Routledge
Taylor & Francis Group

TARGET ARTICLE

Process Overlap Theory: A Unified Account of the General Factor of Intelligence

Kristof Kovacs[a,b] and Andrew R. A. Conway[c]

[a]Center for Research and Development, Eszterházy Károly University, Eger, Hungary; [b]Department of Psychological Methods, University of Amsterdam, Amsterdam, The Netherlands; [c]Division of Behavioral and Organizational Sciences, Claremont Graduate University, Claremont, California

ABSTRACT

The most replicated result in the field of intelligence is the positive manifold, which refers to an all-positive pattern of correlations among diverse cognitive tests. The positive manifold is typically described by a general factor, or g. In turn, g is often identified as general intelligence, yet this explanation is contradicted by a number of results. Here we offer a new account of g: process overlap theory. According to the theory, cognitive tests tap domain-general executive processes, identified primarily in research on working memory, as well as more domain-specific processes. Executive processes are tapped in an overlapping manner across cognitive tests such that they are required more often than domain-specific ones. The theory provides an account of a number of findings on human intelligence. As well, it is formalized as a multidimensional item response model and as a structural model, and the neural mechanisms underlying the proposed overlapping processes are discussed.

g: A Well-Aged Puzzle

Why do people differ in their cognitive abilities? Is there a general intelligence that permeates all human intellectual activity? Or is it more reasonable to postulate specific kinds of talent? After more than a century of research, these questions are still unresolved, and the nature and origin of individual differences in mental abilities remain open to debate.

The most compelling result in this field of study is that people who perform above average on one kind of cognitive test (e.g., vocabulary) tend to perform above average on other kinds of cognitive tests as well (e.g., mental rotation). This pattern of positive correlations was first observed more than a century ago (Spearman, 1904) and is often referred to as the *positive manifold*. Indeed, because mental testing of large samples became common practice, for example, in military and academic contexts, literally hundreds of studies have revealed the positive manifold (Carroll, 1993), making it perhaps the most replicated result in all of psychology.

With the development of factor analysis, a statistical technique that aims to reduce the number of dimensions in large correlation matrices, the empirical observation of the positive correlations among diverse cognitive tests was accounted for by a general factor of intelligence, or g. Factor analysis is considered a data-reduction technique because a relatively small number of factors, or latent variables, identify common sources of variance across tests, which are referred to as manifest variables. In other words, the correlation between two manifest variables can be explained by their connection to a common latent variable. For example, a vocabulary test and a mental rotation test are correlated because they both correlate with the same latent variable "X."

The first factorial model of intelligence (Spearman, 1904) proposed that a single latent variable, g, accounts for all of the positive correlations between measures of mental ability (see Figure 1). The variance in a test not attributable to g was therefore explained by a test specific factor, s.[1] According to this initial theory, the specific factors were orthogonal, each a reflection of unique test content and, necessarily, measurement error. Spearman's idea of a latent causal variable, g, as the underlying reason for the correlations among different cognitive tasks, developed contemporaneously with factor analysis itself.

A general factor is indeed reliably obtained when mental test data are submitted to exploratory factor analysis. Yet the test variance that the general factor could not account for turned out not to be entirely test specific, and some groups of tests, for example, vocabulary and reading comprehension, correlate more strongly with one another than with other groups of tests, for example, mental rotation and spatial navigation. Hence Spearman's view of intelligence was quickly met with criticism and alternative accounts were proposed; the strongest competing model consisted of multiple uncorrelated group factors, representing a set of "Primary Mental Abilities" (Thurstone, 1938; see Figure 2). However, Thurstone's original model was challenged in a similar fashion as he challenged Spearman; the idea of orthogonal factors turned out to be untenable, and their correlations needed to be accounted for by a higher order general factor.

CONTACT Kristof Kovacs ✉ kristof340@gmail.com 🖃 1 Eszterházy ter, H3300, Eger, Hungary.
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/hpli.
[1]In Figures 1 to 5 we use the contemporary notation of unique variance (delta) instead of s.
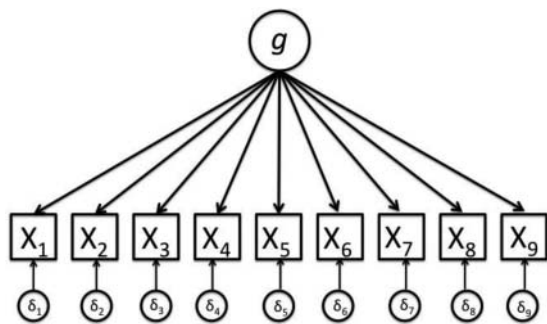
Figure 1. A model depicting Spearman's original conception of a single general factor.

The decades that followed the work of Spearman and Thurstone witnessed numerous studies of individual differences in cognitive ability as well as the development of confirmatory factor analysis (CFA). Contrary to exploratory factor analysis, CFA is a statistical procedure that enables hypothesis testing; one can specify a model of cognitive abilities and test whether observed data corroborate what one would expect based on predictions of the model. These further studies with more advanced methods gravitate toward latent variable models of intelligence that incorporate both a general factor and more specific group factors.

This has been accomplished in two ways: bifactor models and hierarchical models (see Figures 3 and 4). In bifactor models, tests correlate directly with *g* as well as with specific factors, whereas in hierarchical models no test loads directly on *g*. Instead, in hierarchical models domain-general variance is manifested in the correlations between group factors and is ultimately accounted for by the general factor, *g*, at the top level. Thus, contrary to Spearman's original conception "hierarchical *g*" explains correlations among abilities rather than correlations among tests. It arguably does a good job indeed; *g* usually accounts for about 40% (Deary, Penke, & Johnson, 2010) or 50% (Jensen, 1998) of the total variance measured in diverse sets of mental tests administered to sufficiently large samples.

Of course, instead of having uncorrelated first- or second-order factors and a general factor on top of the hierarchy, one could always have correlated first- or second-order factors in the model and no *g* (see Figure 5). Because the higher-order factor model is a nested/constrained version of the oblique first-order factor model, the latter is also usually applicable to describe the positive manifold. But the superficial impression is that the non-*g* model leaves the correlations unexplained, whereas *g*-models do explain them. Or do they?
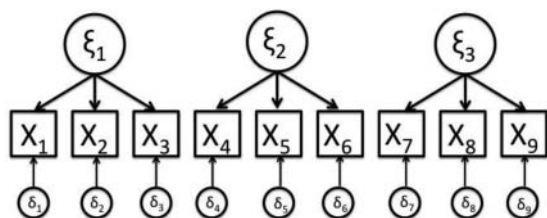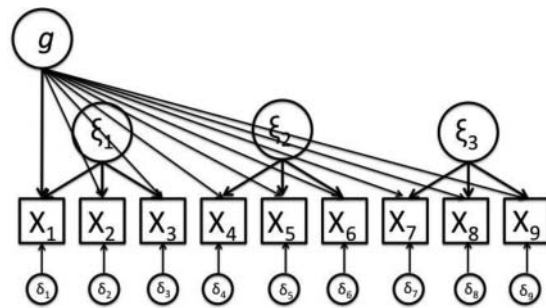
The problem with *g* is simply that still to this day there is no satisfactory consensus about how to interpret it: If there is a casual factor behind *g*, it has not been identified yet. Moreover, it is not only the case that there is controversy about what *g* is; there is substantial confusion about what kind of thing *g*, or indeed what any latent variable, is in the first place (Borsboom, Mellenbergh, & van Heerden, 2003; Conway & Kovacs, 2013).

Here we propose a novel solution to this well-aged puzzle, which we refer to as process overlap theory. The primary aim of process overlap theory is to explain the positive manifold, yet the theory also provides a comprehensive account of established findings on individual differences in intelligence. It is important that process overlap theory explains interindividual differences in behavior in terms of intraindividual psychological processes and neural mechanisms. There have been other approaches, discussed later on, that question the latent cause interpretation of the positive manifold and have offered alternatives. However, in our view, process overlap theory is unique in the sense that it integrates psychometrics, cognitive psychology, and neuroscience.

Such an ambitious integrative approach requires a solid theoretical foundation, which we describe in detail next. To preview, here we consider three axioms, or fundamental premises of the theory:

1. *g* is a necessary consequence of the positive manifold; whenever there are only positive entries in a correlation matrix, it is always possible to extract a single general factor via factor analysis, and this factor will correlate positively with all of the manifest variables or, in the case of hierarchical models, with all of the first- or second-order factors. Of importance, this is not an empirical finding but a mathematical necessity, of which there

Figure 3. A bifactor model of cognitive abilities.

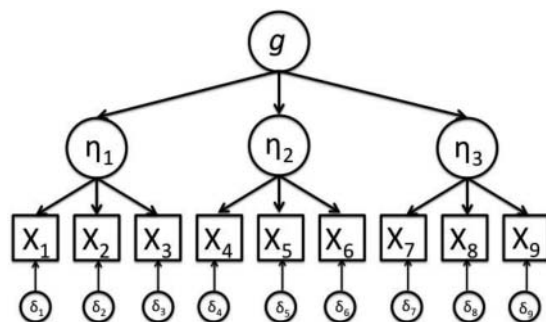**Figure 2.** A model depicting Thurstone's original (but later revised) conception of orthogonal group factors.

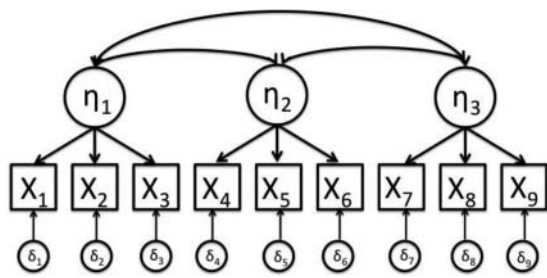**Figure 4.** A hierarchical model of cognitive abilities.

Figure 5. An oblique model of cognitive abilities.

exists adequate algebraic proof (Krijnen, 2004). That is, in a technical sense, *g* is no more, no less, than a reflection of the positive manifold. Hence, "it is always important to remember that it is the positive manifold, not *g* as such, that needs explanation" (Mackintosh, 2011b, p. 165).

2. An ontological stance of entity realism is required if one is to seriously evaluate the theoretical status of latent variables[2] (Borsboom et al., 2003). Theorizing about a latent variable must transcend the world of mathematical abstractions and pinpoint a real entity, which plays a causal role in the correlations among manifest variable—regardless of whether this entity is a process, a set of processes, or some common property/characteristic of processes.

3. Latent variables are *differential* constructs that do not directly translate to within-individual processes or mechanisms (P. C. M. Molenaar & Campbell, 2009; Voelkle, Brose, Schmiedek, & Lindenberger, 2014). Also, latent variables exist because of individual differences, and without variation in mental abilities there would be no latent variables—the last survivor of a meteor collision with Earth would still have cognitive abilities and mental limitations but would not have *g*. Naturally, this stems from the fact that the positive manifold, being a correlation matrix with only positive entries, is itself a between-individual phenomenon. Hence the scope of any explanation of the positive manifold, including but not restricted to latent variables, is not necessarily directly applicable to single individuals.

The structure of the article is as follows. First we discuss the relation between within-individual processes and sources of between-individual variance and provide a critique of the interpretation of *g* as a within-individual construct. A few important characteristics of the general factor that any theory of the positive manifold should probably take into account are surveyed next. The following two sections discuss working memory, first as a within-individual construct and then as a latent variable that is strongly related to variation in fluid reasoning. The reason for discussing working memory is detail is that there is a positive manifold and a general factor obtained in such tasks as well; not only is it strongly related to the positive manifold in

intelligence but it is quite likely that there is a similar explanation of these two positive manifolds.

This is followed by a discussion of goal neglect and prefrontal function, and how they are related to both working memory and fluid reasoning, highlighting the importance of cognitive processes in fluid intelligence that we believe to be crucial in causing the positive manifold. Having surveyed a large bulk of empirical evidence that function as the grounds of our theoretical framework, we turn to outlining process overlap theory as both a cognitive and a structural model of human intelligence, accompanied by a mathematical (psychometric) model. The next section covers studies that employed a network approach to brain functioning; such studies highlight a functional overlap of neural circuitry that corresponds to the overlap of psychological processes hypothesized by our theory. This is followed by a comparison of our theory with previous attempts to explain the positive manifold without a single underlying causal dimension, and we close the article with a few concluding remarks.

The *g* Within?

A parsimonious interpretation of the general factor, based solely on the statistical evidence, is that it represents a single, general ability ("general intelligence" or "general cognitive ability") that manifests itself in all kinds of different tests. However, this is not the only possible explanation of the positive manifold. Thomson (1916) demonstrated that a general factor could appear as the result of a large number of independent, uncorrelated psychological processes, "sampled" by a battery of tests. Thomson's "sampling theory" proposed that every mental test randomly taps a number of "bonds" from a shared pool of neural resources, and the correlation between any two tests is the direct function of the extent of overlap between the bonds, or processes, sampled by different tests.

Because its original formation, there have been statistical elaborations and extensions of the sampling model (Bartholomew, Allerhand, & Deary, 2013; Bartholomew, Deary, & Lawn, 2009; Maxwell, 1972; McFarland, 2012) as well as substantial ones, claiming that the overlap takes place at the genetic (Anderson, 2001) or neural (Hampshire, Highfield, Parkin, & Owen, 2012; Rabaglia, Marcus, & Lane, 2011) level. A developmental account based on mutually beneficial interactions has been proposed that also provides a mathematical explanation of the positive manifold without assuming the causal action of a single general factor (van der Maas et al., 2006). Crucially, with regard to the distinction between sampling models and *g*-models, it has been mathematically demonstrated that "there is no statistical means of distinguishing between the two" (Bartholomew et al., 2009; see also Maxwell, 1972). The conclusion from these studies is that general intelligence, a single common cause of the positive correlations between mental tests, is surely a sufficient, but definitely not a necessary explanation of the positive manifold.

A crucial thing to notice is that the concept of general intelligence interprets *g* as a within-individual mental ability, the involvement of which, in all kinds of cognitive activity, is *causally* responsible for the positive manifold. Therefore, if the concept of general intelligence is correct, then the following

---

[2] More precisely: to evaluate the theoretical status of *reflective* latent variables, see "Process Overlap Theory."

statement is valid: "John used his general intelligence to correctly answer items on both the vocabulary test and the mental rotation test." This, however, is substantially different from the statement: "If John performs better on the vocabulary test than most people, it is likely that he will perform better on the mental rotation test as well," because the latter statement leaves the possibility open that John in fact did not use the same general cognitive ability to solve items in the vocabulary test and the mental rotation test, respectively. Nevertheless, the statistical evidence based on between-subject data validates only the second statement, not the first. To validate the first statement, one has to review other kinds of evidence, and the result is far from convincing.

First, there is a substantial amount of neuropsychological evidence contradicting the idea that people use the same general cognitive ability to perform tests with different content. Damage to different areas of the brain results in the double dissociation of various cognitive abilities. In particular, spatial and verbal abilities can be dissociated this way, as well as fluid reasoning from crystallized abilities (Duncan, Burgess, & Emslie, 1995). Similarly, specific developmental disorders result in impaired spatial abilities, whereas certain verbal skills remain intact, or vice versa (e.g., Vicari, Bellucci, & Carlesimo, 2007; Wang & Bellugi, 1994). This provides strong evidence against the explanation of the positive manifold by a general cognitive ability operating within individuals. For if John excels in both vocabulary and mental rotation because he uses the same single general ability for both, it would not be possible for his performance to deteriorate on only one of these tests following damage to specific areas of his brain. Similarly, there is ample evidence for the dissociation of verbal and spatial tests as a result of various experimental manipulations; such results are also incompatible with the notion that both tap a single general ability (Jonides et al., 1996).

Sex differences can also be a means toward fractionating human intelligence (Mackintosh, 2008); a large number of studies indicate that on average, male and female individuals have somewhat different cognitive profiles, with female participants outperforming male participants in most verbal tests, as well as tests measuring perceptual speed, whereas male participants excel in three-dimensional spatial skills.

Finally, the Flynn-effect, which refers to the secular increase in IQ across generations, also contradicts the within-individual notion of general intelligence. In tests requiring fluid inductive reasoning (see "Understanding g: Characteristic Features," particularly "Figure 1: g and Gf Are Very Strongly Correlated"), such as Raven's Progressive Matrices, the gains per generation have been as high as 15 IQ points, whereas in tests measuring crystallized abilities, such as vocabulary and mental arithmetic, the gains have been negligible; 2–3 IQ points over half a century (Flynn, 2007).

To be fair, g-theories of intelligence could account for all these phenomena by assuming that all fractionation and dissociation occurs only in lower order specific abilities. Because sex differences appear in specific abilities, that argument does indeed seem valid. Similarly, claims have been made that the Flynn-effect is independent of g (e.g., Rushton, 1999), even though this conclusion is controversial (see Flynn, 1998). However, the neuropsychological evidence is harder to dismiss; it appears as if there is simply no place in the brain for general intelligence (see "Overlapping Networks in the Brain" for details). Also, taken together, these converging lines of evidence point to the elusive nature of general intelligence. With all different lines of fractionating evidence taken into account, there is hardly any space left for a general cognitive ability that permeates all human cognition.

It is also important to point out that not all g-theorists equate the general factor with a general ability. Actually, one of the leading g-theorists, Arthur Jensen, opposed such an interpretation: "It is important to understand that g is not a mental or cognitive process or one of the operating principles of the mind, such as perception, learning, or memory" (Jensen, 1998, p. 94–95.). More generally, our emphasis on g being a differential construct is in perfect agreement with his theorizing about the general factor: "A simple distinction between process and factor is that a process could be discovered by observing one person, whereas a factor could be discovered only by observing a number of persons" (Jensen, 1998, p. 95; see also Jensen, 2000).

So how does Jensen, and other g-theorists, interpret g other than a general cognitive ability? They hypothesize that it is a common parameter that influences all of the specific abilities or modules. For instance, Jensen proposed that g reflects individual differences in the speed of mental operations, whereas Eysenck emphasized the role of the efficiency of neural transmission (e.g., Eysenck, 1998). There is indeed valuable contemporary research exploring the link between such phenomena and the general factor; for instance, white matter tract integrity appears to be a promising candidate for such a parameter (Penke et al., 2012). However, even this explains only 10% of the variance in the general factor. Speed and efficiency, even though they surely have explanatory power, only explain a portion of the across-domain variance in mental tests.

Moreover, there are other problems with the theory of mental speed: Among others, attention seems to be responsible for much of the speed–IQ relationship (e.g., Conway, Kane, & Engle, 1999), and it is also most pronounced on psychometric tests of perceptual speed (e.g., Mackintosh & Bennett, 2002). It is not the aim of this article to do justice on the mental speed hypothesis of g, so we stop here by saying that this line of explanation has not been sufficient, and we kindly refer the interested reader to Chapter 3 of Mackintosh's (2011b) textbook for an extensive elaboration on why not

Not a within-individual general cognitive ability, and probably much more than mental speed, the general factor of intelligence remains an unsolved puzzle, and so does the positive manifold. Although several candidates have been offered, there is still no consensual explanation of why there are substantial correlations between cognitive tests that appear to measure very different things.

From a cognitive perspective, the puzzle itself can be summarized as follows: *Why does the variation between people in test performance appear massively domain-general if the abilities they employ to solve such tests are largely domain-specific*? To answer this question, we provide a cognitive account of

item response processes and a corresponding structural model, which are compatible with current research in cognitive psychology and neuroscience as well as with a century of research on the structure of individual differences in intelligence.

Understanding *g*: Characteristic Features

The positive manifold and, consequently, the general factor of intelligence have a number of important characteristics, which process overlap theory attempts to explain. We list four such features of *g*:

Feature 1: g and Gf Are Very Strongly Correlated

The first feature to consider is *g*'s relationship with various group factors, or specific abilities. To fully understand this feature, a brief review of the fluid/crystallized (Gf/Gc) model of intelligence (Cattell, 1971; Horn, 1994) is warranted.[3] The main idea of the model is the distinction between the ability to solve problems in novel situations, regardless of previously acquired knowledge (fluid intelligence or Gf), and the ability to solve problems using already acquired skills or knowledge (crystallized intelligence or Gc). The model includes other group factors as well, the most important of which are Gv (visual-spatial), Gs (speed), and Gr (retrieval from memory). A more recent development is the Cattell–Horn–Carroll (CHC) model (McGrew, 2009), which merges the fluid/crystallized model with Carroll's three-stratum hierarchical model with one crucial difference: the original conception of Gf/Gc did not allow a general factor, whereas CHC does.

A particular appeal of the Gf/Gc model is that the group factors are relatively easy to interpret as within-individual abilities, which can account for correlations at lower levels of the hierarchy, that is, in primary abilities or the mental test scores themselves. Gf is interpreted as fluid reasoning, a thoroughly studied cognitive ability, the neural correlates of which are also identified. Gc, on the other hand, mostly translates to acquired knowledge and/or the amount of formal schooling one has been exposed to (Kan, Kievit, Dolan, & van der Maas, 2011).

Demonstrated first by Gustafsson (1984), and by numerous studies since, the higher order general factor, *g*, is statistically identical to the lower order fluid reasoning factor, Gf, that is, *g* and Gf correlate perfectly. Matzke, Dolan, and Molenaar (2010) reviewed 14 such studies, and even though they emphasized that most of them were underpowered and thus could not have refuted the *g*-Gf identity, the single study with necessary power, as well as two only slightly underpowered studies, equivocally found that the general

factor is identical to the fluid reasoning factor. Moreover, in the remainder of the studies, the correlations between *g* and Gf were between $r = .93$ and $r = .99$ and the fluid reasoning factor had the strongest correlation with *g*, much higher than any other group factor in the CHC model. As well, a perfect correlation between Gf and the lower order factor "inductive reasoning," measured typically by matrix reasoning items and number series was found (Kan et al., 2011), which means that the correlation between *g* and inductive reasoning is perfect or almost perfect as well.

Feature 2: Factor Differentiation

A second important feature of the positive manifold is factor differentiation. Originally discovered by Spearman (1927) who called it the "Law of Diminishing Returns," factor differentiation means that *g* explains more variance at lower levels of mental ability than at higher levels of ability (e.g., Detterman & Daniel, 1989; Kane, Oakland, & Brand, 2006; Molenaar, Dolan, Wicherts, & van der Maas, 2010). Because *g* reflects the strength of the positive manifold, this result means that there are higher cross-domain correlations in samples with lower average ability.

The same phenomenon exists across populations as well; it was recently found that the higher a nation scores on international standardized tests, the less the general factor explains the variance of test scores in that nation (Coyle & Rindermann, 2013). The Flynn-effect is also related to the phenomenon of factor differentiation; the secular gains in IQ are accompanied by a decrease in the average correlation between scores on different intelligence tests and thus a decrease in the variance explained by *g* (Juan-Espinosa, Cuevas, Escorial, & García, 2006; Kane, 2000; Kane & Oakland, 2000; Lynn & Cooper, 1993, 1994; Must, Must, & Raudik, 2003). Even though it has been claimed that the *g* of intelligence is similar to the g (the gravitational constant) in physics (Miele, 2002), factor differentiation, both according to ability within a single cohort and between different cohorts with different levels of ability, demonstrates that *g* is far from being a constant. Instead, the average correlation between diverse tests and thus the domain-generality of the positive manifold varies across time and ability level, and *g* is only informative of the extent of domain-general variance in a given population at a given time.

Feature 3: Complex Tests Correlate Strongly With g

A third important feature is that more complex tests load higher on *g* than less complex tests (Jensen, 1981). This implies that *g* is related to the complexity of cognitive activity. An example is backward digit span, a test in which examinees have to recall digits in reversed order, which has a higher *g* loading than forward digit span, in which digits are recalled in the original order of presentation (Jensen, 1981, 1998).

However, "complexity" is not an explanatory construct that can help our understanding of *g*, nor is it consensual, as there is no necessary agreement between experts about how complex a test is and how complexity differs from difficulty (Mackintosh, 1998). Moreover, there are certainly different "complexities." For example, in a simple continuous performance test, reaction

---

[3]We are aware that there are several important models of intelligence other than the Gf/Gc model (e.g., Johnson & Bouchard, 2005). Yet in practically the entirety of research on working memory and intelligence, as well as on goal neglect and intelligence, Gf-Gc is the model that was applied, and this line of research lays the foundations of our theory. Hence our focus on Gf/Gc is motivated by its proliferation of recent cognitive research on intelligence through providing the comprehensive framework of "fluid reasoning," which is readily interpretable by cognitive psychologists. See, for instance, Blair (2006); Heitz et al. (2006); and Kovacs, Plaisted, and Mackintosh (2006).

time shows a moderate correlation with intelligence but making the continuous performance test more "complex" can enhance the magnitude of the correlation. Three different ways to achieve this enhancement are (a) using the odd-man-out paradigm, in which participants have to select a light that is farther apart from two other lights; (b) showing words instead of lights, and the word that is synonymous to a target word has to be selected; (c) having participants perform a dual task, that is, having them perform a simple reaction time test while information from another test has to be remembered. Although these versions are clearly more complex than the original, they probably invoke rather different cognitive processes.

To explain why "complexity" is related to g, we need to better understand the nature of the cognitive processes involved in more "complex" tests. That is, the nature of "complexity" (or complexities) has to be conceptualized, which we attempt in "Process Overlap Theory."

Feature 4: The Worst Performance Rule

The final g-related phenomenon we consider here is the "worst performance rule," a phrase coined by Larson and Alderton (1990) to describe the finding that worst performance predicts g-loaded measures better than best performance. Larson and Alderton found that the correlation between g and the slowest reaction times was almost twice as large as the correlation between g and the fastest reaction times in a reaction time task. Also, the same effect was found between reaction time and working memory, and the effect was also of the same magnitude. In practice, the worst performance rule means that the difference between the fastest reaction times between high- and low-ability groups is much smaller than the difference between the slowest reaction times. This is consistent with the finding that the correlation between the variability of reaction time and g is as high as the correlation between mean reaction time and g; moreover, the mean and variability of reaction time explain independent parts of the g variance (Jensen, 1992).

Larson and Alderton argued that the worst performance rule is the result of lapses in attention or working memory in people with low cognitive ability. The phenomenon that the difference between high- and low-ability groups is largest in the slowest reaction times and smallest in the fastest reaction times has been found in a number of other studies, some of which used different reaction time tests (e.g., choice vs. simple reaction time). The results demonstrated that the more complex a reaction time test, the stronger the worst performance rule, that is, the larger the slowest reaction times' correlation with intelligence—whereas the correlations between the fastest reaction times and intelligence remained relatively constant (Jensen, 1982; Kranzler, 1992).

Coyle (2001) studied the worst performance rule in a word recall test and found the same effect; the correlation between intelligence and worst performance was significantly larger than it was with best performance. This suggests that this phenomenon is not restricted to reaction time measures. Of importance, Coyle (2003a) repeated a study with an additional group from the top 1 percentile of the intelligence distribution and found no evidence of the worst performance rule in this high-ability group. Also, Coyle (2003b) reviewed studies of the worst

performance rule and concluded that it is the function of the tests' g loading: The difference between the correlations with best and worst performance is larger on tests that are more g loaded.

Overall, these g-related phenomena point to four conclusions:

1. A theory of intelligence must account for the central role of fluid abilities in g.
2. Because the strength of g, and thus of the positive manifold, is population dependent, a new theory must account for why it is stronger in some populations and weaker in others. In particular, it must account for the increasing explanatory power of the general factor at lower levels of ability.
3. Complex tests reveal strong correlations with g. A new theory should, therefore, provide a framework that explains test complexity without falling prey to circular logic.
4. Indices of the worst performance on complex tests reveal strong correlations with g. A new theory should, therefore, focus on the limitations of cognitive processes that result in errors in complex cognitive activity.

