	0.097	*	0.095	*
		R	0.131	**	0.129	**

Note. Variables corrected for age at scan or testing, respectively, and sex, prior to inclusion in model. Associations are Pearson's *r*. Association between *g* and fluid intelligence (Gf): $r = .983$, $p < .001$. L = left; R = right; ACC = Anterior Cingulate Cortex.
$^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

Formative Biology, Reflective g

To test the idea we discussed above, in which formative biological elements produce a reflective *g*, we took two broad-brush measures of the biological contribution to intelligence: intracranial volume (ICV) and a polygenic profile score for educational attainment created from summary data from a recent Genome-Wide Association Study (GWAS; Davies et al., 2016) and modeled their relation with cognitive tests. Again, this was tested in data from the Lothian Birth Cohort 1936.

Using a method similar to Tucker-Drob (2013; see section 1.3.3.), we tested whether ICV and the polygenic score were best modeled having common, independent, or common-plus-independent relations with *g* (in this case indicated by the same four domains of cognitive ability as used in the first empirical test, just discussed, each created from multiple tests). For both biological variables, the parsimonious common-plus-independent pathways model fit better than the common pathways model (*p*s < .02) and no worse than the independent pathways model (*p*s > .65). We combined the models for ICV and for

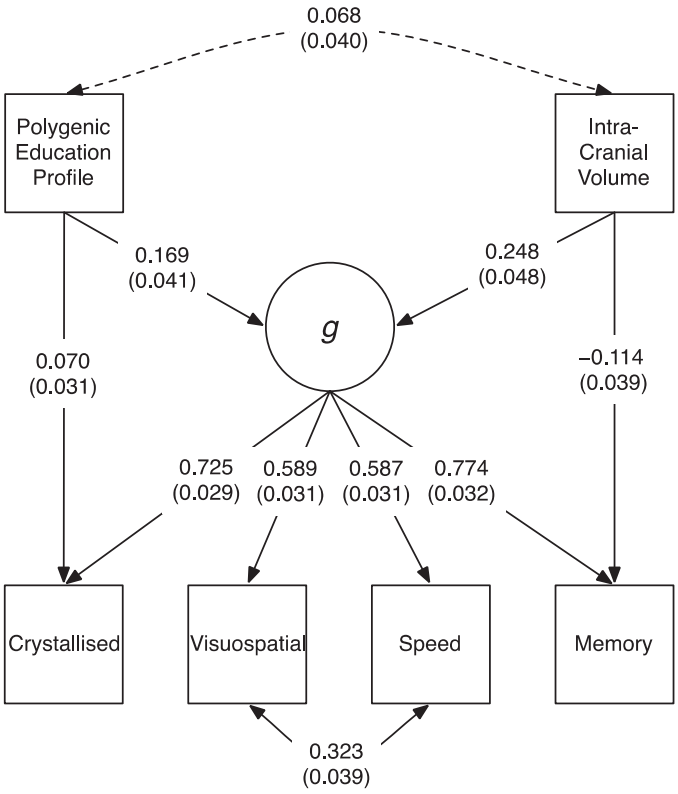


Figure 2. Combined common-plus-independent pathways model of the association of biological factors with *g*. *Note*. Values are standardized regression weights with standard errors in parentheses. The dotted line indicates a nonsignificant path.

the polygenic score as shown in Figure 2. This model had excellent fit to the data, $\chi^2(5) = 11.29$, $p = .046$, root mean square error of approximation = .04, comparative fit index = .99, Tucker-Lewis index = .98). Thus, a well-fitting model could be produced where the biological influences are on *g*, rather than the specific domains alone, though there were additional domain-specific paths as shown in the diagram. Whereas this analysis does not directly test a prediction of process overlap theory, it provides a small-scale example of a useful way to think about *g*: formative (and in this case, uncorrelated) biological elements giving rise to a reflective, psychometric general intelligence.