Working Memory

Working memory is a construct developed by cognitive psychologists to refer to the processes that enable one to hold goal-relevant information in mind, even in the face of concurrent processing and/or distraction. The construct was introduced in a seminal chapter by Baddeley and Hitch (1974). Prior to their work, the dominant theoretical construct used to explain "immediate" memory performance was the short-term store (STS), epitomized by the so-called modal model of memory popular in the late 1960s (Atkinson & Shiffrin, 1968). According to these models, the STS plays a central role in cognitive behavior, essentially serving as a gateway to further information processing.

However, the concept of STS could not account for a number of within-individual phenomena, demonstrated by experimental and neuropsychological studies. Baddeley and Hitch therefore proposed the construct "working memory" that could maintain information in a readily accessible state, consistent with the STS, but could also engage in concurrent processing, as well as maintain access to more information than the limited capacity STS could purportedly maintain. According to this perspective, a small amount of information can be maintained via two domain-specific "slave" storage systems, verbal and spatial, but more information can be processed and accessed via a domain-general central executive (and according to later models, an episodic buffer; see Baddeley, 2000).

Even though the model of working memory was developed to account for intra-individual phenomena, interest soon arose in measuring individual differences in the capacity of this system and, as it happens, such research has greatly furthered our understanding of the limitations of human cognition. It is important to clarify the distinction between working memory and the capacity of working memory. Working memory refers to a complex cognitive system including mechanisms involved in stimulus representation, maintenance, manipulation, and

retrieval, whereas the capacity of working memory refers to the maximum amount of information an individual can maintain in their working memory.

One of the first tests of the capacity of working memory was the reading span test (Daneman & Carpenter, 1980). The test requires subjects to read sentences aloud and remember the last word of each sentence for later recall, thus heavily taxing both the storage and the central executive component of working memory, contrary to memory tasks requiring only storage and retrieval. The number of sentences/words per list varies, typically from two to six or seven.

Another early example is the counting span test (Case, Kurland, & Goldberg, 1982), in which subjects are presented with an array of items, such as blue and red circles and squares, and instructed to count a particular class of items, such as blue squares. After counting aloud, subjects are required to remember the total and are then presented with another array. They again count the number of blue squares aloud and remember the total. After a series of arrays, they are required to recall all the totals in correct serial order. Thus, the storage and recall demands are the same as a simple digit span test, but there is the additional requirement of counting the arrays, which demands controlled attention and therefore disrupts active maintenance of the digits.

A large number of such "complex span tests" have now been developed to measure the capacity of working memory (for a review, see Conway et al., 2005). The crucial point here is that the construction of complex span tests is a theory-driven enterprise. Such tests require subjects to engage in some sort of simple processing task between the presentations of to-be-remembered items. After several items have been presented, the subject is prompted to recall all the to-be-remembered items in correct serial order. Such tests are thought to be valid measures of working memory as proposed by Baddeley and Hitch because they require access to information in the face of concurrent processing.

Simple memory span tests (e.g., digit span, word span, letter span), in contrast to complex memory span tests, do not include an interleaved processing task between the presentation of to-be-remembered items. For example, in digit span, one digit is presented at a time, and after a series of digits the subject is asked to recall the digits in correct serial order.

One of the most important findings from studies investigating complex and simple span tests is that, from an individual differences perspective, complex span is less domain specific than simple span (Turner & Engle, 1989). Kane et al. (2004) administered several verbal and several spatial complex span tests, and the range of correlations across domains was as high as the within-domain correlations among simple span tests, and about two thirds of the covariance among complex span tests was across domains. These results suggest that, although simple span tests appear to be more domain specific, the processes that complex span tests tap beyond the pure storage and retrieval of information appear to be largely domain general. Hence. general factor models fit better for working memory tasks than for simple span tasks (see the next section).

Individual difference studies of working memory reveal the same type of positive manifold common in the intelligence literature; as with batteries of intelligence tests, patterns of convergence and divergence are typically observed amidst the positive manifold. For example, complex span tests with verbal content tend to be more strongly correlated with other verbal tests than with tests with spatial content. Yet the positive manifold is still observed. Because the positive manifold in itself is always sufficient to extract a general factor (see "g: A Well-Aged Puzzle"), it comes as no surprise that a general factor of working memory could be extracted, which is generally referred to as "working memory capacity" (WMC; Conway, Cowan, Bunting, Therriault, & Minkoff, 2002; Conway, Kane, & Engle, 2003; Engle, Tuholski, Laughlin, & Conway, 1999).

In the working memory literature, there is considerable debate about the domain-generality of variation in WMC or, in other words, whether there is a unitary source of variation or multiple sources. The debate bears a striking resemblance to the debate between Spearman and Thurstone. On one side is the more general/unitary view, which assumes that variation is largely caused by domain-general factors, and on the other side is the specificity view, which assumes that variation is largely caused by more specific factors. In the end, the two sides acknowledge the existence of both domain-general and domain-specific sources of variation but they argue about their relative importance.

There are, however, crucial differences between the possible interpretation of the general factor of WMC and the general factor of intelligence. First, as opposed to tests of intelligence, positive correlations between complex span tests have never been a prerequisite of "validity," hence the positive manifold cannot be attributed to test design.[4] Second, working memory researchers cannot interpret this general factor as a unitary, within-individual, domain-general working memory process and/or mechanism that is employed in every working memory task, similarly to how g is often identified with general cognitive ability. Such an interpretation would contradict the very findings that complex span tests were built upon and that define the within-individual construct of working memory as a complex system of domain-general and domain-specific processes. The right question to ask, then, is, *Which component(s) of working memory cause(s) the general variation?*

The answer probably is that WMC reflects individual differences in the executive component of working memory, particularly executive attention and cognitive control (Engle & Kane, 2004; Engle et al., 1999; Kane, Bleckley, Conway, & Engle, 2001; Kane & Engle, 2002). Cognitive control is a construct, synonymous to executive function, used mostly in cognitive neuroscience to refer to the processes, and their neural substrates, that enables top-down, goal-oriented behavior and that describes different functions such as

> sustained activity that is robust to interference; multimodal convergence and integration of behaviorally relevant information; feedback pathways that can exert biasing influences on other structures throughout the brain; and ongoing plasticity that is adaptive to the demands of new tasks. (Miller & Cohen, 2001, p. 182)

This is a natural candidate to explain the cross-domain correlations among complex span tests, as opposed to the within-

---

[4]This is a typical (albeit incorrect; see Mackintosh, 2011b) line of criticism against the importance of the positive manifold.

domain correlations among simple span tests, because the theory of working memory is in fact an overlap-theory: The processes that bridge verbal and spatial tests are the ones that constitute the executive component.

According to this view, the reason for the domain-generality of WMC, as measured by complex span tests, is that complex span tests "reflect primarily general executive processes and secondarily, domain-specific rehearsal and storage processes," whereas simple span tests "reflect domain-specific storage and rehearsal skills and strategies primarily and executive attention processes only secondarily" (Kane, Conway, Hambrick et al., 2007, p. 24). WMC, then, reflects "the ability to engage controlled attention. That is, they reflect the ability to maintain activation to a representation in the face of interference or distraction. Therefore, working memory capacity is not 'capacity' per se, but rather the ability to control activation" (Conway et al., 1999). That is, individuals with greater WMC have better cognitive control processes, such as goal maintenance, selective attention, and interference resolution (inhibition).

There is a great deal of support for this theory. For example, individuals who perform better on complex span tests also perform better on tests of cognitive control, requiring goal maintenance and the inhibition of irrelevant stimuli (Conway, Cowan, & Bunting, 2001; Conway, Tuholski, Shisler, & Engle, 1999; Kane et al., 2001; Kane & Engle, 2003), and are better at resolving proactive interference from previous trials (Bunting, 2006; Kane & Engle, 2000; Unsworth & Engle, 2007). Similarly, individuals who perform better on complex span tests are also more accurate on lure trials in the n-back test (Burgess, Gray, Conway, & Braver, 2011; Gray, Chabris, & Braver, 2003; Kane, Conway, Miura, & Colflesh, 2007).

Research on WMC thus demonstrates that it is domain-general processes of cognitive control that are responsible for across-domain correlations in complex span tests. These processes can be operationally defined as what complex span tests measure beyond the storage and retrieval of information, or more precisely, for instance, in the case of the reading span test, the processes that *we do not engage* when we remember a simple list of words but we *do engage* when we remember a list of words presented as the last word of sentences we read aloud.

So the available evidence points to the role of the central executive component in the positive manifold of WMC. But how should one conceptualize this component? In the original working memory construct,

> the central executive was initially conceived in the vaguest possible terms as a limited capacity pool of general processing resources. … Implicitly, the central executive functioned as a homunculus, a little man who took the important decisions as to how the two slave systems should be used. (Baddeley, 2002, p. 89)

Thus, further research was required to investigate whether the executive component of working memory is "a single coordinated system that serves multiple functions, a true executive, or a cluster of largely autonomous control processes—an executive committee" (Baddeley, 1996, p. 26).

Further research indeed found that this "homunculus" can be fractionated to subcomponents and should not be conceptualized as a single, unitary executive. Many different tests purport to measure executive functioning directly, including

random number generation, Stroop, Tower of Hanoi/London, Stop-signal, Wisconsin Card Sorting Test, and several others. The n-back test, and especially lure trial performance, is also thought to tap executive processes involved in updating and to reflect interference resolution. Research on these tests also indicates a multiplicity of executive processes rather than a unitary central executive. For instance, relatively low correlations have been found between (a) n-back lure trial performance and complex span (Kane, Conway, Miura et al., 2007); (b) complex span, Tower of Hanoi, and Wisconsin Card Sorting (Lehto, 1996); and (c) Tower of Hanoi and random number generation (Miyake, Friedman, Rettinger, Shah, & Hegarty, 2001). Neuroimaging and neuropsychological studies also support the fractionation of executive processes (Dreher & Berman, 2002; Kievit et al., 2014; Parkin, 1998; Robbins, 1996).

A latent variable study of executive functions (Miyake et al., 2000) identified three correlated processes: "(a) shifting between tests or mental sets, (b) updating and monitoring of working memory representations, and (c) inhibition of dominant or pre-potent responses" (p. 54). However, even though the result of some studies are in agreement with the three-component model of executive functions (e.g., Lehto, Juujärvi, Kooistra, & Pulkkinen, 2003), others are inconsistent with it (e.g., McCabe, Roediger, McDaniel, Balota, & Hambrick, 2010; Salthouse, Atkinson, & Berish, 2003; St Clair-Thompson & Gathercole, 2006).

Overall, the emerging view is that there are multiple executive processes involved in the performance of working memory tests and there are multiple and independent sources of variance contributing to variation in test performance. The general factor of WMC does not appear to be linked to a single psychological process. Instead, it reflects multiple domain-general, executive processes that are tapped in an overlapping fashion across a battery of working memory tests.

Working Memory Capacity and Fluid Reasoning (GF)

Because a positive manifold is observed among measures of WMC, as well as measures of intelligence, it is reasonable to ask how these general factors are related. The reading span test, one of the initial complex span tests, was in fact designed to study the extent to which individual differences in WMC predict reading comprehension and reasoning, and results demonstrated that reading span correlated more strongly with the verbal SAT than did a simple word span test (Daneman & Carpenter, 1980).

Subsequent work showed that other complex span tasks that do not involve reading, or even verbal memoranda, also correlate more strongly with verbal SAT and other reasoning tests than do simple memory span tests such as word span, digit span, and letter span, suggesting that the relationship between complex span performance and intelligence is largely domain-general (Kane et al., 2004; Turner & Engle, 1989). Thus, even though within-domain correlations between working memory tests and cognitive tests are generally stronger than cross-domain correlations, complex span tests have shown strong correlations with measures of reasoning in a domain-general fashion: verbal complex span tests predict spatial reasoning tests and vice versa.

(i.e., filtering of irrelevant features) is essential for successful performance.

Another study, using the same rules Carpenter et al. identified, revealed that it is the application of new rules and switching from old ones that drives the correlation between complex span and Gf (Wiley & Jarosz, 2011). Finally, it has been demonstrated that as soon as performance on elementary cognitive tests becomes automatic and therefore does not require controlled attention, the correlation between such tests and Gf decreases (Ackerman, 1988; Rabbitt, 1997).

Although a large number of studies have relied on complex span tests to demonstrate the link between working memory and Gf, there are other tests that purport to measure individual differences in WMC but are based on slightly different operationalizations of the construct. One such method is the visual array comparison test (Luck & Vogel, 1997), in which an array of objects (e.g., colored squares) is briefly presented, followed by a delay interval, then followed by another array of objects that may be the same or different as the previous array. An example of a "different" array would be one in which the color of one square changed from the first array to the second. The examinee must determine whether the second array is the same or different from the first. Performance is nearly perfect when there are fewer than three items in the array but then declines as more items are added, reflecting the capacity of working memory. Such array comparison tests have been shown to correlate quite strongly with tests of fluid intelligence (Chow & Conway, 2015; Cowan et al., 2005; Fukuda, Vogel, Mayr, & Awh, 2010; Shipstead, Redick, Hicks, & Engle, 2012).

Another kind of working memory test requires coordination and transformation; subjects are presented with information and required to manipulate and/or transform that information to arrive at a correct response. An example is letter-number sequencing, a test originally developed for neuropsychological research, which also appears in the most recent versions of the Wechsler Intelligence Scales (Gold, Carpenter, Randolph, Goldberg, & Weinberger, 1997). In this task a series of alternating digits and letters are presented (e.g., K 6 D 3), and the subject is required to recall first the letters in alphabetical order and then the digits in ascending order.

Another widely used coordination and transformation test is alphabet recoding, which requires the subject to perform addition and subtraction using the alphabet, for example, $(C - 2) = A$. The subject is presented with a problem and required to generate the answer. Difficulty is manipulated by varying the number of letters presented, as $(CD - 2) = AB$. Very strong correlations have been found between reasoning ability and a variety of working memory tests that can all be considered in this "coordination and transformation" class (Kyllonen & Christal, 1990; Oberauer, 2004; Oberauer, Süß, Wilhelm, & Wittman, 2003; Süß, Oberauer, Wittmann, Wilhelm, & Schulze, 2002).

An n-back test constitutes yet another kind of working memory test. In an n-back test, the subject is presented with a series of stimuli, one at a time, and must determine if the current stimulus matches the one presented n-back. The stimuli may be verbal, such as letters or words, or visual objects, or spatial locations. Gray et al. (2003) showed that a verbal n-back test was a strong predictor of performance on the Raven's Advanced Progressive Matrices.

Modified versions of simple span tests that transcend simple storage also tap domain-general WM processes and correlate as well with measures of Gf as complex span tests. For instance, simple span tests with long lists correlate as strongly with measures of Gf as complex span tests (Unsworth & Engle, 2006, 2007). Correlations between simple span and Gf also increase if the presentation of stimuli is swift. In a running memory span test (Pollack, Johnson, & Knaff, 1959), subjects are rapidly presented with a very long list of to-be-remembered items, the length of which is unpredictable. At the end of the list, the subject is prompted to recall as many of the last few items as possible.

Cowan et al. (2005) found that running span correlates well with various measures of cognitive ability in children and adults (see also Mukunda & Hall, 1992). Cowan et al. argued that the rapid presentation in the running span task (e.g., four items per second as compared to one item per second in digit span) prevents verbal rehearsal and that any working memory test that prevents well-learned maintenance strategies, such as rehearsal and chunking, will serve as a good predictor of Gf. It is important to note that Cowan does not restrict this interpretation to the running span task: He argued that the critical feature of working memory tasks such as complex span as opposed to short term memory tasks such as digit span is that the former prevent rehearsal, hence they provide a more direct measure of the scope of attention.

In sum, results with working memory tests other than complex span indeed suggest that it is not the dual-task nature of complex span tests (i.e., processing and storage) per se that is necessary for a working memory test to be predictive of Gf; instead, *it is the involvement of executive processes*, achievable in different ways—including but not restricted to dual tasking—that is common to these tasks, and what drives their relation with fluid intelligence.

However, even though all these tests—array comparison, coordination and transformation, n-back, simple span with long lists, and running span—are able to predict Gf, multiple regression analyses indicate that the variance explained by these tests is not entirely the same as the variance explained by complex span tests (Conway, Macnamara, Getz, & Engel de Abreu, 2011; Kane, Conway, Miura et al., 2007). Hence they probably tap overlapping but different executive processes, each of which is differently related to Gf.

Overall, according to the available evidence, *the strong correlation between Gf and working memory is driven by the operation of multiple domain-general cognitive processes* that are required for the performance on tests designed to measure the capacity of working memory and for the performance on test batteries designed to assess fluid intelligence.

Goal Neglect

Further evidence for the association between Gf, WMC, and executive processes comes from studies on goal neglect (Duncan, Emslie, Williams, Johnson, & Freer, 1996; Duncan et al., 2008). In a standard goal-neglect experiment, subjects are presented with two streams of stimuli on a computer screen and are instructed to monitor the appearance of targets in one stream but not in the other. For instance, they might watch two

streams of digits and letters, and they have to read aloud the letters but ignore the digits in one stream and completely ignore the other stream. The task starts with an instruction "watch left" or "watch right," indicating which stream the subjects must watch. Near the end of each trial, subjects see another cue, a + or a – sign, meaning that for the remainder of the task the subject has to watch the right or left stream, respectively. That is, if they are already watching the right stream, a + sign indicates they have to keep watching to the right, whereas a – sign indicates they have to change to the left.

Some subjects regularly fail to follow the goal instructions. Duncan and colleagues (Duncan, 1995; Duncan et al., 1996) termed these errors *goal-neglect*. They found that the correlation between the subjects' ability to effectively switch attention according to the cue strongly correlated with Gf as measured by the Cattell's Culture Fair. Moreover, the relationship was not linear: "Neglect is hardly ever seen among people whose Culture Fair scores are above the population mean but is almost universal at more than one standard deviation below the mean" (Duncan, 1995, p. 725). That is, neglect is almost universal below a fluid IQ of 85 but practically nonexistent above 100.

Also, Duncan concluded that people in the lowest segment of the IQ distribution show symptoms of perseveration similar to those of frontal patients. People with fluid IQ scores under 1 standard deviation below the mean could recall the task requirements after the instruction phase, and just like frontal lobe patients, they were able to correctly recall the instruction at the end of the experiment; they simply failed to maintain the goal throughout the course of the test. Neglect was also sensitive to external prompts, such that when subjects were given trial-by-trial error feedback so that their attention was drawn to the neglected task requirement, those who previously demonstrated goal neglect were able to perform at a normal level. These results demonstrate that goal neglect is due not to people with lower IQ being unable to understand instructions but to their inability to follow them during the task.

Subsequent experiments (Duncan et al., 2008) revealed a few important characteristics of goal neglect. One of these is that goal neglect is unaffected if, instead of + and – signs, more spatially orienting cues, such as arrows pointing to the left or right, are used. Moreover, neglect is determined neither by the attentional demand during task execution nor by readiness to multiple task components. Various experimental modifications of the original goal neglect task, such as increasing the processing demand of the task by increasing the number of letters or numbers to be monitored, or having different instructions simultaneously prepared for different components of the task, had no influence on the extent of goal neglect.

However, a manipulation of the complexity of task instruction, without a corresponding change in the actual real-time demands of the task to be executed, has a strong effect on goal neglect (Duncan et al., 2008). That is, goal neglect reflects a limit in WMC that manifests itself in maintaining representations of task-relevant rules and requirements rather than limits in the actual attentional processing required for the task. This conclusion is further supported by a study (Duncan, Schramm, Thompson, & Dumontheil, 2012) examining a "rule working memory" task. In this new task, participants had to remember a list of complex rules and apply them to stimuli. Duncan et al. (2012) found that performance on this task correlated more strongly with Gf than operation span.

Overall, studies on goal neglect and rule maintenance demonstrate that as task requirements become more complex, and more facts, rules, and instructions have to be stored in working memory while actually performing the task, the more often lapses in goal-related control processes will occur in people with low fluid intelligence.

Process Overlap Theory

We offer a new explanation of the positive manifold, which we refer to as process overlap theory. 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. 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. 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. Such a pattern of overlap of executive and specific processes explains the positive manifold as well as the hierarchical structure of cognitive abilities. In this section we elaborate on this idea as well as its implications.

Process overlap theory is clearly not the first account of the positive manifold that proposes an overlap of psychological processes. In particular, it is in many ways similar to Thomson's sampling theory. However, it is also different in crucial aspects, as becomes apparent from this section and further highlighted in "Comparison with Other Theories." Process overlap theory is also not the first cognitive approach to human intelligence. Yet it is the first cognitive theory that also provides a latent variable model and an item response model (discussed next), as well as an account of the neural mechanisms underlying the proposed overlap of psychological processes (see "Overlapping Networks in the Brain").

Crucially for the theory, the general factor of intelligence seems not to reflect a single, unitary process but instead emerges from a limited number of independent sources. Detterman (1994) demonstrated mathematically, by calculating limits of correlations in different scenarios, that g is the result of a limited number of independent processes, rather than of a single, unitary process or an almost infinitely large number of processes. As well, a large number of studies looked at the correlation between so-called elementary cognitive tasks and intelligence. Summarizing the result of such studies, Detterman (2000) concluded that these elementary tasks do not correlate with one another, yet each task independently correlates with g, and it is only together that they explain a substantial part of the g variance. Similar conclusions were reached by Kranzler and Jensen (1991, 1993).

In fact, in the intelligence literature the expression "0.30 barrier" refers to the fact that although virtually any cognitive task correlates with IQ (in this case, as a proxy for g), the correlation
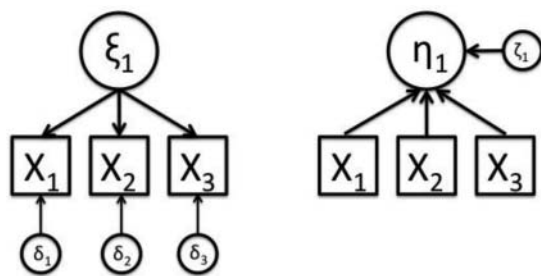
Figure 7. A reflective (left) and a formative (right) model.

is always smaller than 0.30 (see Mackintosh, 2011a). Of importance, this barrier is exceeded by tasks measuring WMC with correlations as high as 0.80. However, WMC arguably reflects executive processes and is therefore hardly elementary. Moreover, as we have seen, WMC itself is the result of a number of independent processes. In fact, according to process overlap theory, WMC correlates with fluid intelligence exactly because it is a multicomponent construct with overlapping processes. 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.

Process overlap theory can be translated to a structural model, similar to the ones depicted in Figures 1 to 5. However, it is different from all those models in a crucial aspect; it challenges the idea that the across-domain correlations between diverse mental tests are caused by an underlying factor. Instead, it proposes that the positive manifold is an emergent property and, consequently, it translates to a formative model with regard to the general factor.

The difference between reflective and formative models is illustrated in Figure 7. The model on the left is a *reflective* model, in which the measurements *reflect* the latent variable. Such a model requires a stance of entity realism with respect to the latent variable, in this case the general factor. For reflective measurement to make sense, one must assume that there is something out there, represented by the construct, and the measures are (imperfect) indicators of this something (Borsboom et al., 2003). In the case of g, it is proposed that g causes the measures as well as the covariance of the measures. According to the theory of general intelligence, g causes the measures because a person's score on the measure, that is, the IQ-test, is determined by his score on the latent variable, that is, g. Consequently, variance in the latent variable determines variance in the manifest variable; hence, the manifest variables' covariance is caused by the latent variable.

In formative models the chain of causation is the opposite. The latent variable emerges *because* of the indicators and not the other way around. In a formative model of g, g is the result, rather than the cause, of the correlations between group factors. Similar formative latent variables are socioeconomic status and general health, which each tap common variance between measures but do not explain it; according to process overlap theory, g is no different (see also van der Maas, Kan, & Borsboom, 2014).[5]

However, at the level of specific abilities, process overlap theory translates into a reflective model. That is, tests indeed reflect specific abilities, which do have ontological reality. Therefore, for the stratum (or strata) below g, process overlap theory is compatible with a standard oblique model, depicted in Figure 5. The only addition is that the specific abilities are not perfectly independent, in the sense that they tap overlapping psychological processes. Consequently, there is no possible categorization of abilities in which the abilities will not be correlated.

Thus, overall, process overlap theory translates to a hybrid structural model: part formative, part reflective. As a reflective causal model it corresponds to the oblique model, but it can also accommodate g as a formative latent variable—the common consequence, rather than the common cause, of the correlation between group factors. This is illustrated in Figure 8 on a simplified model, consisting only of a verbal, a spatial, and a fluid ability factor, and corresponding verbal, visuospatial, and executive processes.

Because process overlap is probably not the only source of the all-positive correlations, this model also accommodates other sources of the general factor, which can range from white matter tract integrity to mutualism, and so on. In the model, this is represented as ζ, the unique variance of g.

The most important difference, then, from g-oriented accounts of the positive manifold is that, whereas reflective general factor theories propose a causal influence of a latent variable, g, on the positive manifold, according to process overlap theory *the positive manifold is an emergent property*, the result of the specific patterns in which item response processes overlap. A crucial aspect of the theory is that it emphasizes the processes responsible for errors in performance on cognitive test items. The processes that are responsible for various aspects of executive attention (goal-monitoring, updating, inhibition of irrelevant stimuli, etc.), and that are incorporated in the more global concept of WMC, reflect limits in domain-general processes that affect performance on a wide range of items.

Therefore, according to process overlap theory, *the processes sampled by different mental test items are not additive*. Each process has its own limitations, and each process has to be functioning at an appropriate level to arrive at a correct answer to a mental test item. Thus, executive processes act as a bottleneck, and they mask individual differences in specific abilities. 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.

The aforementioned aspects of process overlap theory are formalized in an item response model (Equation 1), which provides the probability of a person (p) arriving at a correct answer on a test item (i) that taps component processes (C) from a number of different domains (D). Item response theory is a paradigm of psychometrics for the study of the mathematical relationship between latent traits (abilities, in this case) and test scores. Even though item response theory is primarily used for the construction and scoring of psychometric instruments, including mental ability tests, it also has explanatory applications.

According to process overlap theory, there are distinct within-individual processes (C) tapped by different test items,

---

[5] Formative and reflective measurement is drastically different, but this issue cannot be dealt with in this article. The interested reader is referred to Bagozzi (2007); Edwards (2011); and Howell, Breivik, and Wilcox (2007).
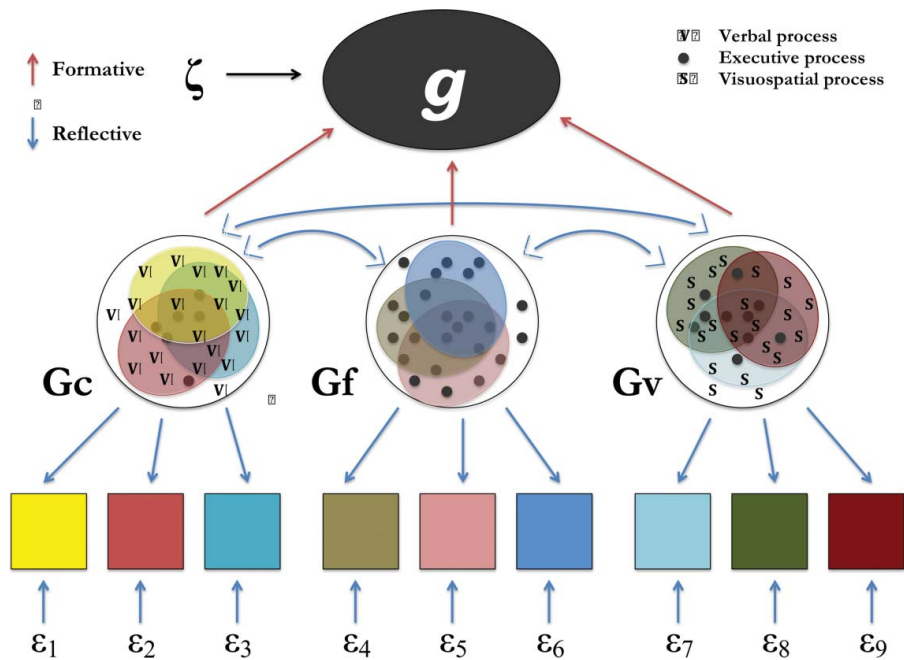
Figure 8. Process overlap theory as a latent variable model.

and these might belong to different cognitive domains (D). Therefore, process overlap theory translates into a *multidimensional* item response theory (MIRT). There are two general kinds of multidimensional models: compensatory and noncompensatory models (for an introduction to MIRT, see Reckase, 2009). In compensatory models, the different dimensions (processes) are combined in a linear, additive manner to produce the probability of solving the item correctly. Therefore a high score on one of the dimensions can *compensate* for a weakness in another.

In noncompensatory models, each dimension is treated separately, and the final probability of solving the item is the product of all of the individual probabilities, as if a single item consisted of a set of independent, unidimensional "subitems," each of which has to be solved correctly in order to arrive at a correct answer. Therefore the probability of solving the item is a nonadditive and nonlinear function of the score on each individual dimension. In such a model, because each dimension has to be passed individually, a low score on any of the dimensions will not be compensated by a high score on another one. Mathematically, the main difference is that in compensatory models it is the *sum* of ability scores that determine the overall probability of success, whereas in noncompensatory models it is their *product*.

$$P(U_{pi} = 1 \mid \Theta_{plm}, a_{il}, b_{il}) = \prod_{l=1}^{D} \frac{e^{\sum_{m=1}^{C} a_{il}(\Theta_{plm} - b_{il})}}{1 + e^{\sum_{m=1}^{C} a_{il}(\Theta_{plm} - b_{il})}}$$

where:

Θ_{plm} = the process score for the pth person on the mth process of the lth domain

a_{il} = the discrimination parameter for the lth domain on the ith item

b_{il} = the difficulty parameter for the lth domain on the ith item

D = number of domains tapped by the item

C = number of processes in the given domain tapped by the item

Again, process overlap theory translates into a hybrid between the two general families of MIRT models. *Within* each cognitive domain (D) the processes are additive, which is reflected by a compensatory model. *Across domains*, however, the model is noncompensatory. The probability of passing each individual dimension (i.e., executive, spatial, verbal, etc.) is calculated, and their overall *product* determines the probability of solving the item. Therefore, if there is a single one of the dimensions involved that the person cannot pass, they will not provide a correct answer—in practice, the model behaves as if the individual cognitive domains are individual and independent obstacles to overcome within the same item.

For example, a person with low-executive "ability" scores will have a low probability of getting an item right, even if the person has high scores on the specific processes that are also tapped by the items. That is, with lower executive functioning, *errors are more likely to be the result of not being able to cope with the executive demands of the task*, regardless of the additional domain-specific components. This nonlinearity is responsible for the bottleneck nature of the overlapping executive processes, which in turn explains why the strength of the positive manifold differs between populations.

For instance, let us assume that the processes that are tapped by the tasks developed by Duncan and colleagues, outlined in the previous section, and that are involved in maintaining task goals in working memory, are tapped along with domain-specific abilities by different tests. Populations that differ in their average level of goal maintenance processes will show marked differences in the extent of domain-general versus

domain-specific variance. The greater the probability of failing on the goal maintenance component, the less individual differences in specific processes matter in arriving at a correct answer in different tests. Therefore, different tests will correlate more strongly, and a general factor will explain more variance. Process overlap theory proposes that this is the cause of factor differentiation.

Yet, according to process overlap theory, the strength of the positive manifold is not the sole function of the population's level of executive functioning; it is also of the extent to which the tests tap executive processes. The more they do, the more probable it is that a person's error will be the result of a failure on the executive dimension(s) of the task, regardless of its burden on other processes, and the person's possible high level on those processes.

Take working memory as a theoretically unambigous example.[6] As we have seen, working memory tasks, such as complex span, require executive processes to a much larger extent than short-term memory tasks, such as simple span. According to process overlap theory, this is exactly why WMC is much more domain-general than short-term memory capacity, that is, why the patterns of variation are more domain-general in complex span than in simple span. In complex span, relative to simple span, errors are more likely to occur as the result of domain-general executive processes, regardless of whether the task is spatial or verbal.

The example of short-term versus working memory also highlights how *complexity* is defined in the context of process overlap theory: It refers to the extent to which a test taps executive/attentional processes. Hence, the reason why tests of fluid reasoning have the highest *g*-loading is the same reason why complex tasks have higher *g*-loadings than less complex tasks; they all tap central executive processes that are involved in a wide variety of mental test performance across domains. This also explains why working memory is strongly related to intelligence in general, and in particular why what working memory tasks measure above and beyond pure storage is most strongly related to fluid reasoning.

Through its emphasis on errors due to ineffective executive processes as well as executive task demands, the theory also accounts for the worst performance rule, because worst performance is often indicative of failures in executive attention processes (Larson & Alderton, 1990). In particular, in the vast majority of studies, the worst performance rule has been identified in reaction time tasks, in which the slowest reaction times are hypothesized to be the result of posterror slowing, which, in turn, reflects response-monitoring and cognitive control (Dutilh et al., 2012).

Overall, the most important aspect of the MIRT model previously proposed is that it formalizes the interplay between a tests' load on the executive system *and* a given population's level of executive functioning in determining the strength of the positive manifold and therefore the amount of variance accounted for by the general factor. This is because the probability of not arriving at a correct solution to a mental test item

due to failures on domain-general rather than domain-specific processes will be a function of both the extent to which a test item taps domain-general executive processes *and* the level of functioning of these domain-general processes in the population studied.

Process overlap theory therefore explains the strength of the positive manifold *in a given population*. This also means that a complete understanding of the within-individual processes that are required to solve an item is not needed in order to explain patterns of individual differences. Figure 9 illustrates this point. The figure shows a matrix-reasoning item, the kind that is typically found in tests of inductive, nonverbal, fluid reasoning that load highly on fluid intelligence (Gf). To solve the item, one has to apply the rule that Carpenter et al. (1990) defined as "distribution of three": The triangles come in three different colors, and the reversed S-s in the middle of the triangles come in three different sizes. Applying this rule to both dimensions gives the correct answer: 1.

Let us imagine that we have a test that comprises dozens of similar items, all of which require the discrimination and interpretation of color in order to map the relation between the figures. If one analyzes the latent dimensions of performance on this test, one is unlikely to find that individual differences in the accuracy of color discrimination, measured by a standard psychophysical task, contribute to variation in the total score.

However, this changes dramatically once the test is administered to a population of completely color-blind people when contrast is equated. If one is not able to discriminate red, green, and yellow, his chances of arriving at a correct answer on this example item is reduced to 33%, because the best response is a guess between Answers 1, 4, and 6—provided, of course, that the person already successfully applied the "distribution of three" rule on the other dimension. In a population where color vision is impaired but still exists, individual differences in color discrimination ability may become important and explain a large portion of the variance in test performance. The point is that, even though color vision is clearly required to solve the task, in normal healthy populations it will not contribute to variation in performance.

Similarly, if one modifies the item so that, instead of three different colors, three blue figures are of slightly different shades, with hardly noticeable differences, variation in the ability to notice such differences will also contribute to individual differences in the performance on the task. In fact, it is such subtle details of test content that determine what a test actually measures:

> Virtually any test can be made into a measure of Gf by raising the requirements for exercising reasoning. Similarly, almost any test can be made into a measure of Gc by increasing the extent to which individual differences in knowledge are assessed. And, by increasing the requirements for speeded performance, almost any test can be made to measure Gs, at least in part. (Horn, 1989)

At the same time, it is important to emphasize that the overlap of cognitive processes tapped by various mental tests is not simply a measurement problem. Of course, the characteristics of the task determine the nature of the processes involved at arriving at a correct solution. For instance, one can design a

---

[6]Working memory serves only as a comprehensible illustration here: Scores on working memory tasks are nondichotomous, and the actual IRT model, described by the preceding equation, is applicable only to dichotomous test scores.
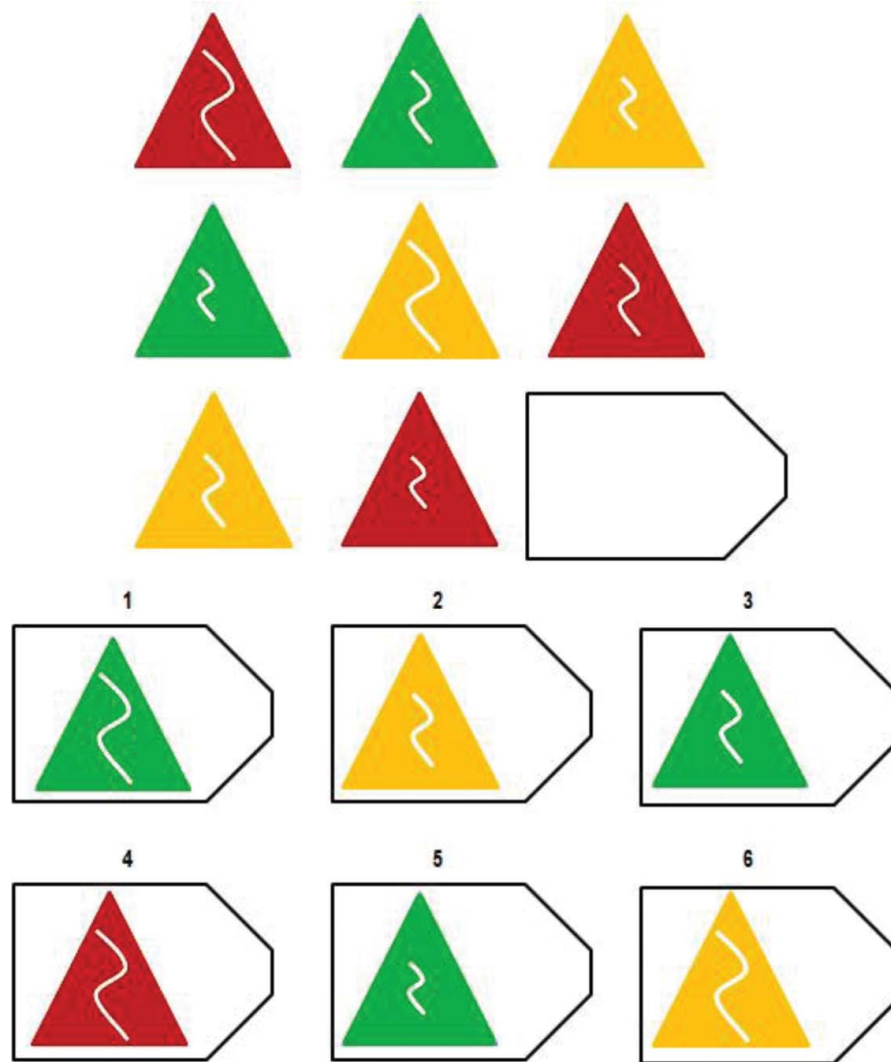
Figure 9. Example item to demonstrate process overlap theory.

spelling test, in which examinees have to decide whether a list of English words appearing on a screen, such as "baccalaureate" or "reconnaissance," are written correctly. Such a task purportedly measures crystallized skills, acquired through formal schooling. Now imagine that each item is mirrored to an axis above the given word. As a result of that, the test would start invoking visual skills. Finally, by adding a strict time constraint to make the correct-incorrect decision, variation in processing speed would start to have a strong role.

However, in practice, the exact opposite is the case. Test developers devote a lot of time and effort to constructing unidimensional measures, tests that purportedly tap a single ability only. That there is still an overlap in executive/attentional processes is more revealing of the psychological nature of such processes than of psychometric test construction.

Crucially, process overlap theory predicts that the psychological processes that determine whether individual differences will be primarily domain-general are not necessarily determined by the cognitive domain the test *purports* to measure. Consider, for instance, the following *number series* item, which is typically categorized as *numerical reasoning* (e.g., in Ackerman et al., 2005). To find the next element in the series, one

has to find the simplest rule according to which the last number(s) can be calculated from the previous one(s).

2, 4, 8, ??

A) 9 B) 12 C) 14 D) 20

When eyeballing the three numbers in the preceding series, the first thing to occur is that they are 2 on the power of 1, 2, and 3. In other words, the subsequent number is always twice the number before, which instinctively provides 16 as the natural continuation of this series. The number 16, however, is not among the possible answers. One must, therefore, find another rule. The correct answer is in fact C, which one can figure out in two ways: (a) the difference between two subsequent numbers increases by two after each element (i.e., 4 – 2 = 2, 8 – 4 = 4, X – 8 = 6?) or (b) the subsequent number equals the sum of the last two elements plus 2. Both rules lead to 14 as the next element (albeit the one following the next differs in the two solutions: 22 and 24, respectively).

What kind of psychological processes contribute to arriving at a correct answer on this item? On a global level, this task requires the ability to find general rules from specific instances, which qualifies as inductive reasoning. Yet on a more refined scale, there are a number of processes at play. Naturally, one

needs to be aware of numbers, as well as basic arithmetic operations. But more important, it also requires *cognitive inhibition*. One has to suppress a dominant response (16) and discard a superficially obvious rule in order to find another one.

Number series items correlate strongly with matrix tests, consisting of items like the one presented in Figure 9. The reason, according to process overlap theory, is the overlap of the psychological processes tapped by the two kinds of tests: Both require inductive reasoning and thus cognitive inhibition. Nevertheless, these two kinds of items are regularly categorized as numerical and figural, respectively, in accordance with the content of the items.[7] In a similar vein, verbal analogies, which also probably tap processes that overlap with the ones required for number series and matrix reasoning, are often categorized as tests of verbal ability.

Naturally, both test makers and test takers need to categorize tests, and at a practical level it does indeed make perfect sense to categorize number series, matrix reasoning, and verbal analogies as numerical, figural, and verbal, respectively. Yet, according to process overlap theory, categorization of tasks according to the kind of material, by domain or content, is not necessarily instrumental in understanding the determinants of individual differences.

The reason why tests of fluid intelligence are particularly successful at measuring the processes responsible for the across-domain correlations between mental tests is that they are more or less free from particular domains. Therefore they are able to reflect "pure" complexity, that is, executive/attentional requirements, which are also present in tests of verbal or spatial reasoning, but in those cases they are tapped alongside with the corresponding domain-specific processes.

This, according to process overlap theory, explains the relation between g and Gf. They are conceptually different, as Gf represents individual differences in fluid reasoning, whereas g does not represent any psychological process. Yet, according to confirmatory factor analysis, they correlate perfectly or almost perfectly. This is because, provided that the general factor was extracted from a large-enough test battery measuring diverse cognitive abilities, which is a key point in obtaining a "good" g (Major, Johnson, & Bouchard, 2011), variation in the specific abilities will be mostly cancelled out, and the variation reflected by g will mostly be the result of individual differences in domain-general processes. Process overlap theory proposes that such processes could mostly, although probably not exclusively, be labeled as executive processes, involved in cognitive control, goal monitoring, inhibition of irrelevant stimuli, and the like.

To sum up this section: Process overlap theory interprets g as a formative construct while accepting a reflective and therefore realist interpretation of specific abilities. It proposes that mental test items tap a number of items in different cognitive domains, and whereas a weakness in a process can be compensated by a strength in another process within the same domain, such compensation is not possible across domains.

Overlapping Networks in the Brain

A comprehensive review of neuroimaging studies, which reviewed imaging studies not only of a wide range of intelligence tests but also of mind games such as chess, found that intelligence is distributed throughout the entire brain (Jung & Haier, 2007). One of the main findings of the publication was that multiple discrete brain regions are associated with intelligence, with no single area activated in all of the studies surveyed. However, the article also demonstrated that the areas active in most studies are typically found in the frontal as well as the parietal lobes.

Another study, focusing on the subscales of the Wechsler Intelligence Scales, demonstrated that the neural correlates of g were to be found in several brain areas, with the strongest relationship in the frontal lobes (Colom, Jung, & Haier, 2006a). Yet another study, which applied the method of correlated vectors (cf. Jensen, 1998) in order to focus specifically on g, reinforced the conclusion that neural correlates of g are distributed across the entire brain, but the majority of them are found in the frontal lobe (Colom, Jung, & Haier, 2006b).

Besides neuroimaging, lesion studies have arrived at a similar conclusion, highlighting distributed brain regions for g but also the importance of prefrontal cortex as well as the white matter tracts connecting it with other areas (Barbey, Colom, & Grafman, 2013; Gläscher et al., 2010). However, as we see, different components of g can be dissociated through frontal damage, because performance on some tests is sensitive to such damage while performance on other kinds of tests remains intact.

Because of a lack of corroborating results, the search for a "*neuro-g*" has met with minimal success. As discussed earlier, the g factors extracted from different batteries are virtually equivalent from a statistical perspective, provided that the batteries are diverse enough. In the light of this, it is remarkable that the search to find the common neural underpinnings of different g factors has failed:

> If two test batteries, for example, are weighted differently with tests of memory, spatial reasoning, verbal ability and the like, different brain correlates of the respective g-factors may emerge, gray matter (GM) correlates of g depend in part on the tests used to derive g. (Haier et al., 2009, p. 137)

A confirmatory modeling approach to the brain correlates of g (Kievit et al., 2012) also found that "neuro g should not be taken to refer to a unidimensional constellation of neural properties identical to g" (p. 7); on the contrary, "neuro-g" is a formative latent property determined by, rather than the cause of or reflected by, neural measures. It indeed appears that "intelligence is a moving target" (Colom et al., 2011). Overall, studies that have focused on g to identify the neural correlates of intelligence have found little consistency, but of equal importance, especially for process overlap theory, is the result that also emerged from such studies, that the prefrontal cortex seems to play an important role.

Instead of g, other studies have focused on specific ability factors, and indeed identified different brain correlates for each factor. For instance, scores on the Wechsler Intelligence Scales have weaker neural correlates in the prefrontal cortex than scores on the Raven's Progressive Matrices, suggesting that the

---

[7]With notable exceptions: Horn (1989), for instance, in his categorization of ability tests according to the Gf-Gc model, put "inductive reasoning, measured using letter series, number series and/or figure series" as the first example of indicators of Gf, "matrices reasoning with visual patterns" comes only second (p. 79).

prefrontal cortex is more strongly related to Gf than to Gc. On the other hand, the temporal lobes were strongly related to Gc but not Gf (Choi et al., 2008). Another study also found that Gc is uniquely correlated with activity in the temporal lobes, whereas Gv, the spatial factor, had nonoverlapping correlates in the frontal and occipital lobes (Colom, Haier, & Head, 2009).

Again, the results of lesion studies corroborate with imaging studies: They also imply different neural substrates for different specific abilities. In classic neuropsychology, the received view was that frontal lobe damage does not impair IQ (e.g., Hebb, 1940; Weinstein & Teuber, 1957) exactly because the clinical tests used in such patients were strongly biased toward crystallized intelligence, Gc. Once the distinction between Gf and Gc is made, it becomes clear that frontal lobe damage severely impairs the former, whereas the latter indeed often remains intact (Duncan, 1995; Duncan et al., 1996).

In the light of such results, it should come as no surprise that the "intelligence" measured by different test batteries gravitating to different specific abilities cannot be universally localized, and the g factors derived from such batteries, despite being statistically indistinguishable, do not have identical neural correlates.

Instead of g, then, let us focus on fluid reasoning (Gf). Again, even though it is statistically identical to g, imaging studies demonstrate their dissociability; whereas g cannot be localized, Gf is linked to the prefrontal (primarlily dorsolateral) and partly to the (primarily posterior) parietal cortex with remarkable consistency. That is, diverse tests tapping fluid reasoning, including matrix items, letter series, or verbal syllogisms, all have similar patterns of activation in the prefronal, and partly in the parietal cortex (for a review, see Kane, 2005). In particular, reasoning problems identical or similar to the ones found in Raven's Progressive Matrices, probably the most typical test of Gf, consistently activate certain areas in the prefrontal cortex (Christoff et al., 2001; Prabhakaran, Smith, Desmond, Glover, & Gabrieli, 1997; Wharton et al., 2000), and this conclusion is further supported by lesion studies (e.g., Waltz et al., 1999). Similar activation was found in other, mostly verbal analogical reasoning tasks (Green, Fugelsang, Kraemer, Shamosh, & Dunbar, 2006; Luo et al., 2003; Volle, Gilbert, Benoit, & Burgess, 2010; Wendelken, Nakhabenko, Donohue, Carter, & Bunge, 2008), pointing to an indifference of content in fluid tasks (see also Duncan et al., 2000).

A lesion study provides further evidence that is in agreement with imaging studies by pointing to the importance of specified areas within the prefrontal and parietal cortex (Woolgar et al., 2010). The study compared different areas of the brain to explore the extent to which, statistically, brain damage in a given area is associated with loss of fluid intelligence on average. Using multiple regression, it found that the same amount of tissue damage that predicts a 1-point loss of fluid IQ, if it occurred elsewhere in the brain, corresponds to a 6.5-point impairment if found in particular prefrontal and parietal regions. Of importance, partial correlations showed that each of the regions studied made an independent contribution to the impairment in fluid intelligence, pointing to the involvement of independent neural mechanisms.

Imaging studies of working memory have identified similar patterns of prefrontal and parietal activation for the central executive as the ones identified for Gf (see Henson, 2001;

Wager & Smith, 2003). A large-scale review concludes that "the central executive maps to mid-lateral prefrontal regions, particularly left and right dorsal lateral prefrontal cortex" (Henson, 2001, p. 166).

Looking at various actual cognitive functions that any mechanism called the "central executive" could be reasonably expected to perform, it has indeed been found that one of the chief functions of the prefrontal cortex is cognitive control (Badre & Wagner, 2004; Botvinick, Braver, Barch, Carter, & Cohen, 2001; Cole & Schneider, 2007; E. K. Miller, 2000; E. K. Miller & Cohen, 2001; Vincent, Kahn, Snyder, Raichle, & Buckner, 2008) or, synonymously, the top-down monitoring of goal-directed behavior (Asplund, Todd, Snyder, & Marois, 2010; Braver & Bongiolatti, 2002; Corbetta & Shulman, 2002; Farooqui, Mitchell, Thompson, & Duncan, 2012; B. T. Miller & D'Esposito, 2005). On a less global scale, the prefrontal cortex is involved in such cognitive operations as task switching (e.g., Derrfuss, Brass, Neumann, & von Cramon, 2005; Sohn, Ursu, Anderson, Stenger, & Carter, 2000) and response inhibition (e.g., Aron, Robbins, & Poldrack, 2004; Chambers et al., 2006), among others.

Once again: In agreement with neuroimaging studies of healthy participants, lesion studies also point to a large commonality between the neural substrate of executive functions and fluid intelligence, and locate this substrate in the prefrontal and posterior parietal areas (Barbey et al., 2012). Yet it is crucial to note that the prefrontal cortex comprises a large portion of the entire cortex and contains a number of distinct subsystems, both cyto-architectonically and functionally. Accordingly, different executive functions probably map on different parts of the prefrontal cortex (e.g., MacDonald, 2000; Stuss & Alexander, 2000).

In particular, some of the studies surveyed earlier in this article have found more medial activation, whereas others registered the activation of lateral areas; some processes seem to induce bilateral activation, whereas the neural substrate of others appears to be unilateral; finally, some studies found the coactivation of the anterior cingulate and/or parietal areas, whereas others have not. However, a recent meta-analysis of 193 imaging studies of different executive processes tapped by various executive tasks was able to identify common activation in what they call a "cognitive control network," comprising the dorsolateral prefrontal cortex, the frontopolar and orbitofrontal cortices, and the anterior cingulate (Niendam, Laird, & Ray, 2012).

To sum up the argument so far: fluid intelligence (Gf), the central executive component of working memory, and various cognitive processes that serve the top-down control of goal-directed behavior have strongly overlapping neural substrates in the prefrontal cortex (for a summary of related evidence, cf. Kane & Engle, 2002).

It is at least as important from the perspective of process overlap theory that the activation of these brain areas is independent of content: These same brain areas of the prefrontal and parietal cortex are involved in different domains of cognition. For instance, Duncan and Owen (2000) reviewed 20 studies that explored brain activation in different types of tasks, the content domain of which included both spatial and verbal tasks. They concluded that the recruitment of frontal areas "is extremely similar from one cognitive demand to another,

suggesting a specific network of prefrontal regions recruited to solve diverse cognitive problems" (Duncan & Owen, 2000, p. 476). The same areas that compose the "cognitive control network" have also been labeled the "multiple demand system" in order to directly refer to the fact that they are involved in diverse cognitive activities (Duncan, 2010).

However, there is a danger of such analysis revealing overlapping activation at the group level even when it does not exist within individuals. This methodological problem was addressed by Fedorenko, Duncan, and Kanwisher (2013), who undertook a study looking at activation overlap within individual subjects. They used seven tasks, including a spatial and a verbal working memory task, a mental arithmetic task, and the Stroop task, and found domain-general activation in the expected frontal and parietal areas at the indivdiual level, too, confirming the results of previous studies that employed group analysis.

A study by Duncan et al. (2000), which purportedly attempted to identify a neural system associated with g but in fact employed tests of fluid reasoning (Gf), found that when high g-(Gf)-loading was contrasted with low g-(Gf)-loading, a pattern of activation in the lateral frontal cortex emerged, and this was the only area commonly activated by spatial and verbal tasks. Another study investigated neuroanatomic overlap of different cognitive abilities and identified specific regions in the frontal lobes that are frequently shared (Colom et al., 2013).

A recent study conducted by Román et al. (2014) took a different perspective: They looked at brain correlates of latent variables at different levels of the "hierarchy of intelligence," that is, in the higher order latent variable model (see "g: A Well-Aged Puzzle"). They found that as one moves upward in the hierarchy from specific factors through group factors to g, the gray matter correlates are smaller and more frontal. The study concluded that "factors capturing the variance common to both specific measures and group factors partial out the specificity present at the measurement level. Interestingly, removing specific variance reveals that frontal regions in the brain are crucial for supporting human intelligence" (p. 3816).

Process overlap theory proposes that as one moves up the hierarchy of abilities, specific component processes gradually disappear, and by the time one gets to the processes directly reflecting g, executive ones are of great importance. Because specific processes have diverse brain correlates, whereas it is mostly frontal regions that are involved in executive processes, the results of Román et al. can be interpreted as the neural equivalent of the psychological explanation proposed by process overlap theory.

Having discussed the domain-general involvement of frontoparietal areas in reasoning tasks, it is important to point out that imaging studies of working memory have also registered the domain-generality of neural activation in the frontal cortex. A meta-analysis of 60 neuroimaging studies (Wager & Smith, 2003) found that the fractionation of working memory according to content was limited to the posterior areas: No fractionation was found in the frontal cortex according to content domains. More precisely: They found that the central executive could be further fractionated as well, but according to processes rather than the type of material.

Because the significance of complex span tasks has been emphasized throughout this article, an imaging study of complex span tasks is particularly interesting, especially because it directly looked for the common neural underpinnings of spatial and verbal complex span, applying a novel methodology that uses both within-domain and across-domain conditions, as well as contrasting complex span with both pure storage and pure processing (Chein, Moore, & Conway, 2011). The study indeed demonstrated the domain-general activation of the prefrontal cortex, the posterior parietal cortex, and the anterior cingulate in complex span tasks.