Conclusion

We applaud Kovacs and Conway's detailed synthesis. They address the greatest (though still most mysterious) empirical discovery and regularity in psychology: the positive correlations among diverse mental tests. They combine biology, cognitive neuroscience, and psychometrics in an attempt to understand the positive correlations. They recognize the value of the ideas of Thomson, a figure who has been relatively ignored and to whom we in Edinburgh owe so much; we thank them for their article in so far as it is a celebratory rediscovery of Thomson's (1916) theory, 100 years since his first throw (literally, of dice, in his slippers) at an alternative to Spearman's *g*. We trust that our at times seemingly crotchety remarks will be taken in an encouraging spirit: Kovacs and Conway's ideas made us engage our fluid and crystallized intelligence to think hard with both some novel and more familiar materials. In many places in the target article we wanted to ask questions and hear more from them. Perhaps our disagreements boil down to our putting more emphasis on what they call "ζ, the unique variance of *g*," than they do, and our skepticism that their cognitive processes are "ontologically real" (whatever that apparent pleonasm means). Now, though, because we've been banging on about the importance of empirical work on the biology of *g*, we had better get back to it.

Funding

Ian Deary and Stuart Ritchie are supported by the University of Edinburgh Centre for Cognitive Ageing and Cognitive Epidemiology, part of the cross-Council Lifelong Health and Wellbeing Initiative (MR/K026992/1). Funding from the Biotechnology and Biological Sciences Research Council and Medical Research Council is gratefully acknowledged. Ian Deary is also supported by the Medical Research Council's Dementias Platform UK (MR/L015382/1). Simon Cox is supported by the Medical Research Council (MR/M01311/1). The Lothian Birth Cohort 1936 (LBC1936) is supported by Age UK (Disconnected Mind project). We thank the LBC1936 participants and team for providing data for the empirical examples here.

References

Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2005). Working memory and intelligence: The same or different constructs? *Psychological Bulletin*, *131*, 30–60.

Anderson, M. (1992). *Intelligence and development: A cognitive theory*. Oxford, UK: Blackwell.

Andrés, P., Guerrini, C., Phillips, L. H., & Perfect, T. J. (2008). Differential effects of aging on executive and automatic inhibition. *Developmental Neuropsychology*, *33*, 101–123.

Bartholomew, D. J., Allerhand, M., & Deary, I. J. (2013). Measuring mental capacity: Thomson's bonds model and Spearman's g-model compared. *Intelligence*, *41*, 222–233.

Bartholomew, D. J., Deary, I. J., & Lawn, M. (2009). A new lease of life for Thomson's bonds model of intelligence. *Psychological Review*, *116*, 567–579.

Burt, C. (1940). *The factors of the mind*. London, UK: University of London Press.

Carpenter, P. A., Just, M. A., & Shell, P. (1990). What one intelligence test measures: A theoretical account of processing in the Raven's Progressive Matrices Test. *Psychological Review*, *97*, 404–431.

Carroll, J. B. (1993). Human cognitive abilities: A survey of factor analytic studies. Cambridge, UK: Cambridge University Press.

Colom, R., & Thompson, P. M. (2011). Understanding Human Intelligence by Imaging the Brain. In T. Chamorro-Premuzic, S. von Stumm, & A. Furnham (Eds) *The Wiley-Blackwell handbook of individual differences* (pp. 330–352) West Sussex, UK: Wiley Blackwell.

Davies, G., Marioni, R. E., Liewald, D. C., Hill, W. D., Hagenaars, S. P., Harris, S. E., & Deary, I. J. (2016). Genome-wide association study of cognitive functions and educational attainment in UK Biobank (N=112 151). *Molecular Psychiatry*. Advance online publication. http://dx.doi.org/10.1038/mp.2016.45.

Deary, I. J. (2000). Looking down on human intelligence: From psychometrics to the brain. Oxford, UK: Oxford University Press.

Deary, I. J., Gow, A. J., Pattie, A., & Starr, J. M. (2012). Cohort profile: The Lothian birth cohorts of 1921 and 1936. *International Journal of Epidemiology*, *41*, 1576–1584.

Deary, I. J., Gow, A. J., Taylor, M. D., Corley, J., Brett, C., Wilson, V., & Starr, J. M. (2007). The Lothian Birth Cohort 1936: A study to examine influences on cognitive ageing from age 11 to age 70 and beyond. *BMC Geriatrics*, *7*, 28.

Dennett, D. C. (1995). *Darwin's dangerous idea*. London, UK: Allen Lane.

Desikan, R. S., Segonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., & Killiany, R. J. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage*, *31*, 968–980.

Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, *64*, 115–168.