More recent studies employing a network perspective have also concluded that the prefrontal cortex is often coactivated with brain areas involved in domain-specific cognition. The network approach to brain functioning is an emerging paradigm in cognitive neuroscience, based on the recognition that neural computations involved in cognition are not performed by isolated brain areas but rather are the result of networks of interconnected areas (Bressler & Menon, 2010; He & Evans, 2010; Heuvel & Pol, 2010; Sporns, 2002). Therefore, the study and graph theoretical modeling of the structural and functional connectivity of the human brain—*the human connectome* (Toga, Clark, & Thompson, 2012)—is the central aim of research in the network approach.

Network analysis of the human brain has revealed that it can be characterized as a "*small world network*," that is, a network consisting of local clusters of strongly interconnected nodes but also of short paths that link the individual clusters (Achard, Salvador, Whitcher, Suckling, & Bullmore, 2006; Bassett & Bullmore, 2006). This architecture, which is both modular and strongly interconnected, makes brain wiring economical and highly efficient at the same time. The specialized modules are connected with the aid of *connector hubs*: sets of highly connected and central nodes with diverse and long-range connections that function as global interlinks or bridges between the individual modules or clusters, ensuring short overall path length and thus high efficiency (Sporns, Honey, & Kötter, 2007).

Of importance, "most studies identified hubs among parietal and prefrontal regions, providing a potential explanation for their well-documented activation by many cognitive functions" (Bullmore & Sporns, 2009, p. 190), and "studies on the network of areas of the primate and human cerebral cortex showed that the PFC, especially the dorsolateral part (PFC DL) is an important hub region where information from different functional brain systems are integrated" (Négyessy, Bányai, Nepusz, & Bazsó, 2012, p. 39). Négyessy et al. (2012) also documented that in the imaging literature the single area identified most often is the prefrontal cortex, and they performed network analysis to demonstrate that this is not the result of the selectivity of eresearchers but an inevitable consequence of cortical processing.

Apparently, the same regions that were identified in tradtional studies as the overlapping neural substrate of executive processes, working memory, and fluid reasoning are referred to as the "frontoparietal control system" in network neuroscience as well (Spreng, Sepulcre, Turner, Stevens, & Schacter, 2013; Vincent et al., 2008). Being one of the most connected networks

of the brain (Cole, Pathak, & Schneider, 2010), this system is attributed with functions of regulating other subnetworks.

It is remarkable from an individual differences perspective that of all brain networks the frontoparietal network has the largest variability in functional connectivity, larger than any other network in the brain (Mueller et al., 2013). Moreover, several studies have demonstrated that variation in the global connectivity of these regions correlates with intelligence as well as cognitive control (Cole & Yarkoni, 2012; Heuvel & Stam, 2009; Santarnecchi & Galli, 2014; Song et al., 2008).

There are two further results in cognitive neuroscience that are highly relevant with regard to process overlap theory. First, a number of studies found that the same frontal areas that function as hubs are capable of serial processing only, and therefore they severely limit the capacity of different domain-specific cognitive systems: "The prefrontal and dorsal medial frontal cortex [function] as a frontal lobe network recruited to meet a wide variety of cognitive demands, making this system well suited to act as a central, amodal bottleneck of information processing" (Dux, Ivanoff, Asplund, & Marois, 2006). These areas are therefore primary candidates for being the neural substrate of capacity limits (Dux et al., 2006; Koechlin & Hyafil, 2007; Marois & Ivanoff, 2005; Tombu et al., 2011), probably strongly affecting working memory and intelligence. Because process overlap theory focuses on the limitations of executive processes as the cause of both the positive manifold and factor differentiation, this provides a direct link from the theory's psychological hypothesis to its possibly underlying neural mechanism.

Second, process overlap theory proposes that the interaction between the level of executive processes and the executive demands of the task is of critical importance with regard to the strength of the positive manifold. Hence it is of great significance that activation in relevant regions appears to be a function of both the level of ability and the executive demands of a given task. The vast majority of the studies just cited, documenting prefrontal and parietal activation for executive processes, working memory, and fluid intelligence, demonstrated increased activation as a function of the demand for the construct in question. As well, several studies found an increase in activation that is inversely related to the participants' level on the construct. Kane's (2005) review of prefrontal involvement in fluid reasoning concludes that "PFC is recruited to solve inductive reasoning problems under worst-case conditions, such as when problems are most difficult or when one has reduced fluid abilities" (p. 156).

Unfortunately, simultaneous tests for brain activity as the function of performance within a task and as the function of differences between tasks are largely missing from the literature. That is, activation differences due to task complexity and activation differences due to variation in individuals' ability are not clearly differentiated. Therefore a current study, which addresses exactly this question, is particularly interesting. Kievit et al. (2016) employed a modern psychometric approach to neuroimaging to test for overlapping brain correlates of difficulty and ability parameters in fluid reasoning tasks. Using a conjunction analysis, they found three regions the activation of which depended on difficulty and ability: bilateral angular gyri, bilateral precuneus, and the left superior frontal gyrus. This demonstrates that the regions that are registered in between-subject designs (of differing fluid intelligence) are the same ones that are registered in within-subject designs (of increasing difficulty in fluid tasks); again, this seems to point to the neural underpinning of the interaction proposed by process overlap theory.

To summarize this section: According to process overlap theory, the positive manifold is caused by the overlap of executive processes that are involved in both working memory and intelligence. The present state of research in neuroscience demonstrates that the neural correlates of such processes are (a) indeed involved in working memory and intelligence, and (b) indeed activated in an overlapping fashion that is in agreement with the tenets of the theory, and finally: (c) the frontal lobe is strongly connected to other, more specialized parts of the brain. In other words, the overlap the theory proposes appears to actually take place in the human brain.

Comparison With Other Theories

There have been enormous theoretical endeavors in the field of human intelligence, mostly focusing on the nature, structure, or interpretation of the concept itself. Even a simple elaboration of these accounts is beyond the possibilities or aims of this article. We have provided an explanation of the positive manifold and a number or strongly related phenomena. In this section, therefore, we compare only process overlap theory to accounts of the same empirical phenomena and not to theories of intelligence in the broad sense. Similarly, because process overlap theory is not a taxonomy of the structure of variation in human abilities, no comparison to such taxonomies (like the CHC, McGrew, 2009; or the VPR, Johnson & Bouchard, 2005, model) is provided.

The first theory to consider is, of course, *g*-theory, the idea that different IQ-tests correlate because they all measure the same latent variable, which can be interpreted as either general intelligence or a parameter affecting all cognitive operations. Because this idea is thoroughly criticized in the first part of the article, we find it unnecessary to further elaborate on why process overlap theory is more plausible than this account.

The second is Thomson's sampling theory, which proposes that the correlation between any two mental tests is the function of the number of shared "bonds" the tests sample. Thomson demonstrated that this principle is sufficient to produce the positive manifold, without postulating a general factor. This account has a lot in common with process overlap theory, especially with regard to higher order, more general processes versus lower order, more specific processes:

> The mind, in carrying out any activity such as a mental test, has two levels at which it can operate. The elements of activity at the lower level are entirely specific, but those at the higher level are such that they may come into play in different activities. Any activity is a sample of these elements. (Thomson, 1916, p. 341)

In fact, from a broad perspective, process overlap theory can be considered a modern sampling theory.

The continuation of the preceding paragraph, however, already highlights a crucial difference: "The elements are assumed to be additive like dice, and each to act on the 'all or

none' principle, not being in fact further divisible"[8] (Thomson, 1916, p. 341). Contrary to this assumption, process overlap theory proposes a nonadditive overlap of psychological processes. In particular, the executive/attentional processes that typically overlap with domain-specific ones function as a bottleneck: Failure to pass the executive demands of a test renders individual differences in specific processes unimportant for overall performance. As a consequence, the correlation between tests is not simply the function of the sheer number of overlapping processes in relation to the total number of activated processes, as in Thomson's account.

The two accounts also differ markedly in their view of brain functioning. The bonds theory subscribed to a version of contemporary views on "equipotentiality," denying the localization of brain function. In fact, Thomson argued that the human brain consists of a myriad of bonds and assumed that the sampling process is completely random, with tests differing only in the number of bonds they sample. Process overlap theory, on the other hand, draws heavily on results from neuroscience that have been obtained since Thomson's time, and which demonstrate that executive processes are primarily seated in the prefrontal cortex and that this area of the brain is the one most heavily interconnected with other areas.

This is an important difference, as it directly addresses two valid criticisms of the sampling model (summarized in Eysenck, 1987, and van der Maas et al., 2006). First, it logically follows from the sampling model that the more bonds a test samples, the higher its average correlation with all other tests, because it is more likely to randomly share bonds sampled by other tests. This means that a test's g loading is the sole function of the number of bonds sampled by the test. However, a number of tests, which supposedly measure a narrow range of "bonds," load highly on g. Yet, according to process overlap theory, g loadings depend on the involvement of executive processes seated primarily in the prefrontal cortex rather than on the number of processes measured.

The second criticism is even more directly related to the brain: It has been cited as falsifying evidence against the sampling model that brain damage can lead to specific impairments, whereas its conception on brain functioning determines the bonds theory to predict general impairments. Again, according to process overlap theory it is damage to the neural substrate of overlapping executive process that is relevant in predicting the generality of the impairment rather than the total amount of damage.

There is a third criticism against the sampling model, which is particularly informative in highlighting the difference between Thomson's account and process overlap theory: "Some seemingly completely unrelated tests, such as visual and memory scan tasks, are consistently highly correlated, whereas related tests, such as forward and backward digit span, are only modestly correlated" (van der Maas et al., 2006, p. 843.)

Because process overlap theory, as opposed to sampling, does not propose additive processes, it does not predict a linear

relationship between the size of the correlation and the extent of the overlap relative to the total number of activated processes. Instead, it predicts that the size of the correlation will be a function of the overlap of domain-general executive processes. Therefore the third criticism is not relevant for process overlap theory. In particular, whereas forward digit span measures only the storage and retrieval of digits, backward digit span also taps executive processes involved in fluid reasoning (Kovacs et al., 2016). With regard to visual and memory scan tasks: They correlate strongly exactly because both are good measures of the executive component of working memory.

Anderson (2001) provided an account of the general factor similar to the one provided by Thomson, but here the overlap of elements takes place at the level of genes. He argued that any cognitive task requires the coordinated functioning of distributed neurons, and because the development of these neurons depends on a large number of genes, "any two cognitive tasks of the type used for IQ tests will share some fraction of their genetic determinants" (p. 368).

Assuming that each locus has an independent and equal effect on behavioral variance, Anderson (2001) claimed that the overlapping genetic components cause the positive manifold: "Any two traits with shared components will have a positive correlation" (p. 369). Indeed, this account is very similar to the one proposed by Thomson, even to the equation predicting the size of the correlation based on the number of shared genes.[9] Therefore, the reasons why process overlap theory is more empirically plausible than the sampling model are also relevant to Anderson's account. Moreover, we disagree about the optimal level of explanation. It is not genes but psychological processes that are involved in cognitive behavior, hence we need an understanding of the nature of psychological processes as a proximate cause for the positive manifold.

The third theoretical account of the positive manifold that we wish to discuss is the mutualism model, a developmental account of the positive manifold that proposes positive reciprocal interactions between cognitive processes during development (van der Maas et al., 2006). The model describes the development of intelligence as the emergence of a complex dynamical network through the mutually beneficial interaction of modules or processes. According to this model, individual differences in cognitive abilities are uncorrelated at the beginning of development and start to correlate only because of such interactions.

The mutualism model bears many similarities to process overlap theory. It also explains g without postulating a single, general ability; it also rejects the reflective interpretation of g; and the explanation also relies on the interaction of separate processes. At the same time, van der Maas and colleagues proposed the functional independence of cognitive processes in mental test performance in their model while arguing that the positive manifold is the result of mutual interactions between cognitive processes only during development. That is, whereas in the mutualism model the interaction between processes takes place during development only, process overlap theory claims

---

[8]This assumption by Thomson (1916) was, in fact, more practical than substantive: "Note that I do not for one moment suggest that psychological 'factors,' if they exist, can be added together like dice: I merely intend to apply Professor Spearman's formulae to dice throwing" (p. 275).

[9]Even though the article does not refer to Thomson or to the concept of sampling.

that such interaction takes place when people solve mental tests.

The central assumption of the mutualism model is that learning in one cognitive domain positively affects development in other domains. In our opinion, even though the mathematical scaffolding of the mutualism model is greatly sophisticated and appealing, this assumption may need further empirical grounding. In particular, the strong cognitive transfer across domains that it proposes seems somewhat implausible. Furthermore, mutualism predicts that in adults, elementary cognitive tasks will be correlated, and as we have seen, this is not the case. Finally, combining evidence from psychometrics, experimental cognitive psychology, and neuroscience, process overlap theory is arguably based on more converging evidence.

On the other hand, because cognitive transfer probably occurs more easily within than across domains, mutualism appears as a very plausible explanation of how specific psychological processes get organized into clusters of abilities, represented by broad group factors—more so than of the correlations between the group factors themselves or between tests tapping different domains. Therefore, it might be possible to reconcile the two accounts, as it is quite likely that some processes indeed interact during development but not later in life, whereas others interact during actual problem solving.

The final theoretical account of the positive manifold to discuss is Detterman's (1987) system theory of intelligence. It argues that human intelligence functions as a complex system composed of smaller parts, and a global rating of cognitive functioning, such as IQ, does not reflect its constituents. In his conception of intelligence, Detterman borrowed two central concepts from system theory: *wholeness* and *centrality*. Wholeness refers to the interrelatedness of different parts of the system, and centrality means the extent to which a single part of the system influences the operation of the entire system.

Detterman (1987) argued that "the amount of variance accounted for by the first principal component is considered to be a measure of system wholeness for the variables measured" (p. 6). Therefore, the identification of individual components of the system results in processes that do not correlate. Moreover, according to Detterman, "a measure of wholeness, which I regard the first principal component to be, says nothing about centrality" (p. 7).

We completely subscribe to Detterman's basic theoretical approach and his conception of intelligence as a complex system with many independent components. In fact, it is quite easy to integrate the two theories. Employing his system terminology, process overlap theory emphasizes the centrality of executive processes rather than system wholeness as the main reason for the emergence of the positive manifold. The empirical evidence points to such executive processes overlapping with domain-specific ones in cognitive activity rather than to every process being related to every other process, as would be the case if intelligence were a system with very high wholeness.

Conclusion

Process overlap theory builds on available knowledge from psychometrics, cognitive psychology, and neuroscience to explain patterns of variation in mental abilities. As such, it is not a

taxonomy of human cognitive abilities and more than a latent variable model: It is a theoretical account that specifies the within-individual item response processes that are responsible for the positive manifold in intelligence. Besides the positive manifold, the theory explains a number of related phenomena: factor differentiation, the decrease of across-domain variance as a result of the Flynn effect, the identity or near-identity of Gf and *g* from an individual differences perspective, and the worst performance rule.

The theory proposes that the positive manifold, and thus *g*, will emerge from a battery of tasks that tap various important domain-general processes in an overlapping fashion. In particular, executive processes, seated primarily in the prefrontal and partly in the parietal cortex, overlap more with domain-specific processes in mental test performance than such specific processes overlap with one another. To arrive at a correct answer on a mental test item, one has to pass each tapped "dimension"; therefore, individual differences in executive processes function as a bottleneck for variation in specific processes. As a consequence, complex tasks requiring substantial executive processing, as well as errors in tasks requiring attention, are the most indicative of the domain-generality of the positive manifold.

It is important to note that the prefrontal cortex is not the seat of a unitary central executive, nor is executive function unitary from a psychological point of view. Hence *there need not be a single psychological process tapped by all intelligence tests* to obtain the positive manifold. Instead, a set of executive processes function as a "bridge" connecting more specialized networks of cognitive processes. Accordingly, process overlap theory's interpretation of double dissociation results in the light of the positive manifold is that cognition is not characterized by independent encapsulated processes or "modules" but instead by multiple sets of processes that are engaged in an overlapping fashion by cognitive operations.

Process overlap theory does not question the existence of psychometric *g*. In fact, it is not even logically possible to admit the existence of the positive manifold but not of a general factor, because the latter is a necessary algebraic consequence of the former. What is discarded is "*psychological g*": the interpretation of psychometric *g* as a psychological construct. There is no psychological process that corresponds to psychometric *g*. Instead, *g* is conceptualized as a formative variable: It emerges *because* of the positive manifold rather than *explaining* it.

Thus, it is imperative not to interpret process overlap theory as if it identified *g* with executive functions—with a few possible mediators like fluid reasoning and working memory. The theory indeed says that working memory and fluid intelligence are hugely overlapping constructs and that the overlap is caused by executive functions but *g* is not interpreted as a psychological construct of any kind. Instead, it is characterized as an emergent property, a result of how processes overlap to produce cognitive activity required by mental tests.

Also, even though our reading of the evidence is that such a functional overlap can account for the bulk of the domain-general variance that can be described with psychometric *g*, we are ready to acknowledge that there might be other sources contributing to the positive manifold. Mutualism is a likely candidate (see "Comparison with Other Theories"), and so is

associative learning (e.g., Kaufman, DeYoung, Gray, Brown, & Mackintosh, 2009).

Besides explaining a large number of empirical phenomena, process overlap theory also makes a number of unique predictions. First, if the theory is correct, differentiation should occur in working memory as it occurs in intelligence. That is, correlations between verbal and spatial working memory tasks should be stronger below the population mean than above, and such differentiation should be more characteristic of working memory than of short-term memory. Second, there is a controversy surrounding age-differentiation, the assumption that the positive manifold is stronger in younger children. The available results are inconclusive, largely because the batteries and age groups are created in an arbitrary manner. 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 different developmental and aging patterns, and there is large individual variation in the maturation and aging process itself.

Finally, process overlap theory and sampling provide different predictions for neuroscience. Thomson postulated a large number of domain-general bonds that are randomly sampled by different cognitive demands, and the more bonds sampled the higher they correlate with the general factor. Therefore, according to original sampling models, g loadings should correlate with the number of activated clusters in the brain, regardless of their location. Process overlap theory, on the other hand, predicts that g loading should be a function of the involvement of particular areas of the brain rather than total activation. We hope that the theory will inspire substantial empirical research and data-driven development in the fascinating field of human intelligence.

Acknowledgments

Funding

References

Achard, S., Salvador, R., Whitcher, B., Suckling, J., & Bullmore, E. (2006). A resilient, low-frequency, small-world human brain functional network with highly connected association cortical hubs. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 26(1), 63–72. doi:10.1523/JNEUROSCI.3874-05.2006

Ackerman, P. L. (1988). Determinants of individual differences during skill acquisition: Cognitive abilities and information processing. *Journal of Experimental Psychology: General*, 117, 288–318. doi:10.1037/0096-3445.117.3.288

Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2005). Working memory and intelligence: The same or different constructs? *Psychological Bulletin*, 131, 30–60. doi:10.1037/0033-2909.131.1.30

Anderson, B. (2001). g as a consequence of shared genes. *Intelligence*, 29 (5), 367–371. doi:10.1016/S0160-2896(01)00069-1

Aron, A. R., Robbins, T. W., & Poldrack, R. A. (2004). Inhibition and the right inferior frontal cortex. *Trends in Cognitive Sciences*, 8(4), 170–177. doi:10.1016/j.tics.2004.02.010

Asplund, C. L., Todd, J. J., Snyder, A. P., & Marois, R. (2010). A central role for the lateral prefrontal cortex in goal-directed and stimulus-driven attention. *Nature Neuroscience*, 13(4), 507–512. doi:10.1038/nn.2509

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. *Psychology of Learning and Motivation*, 2, 89–105. doi:10.1016/S0079-7421(08)60422-3

Baddeley, A. D. (1996). Exploring the central executive. *The Quarterly Journal of Experimental Psychology Section A*, 49(1), 5–28. doi:10.1080/713755608

Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, 4(11), 417–423. doi:10.1016/S1364-6613(00)01538-2

Baddeley, A. D. (2002). Is working memory still working? *European Psychologist*, 7(2), 85–97. doi:10.1027/1016-9040.7.2.85

Baddeley, A. D., & Hitch, G. (1974). Working memory. *Psychology of Learning and Motivation*, 8, 47–89. doi:10.1016/S0079-7421(08)60452-1

Badre, D., & Wagner, A. D. (2004). Selection, integration, and conflict monitoring; assessing the nature and generality of prefrontal cognitive control mechanisms. *Neuron*, 41(3), 473–487.

Bagozzi, R. P. (2007). On the meaning of formative measurement and how it differs from reflective measurement: Comment on Howell, Breivik, and Wilcox (2007). *Psychological Methods*, 12(2), 229–237; discussion 238–245. doi:10.1037/1082-989X.12.2.229

Barbey, A. K., Colom, R., & Grafman, J. (2013). Dorsolateral prefrontal contributions to human intelligence. *Neuropsychologia*, 51(7), 1361–1369. doi:10.1016/j.neuropsychologia.2012.05.017

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: A Journal of Neurology*, 135(Pt. 4), 1154–1164. doi:10.1093/brain/aws021

Bartholomew, D. J., Allerhand, M., & Deary, I. J. (2013). Measuring mental capacity: Thomson's Bonds model and Spearman's g-model compared. *Intelligence*, 41(4), 222–233. doi:10.1016/j.intell.2013.03.007

Bartholomew, D. J., Deary, I. J., & Lawn, M. (2009). A new lease of life for Thomson's bonds model of intelligence. *Psychological Review*, 116(3), 567–579. doi:10.1037/a0016262

Bassett, D. S., & Bullmore, E. (2006). Small-world brain networks. *The Neuroscientist : A Review Journal Bringing Neurobiology, Neurology and Psychiatry*, 12(6), 512–523. doi:10.1177/1073858406293182

Blair, C. (2006). How similar are fluid cognition and general intelligence? A developmental neuroscience perspective on fluid cognition as an aspect of human cognitive ability. *The Behavioral and Brain Sciences*, 29(2), 109–125; discussion 125–160. doi:10.1017/S0140525×06009034

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2003). The theoretical status of latent variables. *Psychological Review*, 110(2), 203–219.

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, 108(3), 624–652.

Braver, T. S., & Bongiolatti, S. R. (2002). The role of frontopolar cortex in subgoal processing during working memory. *NeuroImage*, 15(3), 523–536. doi:10.1006/nimg.2001.1019

Bressler, S., & Menon, V. (2010). Large-scale brain networks in cognition: Emerging methods and principles. *Trends in Cognitive Sciences*, 14, 277–290. doi:10.1016/j.tics.2010.04.004

Bullmore, E., & Sporns, O. (2009). Complex brain networks: Graph theoretical analysis of structural and functional systems. *Nature Reviews. Neuroscience*, 10(3), 186–198. doi:10.1038/nrn2575

Bunting, M. (2006). Proactive interference and item similarity in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32(2), 183–196. doi:10.1037/0278-7393.32.2.183

Burgess, G. C., Gray, J. R., Conway, A. R. A., & Braver, T. S. (2011). Neural mechanisms of interference control underlie the relationship between
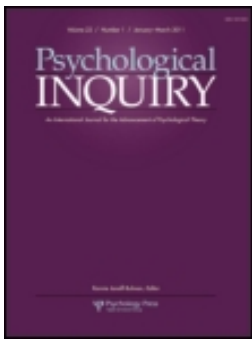
fluid intelligence and working memory span. *Journal of Experimental Psychology. General*, 140(4), 674–692. doi:10.1037/a0024695

Carpenter, P. A., Just, M. A., & Shell, P. (1990). What one intelligence test measures: A theoretical account of the processing in the Raven Progressive Matrices Test. *Psychological Review*, 97(3), 404–431.

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

Case, R., Kurland, D. M., & Goldberg, J. (1982). Operational efficiency and the growth of short-term memory span. *Journal of Experimental Child Psychology*, 33(3), 386–404. doi:10.1016/0022-0965(82)90054-6

Cattell, R. B. (1971). *Abilities: Their structure, growth, and action*. Boston, MA: Houghton Mifflin.

Chambers, C. D., Bellgrove, M. A., Stokes, M. G., Henderson, T. R., Garavan, H., Robertson, I. H., … Mattingley J. B. (2006). Executive "brake failure" following deactivation of human frontal lobe. *Journal of Cognitive Neuroscience*, 18(3), 444–455. doi:10.1162/089892906775990606

Chein, J. M., Moore, A. B., & Conway, A. R. A. (2011). Domain-general mechanisms of complex working memory span. *NeuroImage*, 54(1), 550–559. doi:10.1016/j.neuroimage.2010.07.067

Choi, Y. Y., Shamosh, N. A., Cho, S. H., DeYoung, C. G., Lee, M. J., Lee, J.-M., … Lee, K. H. (2008). Multiple bases of human intelligence revealed by cortical thickness and neural activation. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 28(41), 10323–10329. doi:10.1523/JNEUROSCI.3259-08.2008

Chow, M., & Conway, A. R. A. (2015). The scope and control of attention: Sources of variance in working memory capacity. *Memory & Cognition*, 43(3), 325–339. doi:10.3758/s13421-014-0496-9

Christoff, K., Prabhakaran, V., Dorfman, J., Zhao, Z., Kroger, J. K., Holyoak, K. J., & Gabrieli, J. D. (2001). Rostrolateral prefrontal cortex involvement in relational integration during reasoning. *NeuroImage*, 14(5), 1136–1149. doi:10.1006/nimg.2001.0922

Chuderski, A. (2015). The broad factor of working memory is virtually isomorphic to fluid intelligence tested under time pressure. *Personality and Individual Differences*, 85, 98–104. doi:10.1016/j.paid.2015.04.046

Cole, M. W., Pathak, S., & Schneider, W. (2010). Identifying the brain's most globally connected regions. *NeuroImage*, 49(4), 3132–3148. doi:10.1016/j.neuroimage.2009.11.001

Cole, M. W., & Schneider, W. (2007). The cognitive control network: Integrated cortical regions with dissociable functions. *NeuroImage*, 37(1), 343–360. doi:10.1016/j.neuroimage.2007.03.071

Cole, M., & Yarkoni, T. (2012). Global connectivity of prefrontal cortex predicts cognitive control and intelligence. *The Journal of Neuroscience*, 32(26), 8988–8999. doi:10.1523/JNEUROSCI.0536-12.2012

Colom, R., Burgaleta, M., Martínez, K., Román, F. J., Álvarez-Linera, J., Quiroga, M. Á., … Privado, J. (2011). Consistency of intelligence in the brain. In *International Society for the Study of Individual Differences (ISSID) 2011 meeting*.

Colom, R., Burgaleta, M., Román, F. J., Karama, S., Álvarez-linera, J., Abad, F. J., … Haier R. J. (2013). Neuroanatomic overlap between intelligence and cognitive factors: Morphometry methods provide support for the key role of the frontal lobes. Paper presented at International Society for the Study of Individual Differences (ISSID), Barcelona, Spain, July 22–25. doi:10.1016/j.neuroimage.2013.01.032

Colom, R., Haier, R., & Head, K. (2009). Intelligence Gray matter correlates of fluid, crystallized, and spatial intelligence: Testing the P-FIT model. *Intelligence*, 37(2), 124–135. doi:10.1016/j.intell.2008.07.007

Colom, R., Jung, R. E., & Haier, R. J. (2006a). Distributed brain sites for the g-factor of intelligence. *NeuroImage*, 31(3), 1359–65. doi:10.1016/j.neuroimage.2006.01.006

Colom, R., Jung, R. E., & Haier, R. J. (2006b). Finding the g-factor in brain structure using the method of correlated vectors. *Intelligence*, 34(6), 561–570. doi:10.1016/j.intell.2006.03.006

Conway, A. R. A., Cowan, N., & Bunting, M. F. (2001). The cocktail party phenomenon revisited: The importance of working memory capacity. *Psychonomic Bulletin & Review*, 8(2), 331–335. doi:10.3758/BF03196169

Conway, A. R. A., Cowan, N., Bunting, M. F., Therriault, D. J., & Minkoff, S. R. (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence*, 30(2), 163–183. doi:10.1016/S0160-2896(01)00096-4

Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, 12(5), 769–786. doi:10.3758/BF03196772

Conway, A. R. A., Kane, M. J., & Engle, R. W. (1999). Is Spearman's g determined by speed or working memory capacity? Book review of Jensen on Intelligence-g-Factor. *PSYCOLOQUY*, 10(074).

Conway, A. R. A., Kane, M. J., & Engle, R. W. (2003). Working memory capacity and its relation to general intelligence. *Trends in Cognitive Sciences*, 7(12), 547–552. doi:10.1016/j.tics.2003.10.005

Conway, A. R. A., & Kovacs, K. (2013). Individual differences in intelligence and working memory. *Psychology of Learning and Motivation*, 58, 233–270. doi:10.1016/B978-0-12-407237-4.00007-4

Conway, A. R. A., Macnamara, B., Getz, S., & Engel de Abreu, P. (2011). Working memory and fluid intelligence: A multi-mechanism view. In R. Sternberg & S. Kaufman (Eds.), *Cambridge handbook of intelligence* (pp. 394–418). New York, NY: Cambridge University Press.

Conway, A. R. A., Tuholski, S. W., Shisler, R. J., & Engle, R. W. (1999). The effect of memory load on negative priming: An individual differences investigation. *Memory & Cognition*, 27(6), 1042–1050. doi:10.3758/BF03201233

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews. Neuroscience*, 3(3), 201–215. doi:10.1038/nrn755

Cowan, N., Elliott, E. M., Saults, S. J., Morey, C. C., Mattox, S., Hismjatullina, A., & Conway, A. R. A. (2005). On the capacity of attention: Its estimation and its role in working memory and cognitive aptitudes. *Cognitive Psychology*, 51, 42–100. doi:10.1016/j.cogpsych.2004.12.001

Coyle, T. R. (2001). IQ is related to the worst performance rule in a memory task involving children. *Intelligence*, 29(2), 117–129. doi:10.1016/S0160-2896(00)00044-1

Coyle, T. R. (2003a). IQ, the worst performance rule, and Spearman's law: A reanalysis and extension. *Intelligence*, 31(5), 473–489. doi:10.1016/S0160-2896(02)00175-7

Coyle, T. R. (2003b). A review of the worst performance rule: Evidence, theory, and alternative hypotheses. *Intelligence*, 31(6), 567–587. doi:10.1016/S0160-2896(03)00054-0

Coyle, T. R., & Rindermann, H. (2013). Spearman's Law of Diminishing Returns and national ability. *Personality and Individual Differences*, 55(4), 406–410. doi:10.1016/j.paid.2013.03.023

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, 19(4), 450–466. doi:10.1016/S0022-5371(80)90312-6

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

Derrfuss, J., Brass, M., Neumann, J., & von Cramon, D. Y. (2005). Involvement of the inferior frontal junction in cognitive control: Meta-analyses of switching and Stroop studies. *Human Brain Mapping*, 25(1), 22–34. doi:10.1002/hbm.20127

Detterman, D. K. (1987). Theoretical notions of intelligence and mental fretardation. *American Journal of Mental Deficiency*, 92, 2–11.

Detterman, D. K. (1994). Theoretical possibilities: The relation of human intelligence to basic cognitive abilities. In D. K. Detterman (Ed.), *Current topics in human intelligence volume 4: Theories of intelligence* (pp. 85–115). Norwood, NJ: Ablex.

Detterman, D. K. (2000). General intelligence and the definition of phenotypes. In G. R. Bock, J. A. Goode, & K. Webb (Eds.), *The nature of intelligence* (pp. 136–143). New York, NY: Wiley & Sons.

Detterman, D. K., & Daniel, M. H. (1989). Correlations of mental tests with each other and with cognitive variables are highest for low IQ groups. *Intelligence*, 13(4), 349–359. doi:10.1016/S0160-2896(89)80007-8

Dreher, J.-C., & Berman, K. F. (2002). Fractionating the neural substrate of cognitive control processes. *Proceedings of the National Academy of Sciences of the United States of America*, 99, 14595–14600. doi:10.1073/pnas.222193299

Duncan, J. (1995). Attention, intelligence, and the frontal lobes. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 721–733). Cambridge, MA: MIT Press.

Duncan, J. (2010, February). The multiple-demand (MD) system of the primate brain: Mental programs for intelligent behaviour. *Trends in Cognitive Sciences*, pp. 172–179. doi:10.1016/j.tics.2010.01.004

Duncan, J., Burgess, P., & Emslie, H. (1995). Fluid intelligence after frontal lobe lesions. *Neuropsychologia*, 33(3), 261–268. doi:10.1016/0028-3932(94)00124-8

Duncan, J., Emslie, H., Williams, P., Johnson, R., & Freer, C. (1996). Intelligence and the frontal lobe: The organization of goal-directed behavior. *Cognitive Psychology*, 30(3), 257–303. doi:10.1006/cogp.1996.0008

Duncan, J., & Owen, A. M. (2000). Common regions of the human frontal lobe recruited by diverse cognitive demands. *Trends in Neurosciences*, 23(10), 475–483. doi:10.1016/S0166-2236(00)01633-7

Duncan, J., Parr, A., Woolgar, A., Thompson, R., Bright, P., Cox, S., … Nimmo-Smith, I. (2008). Goal neglect and Spearman's g: Competing parts of a complex task. *Journal of Experimental Psychology. General*, 137, 131–148. doi:10.1037/0096-3445.137.1.131

Duncan, J., Schramm, M., Thompson, R., & Dumontheil, I. (2012). Task rules, working memory, and fluid intelligence. *Psychonomic Bulletin & Review*, 19(5), 864–870. doi:10.3758/s13423-012-0225-y

Duncan, J., Seitz, J., Kolodny, J., Bor, D., Herzog, H., Ahmed, A., … Emslie, H. (2000, July). A neural basis for general intelligence. *Science*, 289, 457–460.

Dutilh, G., Vandekerckhove, J., Forstmann, B. U., Keuleers, E., Brysbaert, M., & Wagenmakers, E.-J. (2012). Testing theories of post-error slowing. *Attention, Perception, & Psychophysics*, 74, 454–465. doi:10.3758/s13414-011-0243-2

Dux, P. E., Ivanoff, J., Asplund, C. L., & Marois, R. (2006). Isolation of a central bottleneck of information processing with time-resolved FMRI. *Neuron*, 52(6), 1109–1120. doi:10.1016/j.neuron.2006.11.009

Edwards, J. R. (2011). The Fallacy of Formative Measurement. *Organizational Research Methods*, 14(2), 370–388. doi:10.1177/1094428110378369

Engle, R. W., & Kane, M. J. (2004). Executive attention, working memory capacity, and a two-factor theory of cognitive control. *The Psychology of Learning and Motivation*, 44, 145–199.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology. General*, 128(3), 309–331.

Eysenck, H. J. (1987). Thomson's "bonds" or Spearman's "energy": Sixty years on. *Mankind Quarterly*, 27, 259–274.

Eysenck, H. J. (1998). *Intelligence: A new look*. New Brunswick, NJ: Transaction.

Farooqui, A., Mitchell, D., Thompson, R., & Duncan, J. (2012). Hierarchical organization of cognition reflected in distributed frontoparietal activity. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, 32(48), 17373–17381. doi:10.1523/JNEUROSCI.0598-12.2012

Fedorenko, E., Duncan, J. & Kanwisher, N. (2013). Broad domain generality in focal regions of frontal and parietal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 110, 16616–16621. doi:10.1073/pnas.1315235110/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1315235110

Flynn, J. R. (1998). Reply to Rushton: A gang of gs overpowers factor analysis. *Personality and Individual Differences*, 26, 391–393.

Flynn, J. R. (2007). *What Is intelligence?: Beyond the Flynn Effect*. Cambridge, UK: Cambridge University Press.

Fukuda, K., Vogel, E., Mayr, U., & Awh, E. (2010). Quantity, not quality: The relationship between fluid intelligence and working memory capacity. *Psychonomic Bulletin & Review*, 17, 673–679. doi:10.3758/17.5.673

Gläscher, 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:10.1073/pnas.0910397107

Gold, J. M., Carpenter, C., Randolph, C., Goldberg, T. E., & Weinberger, D. R. (1997). Auditory Working Memory and Wisconsin Card Sorting Test performance in schizophrenia. *Archives of General Psychiatry*, 54, 159–165. doi:10.1001/archpsyc.1997.01830140071013

Gray, J. R., Chabris, C. F., & Braver, T. S. (2003). Neural mechanisms of general fluid intelligence. *Nature Neuroscience*, 6(3), 316–322. doi:10.1038/nn1014

Green, A. E., Fugelsang, J. a, Kraemer, D. J. M., Shamosh, N. A., & Dunbar, K. N. (2006). Frontopolar cortex mediates abstract integration in analogy. *Brain Research*, 1096(1), 125–137. doi:10.1016/j.brainres.2006.04.024

Gustafsson, J.-E. (1984). A unifying model for the structure of intellectual abilities. *Intelligence*, 8(3), 179–203. doi:10.1016/0160-2896(84)90008-4

Haier, R. J., Colom, R., Schroeder, D. H., Condon, C. A., Tang, C., Eaves, E., & Head, K. (2009). Intelligence Gray matter and intelligence factors: Is there a neuro-g? *Intelligence*, 37(2), 136–144. doi:10.1016/j.intell.2008.10.011

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

He, Y., & Evans, A. (2010). Graph theoretical modeling of brain connectivity. *Current Opinion in Neurology*, 23(4), 341–350. doi:10.1097/WCO.0b013e32833aa567

Hebb, D. O. (1940). Human behavior after extensive bilateral removal from the frontal lobes. *Archives of Neurology And Psychiatry*, 44(2), 421. doi:10.1001/archneurpsyc.1940.02280080181011

Heitz, R. P., Redick, T. S., Hambrick, D. Z., Kane, M. J., Conway, A. R. A., & Engle, R. W. (2006). Working memory, executive function, and general fluid intelligence are not the same. *Behavioral and Brain Sciences*, 29(02), 135–136. doi:10.1017/S0140525×06319036

Henson, R. N. A. (2001). Neural working memory. In J. Andrade (Ed.), *Working memory in perspective* (pp. 151–173). New York: Psychology Press.

Heuvel, M. P. Van Den, & Pol, H. E. H. (2010). Exploring the brain network: A review on resting-state fMRI functional connectivity. *European Neuropsychopharmacology*, 20(8), 519–534. doi:10.1016/j.euroneuro.2010.03.008

Heuvel, M. van den, & Stam, C. (2009). Efficiency of functional brain networks and intellectual performance. *The Journal of Neuroscience*, 29, 7619–7624. doi:10.1523/JNEUROSCI.1443-09.2009

Horn, J. L. (1989). Cognitive diversity: A framework of learning. In P. L. Ackerman, R. J. Sternberg, & R. Glaser (Eds.), *Learning and individual differences* (pp. 61–116). New York, NY: W.H. Freeman.

Horn, J. L. (1994). Theory of fluid and crystallized intelligence. In R. Sternberg (Ed.), *Encyclopedia of Human Intelligence* (pp. 443–451). New York: MacMillan Reference Library.

Howell, R. D., Breivik, E., & Wilcox, J. B. (2007). Reconsidering formative measurement. *Psychological Methods*, 12(2), 205–218. doi:10.1037/1082-989X.12.2.205

Jensen, A. R. (1981). Reaction time and intelligence. In M. Friedman, J. P. Das, & N. O'Connor (Eds.), *Intelligence and learning.* (pp. 39–50). New York: Plenum.

Jensen, A. R. (1982). Reaction time and psychometric g. In H. J., Eysenck (Ed.), *A model for intelligence* (pp. 93–132). New York: Springer Verlag.

Jensen, A. R. (1992). The importance of intraindividual variation in reaction time. *Personality and Individual Differences*, 13(8), 869–881. doi:10.1016/0191-8869(92)90004-9

Jensen, A. R. (1998). *The g factor: The science of mental ability. East* (Vol. 17). Westport, CT: Praeger.

Jensen, A. R. (2000). "The g factor" is about variance in mental abilities, not a cognitive theory of mental structure. Reply to Anderson on Jensen on Intelligence-g-Factor. *PSYCOLOQUY*, 11(041).

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:10.1016/j.intell.2004.12.002

Jonides, J., Reuter-Lorenz, P. A., Smith, E. E., Awh, E., Barnes, L. L., Drain, M., & Al, E. (1996). Verbal and spatial working memory in humans. *Psychology of Learning and Motivation*, 35, 43–88.

Juan-Espinosa, M., Cuevas, L., Escorial, S., & García, L. F. (2006). The differentiation hypothesis and the Flynn effect. *Psicothema*, 18(2), 284–287.

Jung, R. E., & Haier, R. J. (2007). The Parieto-Frontal Integration Theory (P-FIT) of intelligence: Converging neuroimaging evidence. *The Behavioral and Brain Sciences*, 30(2), 135–154; discussion 154–187. doi:10.1017/S0140525×07001185

Kan, K.-J., Kievit, R. A., Dolan, C., & van der Maas, H. L. J. (2011). On the interpretation of the CHC factor Gc. *Intelligence*, 39(5), 292–302. doi:10.1016/j.intell.2011.05.003

Kane, H. (2000). A secular decline in Spearman's g: Evidence from the WAIS, WAIS-R and WAIS-III. *Personality and Individual Differences*, 29, 561–566.

Kane, H. D., Oakland, T. D., & Brand, C. R. (2006). Differentiation at higher levels of cognitive ability: Evidence from the United States. *The Journal of Genetic Psychology*, *167*(3), 327–341. doi:10.3200/GNTP.167.3.327-341

Kane, H., & Oakland, T. D. (2000). Secular declines in Spearman's g: Some evidence from the United States. *The Journal of Genetic Psychology*, *161*(3), 337–345. doi:10.1080/00221320009596716

Kane, M., & Engle, R. (2000). Working-memory capacity, proactive interference, and divided attention: Limits on long-term memory retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(2), 336–358. doi:10.10371/0278-7393.26.2.336

Kane, M. J. (2005). Full frontal fluidity. In O. Wilhelm & R. Engle (Eds.), *Handbook of understanding and measuring intelligence* (pp. 141–165). Thousand Oaks, CA: Sage.

Kane, M. J., Bleckley, M. K., Conway, A. R. A., & Engle, R. W. (2001). A controlled-attention view of working-memory capacity. *Journal of Experimental Psychology: General*, *130*(2), 169–183. doi:10.1037/0096-3445.130.2.169

Kane, M. J., Conway, A. R. A., Miura, T. K., & Colflesh, G. J. H. (2007). Working memory, attention control, and the N-back task: A question of construct validity. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *33*(3), 615–622. doi:10.1037/0278-7393.33.3.615

Kane, M. J., Conway, A. R. A., Hambrick, D. Z., & Engle, R. W. (2007). Variation in working memory as variation in executive attention and control. In A. R. A. Conway, C. Jarrold, M. J. Kane, A. Miyake & J. N. Towse (Eds.),*Variation in working memory*. New York: Oxford University Press.

Kane, M. J., & Engle, R. W. (2002). The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: An individual-differences perspective. *Psychonomic Bulletin & Review*, *9*(4), 637–671.

Kane, M. J., & Engle, R. W. (2003). Working-memory capacity and the control of attention: The contributions of goal neglect, response competition, and task set to Stroop interference. *Journal of Experimental Psychology: General*, *132*(1), 47–70. doi:10.1037/0096-3445.132.1.47

Kane, M. J., Hambrick, D. Z., & Conway, A. R. A. (2005). Working memory capacity and fluid intelligence are strongly related constructs: Comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin*, *131*, 66–71; author reply 72–75. doi:10.1037/0033-2909.131.1.66

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: A latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology. General*, *133*(2), 189–217. doi:10.1037/0096-3445.133.2.189

Kaufman, S. B., DeYoung, C. G., Gray, J. R., Brown, J., & Mackintosh, N. (2009). Associative learning predicts intelligence above and beyond working memory and processing speed. *Intelligence*, *37*(4), 374–382. doi:10.1016/j.intell.2009.03.004

Kievit, R. A., Davis, S. W., Mitchell, D. J., Taylor, J. R., Duncan, J., & Henson, R. N. A. (2014). Distinct aspects of frontal lobe structure mediate age-related differences in fluid intelligence and multitasking. *Nature Communications*, *5*, 56–58. doi:10.1038/ncomms6658

Kievit, R. A., Rooijen, H. van, Wicherts, J. M., Waldorp, L. J., Kan, K.-J., Scholte, H. S., & Borsboom, D. (2012). Intelligence and the brain: A model-based approach. *Cognitive Neuroscience*, *3*(2), 89–97. doi:10.1080/17588928.2011.628383

Kievit, R. A., Scholte, H. S., Waldorp, L. J. & Borsboom, D. (2016). Inter- and intra-individual differences in fluid reasoning show distinct cortical responses. *bioRxiv*, preprint. doi:10.1101/039412

Koechlin, E., & Hyafil, A. (2007). Anterior Prefrontal Function and the Limits of Human Decision-Making. *Science*, *318*(5850), 594–598.

Kovacs, K. (2009). A component process account of the general factor of intelligence. Unpublished PhD thesis. University of Cambridge.

Kovacs, K., Plaisted, K. C., & Mackintosh, N. J. (2006). Difficulties differentiating dissociations. *Behavioral and Brain Sciences*, *29*(02), 138–139. doi:10.1017/S0140525×06349035

Kovacs, K., Rozsa, S., Ko, N., Meszaros, A., Mlinko, R., Kuncz, E., Nagyne Rez, I., Lanyine Engelmayer, A. (2016). Forward and backward digit span measure different cognitive abilities. Manuscript submitted for publication.

Kranzler, J. H. (1992). A test of Larson and Alderton's (1990) worst performance rule of reaction time variability. *Personality and Individual Differences*, *13*(3), 255–261. doi:10.1016/0191-8869(92)90099-B

Kranzler, J. H., & Jensen, A. R. (1991). The nature of psychometric g: Unitary process or a number of independent processes? *Intelligence*, *15*(4), 397–422. doi:10.1016/0160-2896(91)90003-V

Kranzler, J. H., & Jensen, A. R. (1993). Psychometric g is still not unitary after eliminating supposed "impurities": Further comment on Carroll. *Intelligence*, *17*(1), 11–14. doi:10.1016/0160-2896(93)90033-2

Krijnen, W. P. (2004). Positive loadings and factor correlations from positive covariance matrices. *Psychometrika*, *69*(4), 655–660. doi:10.1007/BF02289861

Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity?! *Intelligence*, *14*, 389–433. doi:10.1016/S0160-2896(05)80012-1

Larson, G. E., & Alderton, D. L. (1990). Reaction time variability and intelligence: A "worst performance" analysis of individual differences. *Intelligence*, *14*(3), 309–325. doi:10.1016/0160-2896(90)90021-K

Lehto, J. (1996). Are executive function tests dependent on working memory capacity? *The Quarterly Journal of Experimental Psychology A*, *49*, 29–50. doi:10.1080/027249896392793

Lehto, J. E., Juujärvi, P., Kooistra, L., & Pulkkinen, L. (2003). Dimensions of executive functioning: Evidence from children. *British Journal of Developmental Psychology*, *21*(1), 59–80. doi:10.1348/026151003321164627

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*, 279–281. doi:10.1038/36846

Luo, Q., Perry, C., Peng, D., Jin, Z., Xu, D., Ding, G., & Xu, S. (2003). The neural substrate of analogical reasoning: An fMRI study. *Brain Research. Cognitive Brain Research*, *17*(3), 527–534.

Lynn, R., & Cooper, C. (1993). A secular decline in Spearman's g in France. *Learning and Individual Differences*, *5*(1), 43–48.

Lynn, R., & Cooper, C. (1994). A secular decline in the strength of Spearman'sg in Japan. *Current Psychology*, *13*(1), 3–9.

MacDonald, A. W. (2000). Dissociating the role of the dorsolateral prefrontal and anterior cingulate cortex in cognitive control. *Science*, *288*(5472), 1835–1838. doi:10.1126/science.288.5472.1835

Mackintosh, N., & Bennett, E. (2002). IT, IQ and perceptual speed. *Personality and Individual Differences*, *32*(4), 685–693. doi:10.1016/S0191-8869(01)00069-1

Mackintosh, N. J. (1998). *IQ and human intelligence (first edition)*. Oxford, UK: Oxford University Press.

Mackintosh, N. J. (2008). Sex differences and IQ. *Journal of Biosocial Science*, *28*(04), 558–571. doi:10.1017/S0021932000022586

Mackintosh, N. J. (2011a). History of theories and measurement of intelligence. In R. Sternberg & S. Kaufman (Eds.), *Cambridge handbook of intelligence* (pp. 3–19). New York, NY: Cambridge University Press.

Mackintosh, N. J. (2011b). *IQ and human intelligence*. Oxford, UK: Oxford University Press.

Major, J. T., Johnson, W., & Bouchard, T. J. (2011). The dependability of the general factor of intelligence: Why small, single-factor models do not adequately represent g. *Intelligence*, *39*(5), 418–433. doi:10.1016/j.intell.2011.07.002

Marois, R., & Ivanoff, J. (2005). Capacity limits of information processing in the brain. *Trends in Cognitive Sciences*, *9*(6), 296–305. doi:10.1016/j.tics.2005.04.010

Matzke, D., Dolan, C. V., & Molenaar, D. (2010). The issue of power in the identification of "g" with lower-order factors. *Intelligence*, *38*(3), 336–344. doi:10.1016/j.intell.2010.02.001

Maxwell, A. E. (1972). Factor analysis: Thomson's sampling theory recalled. *British Journal of Mathematical and Statistical Psychology*, *25*(1), 1–21. doi:10.1111/j.2044-8317.1972.tb00474.x

McCabe, D. P., Roediger, H. L., McDaniel, M. A., Balota, D. A., & Hambrick, D. Z. (2010). The relationship between working memory capacity and executive functioning: Evidence for a common executive attention construct. *Neuropsychology*, *24*(2), 222–243. doi:10.1037/a0017619

McFarland, D. J. (2012). A single g factor is not necessary to simulate positive correlations between cognitive tests. *Journal of Clinical*

and Experimental Neuropsychology, 34(4), 378–384. doi:10.1080/13803395.2011.645018

McGrew, K. S. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence*, 37(1), 1–10. doi:10.1016/j.intell.2008.08.004

Miele, F. (2002). *Intelligence, race, and genetics. Conversations with Arthur R. Jensen.* Boulder, CO: Westview Press.

Miller, B. T., & D'Esposito, M. (2005). Searching for "the top" in top-down control. *Neuron*, 48(4), 535–538. doi:10.1016/j.neuron.2005.11.002

Miller, E. K. (2000). The prefrontal cortex and cognitive control. *Nature Reviews. Neuroscience*, 1(1), 59–65. doi:10.1038/35036228

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, 24, 167–202. doi:10.1146/annurev.neuro.24.1.167

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "Frontal Lobe" tasks: A latent variable analysis. *Cognitive Psychology*, 41, 49–100. doi:10.1006/cogp.1999.0734

Miyake, A., Friedman, N. P., Rettinger, D. A., Shah, P., & Hegarty, M. (2001). How are visuospatial working memory, executive functioning, and spatial abilities related? A latent-variable analysis. *Journal of Experimental Psychology: General*, 130, 621–640. doi:10.1037/0096-3445.130.4.621

Molenaar, D., Dolan, C. V., Wicherts, J. M., & van der Maas, H. L. J. (2010). Modeling differentiation of cognitive abilities within the higher-order factor model using moderated factor analysis. *Intelligence*, 38(6), 611–624. doi:10.1016/j.intell.2010.09.002

Molenaar, P. C. M., & Campbell, C. G. (2009). The New Person-Specific Paradigm in Psychology. *Current Directions in Psychological Science*, 18(2), 112–117. doi:10.1111/j.1467-8721.2009.01619.x

Mueller, S., Wang, D., Fox, M. D., Yeo, B. T. T., Sepulcre, J., Sabuncu, M. R., … Liu, H. (2013). Individual variability in functional connectivity architecture of the human brain. *Neuron*, 77(3), 586–595. doi:10.1016/j.neuron.2012.12.028

Mukunda, K. V., & Hall, V. C. (1992). Does performance on memory for order correlate with performance on standardized measures of ability? A meta-analysis. *Intelligence*, 16(1), 81–97. doi:10.1016/0160-2896(92)90026-N

Must, O., Must, A., & Raudik, V. (2003). The secular rise in IQs: In Estonia, the Flynn effect is not a Jensen effect. *Intelligence*, 31(5), 461–471. doi:10.1016/S0160-2896(03)00013-8

Négyessy, L., Bányai, M., Nepusz, T., & Bazsó, F. (2012). What makes the prefrontal cortex so appealing in the era of brain imaging? A network analytical perspective. *Acta Biologica Hungarica*, 63, 38–53. doi:10.1556/ABiol.63.2012.Suppl.1.5

Niendam, T., Laird, A., & Ray, K. (2012). Meta-analytic evidence for a superordinate cognitive control network subserving diverse executive functions. *Cognitive, Affective, & Behavioral Neuroscience*, 12(2), 241–268. doi:10.3758/s13415-011-0083-5.Meta-analytic

Oberauer, K. (2004). The measurement of working memory capacity. In O. Wilhelm & R. W. Engle (Eds.), *Handbook of understanding and measuring intelligence* (pp. 393–408). London, UK: Sage.

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. doi:10.1037/0033-2909.131.1.61

Oberauer, K., Süß, H.-M., Wilhelm, O., & Wittman, W. W. (2003). The multiple faces of working memory: Storage, processing, supervision, and coordination. *Intelligence*, 60, 167–193. doi:10.1016/S0160-2896(02)00115-0

Parkin, A. (1998). The central executive does not exist. *Journal of the International Neuropsychological Society*, 4, 518–22. doi:10.1017/s1355617798005128

Penke, L., Maniega, S. M., Bastin, M. E., Valdés Hernández M. C., Murray, C., Royle, N. A., … Deary, I. J. (2012). Brain white matter tract integrity as a neural foundation for general intelligence. *Molecular Psychiatry*, 17(10), 1026–1030. doi:10.1038/mp.2012.66

Pollack, I., Johnson, L. B., & Knaff, P. R. (1959). Running memory span. *Journal of Experimental Psychology*, 57, 137–146. doi:10.1037/h0046137

Prabhakaran, V., Smith, J. A, Desmond, J. E., Glover, G. H., & Gabrieli, J. D. (1997). Neural substrates of fluid reasoning: An fMRI study of neocortical activation during performance of the Raven's Progressive Matrices Test. *Cognitive Psychology*, 33(1), 43–63. doi:10.1006/cogp.1997.0659

Rabaglia, C., Marcus, G., & Lane, S. (2011). What can individual differences tell us about the specialization of function? *Cognitive Neuropsychology*, 28, 37–41.

Rabbitt, P. M. A. (1997). Introduction: Methodologies and models in the study of executive function. In P. M. A. Rabbitt (Ed.), *Methodology of frontal and executive function* (pp. 1–38). East Sussex, UK: Psychology Press.

Reckase, M. D. (2009). *Multidimensional item response theory*. New York, NY: Springer-Verlag.