Johnson, W., Logie, R. H., & Brockmole, J. R. (2010). Working memory tasks differ in factor structure across age and cohorts: Implications for dedifferentiation. *Intelligence*, *38*, 513–528.

Kievit, R. A., Davis, S. W., Mitchell, D. J., Taylor, J. R., Duncan, J., Cam-CAN Research Team, & Henson, R. N. (2014). Distinct aspects of frontal lobe structure mediate age-related differences in fluid intelligence and multitasking. *Nature Communications*, *5*, 5658.

Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity?! *Intelligence*, *14*, 389–433.

Luciano, M., Posthuma, D., Wright, M. J., de Geus, E. J. C., Smith, G. A., Geffen, G. M., & Martin, N. J. (2005). Perceptual speed does not cause intelligence, and intelligence does not cause perceptual speed. *Biological Psychology*, *70*, 1–8.

MacPherson, S. E., Della Sala, S., Cox, S. R., Girardi, A., & Iveson, M. H. (2015). *Handbook of frontal lobe assessment*. Oxford, UK: Oxford University Press.

Oberauer, K., Schulze, R., Wilhelm, O., & Süss, H.-M. (2005). Working memory and intelligence—Their correlation and their relation: Comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin*, *131*, 61–65; author reply 72–75.

Penke, L., Maniega, S. M., Bastin, M. E., Hernandez, M. C. V., Murray, C., Royle, N. A., & Deary, I. J. (2012). Brain white matter integrity as a neural foundation for general intelligence. *Molecular Psychiatry*, *17*, 1026–1030.

Plomin, R., & Spinath, F. M. (2002). Genetics and general cognitive ability (*g*). *Trends in Cognitive Sciences*, *6*, 169–176.

Ritchie, S. J., Booth, T., Hernández, M. D. C. V., Corley, J., Maniega, S. M., Gow, A. J., & Deary, I. J. (2015). Beyond a bigger brain: Multivariable structural brain imaging and intelligence. *Intelligence*, *51*, 47–56.

Spearman, C. (1904). "General Intelligence" objectively determined and measured. *American Journal of Psychology*, *15*, 201–293.

Spearman, C. (1927). *The abilities of man*. London, UK: Macmillan.

Stauffer, J. M., Ree, M. J., & Carretta, T. R. (1996). Cognitive-components tests are not much more than *g*: An extension of Kyllonen's analyses. *The Journal of General Psychology*, *123*, 193–205.

Thomson, G. H. (1916). A hierarchy without a general factor. *British Journal of Psychology*, *8*, 271–281.

Thomson, G. H. (1939). *The factorial analysis of human ability*. London, UK: University of London Press.

Thomson, G. H. (1939/1940). The factorial analysis of ability: Agreement and disagreement in factor analysis. A summing up. *British Journal of Psychology*, *30*, 105–108.

Thomson, G. H. (1947). Charles Spearman. 1863–1945. *Obituary Notices of Fellows of the Royal Society*, *5*, 373–385.

Tucker-Drob, E. M. (2009). Differentiation of cognitive abilities across the life span. *Developmental Psychology*, *45*, 1097–1118.

Tucker-Drob, E. M. (2013). How many pathways underlie socioeconomic differences in the development of cognition and achievement? *Learning and Individual Differences*, *25*, 12–20.

Tucker-Drob, E. M., Briley, D. A., Starr, J. M., & Deary, I. J. (2014). Structure and correlates of cognitive aging in a narrow age cohort. *Psychology and Aging*, *29*, 236–249.

Vernon, P. A., & Weese, S. E. (1993). Predicting intelligence with multiple speed of information-processing tests. *Personality and Individual Differences*, *14*, 413–419.

Wardlaw, J. M., Bastin, M. E., Valdés Hernández, M. C., Muñoz Maniega, S., Royle, N. A., Morris, Z., & Deary, I. J. (2011). Brain aging, cognition in youth and old age and vascular disease in the Lothian Birth Cohort 1936: Rationale, design and methodology of the imaging protocol. *International Journal of Stroke*, *6*, 547–559.

Zald, D. H. (2007). Orbital versus dorsolateral prefrontal cortex: Anatomical insights into content versus process differentiation models of the prefrontal cortex. *Annals of the New York Academy of Sciences*, *1121*, 395–406.