Robbins, T. W. (1996). Dissociating executive functions of the prefrontal cortex. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 351, 1463–1470; discussion 1470–1471. doi:10.1098/rstb.1996.013

Román, F. J., Abad, F. J., Escorial, S., Burgaleta, M., Martínez, K., Alvarez-Linera, J., … Colom, R. (2014). Reversed hierarchy in the brain for general and specific cognitive abilities: A morphometric analysis. *Human Brain Mapping*, 35, 3805–3818. doi:10.1002/hbm.22438

Rushton, J. (1999). Secular gains in IQ not related to the g factor and inbreeding depression Ð unlike Black-White differences: A reply to Flynn. *Personality and Individual Differences*, 26, 381–389.

Salthouse, T. A., Atkinson, T. M., & Berish, D. E. (2003). Executive functioning as a potential mediator of age-related cognitive decline in normal adults. *Journal of Experimental Psychology. General*, 132(4), 566–594. doi:10.1037/0096-3445.132.4.566

Santarnecchi, E., & Galli, G. (2014). Efficiency of weak brain connections support general cognitive functioning. *Human Brain*, 35, 4566–4582. doi:10.1002/hbm.22495

Shipstead, Z., Redick, T. S., Hicks, K. L., & Engle, R. W. (2012). The scope and control of attention as separate aspects of working memory. *Memory*, 20, 608–662. doi:10.1080/09658211.2012.691519

Sohn, M. H., Ursu, S., Anderson, J. R., Stenger, V. A., & Carter, C. S. (2000). The role of prefrontal cortex and posterior parietal cortex in task switching. *Proceedings of the National Academy of Sciences of the United States of America*, 97(24), 13448–13453. doi:10.1073/pnas.240460497

Song, M., Zhou, Y., Li, J., Liu, Y., Tian, L., Yu, C., & Jiang, T. (2008). Brain spontaneous functional connectivity and intelligence. *Neuroimage*, 41, 1168–1176. doi:10.1016/j.neuroimage.2008.02.036

Spearman, C. (1904). "General intelligence," Objectively determined and measured. *The American Journal of Psychology*, 15(2), 201–292. doi:10.2307/1412107

Spearman, C. (1927). *The abilities of Man: Their nature and measurement.* New York: Macmillan.

Sporns, O. (2002). Network analysis, complexity, and brain function. *Complexity*, 8(1), 56–60.

Sporns, O., Honey, C., & Kötter, R. (2007). Identification and classification of hubs in brain networks. *PloS One*, 10. doi:10.1371/Citation

Spreng, R. N., Sepulcre, J., Turner, G. R., Stevens, W. D., & Schacter, D. L. (2013). Intrinsic architecture underlying the relations among the default, dorsal attention, and frontoparietal control networks of the human brain. *Journal of Cognitive Neuroscience*, 25(1), 74–86. doi:10.1162/jocn_a_00281

St Clair-Thompson, H. L., & Gathercole, S. E. (2006). Executive functions and achievements in school: Shifting, updating, inhibition, and working memory. *Quarterly Journal of Experimental Psychology*, 59(4), 745–759. doi:10.1080/17470210500162854

Stuss, D. T., & Alexander, M. P. (2000). Executive functions and the frontal lobes: A conceptual view. *Psychological Research*, 63(3–4), 289–298.

Süß, H. M., Oberauer, K., Wittmann, W. W., Wilhelm, O., & Schulze, R. (2002). Working-memory capacity explains reasoning ability—And a little bit more. *Intelligence*, 30, 261–288. doi:10.1016/S0160-2896(01)00100-3

Thomson, G. H. (1916). A hierarchy without a general factor. *British Journal of Psychology, 1904-1920*, 8, 271–281. doi:10.1111/j.2044-8295.1916.tb00133.x

Thurstone, L. L. (1938). *Primary mental abilities*. Chicago, IL: University of Chicago Press.

Toga, A., Clark, K., & Thompson, P. (2012). Mapping the human connectome. *Neurosurgery*, 71(1), 1–5.

Tombu, M. N., Asplund, C. L., Dux, P. E., Godwin, D., Martin, J. W., & Marois, R. (2011). A unified attentional bottleneck in the human brain. *Proceedings of the National Academy of Sciences of the United States of America*, 108(33), 13426–13431. doi:10.1073/pnas.1103583108

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language*, 28(2), 127–154. doi:10.1016/0749-596X(89)90040-5

Unsworth, N., & Engle, R. W. (2006). Simple and complex memory spans and their relation to fluid abilities: Evidence from list-length effects. *Journal of Memory and Language*, 54, 68–80. doi:10.1016/j.jml.2005.06.003

Unsworth, N., & Engle, R. W. (2007). The nature of individual differences in working memory capacity: Active maintenance in primary memory and controlled search from secondary memory. *Psychological Review*, 114(1), 104–132. doi:10.1037/0033-295X.114.1.104

van der Maas, H. L. J., Dolan, C. V., Grasman, R. P. P. P., Wicherts, J. M., Huizenga, H. M., & Raijmakers, M. E. J. (2006). A dynamical model of general intelligence: The positive manifold of intelligence by mutualism. *Psychological Review*, 113(4), 842–861. doi:10.1037/0033-295X.113.4.842

van der Maas, H. L. J., Kan, K.-J., & Borsboom, D. (2014). intelligence is what the intelligence test measures. Seriously. *Journal of Intelligence*, 2(1), 12–15. doi:10.3390/jintelligence2010012

Vicari, S., Bellucci, S., & Carlesimo, G. A. (2007). Visual and spatial working memory dissociation: Evidence from Williams syndrome. *Developmental Medicine & Child Neurology*, 45(4), 269–273. doi:10.1111/j.1469-8749.2003.tb00342.x

Vincent, J. L., Kahn, I., Snyder, A. Z., Raichle, M. E., & Buckner, R. L. (2008). Evidence for a frontoparietal control system revealed by intrinsic functional connectivity. *Journal of Neurophysiology*, 100(6), 3328–3342. doi:10.1152/jn.90355.2008

Voelkle, M. C., Brose, A., Schmiedek, F., & Lindenberger, U. (2014). Toward a unified framework for the study of between-person and within-person structures: Building a bridge between two research paradigms. *Multivariate Behavioral Research*, 49(3), 193–213. doi:10.1080/00273171.2014.889593

Volle, E., Gilbert, S. J., Benoit, R. G., & Burgess, P. W. (2010). Specialization of the rostral prefrontal cortex for distinct analogy processes. *Cerebral Cortex (New York, N.Y. : 1991)*, 20(11), 2647–2659. doi:10.1093/cercor/bhq012

Wager, T. O. R. D., & Smith, E. E. (2003). Neuroimaging studies of working memory: A meta-analysis. *Cognitive, Affective, & Behavioral Neuroscience*, 3(4), 255–274.

Waltz, J. A., Knowlton, B. J., Holyoak, K. J., Boone, K. B., Mishkin, F. S., de Menezes Santos, M., … Miller, B. L. (1999). A System for Relational Reasoning in Human Prefrontal Cortex. *Psychological Science*, 10(2), 119–125. doi:10.1111/1467-9280.00118

Wang, P. P., & Bellugi, U. (1994). Evidence from two genetic syndromes for a dissociation between verbal and visual-spatial short-term memory. *Journal of Clinical and Experimental Neuropsychology*, 16(2), 317–322. doi:10.1080/01688639408402641

Weinstein, S., & Teuber, H.-L. (1957). Effects of penetrating brain injury on intelligence test scores. *Science*, 125, 1036–1037. doi:10.1126/science.125.3256.1036

Wendelken, C., Nakhabenko, D., Donohue, S. E., Carter, C. S., & Bunge, S. A. (2008). "Brain is to thought as stomach is to ??": Investigating the role of rostrolateral prefrontal cortex in relational reasoning. *Journal of Cognitive Neuroscience*, 20(4), 682–693. doi:10.1162/jocn.2008.20055

Wharton, C. M., Grafman, J., Flitman, S. S., Hansen, E. K., Brauner, J., Marks, A., & Honda, M. (2000). Toward neuroanatomical models of analogy: A positron emission tomography study of analogical mapping. *Cognitive Psychology*, 40(3), 173–197. doi:10.1006/cogp.1999.0726

Wiley, J., & Jarosz, A. (2011). New rule use drives the relation between working memory capacity and Raven's Advanced Progressive Matrices. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 37(1), 256–263. doi:10.1037/a0021613

Woolgar, A., Parr, A., Cusack, R., Thompson, R., Nimmo-Smith, I., Torralva, T., … Duncan, J. (2010). Fluid intelligence loss linked to restricted regions of damage within frontal and parietal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 107(33), 14899–14902. doi:10.1073/pnas.1007928107

Process Overlap and g Do Not Adequately Account for a General Factor of Intelligence

Phillip L. Ackerman

Published online: 02 Aug 2016.

Submit your article to this journal ↗

View related articles ↗

View Crossmark data ↗

COMMENTARIES

Process Overlap and *g* Do Not Adequately Account for a General Factor of Intelligence

Phillip L. Ackerman

School of Psychology, Georgia Institute of Technology, Atlanta, Georgia

Reasoning problem:

Proposition: Process measures correlate with a general factor of intelligence.

Conclusion: Therefore, the general factor of intelligence is defined by process measures.

Question: True or false?

Process overlap theory (Kovacs & Conway, this issue) provides an interesting account of basic elements of *some* mental processes—especially those that include working memory tasks and abstract reasoning tests (such as Raven's Progressive Matrices). The theory appears to be internally consistent within the boundaries that are described in Kristof Kovacs and Andrew R. A. Conway's target article. Yet there is something highly unsatisfactory when the theory is presented as accounting for general fluid intelligence (Gf), a general factor of intelligence or even general intelligence, as represented in the article.

Kovacs and Conway describe a variety of theories of the structure of intelligence and provide a good review of one approach to intelligence. But the approach is almost entirely devoid of any consideration of the much larger field of intelligence research and application. The reasons for this are complex, but some history is important to review.

In the early 1900s, two approaches to human intelligence were proposed, one by Spearman and the other by Binet. In 1904, Spearman set out his theory of General Intelligence (g). He offered several propositions about *g* in that original paper, and others subsequently. What *g* actually was, was not entirely clear in his early statements. The foundation of *g* that Spearman derived empirically in 1904 was a hodgepodge of academic performances (Classics, *g*-loading of .99; Common Sense, *g*-loading of .98) and psychophysical abilities (Pitch Discrimination, *g*-loading of .94; Spearman, 1904, p. 276). Later, however, Krueger and Spearman (1907) pointed to the Ebbinghaus Completion Test as the single best test of *g* (Ebbinghaus, 1896–1897). In fact, in the 1907 paper, they reported the g-loading of the Ebbinghaus Completion Test was .97! The Ebbinghaus Completion Test is a measure of verbal fluency and memory (e.g., see Ackerman, Beier, & Bowen, 2000). Much later (in the 1930s), Spearman decided that the spatial inductive reasoning test developed by Penrose and Raven was an ideal measure of *g* (Spearman, 1938). Spearman showed little interest in the *external validity* of intelligence assessments. That is, prediction of criteria such as school success or occupational success was not

a concern for the validity or utility of his theory of general intelligence.

Binet, in contrast, had as his central focus, the use of intelligence assessments to predict failure in educational situations. Although Binet rightly pointed out the importance of an individual's knowledge in predicting academic success (Cattell would later refer to knowledge as subsumed under Crystallized Intelligence [Gc]), Binet largely rejected including assessments of knowledge in his intelligence scales, for the practical reason that he believed that such tests would place lower socioeconomic status students at a disadvantage to students from more enriched backgrounds. Even without extensive intellectual knowledge assessments, the Binet and Simon (1905/1961) scales were quite successful in predicting which students would not succeed in the normal classroom environment. There was a wide variety of components to these scales, ranging from counting, naming colors and the days of the week, constructing sentences, long-term and short-term memory, defining words, identifying similarities, and so on. The main criteria for retention of particular items/scales were (a) age differentiation (older students, on average, performed better than younger students) and (b) relevance to predicting academic performance. The scales became a nearly instant success. Within a short time, numerous translations and adaptations of the Binet–Simon scales appeared in the United States and elsewhere, such as the versions by Goddard, by Kuhlman, and later the Stanford–Binet by Terman (1916). The Binet scales became the benchmark against which most newer intelligence tests were assessed. The contents of general intelligence tests used today are markedly similar to those that Binet used/developed more than 100 years ago. Also, although Binet's tests were aimed at children between ages 3 and 13, extensions of the tests for adult assessment are also in wide use. Even Wechsler's Adult Intelligence Scales have substantial overlap with those in the Binet scales (Boake, 2002). Spearman (1930) denigrated Binet's assessments, partly because Binet's tests were fundamentally based on the notion of intellectual development during childhood (i.e., age differentiation), whereas Spearman's theory called for fixed intelligence, unchanging during development. Spearman's proposition is clearly not tenable—the average 10-year-old is much more capable than the average 5-year-old on any intellectually demanding task one could administer. Even studies with tests of abstract reasoning, such as the Penrose and Raven (1936) progressive matrices test, show higher average

scores for older children and adults, compared to young children.

In the decades that followed Spearman's theory of intelligence and Binet's intelligence scales, developments in these two areas often interacted. Although some researchers attempted to develop models of the structure of intellectual abilities that include Spearman's *g*, other more recent adherents of Spearman's theory have searched for the fundamental sources of *g*. In contrast, many of Binet's followers have attempted to improve intelligence assessments in terms of completeness and reliability, but ultimately such improvements were measured against the external validity of the assessments. Those who followed Spearman in attempting to measure *g* for application purposes were often substantially disappointed by the lack of validity shown by Raven's Progressive Matrices for real-world criteria, such as job performance (e.g., see Vernon & Parry, 1949).

In many ways, though, the connection has been broken between those who take Spearman's approach as the means toward understanding intelligence and those who actually develop and apply intelligence tests in the real world. In the past 30 years or so, numerous experimental psychologists have focused entirely on *only* the kinds of tests that Spearman advocated late in his career (e.g., the abstract reasoning tests that require little or no content knowledge), and they ignore both domain knowledge (Gc), and external validity concerns.

Even if we take on face value the fact that Kovacs and Conway adopt Vernon's (1950) notion that *g* might account for about 40% of the variance in general intelligence, how can an "adequate" theory of intelligence fail to account for roughly 60% of the variance in general intelligence? Moreover, how can an adequate theory be so firmly disconnected from any considerations of external validity?

Additional Issues

There are some additional concerns about the approach proposed by Kovacs and Conway (this issue), which I mention only briefly here.

1. *g*/Gf/Gc distinction. In Spearman's view, *g* was innate and fixed. Cattell's view of Gf theoretically diverged from Spearman's g, in that Gf "increases until adolescence and then slowly declines" (Cattell, 1943, p. 178). Although Cattell suggested that Gc represents intelligence that is educational or experiential, there is more than ample evidence that both Gf and Gc are developed through education and experience (e.g., see Ackerman & Lohman, 2003; Ceci, 1991; Snow, 1982, 1996). If a child does not go to school, for example, his or her performance on *g*-type tests will be seriously impaired (Ferguson, 1954). The effect of education on *g* does not appear to be accounted for in the process theory representation provided by Kovacs and Conway—indeed, the subject of education in the context of intelligence does not appear in the article at all.

2. Brunswik Symmetry. This is a concept introduced by Wittmann and Suß (1999), and it pertains to the match/mismatch between the breadth of predictors and criterion measures. The measures advocated by Kovacs and Conway appear to conflate "complexity" with "breadth"—that is, although many working memory and abstract reasoning tests are complex, they fail to represent the kind of breadth that is inherent in the general intellectual ability factor.

3. Indifference of the indicator. One of Spearman's main propositions was the "indifference of the indicator." That is, when it comes to estimating *g*, it "for the purpose of indicating the amount of *g* possessed by a person, any test will do just as well as any other, provided only that its correlation with *g* is equally high" (Spearman, 1927, p. 197). Process-based theories, such as the one proposed by Kovacs and Conway, fail in that they don't consider that this proposition is highly questionable—they implicitly assuming that content-free or content-reduced tests are the only tests that need to be examined, to understand *g*. That is, given that for example, the Ebbinghaus Completion Test has a higher *g*-loading than even the Raven's test, their theory should work as well explaining performance on the Completion Test—which most investigators might argue has only modest overlap with working memory tests. For example, with data from an experiment with college students, Ackerman, Beier, and Boyle (2002) found that a version of the Completion test correlated $r = .62$ with a general ability composite, in comparison to the g-loading of Raven's Progressive Matrices ($r = .58$) and the g-loading of a battery of working memory tests ($r = .55$). The Completion Test scores only correlated $r = .39$ with the battery of working memory tests. A challenging test for the process theory would be to take it beyond relatively content-free tests and place it firmly within a broader context of general intelligence (or *g*) as exemplified by verbal fluency abilities.

4. Correlation and causation. There are two things to say about this issue. First, "accounting for" variance via correlations can be informative, but correlations, in and of themselves, do not provide proof of causal relations. There are myriad alternate theories that can account for the correlations among ability measures, in addition to the process overlap theory. The validity of any theory is more likely to result from new predictions that, in one way or another, diverge from those made by competing theories (e.g., see Lakatos, 1970). The second point is that even accounting for large portions of variance does not, in itself, provide greater support for the theory; for example, human height and weight are highly correlated ($r > .90$; Diverse Populations Collaborative Group, 2005), but they are *not* the same thing.

Conclusion

As can probably be inferred from the preceding text, it is my assertion that the correct answer to the reasoning test question posed at the beginning of this article is "False." Process measures do correlate with *g*, but they do not adequately represent general intelligence. The general intelligence factor and general intelligence are much more than what is assessed by process measures, and the proof can be found in the fact that omnibus intelligence tests, such as the Stanford–Binet and the Wechsler

tests, do a much better job of predicting real-world criteria of individual differences in intellectually demanding tasks. Until such a time when process measures can even approach the validity of these intelligence tests, it is not reasonable to say that one has developed an adequate theory of the general intelligence factor.

References

Ackerman, P. L., Beier, M. E., & Bowen, K. R. (2000). Explorations of crystallized intelligence: Completion tests, cloze tests and knowledge. *Learning and Individual Differences: A Multidisciplinary Journal in Education, 12*, 105–121.

Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2002). Individual differences in working memory within a nomological network of cognitive and perceptual speed abilities. *Journal of Experimental Psychology: General, 131*, 567–589.

Ackerman, P. L. & Lohman, D. F. (2003). Education and *g*. In H. Nyborg (Ed.), *The scientific study of general intelligence—Tribute to Arthur R. Jensen* (pp. 275–292). Amsterdam, the Netherlands: Pergamon/Elsevier Science.

Binet, A., & Simon, T. (1961). New methods for the diagnosis of the intellectual level of subnormals (E. S. Kite, Trans.). In J. J. Jenkins & D. G. Paterson (Eds.), *Studies of individual differences: The search for intelligence* (pp. 90–96). New York, NY: Appleton-Century-Crofts. (Reprinted from New methods for the diagnosis of the intellectual level of subnormals. *L'Année Psychologique, 11*, 191–244, 1905.)

Boake, C. (2002). From the Binet–Simon to the Wechsler–Bellevue: Tracing the history of intelligence testing. *Journal of Clinical and Experimental Neuropsychology, 24*, 383–405.

Cattell, R. B. (1943). The measurement of adult intelligence. *Psychological Bulletin, 40*, 153–193.

Ceci, S. J. (1991). How much does schooling influence general intelligence and its cognitive components? A reassessment of the evidence. *Developmental Psychology, 27*, 703–722.

Diverse Populations Collaborative Group. (2005). Weight-height relationships and body mass index: Some observations from the diverse populations collaboration. *American Journal of Physical Anthropology, 128*, 220–229.

Ebbinghaus, H. (1896–1897). Über eine neue Methode zur Prüfung geistiger Fähigkeiten und ihre Anwendung bei Schulkindern [On a new method for testing mental abilities and its use with school children]. *Zeitschrift für Psychologie und Psysiologie der Sinnesorgane, 13*, 401–459.

Ferguson, G. A. (1954). On learning and human ability. *Canadian Journal of Psychology, 8*, 95–112.

Krueger, F., & Spearman, C. (1907). Die Korrelation zwischen verschiedenen geistigen Leistungsfahigkeiten, *Zeitschrift fur Psychologie (Leipzig), 44*, 50–114.

Lakatos, I. (1970). Falsification and the methodology of scientific research programmes. In I. Lakatos & A. Musgrave (Eds.), *Criticism and the growth of knowledge* (pp. 91–92, 94–125, 127–138, 189–195). New York, NY: Cambridge University Press.

Penrose, L. S., & Raven, J. C. (1936). A new series of perceptual tests: Preliminary communication. *British Journal of Medical Psychology, 16*, 97–105.

Snow, R. E. (1982). The training of intellectual aptitude. In D. K. Detterman & R. J. Sternberg (Eds.), *How and how much can intelligence be increased* (pp. 1–37). Norwood, NJ: Ablex.

Snow, R. E. (1996). Aptitude development and education. *Psychology, Public Policy, and Law, 2*, 536–560.

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

Spearman, C. (1927). *The nature of "intelligence" and the principles of cognition*. London, UK: Macmillan.

Spearman, C. (1930). Disturbers of tetrad differences. Scales. *Journal of Educational Psychology, 21*, 559–573.

Spearman, C. (1938). Measurement of intelligence. *Scientia, Milano, 64*, 75–82.

Terman, L. M. (1916). *The measurement of intelligence*. Boston, MA: Houghton Mifflin.

Vernon, P. E. (1950). *The structure of human abilities*. New York, NY: Wiley & Sons.

Vernon, P. E., & Parry, J. B. (1949). *Personnel selection in the British Forces*. London, UK: University of London Press.

Wittmann, W. W., & Süß, H.-M. (1999). Investigating the paths between working memory, intelligence, knowledge, and complex problem-solving performances via Brunswik Symmetry. In P. L. Ackerman, P. C. Kyllonen, & R. D. Roberts (Eds.), *Learning and individual differences: Process, trait, and content determinants* (pp. 77–108). Washington, DC: American Psychological Association.

Bridge Over Troubled Water: Commenting on Kovacs and Conway's Process Overlap Theory

Roberto Colom, Adam Chuderski & Emiliano Santarnecchi

Published online: 02 Aug 2016.

Submit your article to this journal ⎘

View related articles ⎘

View Crossmark data ⎘

Routledge
Taylor & Francis Group

Bridge Over Troubled Water: Commenting on Kovacs and Conway's Process Overlap Theory

Roberto Colom[a], Adam Chuderski[b], and Emiliano Santarnecchi[c]

[a]Department of Psychology, Universidad Autónoma de Madrid, Madrid, Spain; [b]Institute of Philosophy, Jagiellonian University in Krakow, Krakow, Poland; [c]Berenson-Allen Center for Non-Invasive Brain Stimulation, Department of Cognitive Neurology, Harvard Medical School, Boston, Massachusetts

Introduction

Scientific theories build bridges connecting available evidence in novel ways. They are "the glue that holds scientific observations together, for they summarize the chains of cause and effect (that help) to understand how the world works" (Hunt, 2011, p. 65). Scientific theories consider measurable variables; are objective; account for data; and, crucially, are never confirmed. We admire theories still playing the scientific game because they have not been refuted. But this must be for good reasons, not for bad ones.

Let us say from the outset that we applaud Kristof Kovacs and Andrew Conway's (this issue) effort for introducing a theoretical proposal presented under the rubric process overlap theory (POT). As properly underscored by Johnson (2013), conceptual tools are essential for making sense of both the huge amounts of already available data and the new findings derived from neuroscience and molecular genetics.

POT is aimed at connecting evidence derived from psychometrics, cognitive psychology, and neuroscience in an attempt to explain one of the most replicated findings in science, namely, the positive manifold: When individuals picked at random from the general population complete varied mental tests, those with better scores in a given test tend to have better scores in the remaining tests (and vice versa). The statistical analysis of a correlation matrix comprising these mental tests produces a general factor (g). No matter how this information is analyzed, at the end of the day we will distill g if there is a truly positive manifold in the data.

The interest about the nature of g dates back to the researcher who discovered this empirical fact, Charles Spearman (1904, 1927). The comprehensive book by Jensen (1998) reviews the major and minor topics related to g and discusses extensively the positive manifold at the psychometric, cognitive, biological, and genetic levels. In this regard, several points can be highlighted after his review: (a) g results from the common source of individual differences observed after resolving a variety of mental tests; (b) some tests are better measures of g than others, but the superficial characteristics of the former do not help to characterize g, (c) individual differences in cognitive abilities are remarkably greater than ability differences within a given individual; (d) psychometric g cannot be interpreted as a

cognitive process or a brain feature; and (e) g might be compared with a computer's CPU.

One question of paramount relevance relates to the unitary nature of g. In this regard, Kranzler and Jensen (1991) analyzed a large battery of intelligence tests and elementary cognitive tasks failing to support the proposal that a unitary process underlies the general factor (g). However, Carroll (1991) reanalyzed their data arriving at the conclusion that "it seems parsimonious to assume that g is unitary and represents a single entity … that influences a great variety of behaviors and performances, including speed and efficiency of information processing" (p. 434). If g is not unitary, statistical analyses should reveal several high-order factors, but this is hardly the case.

Bridges must be solid enough to resist earthquakes. As we show here, POT is a suggestive and courageous bridge built over troubled water. Let's see two examples before moving forward.

According to the Spearman Law of Diminishing Returns, g explains more variance at lower levels than at higher levels of cognitive ability. Following Kovacs and Conway (this issue), this factor differentiation shows that "g is far from being a constant …. The domain-generality of the positive manifold varies across ability level" (p. 155). However, there is no Spearman Law of Diminishing Returns effect for the relationship between fluid intelligence (Gf) and working memory capacity (WMC; Gignac & Weiss, 2015; Kroczek, Ociepka, & Chuderski, 2016). Furthermore, Abad, Colom, Juan-Espinosa, and García (2003) observed that strong differentiation effects are found when crystallized batteries are analyzed (Detterman & Daniels, 1989; Lynn, 1992). However, the analysis of fluid batteries reveals meager or null differentiation effects (Deary et al., 1996; Fogarty & Stankov, 1995). Their own study, analyzing 4,253 individuals, compared Gf and crystallized intelligence (Gc) batteries, finding a very weak effect for the former and a remarkable effect for the latter. They suggested that the higher correlations observed in the bottom half of the intelligence distribution might be a by-product of educational differences separating the low and high IQ bands, not a genuine intelligence effect.

The second example refers to "goal neglect." This cognitive function is thought to reflect a limit in WMC. Individuals must

maintain representations of the relevant information for successfully completing a given task. Increased numbers of items (instructions, rules, etc.) facilitate more lapses in the control processes required for goal management. This sounds like a simple short-term storage limitation. Nevertheless, this is not our line of reasoning here. As noted by Kovacs and Conway (this issue), goal neglect is "almost universal below a [Gf] of 85 but practically nonexistent above 100" (p. 161). If there are not individual differences in goal neglect above this threshold, but there are still 3 to 4 standard deviations in Gf beyond this point, lapses in goal-related control processes cannot explain individual differences in Gf across the population distribution.

POT Emergent *g*

Departing from the parsimonious unitary view, Kovacs and Conway seem to support the main thesis that *g* "emerges" from a set of independent cognitive processes. *g* is a result, not a cause. Specifically, it is argued that executive processes (such as updating, inhibition, or switching) underlie and limit performance across varied cognitive challenges, including intelligence tests and working memory tasks. Of importance, Kovacs and Conway highlight the recognized "indifference of the indicator" just noted: The superficial characteristics of tests and tasks do not help to understand "the determinants of individual differences" (p. 166).

Again, there is a noteworthy long tradition on intelligence research regarding the ontological status of *g*. Statistical analyses may produce one single unitary factor, but is it "real" (Horn & Cattell, 1966)? *g* may or may not drive broad cognitive abilities. The obtained higher order factor could be a simple statistical reflection of variance shared by lower order factors/abilities. In POT's terminology, *g* might be a "formative factor."

Why *g* and Gf can be very strongly correlated without any causal power of *g* has been elegantly explained by Cattell's (1971) investment theory: This is inevitable because, although *g* represents variance shared by Gf and Gc, Gc itself results from the investment of Gf across acculturation and long-term learning processes. However, the Gf–Gc correlation is far from perfect, because Gf is relevant for Gc at the time of learning (historical Gf differing from the current Gf level). So one could easily extract *g* out of Gf and Gc, but still *g* is irrelevant for theory.

Although the investment theory provides one straightforward explanation for the "emergence" of *g*, Kovacs and Conway's proposal is unclear (*g* as "emergent property"). There are disparate conceptions of "emergent property" (Stephan, 1999): weak, strong, synchronic, diachronic, epistemological, ontological, and so forth. POT fails to specify in what way *g* "emerges" from broad cognitive abilities. Bunge (2003) underscored that if anything was to be an "emergent property" of a system, it should constitute either the qualitative novelty in that system (ontologically emergent property assumed by theories of strong emergence) or should be unpredictable from lower levels of the system (epistemologically emergent property assumed by theories of weak emergence), or both.

However, POT's *g* does not fit any of these features, and therefore it is hard to see how it is emergent in the strict scientific sense. POT's emergent *g* follows a kind of magic process. As noted by Hunt (2011), theoretical statements are not unique to science. Cause-and-effect relationships are specified by theories, but these theories may or may not be scientific. Speaking of "emergent properties" without the required details just pushes the discussion toward a slippery slope.

Fluid Intelligence, Working Memory, Executive Control, … and the Reliable Short-Term Maintenance of the Relevant Information

After noting that research has identified one single WMC, because a positive manifold also underlies available measures of this capacity, Kovacs and Conway suggest that both positive manifolds (for *g* and for WMC) might share their explanatory variable(s). They ask, "Which component(s) of working memory cause(s) the general variation?" (p. 157). Their answer opens a dangerous door: cognitive control in general and executive attention in particular. People showing high WMC do have better executive control processes.

They also assume that (a) complex span tasks show stronger correlations with fluid intelligence than simple span tasks, and (b) executive processes account for the substantial correlation between WMC and Gf. This relationship "is driven by the operation of multiple domain-general cognitive processes … identified as executive processes in cognitive psychology" (p. 160). The problems composing standardized tests recruit both domain-general and domain-specific processes. The positive manifold is simply the result of the overlap between both types of processes. Therefore, *g* does not reflect one single and unitary process. It presumably emerges from a number of independent sources.

However, this is mainly based on outdated empirical evidence. The last decade of research regarding the analyses of the cognitive correlates of Gf and WMC has introduced a serious challenge to the relevance of executive control. As we see, their review of the evidence is selective and, worse, their inferences are incorrect.

Crucially, with regard to the second point, POT assumes

> If executive control tasks (A) correlate with working memory tasks (B)
>
> And if working memory tasks (B) correlate with fluid intelligence tests (C)
>
> Then A drives the correlation between B and C.

However, the relations are not transitive here: Variance in B may reflect variance in A, as well as variance in another factor (D), which (unlike A) is also reflected in C. Therefore, regardless of the A–B correlation, the B–C correlation may be driven by D and data showing that some executive control tasks are related to complex span tasks fails to support POT.

Studies addressing the relationship between executive tasks and Gf are consistent with the conclusion that it is weak and unstable. In this regard, Friedman et al. (2006) did not find significant associations between inhibition and Gf, as well as between the latter and switching. Updating was related to Gf, but tasks tapping this executive function are highly confusing because they merge executive control requirements and general maintenance mental processes (Oberauer, Süß, Wilhelm, & Sander, 2007).

Further evidence reveals null relationships between Gf and shifting (Oberauer, Süß, Wilhelm, & Wittmann, 2008) and moderate correlations, at the latent level, between Gf and inhibition ($r = .39$)—much weaker than the observed between Gf and short-term storage ($r = 1.0$; Martinez et al., 2011). Only 7% of variance in Gf was univocally attributed to attention control in the Unsworth and Spillers (2010) study. Chuderski, Taraday, Nęcka, and Smoleń (2012) showed that inhibition, interference resolution, and attention control were not related to Gf once short-term storage was controlled for (see also Shipstead, Lindsay, Lakey, & Engle, 2014).

It is important to note that available evidence no longer supports the first point: Complex span tasks *are not* stronger predictors of Gf than simple span tasks. Early reports suggesting a large difference between complex and simple span tasks (Engle, Tuholsky, Laughlin, & Conway, 1999) were flawed as a result of the incorrect tasks' scoring (absolute scoring with less than ideal psychometric properties). Once properly scored, simple span tasks are equally strong predictors of Gf (Unsworth & Engle, 2007; Shipstead, Harrison, & Engle, 2015). There are several research reports yielding similar findings (Chuderski, 2014; Colom, Abad, Quiroga, Shih, & Flores-Mendoza, 2008; Colom, Rebollo, Abad, & Shih, 2006; Martinez et al., 2011).

Although studies rejecting the executive control view are ignored, some others are improperly highlighted. The Kane et al. (2004) study is a key example: "Although simple span tests appear to be more domain specific, the processes that complex span tests tap beyond the pure storage and retrieval of information appear to be largely domain general" (Kovacs & Conway, this issue, p. 157). Their general structural model relating Short-Term Memory (STM), WMC, and reasoning suffered from the frequent multicollinearity problem. Surprisingly, an alternative conceptual approach to that employed by Engle et al. (1999) and Conway, Cowan, Bunting, Therriault, and Minkoff (2002)—namely, the common variance between WMC and STM was thought to reflect primarily storage, whereas the residual WMC variance was thought to reflect primarily executive control processes—was adopted by Kane et al.: One general executive attention factor, with loadings from all the span variables, was thought to reflect the domain-general executive variance shared by all the span tasks (!). This is really hard to buy into, especially because their STM tasks were carefully designed to measure short-term storage only. If the additional processing demand is removed from WMC tasks for modeling STM tasks, assuming that one general factor defined by both kinds of tasks represents executive control seems a post hoc movement, to say the least. Finally, Gf was predicted by the executive attention factor (.52) and by one orthogonal STM spatial factor (.54). This hardly supports the view that executive-attention processes drive primarily the predictive utility of memory span measures (see Colom, Abad, Rebollo, & Shih, 2005, for further arguable details).

The reanalyses of key studies by Colom, Rebollo, et al. (2006)—including Engle et al. (1999); Miyake, Friedman, Rettinger, Shah, and Hegarty (2001); Conway et al. (2002); Bayliss, Jarrold, Gunn, and Baddeley (2003); and Kane et al. (2004)—led to several conclusions: (a) Individual differences in memory span tasks are strongly explained by some general component, four times more important than the observed specific components; (b) complex span tasks cannot be clearly distinguished from simple span tasks; (c) their shared component might be responsible for their association with cognitive abilities (including Gf); and (d) this shared component can be identified with "simple" short-term storage capacity. In passing, we note that the theoretical interpretation of the findings observed across the reanalyzed studies ("there are cognitive systems more prone to coping successfully with several diverse challenges … concurrent processing requirements leave less capacity for temporary storage of information, which diminishes the reliability of the stored information, which in turn is responsible for the behavioral effects observed in memory span tasks" (Colom, Rebollo et al., 2006, p. 170)) was seen as consistent with the Carroll's (1991) unitary view discussed in the Introduction section.

Relatedly, the Wiley, Jarosz, Cushen, and Colflesh's (2011) finding that "it is the application of new rules and switching from old ones that drives the correlation between complex span and Gf" (p. 30) has been refuted by at least three other sophisticated recent studies (Harrison, Shipstead, & Engle, 2015; Little, Lewandowsky, & Craig, 2014; Smoleń & Chuderski, 2015). New-rule and old-rule items of the Raven test equally strongly predict WMC, contrary to POT predictions.

Moving from correlations to mechanisms, POT proposes that Gf (remember: variance in abstract reasoning on novel problems) strongly relies on executive control. In support of this statement, Kovacs and Conway (this issue) cite the classic model by Carpenter, Just, and Shell (1990). However, this model is misrepresented. The requirement of maintaining more rules was not treated by Carpenter et al. as cognitive control but as sheer storage capacity. "Control" here was related to the ability to coordinate inference processes, backtrack from unsuccessful hypotheses, and so forth. Such ability is a kind of strategic control (goal management) departing from simple executive processes. Notably, there are computational models of Raven and analogy tests that do not rely significantly on executive processes (Hummel & Holyoak, 1997, 2003; Kunda, McGreggor, & Goel, 2013; Lovett, Forbus, & Usher, 2010; Rasmussen & Eliasmith, 2014; Wilson, Halford, Gray, & Phillips, 2001). The crucial feature of these models is the maximum number of elements/size of a relation (relevant to the solution) that can be stored. The simulation of individual differences in reasoning (figural analogies) by Chuderski and Andrelczyk (2015) has shown the key role of the number of distinct role-filler bindings that can be maintained simultaneously.

It is true that "individuals with greater WMC have better cognitive control processes" (Kovacs & Conway, this issue, p. 157). However, these individuals also show further cognitive advantages, such as learning (Chuderski, 2013; Kaufman, DeYoung, Gray, Brown, & Mackintosh, 2009), long-term memory retrieval (Unsworth & Spillers, 2010; Unsworth, Spillers, & Brewer, 2010), short-term memory capacity (Chuderski, 2014; Colom et al., 2008; Cowan, Fristoe, Elliott, Brunner, & Saults, 2006; Shipstead et al., 2015), or detection of simple perceptual patterns (Chuderski, 2014; Oberauer et al., 2008). We might infer that complex span tasks correlate with Gf because of some of these processes, beyond executive control. Indeed, Unsworth, Fukuda, Awh, and Vogel's (2014) study suggests that Gf variance must be explained considering several cognitive factors

beyond executive control (the ability for selecting and maintaining items in the face of distractions), namely, capacity (the ability to actively maintain distinguishable items in the short term) and secondary memory (encode items into the intermediate memory and recover those relevant for the task at hand):

> Working memory limitations can arise for a number of reasons. … [These limitations] can be multifaceted and can help to resolve discrepancies in the literature where some studies find evidence for the importance of deficits in one, whereas other studies find evidence for the importance of another. (p. 21)

Kovacs and Conway's proposal neglects existing evidence showing that executive control is (almost) irrelevant for explaining individual differences in Gf. Gf and WMC can be explained by the number of items (and/or the number of bindings between items) that allows encoding the structure/solution in working memory (Cowan et al., 2006; Halford, Andrews, & Wilson, 2014; Halford, Cowan, & Andrews, 2007; Martínez et al., 2011; Oberauer et al., 2007). Therefore, the overlap may be explained by a common capacity ("abstract working memory"; Cowan et al., 2011) or the ability to construct and maintain arbitrary bindings ("relational integration"; Oberauer et al., 2007).

At the latent variable level, memory span factors (short-term memory, WMC, and executive updating) are hardly distinguishable from Gf (Martínez et al., 2011). From a theoretical perspective, short-term storage is the cognitive component shared by these span measures (Colom et al., 2008; Colom, Rebollo, et al., 2006; Hornung, Brunner, Reuter, & Martin, 2011; Krumm et al., 2009; Shahabi, Abad, & Colom, 2014), and therefore individual differences in Gf can be explained by basic mental processes underlying memory span, namely, encoding, maintenance, and retrieval (Jonides et al., 2008). Shared capacity limitations might derive from the number of items that can be kept active in the short term or the number of relationships between elements that can be kept active during the reasoning process necessary for solving problems composing standard intelligence tests (Halford et al., 2014; Halford et al., 2007). The limitations shared by intelligence and memory span might be based on the ability to build and keep relevant connections in the short term.

After the simultaneous consideration of a varied set of candidate relevant constructs, Martínez et al. (2011) concluded that cognitive processes theoretically involved in working memory and executive updating, inhibition, and processing speed fail to add significant information for answering the question regarding the basic processes underlying fluid reasoning. Chuderski et al. (2012) also supported the crucial role of storage capacity regarding the mechanisms underlying the correlation between WMC and Gf. Executive control processes (attention, interference resolution, and inhibition) failed to contribute in some relevant way to this relationship.

As noted in the Introduction section, scientific theories account for data by building bridges. POT fails to connect highly relevant recent data, which is not good for a healthy theory. Looking for converging evidence only is like being in a courtroom playing the role of the defender. But the proper contrast of a theory requires playing the role of the prosecutor. Kovacs and Conway fail to do that, and, indeed, POT may be seen more as an unfalsifiable general framework than as a precise theory of the human intellect.

Fluid Intelligence and Executive Processes in the Brain

POT ignores crucial findings observed in cognitive research for selectively assembling supporting evidence regarding the role of executive control processes for explaining the strong connection between Gf and WMC, as described previously. As we see in this section, one similar strategy is applied for selecting evidence accumulated by neuroscience research.

The recent meta-analysis by Basten, Hilger, and Fiebach (2015) underscores the lack of overlap between structural and functional correlates of cognitive ability differences (consistent with Colom's (2007, 2014) analysis of the evidence), noting that a brain region might be activated for copying with a cognitive challenge (*task approach*), but individual differences in this activation might not be associated with individual differences in cognitive ability (*individual differences approach*). Therefore, the relevant question is, Do individual differences in brain features predict cognitive ability differences? The well-known P-FIT model combines the task and individual differences approaches (Jung & Haier, 2007), whereas Basten et al. (2015) considered only the latter in their meta-analysis.

For their meta-analysis, 457 subjects and 415 data points provided structural data, whereas 464 subjects and 151 data points provided functional data. Structural studies supported the relevance of frontal, temporal, and occipital regions, whereas functional results supported the relevance of parietal and frontal regions. The parietal cortex was not relevant in structural studies. The main conclusion was that structural and functional correlates of cognitive ability differences couldn't be located in overlapping brain regions. The whole picture hardly supports the crucial role of the PFC, as underscored by POT.

Kovacs and Conway assume that Gf is linked to the dorsolateral prefrontal cortex with remarkable consistency. This brain structure supports cognitive control, and therefore, again, executive functioning is highlighted. From here, they move to the relevance of the multiple demand system (MD) defining one cognitive control network (Duncan, 2010).

However, the MD approach has been questioned. Let's see one example. Hampshire et al.'s (2012) report was devoted to demonstrating that the *g* factor is an invalid psychological construct, because general intelligence requires at least two orthogonal components defined by different (MD) brain networks. In their study, factor analysis was applied to brain image voxels to find clusters interpreted as brain networks. One strong, unrotated general factor and several further factors were identified. However, the obtained factors were arbitrarily submitted to a varimax rotation, which impose their independence. The final solution was the basis of their theoretical interpretation regarding the independence of two brain networks. As noted by Haier, Karama, Colom, Jung, and Johnson (2014) in their critical review, "the unrotated factor solution that shows the strong general factor reflects brain organization to the same degree as the statistically independent factors … and could well be a neuro-*g*."

Colom, Jung, and Haier (2006) noted similar questionable interpretations in Duncan et al.'s (2000) report addressing the neural basis for *g* (discussed by Kovacs and Conway, this issue). A close examination of their study reveals support for a distributed model: Results obtained from the most highly *g*-loaded

items in their analysis reveal activations in frontal, parietal, and occipital brain regions. In their own study, Colom, Jung, et al. (2006) analyzed a set of intelligence measures showing a broad spectrum of *g*-loadings (range = 0.23–0.90), finding that increasing *g*-loading is associated with greater gray matter volume in several areas distributed throughout the brain. Although frontal areas clearly showed the most relationship to *g*, many other areas also were related.

Therefore, there is life beyond the frontal lobes and executive processes. The lesion study by Barbey, Colom, Paul, and Grafman (2014) is a nice example. Here, the simultaneous relationships between several factors representing working memory–related cognitive processes (verbal working memory, spatial working memory, manipulation, and monitoring) and Gf were analyzed. As shown in Figure 1, one remarkable overlap was found for spatial working memory and Gf. However, one executive factor defined by versions of the n-back task, varying in their cognitive complexity, showed a meager overlap with Gf. Furthermore, the frontal lobes were not the only stars in the sky regarding Gf.

Network approaches also depart from this "frontal/control perspective." Let's see two examples. Pineda-Pardo, Martínez, Román, and Colom (2016) analyzed local and global efficiency indices within a structural network defined by regions composing the P-FIT model (Jung & Haier, 2007). Variations on these network indices were related to individual differences in WMC and Gf. Parietal and frontal regions were found key for maintenance of the analyzed network structural integrity. The middle frontal gyrus, the superior frontal gyrus, and the precuneus showed large connectivity values, and, of importance, penalizing these parietal and frontal regions degraded network efficiency along with the observed relationships with cognitive ability differences.

From another perspective, the resting-state fMRI graph-analysis by Santarnecchi, Rossi, and Rossi (2015) showed enhanced brain resilience to targeted and random attacks in individuals with high intelligence scores. It is important to note that this increased resilience was based on a greater distributed processing capacity, including cortical regions supporting memory and language:

> By implying intelligence as responsible for a more widespread and efficient brain resource allocation at rest, our results support previous observations of a positive spatial correlation between intelligence level and brain volumes—mostly encompassing frontal, parietal, and occipital lobes, contrasting the idea of prefrontal cortices as primary brain sites related to intelligence level. (p. 305)

Furthermore, two other investigations correlating brain spontaneous activity with individual variability in
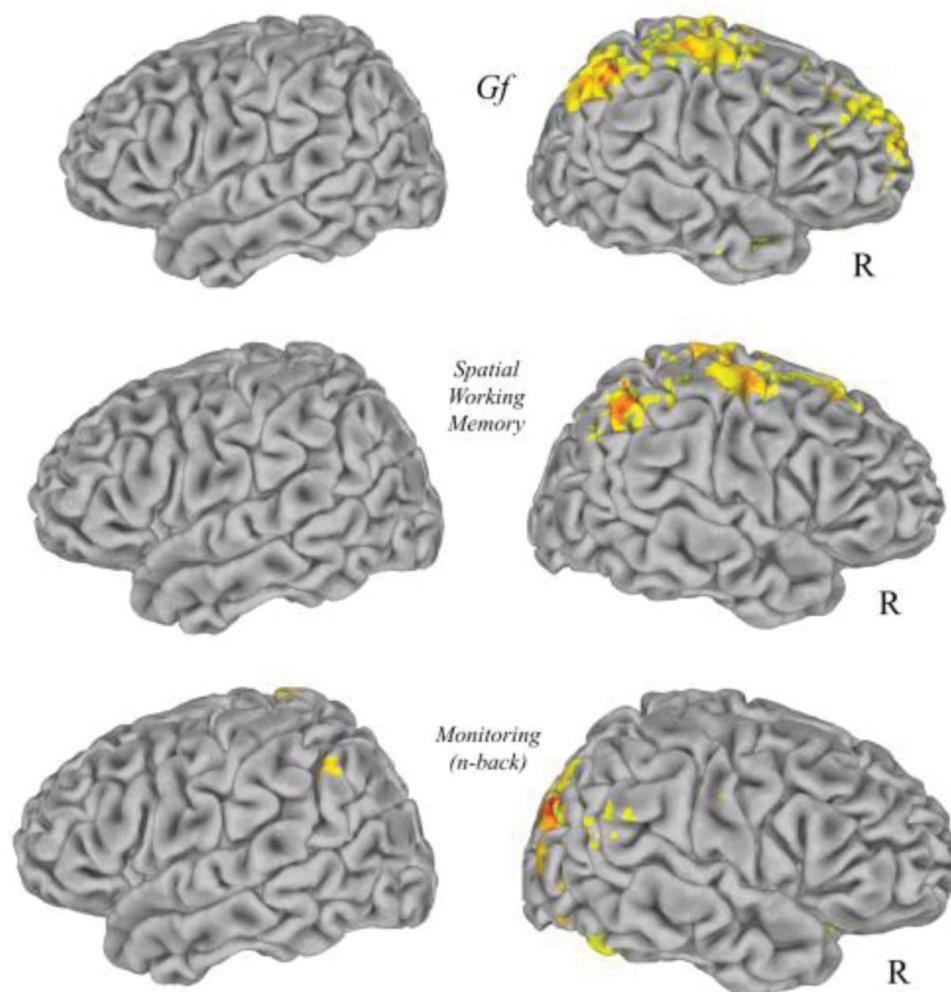

Figure 1. Lesion results observed in the Barbey et al. (2014) study for fluid intelligence (Gf), spatial working memory, and monitoring (*N* = 158).

intelligence measures have underscored the relevance of whole-brain activity, with particular contributions by the parietal, occipital and temporal lobes. Looking at the role of strong and weak brain functional connections, Santarnecchi, Galli, Polizzotto, Rossi, and Rossi (2014) documented a pivotal role for a widespread network of weak connections encompassing the two hemispheres, suggesting the importance of flexible, long-range connections in determining individual cognitive profiles. In a subsequent study focusing on the connectivity profile of homologues brain regions, a reduction in the symmetry of the connectivity profile of multiple brain regions has been identified as a positive predictor of higher fluid and crystallized intelligence, with significant differences regarding low-level processing regions such as the visual cortex, temporal lobe, and somatosensory cortex (Santarnecchi, Tatti, Rossi, Serino, & Rossi, 2015) and no crucial role for the variability in the PFC.

Beyond fMRI studies, the role for g/Gf/WMC of brain regions both within and beyond PFC was shown by means of transcranial electrical stimulation studies, a noninvasive brain stimulation technique allowing modulating brain cortico-spinal excitability as well as enhancing brain oscillatory patterns in a frequency-specific manner (Frohlich & McCormick, 2010; Santarnecchi et al., 2015). Specifically, although studies targeting the left middle prefrontal gyrus using transcranial alternating current stimulation have shown improvement in Gf-related reaction times and

efficiency during stimulation (Santarnecchi et al., 2013) with no effects on WM performance (Santarnecchi et al., 2016), several others reported increases in Gf test scores (Pahor & Jaušovec, 2014) and WM task recall (Jaušovec & Jaušovec, 2014; Jaušovec, Jaušovec, & Pahor, 2014; Polanía, Nitsche, Korman, Batsikadze, & Paulus, 2012) when the parietal cortex was stimulated (and some even reported null effects for the frontal cortex, e.g., Jaušovec et al., 2014). Moreover, additional evidence for the "beyond PFC" argument comes from a study by Polania et al. (2012) using transcranial alternating current stimulation in the theta frequency band, showing improvement of WM performance after synchronous (i.e., in phase)—but not asynchronous—oscillatory theta (6 Hz) stimulation of a left fronto-parietal network. This plausibly reinforces the idea that the coordinated activity of neuronal populations belonging to a distributed network (also possibly expanding subcortically and in the controlateral hemisphere) might represent the basic neurophysiological substrate of higher order cognition.

We close this section reporting the findings of a recent meta-analysis by Santarnecchi et al. (2016) including fluid intelligence, inhibition, switching/flexibility, and updating (Figure 2). Clearly, capacity-related brain responses *are not* confined before the central sulcus. There is definitely a distributed network involved for both Gf and executive control: High-complexity trials did elicit widespread activations encompassing the
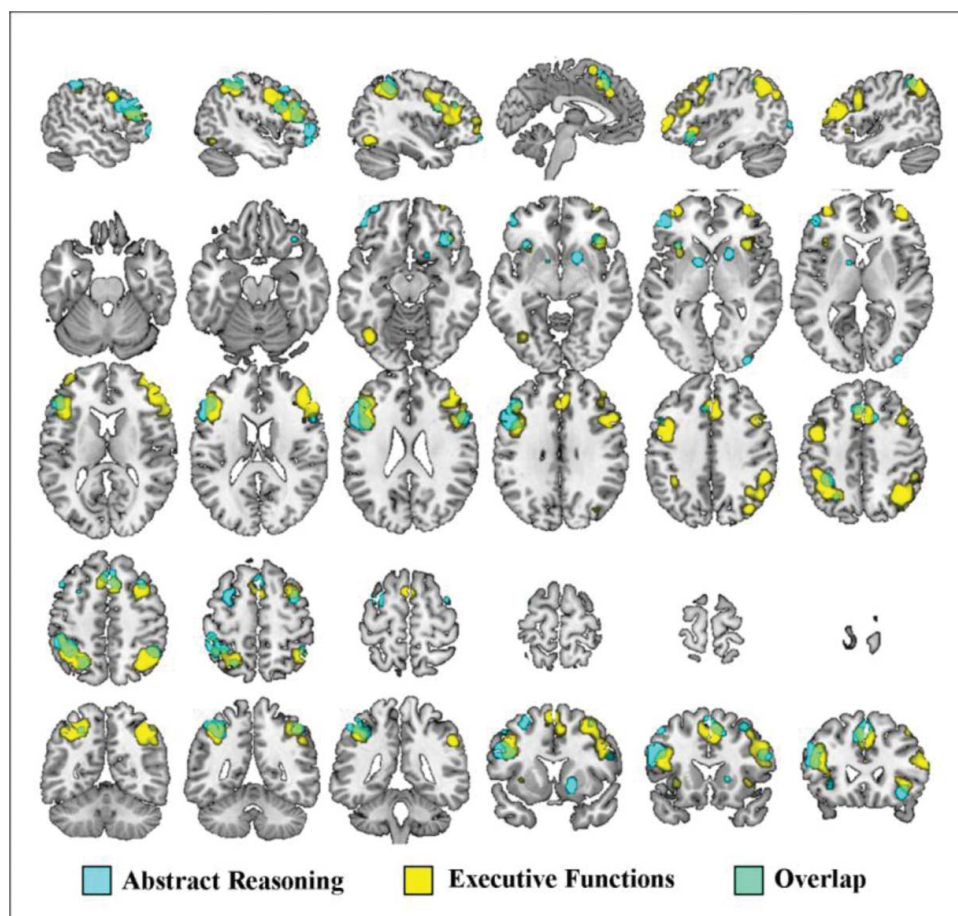

Figure 2. Meta-analysis results for fluid intelligence and executive processes (adapted and modified from Santarnecchi et al., 2016).

Figure 3. "X Marks The Spot" (*Indiana Jones and The Last Crusade,* 1989).

PFC but also including subcortical and occipito-temporal regions.

Conclusion: X Marks the Spot

Indiana Jones was looking for his missing and beloved father in Venice. Marcus Brody and Elsa—a Nazi spy—were with him in one of the numerous churches of the unique city. When the woman leaves the room in order to ask permission to stay in the building, the men use their private time to look at the notebook Dad sent to Indy. Indy finds a drawing representing the main stained glass window placed in front. He notes that there are printed numbers in the glass and in the columns, but he cannot find number 10. They look from the ground to the left, to the right, up, down, and so forth. Nothing. But then Indy notices one spiral staircase, climbs it, and finds the missed number in the floor: "*X marks the spot*" (Figure 3).

We began our comment noting that scientific theories build bridges connecting observations. In this regard, they account for available data better than rival theories and, ideally, predict future discoveries. POT attempts to do that, and the authors proposing this theory must be recognized. It is easier to find failures in an already build structure than put together the pieces giving rise to it. Therefore, we appreciate the effort invested by Kovacs and Conway.

However, it was our work here to enumerate and briefly discuss what we think are problems suffered by the proposed theory. In essence, we found that POT attempts to integrate psychometric, cognitive, and neuroscience evidence, but in a selective way, which makes their bridge shaky and unsatisfying. The positive manifold is thought to be emergent from the operation of varied mental processes, but Kovacs and Conway's conception of "emergent property" is far from straightforward. In addition, it seems, at a first glance, that varied mental processes are crucial for the theory, but then executive control is distinguished as the most brilliant star in the sky.

POT rejects unitary views trying to account for the positive manifold and, therefore, the general factor of intelligence (*g*). However, following Carroll (1991), it would be reasonable to preserve the unitary view (at least because of its higher parsimony) instead of accepting models based on the interaction between a large set of independent processes (until further notice). The unitary and multiple/sampling views cannot be distinguished mathematically. As Kovacs and Conway (this issue) acknowledge, "A single common cause of the positive correlations between mental tests, is surely a sufficient … explanation of the positive manifold" (p. 153). It may or may not be necessary, but time and further research will tell.

This research suggests that, for instance, intelligence is a moving target in the brain (Martínez et al., 2015). Available evidence might depart from the view that there is a place in the brain for general intelligence, and even that only "places" are relevant, as temporal properties of neural processing like synchronization and coordination might also play an important role (see Cohen, 2011), as well as network-level dynamics promoting evolvability, robustness, and plasticity (Csemerly, 2015).

As in the scene where we saw Indy, Marcus, and Elsa reading numbers across disparate places in the building, perhaps we still miss the X marking the spot. Maybe there is no place in the brain for general intelligence, because the brain itself is the place. And we only have a single brain.

References

Abad, F. J., Colom, R., Juan-Espinosa, M., & García, L. F. (2003). Intelligence differentiation in adult samples. *Intelligence*, 31, 157–166.

Barbey, A. K., Colom, R., Paul, E. J., & Grafman, J. (2014). Architecture of fluid intelligence and working memory revealed by lesion mapping. *Brain Structure & Function*, 219, 485–494.

Basten, U., Hilger, K., & Fiebach, C. J. (2015). Where smart brains are different: A quantitative meta-analysis of functional and structural brain imaging studies on intelligence. *Intelligence*, 51, 10–27.

Bayliss, D. M., Jarrold, C., Gunn, D. M., & Baddeley, A. (2003). The complexities of complex span: Explaining individual differences in working memory in children and adults. *Journal of Experimental Psychology: General*, 132, 71–92.

Bunge, M. (2003). *Emergence and convergence: Qualitative novelty and the unity of knowledge*. Toronto, Canada: Toronto University Press.

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

Carroll, J. B. (1991). No demonstration that g is not unitary, but there's more to the story: Comment on Kranzler and Jensen. *Intelligence*, 15 (4), 423–436.

Cattell, R. B. (1971). *Abilities: Their structure, growth, and action*. Boston, MA: Houghton Mifflin.

Chuderski, A. (2013). When are fluid intelligence and working memory isomorphic and when are they not? *Intelligence*, 41, 244–262.

Chuderski, A. (2014). Relational integration task explains fluid reasoning above and beyond other working memory tasks. *Memory & Cognition*, 42, 448–463.

Chuderski, A., & Andrelczyk, K. (2015). From neural oscillations to complex cognition: Simulating the effect of the theta-to-gamma cycle length ratio on analogical reasoning. *Cognitive Psychology*, 76, 78–102.

Chuderski, A., Taraday, M., Nęcka, E., & Smoleń, T. (2012). Storage capacity explains fluid intelligence while executive control does not. *Intelligence*, 40, 278–295.

Cohen, M. X. (2011). It's about time. *Frontiers in Human Neuroscience*, 5, Article 2.

Colom, R. (2007). Intelligence? What intelligence? *Behavioral and Brain Sciences*, 30, 155–156.

Colom, R. (2014). From the Earth to the brain. *Personality and Individual Differences*, 61–62, 3–6.

Colom, R., Abad, F. J., Quiroga, M. A., Shih, P. C., & Flores-Mendoza, C. (2008). Working memory and intelligence are highly related constructs, but why? *Intelligence*, 36, 584–606.

Colom, R., Abad, F. J., Rebollo, I., & Shih, P. C. (2005). Memory span and general intelligence: A latent-variable approach. *Intelligence*, 33, 623–642.

Colom, R., Jung, R. E., & Haier, R. J. (2006). Distributed brain sites for the *g*-factor of intelligence. *Neuroimage*, 31, 1359–1365.

Colom, R., Rebollo, I., Abad, F. J., & Shih, P. C. (2006). Complex span tasks, simple span tasks, and cognitive abilities: A re-analysis of key studies. *Memory & Cognition*, 34, 158–171.

Conway, A. R. A., Cowan, N., Bunting, M., Therriault, D., & Minkoff, S. (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence*, 30, 163–183.

Cowan, N., Fristoe, N. M., Elliott, E. M., Brunner, R. P., & Saults, J. S. (2006). Scope of attention, control of attention, and intelligence in children and adults. *Memory & Cognition*, 34, 1754–1768.

Cowan, N., Li, D., Moffitt, A., Becker, T. M., Martin, E. A., Saults, J. S., & Christ, S. E. (2011). A neural region of abstract working memory. *Journal of Cognitive Neuroscience*, 23, 2852–2863.

Csemerly, P. (2015). *Plasticity-rigidity cycles: A general adaptation mechanism* [Preprint]. http://arxiv.org/abs/1511.01239

Deary, I. J., Egan, V., Gibson, G. J., Austin, E. J., Brand, C. R., & Kellaghan, T., (1996). Intelligence and the differentiation hypothesis. *Intelligence*, 23, 105–132.

Detterman, D. K., & Daniels, M. H. (1989). Correlations of mental tests with each other and with cognitive abilities are highest for lower IQ groups. *Intelligence*, 13, 349–359.

Duncan, J. (2010). The multiple-demand (MD) system of the primate brain: Mental programs for intelligent behavior. *Trends in Cognitive Sciences*, 14, 4, 172–179.

Duncan, J., Seitz, R. J., Kolodny, J., Bor, D., Herzog, H., Ahmed, A., & Emslie, H. (2000). A neural basis for general intelligence. *Science*, 289, 457–460.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent variable approach. *Journal of Experimental Psychology: General*, 128, 309–331.

Fogarty, G. J., & Stankov, L. (1995). Challenging the "Law of Diminishing Returns." *Intelligence*, 21, 157–174.

Friedman, N. P., Miyake, A., Corley, R. P., Young, S. E., DeFries, J. C., & Hewitt, J. K. (2006). Not all executive functions are related to intelligence. *Psychological Science*, 17, 172–179.

Frohlich, F., & McCormick, D. A. (2010). Endogenous electric fields may guide neocortical network activity. *Neuron*, 67, 129–143.

Gignac, G. E., & Weiss, L. G. (2015). Digit Span is (mostly) related linearly to general intelligence: Every extra bit of span counts. *Psychological Assessment*, 27, 1312–1323.

Haier, R. J., Karama, S., Colom, R., Jung, R. E., & Johnson, W. (2014). A comment on "fractionating intelligence" and the peer review process. *Intelligence*, 46, 323–332.

Halford, G. S., Andrews, G., & Wilson, W. H. (2014). Relational processing in reasoning: The role of working memory. In A. Feeney & V. Thompson (Eds.), *Reasoning as memory* (pp. 34–52). Hove, UK: Psychology Press.

Halford, G. S., Cowan, N., & Andrews, G. (2007). Separating cognitive capacity from knowledge: A new hypothesis. *Trends in Cognitive Sciences*, 11, 236–242.

Hampshire, A., Highfield, R. R., Parkin, B. L., & Owen, A. M. (2012). Fractionating human intelligence. *Neuron*, 76, 6, 1225–1237.

Harrison, T. L., Shipstead, Z., & Engle, R. W. (2015). Why is working memory capacity related to matrix reasoning tasks. *Memory & Cognition*, 43(3), 389–396.

Horn, J., & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized intelligence. *Journal of Educational Psychology, 57*, 253–270.

Hornung, C., Brunner, M., Reuter, R. A., & Martin, R. (2011). Children's working memory: Its structure and relationship to fluid intelligence. *Intelligence*, 39, 210–221.

Hummel, J. E., & Holyoak, K. J. (1997). Distributed representations of structure: A theory of analogical access and mapping. *Psychological Review*, 104, 427–466.

Hummel, J. E., & Holyoak, K. J. (2003). A symbolic-connectionist theory of relational inference and generalization. *Psychological Review*, 110, 220–264.

Hunt, E. B. (2011). *Human intelligence*. Cambridge, UK: Cambridge University Press.

Indiana Jones and The Last Crusade. (1989). Lucasfilm Ltd.

Jaušovec, N., & Jaušovec, K. (2014). Increasing working memory capacity with theta transcranial alternating current stimulation (tACS). *Biological Psychology*, 96, 42–47.

Jaušovec, N., Jaušovec, K., & Pahor, A. (2014). The influence of theta transcranial alternating current stimulation (tACS) on working memory storage and processing functions. *Acta Psychologica*, 146, 1–6.

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

Johnson, W. (2013). Whither intelligence research? *Journal of Intelligence*, 1, 25–35.

Jonides, J., Lewis, R. L., Nee, D. E., Lustig, C. A., Berman, M. G., & Moore, K. S. (2008). The mind and brain of short-term memory. *Annual Review of Psychology*, 59, 193–224.

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

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: A latent variable approach to verbal and visuo-spatial memory span and reasoning. *Journal of Experimental Psychology: General*, 133, 189–217.

Kaufman, S. B., DeYoung, C. G., Gray, J. R., Brown, J., & Mackintosh, N. (2009). Associative learning predicts intelligence above and beyond working memory and processing speed. *Intelligence*, 37, 374–382.

Kranzler, J. H., & Jensen, A. R. (1991). The nature of psychometric *g*: Unitary processes or a number of independent processes? *Intelligence*, 15, 397–422.

Kroczek, B., Ociepka, M., & Chuderski, A. (2016). No Spearman's Law of Diminishing Returns for the working memory and intelligence relationship. *Polish Psychological Bulletin, 47*, 73–80.

Krumm, S., Schmidt-Atzert, L., Buehner, M., Ziegler, M., Michalczyk, K., & Arrow, K. (2009). Storage and non-storage components of working memory predicting reasoning: A simultaneous examination of a wide range of ability factors. *Intelligence*, 37, 347–364.

Kunda, M., McGreggor, K., & Goel, A. K. (2013). A computational model for solving problems from the Raven's progressive matrices intelligence test using iconic visual representations. *Cognitive Systems Research*, 22–23,47–66.

Little, D. R., Lewandowsky, S., & Craig, S. (2014). Working memory capacity and fluid abilities: The more difficult the item, the more more is better. *Frontiers in Psychology*, 5, Article 239.

Lovett, A., Forbus, K., & Usher, J. (2010). A structure-mapping model of Raven's progressive matrices. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd annual conference of the Cognitive Science Society* (pp. 2761–2766). Austin, TX: Cognitive Science Society.

Lynn, R. (1992). Does Spearman's *g* decline at high levels? Some evidence from Scotland. *Journal of Genetic Psychology*, 153, 229–230.

Martínez, K., Burgaleta, M., Román, F. J., Escorial, S., Shih, P. C., Quiroga, M. A., & Colom, R. (2011). Can fluid intelligence be reduced to "simple" short-term storage? *Intelligence*, 39, 473–480.

Martínez, K., Madsen, S., Joshi, A. A., Shantanu, J., Román, F. J., Villalón-Reina, J., & Colom, R. (2015). Reproducibility of brain-cognition relationships using three cortical surface based protocols: An exhaustive analysis based on cortical thickness. *Human Brain Mapping*, 36, 3227–3245.

Miyake, A., Friedman, N. P., Rettinger, D. A., Shah, P., & Hegarty, M. (2001). How are visuospatial working memory, executive functioning, and spatial abilities related? A latent-variable analysis. *Journal of Experimental Psychology: General*, 130, 621–640.

Oberauer, K., Süß, H-M., Wilhelm, O., & Sander, N. (2007). Individual differences in working memory capacity and reasoning ability. In A. R. A. Conway, C. Jarrold, M. J. Kane, A. Miyake, & J. N. Towse (Eds.), *Variation in working memory* (pp. 49–75). Oxford, UK: Oxford University Press.

Oberauer, K., Süß, H-M., Wilhelm, O., & Wittman, W. W. (2008). Which working memory functions predict intelligence? *Intelligence*, 36, 641–652.

Pahor, A., & Jaušovec, N. (2014). The effects of theta transcranial alternating current stimulation (tACS) on fluid intelligence. *International Journal of Psychophysiology*, 93, 322–331.

Pineda-Pardo, J. A., Martínez, K., Román, F. J., & Colom, R. (2016). Structural efficiency within a parieto-frontal network and cognitive differences. *Intelligence*, 54, 105–116.

Polanía, R., Nitsche, M. A., Korman, C., Batsikadze, G., & Paulus, W. (2012). The importance of timing in segregated theta phase-coupling for cognitive performance. *Current Biology*, 22, 1314–1318.

Rasmussen, D., & Eliasmith, C. (2014). A spiking neural model applied to the study of human performance and cognitive decline on Raven's Advanced Progressive Matrices. *Intelligence*, 42, 53–82.

Santarnecchi, E., Brem, A. K., Levenbaum, E., Thompson, T., Kadosh, R. C., & Pascual-Leone, A. (2015). Enhancing cognition using transcranial electrical stimulation. *Current Opinion in Behavioural Sciences*, 171–178.

Santarnecchi, E., Galli, G., Polizzotto, N. R., Rossi, A., & Rossi, S. (2014). Efficiency of weak brain connections support general cognitive functioning. *Human Brain Mapping*, 35, 4566–4582.

Santarnecchi, E., Muller, T., Rossi, S., Sarkar, A., Polizzotto, N. R., Rossi, A., & Kadosh, R. C. (2016). Individual differences and specificity of prefrontal gamma frequency-tACS on fluid intelligence capabilities. *Cortex*, 75, 33–43.

Santarnecchi, E., Polizzotto, N. R., Godone, M., Giovannelli, F., Feurra, M., Matzen, L., & Rossi, S. (2013). Frequency-dependent enhancement of fluid intelligence induced by transcranial oscillatory potentials. *Current Biology*, 23, 1449–1453.

Santarnecchi, E., Rossi, S., & Rossi, A. (2015). The smarter, the stronger: intelligence level correlates with brain resilience to systematic insults. *Cortex*, 64, 293–309.

Santarnecchi, E., Tatti, E., Rossi, S., Serino, V., & Rossi, A. (2015). Intelligence-related differences in the asymmetry of spontaneous cerebral activity. *Human Brain Mapping*, 36, 3586–3602.

Shahabi, S. R., Abad, F. J., & Colom, R. (2014). Short-term storage is a stable predictor of fluid intelligence whereas working memory capacity and executive function are not: A comprehensive study with Iranian schoolchildren. *Intelligence*, 44, 134–141.

Shipstead, Z., Harrison, T. L., & Engle, R. W. (2015). Working memory capacity and the scope and control of attention. *Attention, Perception, & Psychophysics*, 77, 1863–80.

Shipstead, Z., Lindsey, D. R. B., Lakey, C. E., & Engle, R. W. (2014). The mechanisms of working memory capacity: Primary memory, secondary memory, and attention control. *Journal of Memory and Language*, 72, 116–141.

Smoleń, T., & Chuderski, A. (2015). The quadratic relationship between difficulty of intelligence test items and their correlations with working memory. *Frontiers in Psychology*, 6, Article 1270.

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.

Stephan, A. (1999). Varieties of emergentism. *Evolution and Cognition*, 5, 49–59.

Unsworth, N., & Engle, R. W. (2007). On the division of short-term and working memory: An examination of simple and complex span and their relation to higher order abilities. *Psychological Bulletin*, 133, 1038–1066.

Unsworth, N., Fukuda, K., Awh, E., & Vogel, E. K. (2014). Working memory and fluid intelligence: Capacity, attention control, and secondary memory retrieval. *Cognitive Psychology*, 71, 1–26.

Unsworth, N., & Spillers, G. J. (2010). Working memory capacity: Attention, memory, or both? A direct test of the dual-component model. *Journal of Memory and Language*, 62, 392–406.

Unsworth, N., Spillers, G. J., & Brewer, G. A. (2010). The contributions of primary and secondary memory to working memory capacity: An individual differences analysis of immediate free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36, 240–247.

Wiley, J., Jarosz, A. F., Cushen, P. J., & Colflesh, G. J. H. (2011). New rule use drives the relation between working memory capacity and Raven's advanced progressive matrices. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 37, 256–263.

Wilson, W. H., Halford, G. S., Gray, B., & Phillips, S. (2001). The STAR-2 model for mapping hierarchically structured analogs. In D. Gentner, K. Holyoak, & B. Kokinov (Eds.), *The analogical mind* (pp. 59–124). Cambridge, MA: MIT Press.

Process Overlap Theory and First Principles of Intelligence Testing

Nelson Cowan

Published online: 02 Aug 2016.

Submit your article to this journal ↗

View related articles ↗

View Crossmark data ↗

Routledge
Taylor & Francis Group

Process Overlap Theory and First Principles of Intelligence Testing

Nelson Cowan

Department of Psychological Sciences, University of Missouri–Columbia, Columbia, Missouri

The purpose of this comment is to put the process overlap theory of Kristof Kovacs and Andrew Conway (this issue) in the broadest possible context. I briefly discuss the nature of intelligence testing and then relate it to the theory under consideration before making a few concluding comments.

Nature of Intelligence Testing

Intelligence testing was supposed to be a means to determine who is most capable at school or work. School itself consists of educated guesses about what skills would be most important to teach and assess as precursors to a successful adult life, including but perhaps not limited to work. In the workplace, if the task involved is specific, the test can be commensurately specific. If the workplace includes a variety of jobs with a variety of complexity levels, the intelligence test justifiably comes into play to try to assess who is best able to learn a new skill and who has the best bank of knowledge applicable to a wide range of situations.

The criteria for intelligence tests ideally are success at school or at work. Usually, however, these criteria are hard to come by, a situation that has limited test development. Sometimes there is a good proxy that can be used as a criterion; in child testing, for example, the skills that increase with age in the typical child have served as good proxies. The reason is that it is a safe bet that if a child resembles the average child who is older (younger) than him- or herself, that particular child is relatively intelligent (unintelligent).

Still, a lot is left to be desired in intelligence testing because of compromises made in the name of practicality. We all know that social and emotional skills and wisdom in decision making are quite important in the workplace, as is creativity, though all of these have proven difficult to test in the conventional sit-down situation, and perhaps for that reason have been omitted from most conventional tests of intelligence.

Process Overlap Theory in the Bigger Picture of Things

The authors' proposed theory is one example of what I see as the most important trend in intelligence testing since its inception in the late 1800s. At that point, test developers had hunches about what kinds of material to include in tests, resulting in a range of different, sometimes quirky, kinds of test items. Whatever "worked" was kept in the test, and items that

did not predict anything important were excluded. With the hindsight of about 150 years of experimental psychology, though, it has become possible to make more focused predictions about what kinds of test items will be most diagnostic. Moreover, theories of the mental structure related to the tests can be based on this knowledge. The process overlap theory tries to capitalize on this research base, in particular from cognitive psychology and cognitive neuroscience. It is hoped that the theory will consequently be of use in (a) guiding the kinds of test questions that would be most important to add; and (b) predicting performance on kinds of tests that are not even included within intelligence tests—at least not yet.

Considered most broadly, the key types of intelligence test items might be those that help to answer the question of the extent to which, observing the person in question, there is "anybody home" in there, and whether it's someone who could be useful in a work or school situation. As one such essential, high-level ability, working memory capacity indicates the amount of information that can be held in mind, which is related to the complexity of ideas that can be put across to the individual successfully. As a simple example relevant to young children, which I present because it is an easy example to explain, understanding of the meaning of the word *tiger* requires keeping in mind that it is a kind of cat, that it is large, and that it has stripes (or else, overlooking one of these characteristics, one could be talking about a zebra, a house cat, or a lion, respectively).

As a more complex and intertwined set of essential, high-level abilities, *executive functions* include various self-management skills that, applied to the workplace, might be needed in order to ensure that one can say what needs to be said (provided that one knows what that information is); avoid saying something at all, if it would be clearly unwise to do so; keep in mind the context in which one is working; avoid making statements without taking into account the feelings of coworkers; switch rapidly from one task to another when that is necessary; observe one's own behavior enough to know when to avoid harmful distractions; and so on. When a person puts those higher level executive skills to good use, then we indeed feel that someone is "at home" in there, and it may well be someone we would want in our workplace.

As the authors note, though, it would be a mistake to insist that these higher level management skills are all that a person needs. If the person is out of his or her element, there might be

knowledge missing so that the higher level skills cannot be well applied. Although most skills that have been tested tend to correlate fairly highly with one another, some people do seem to have more facility with, say, verbal materials than with spatial items, or vice versa. There are no doubt other individual specialties. The process overlap theory does a good job of pointing out that these skills are individually important but that the working memory and executive function skills serve as bottlenecks for all of them. As an analogy, a restaurant can make excellent food of various types, but the food quality doesn't matter unless the waiters are able to seat you and serve the food in a timely manner, before it gets cold or you have to leave.

If we had better test criteria, theories of intelligence would reveal other bottlenecks. For example, there are various sociopaths who function well on executive skills and have a lot of general knowledge and learning ability but whom you would not want in the workplace because of a personal defect in terms of antisocial motivation. Another kind of person not helpful in the workplace is one undergoing a sustained, debilitating depression that cannot soon be cured. Such key elements of the mind are omitted from the tests, and some of them are considered inappropriate for the tests (e.g., too personal, insensitive to cultural differences, considered medical disabilities to be accommodated). Therefore, they evade the theories based on the tests, including the authors' theory. What kind of test might allow us to determine who not only has fluid intelligence, working memory, and executive function but also who among the capable individuals are the ones most likely to put their talents to effective use? That kind of additional bottleneck occupies the minds of college admissions board members, who therefore heavily consider things like essays and extracurricular activities.

Concluding Remarks

The authors seem to have a good theory of intelligence tests, founded in the extant research on what factors predominate when a problem has to be solved. Individual interests and specific skills in a particular type of material can be important but cannot shine through without adequate memory and executive processes.

We must keep in mind for the future of intelligence testing that, at present, theories such as this one come across as theories founded on arbitrarily constructed tests. For improvement in the utility of the tests, we need to consider what additional human characteristics are important in determining who will make the most of an opportunity and who will waste it. When such tests are conceived, the process overlap theory may become a building block of a more general theory in which the boundary between intelligence and personality is pretty much blurred. For that to happen, of course, we will have to revisit issues about the purposes of the tests and the ethical constraints that should be placed on them.

Funding

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

Published online: 02 Aug 2016.

Submit your article to this journal 🗗

View related articles 🗗

View Crossmark data 🗗

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

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 neuro-anatomy, 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.

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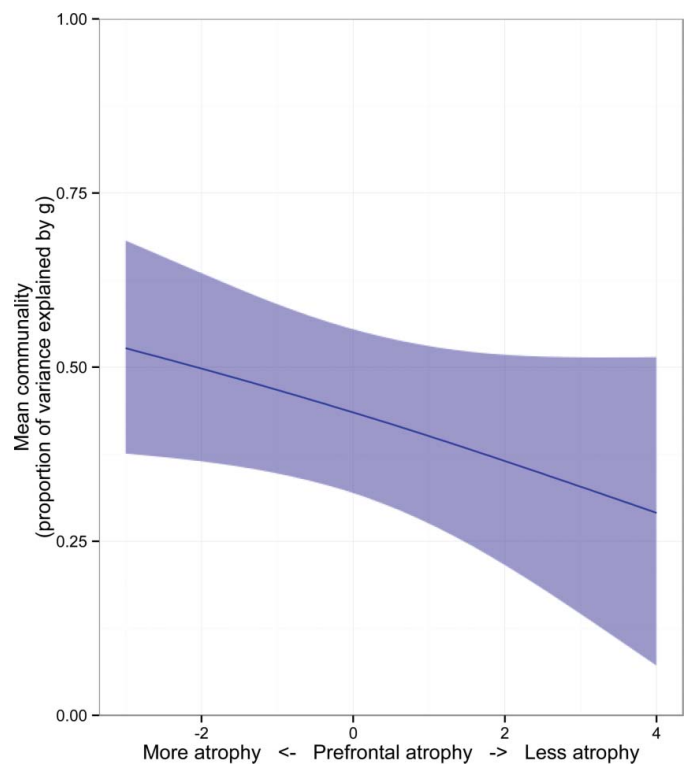
Figure 1. Test of prefrontal differentiation. *Note.* Shaded area around the mean communality line is the 95% confidence interval. The *x*-axis is in standard deviation units.

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

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.
*$p < .05$. **$p < .01$. ***$p < .001$.

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.

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 ($ps < .02$) and no worse than the independent pathways model ($ps > .65$). We combined the models for ICV and for
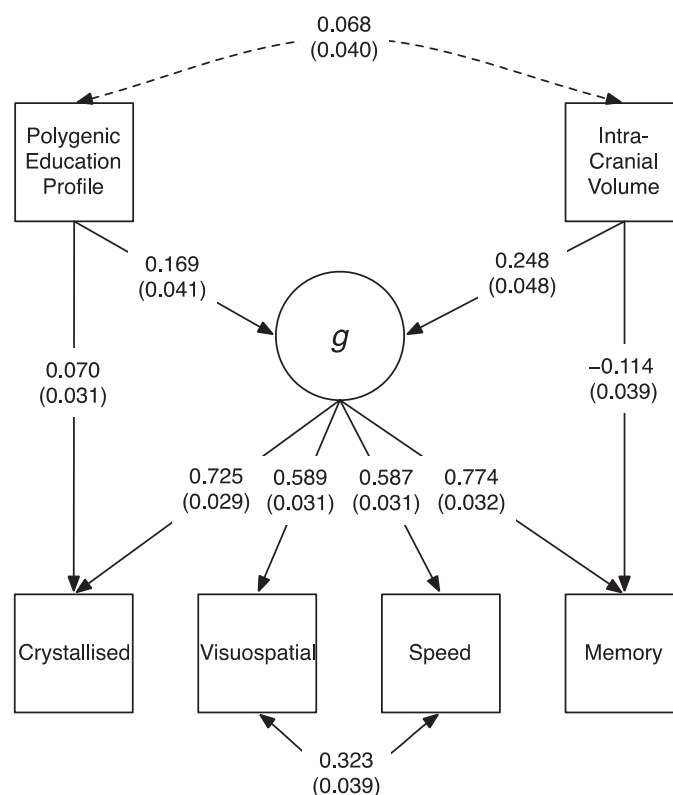

Figure 2. Combined common-plus-independent pathways model of the association of biological factors with *g*. *Note.* Values are standardized regression weights with standard errors in parentheses. The dotted line indicates a nonsignificant path.

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

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.

Process Overlap and System Theory: A Simulation of, Comment on, and Integration of Kovacs and Conway

D. K. Detterman, E. Petersen & M. C. Frey

Published online: 02 Aug 2016.

Submit your article to this journal ⬈

View related articles ⬈

View Crossmark data ⬈

Routledge
Taylor & Francis Group

Process Overlap and System Theory: A Simulation of, Comment on, and Integration of Kovacs and Conway

D. K. Detterman[a], E. Petersen[a], and M. C. Frey[b]

[a]Psychological Science, Case Western Reserve University, Cleveland, Ohio; [b]Psychology, Otterbein University, Westerville, Ohio

Kristof Kovacs and Andrew Conway (this issue) have written an exceptional article that accomplishes two things. First, it identifies what they believe to be the most important processes for human intelligence as domain-general executive processes largely derived from working memory research. Second, it presents a theory called process overlap theory and postulates that these executive processes are used in an overlapping manner far more often than domain-specific processes. This commentary addresses only the second accomplishment, overlap theory.

Kovacs and Conway suggest that many of the findings associated with general intelligence can be predicted from the structural characteristics of the system they propose. As they point out, their theory is closely related to a theory proposed by Detterman (1986, 1987, 1994, 2000) called system theory. Indeed, the principal difference between what they propose and system theory is the additional condition that central elements overlap or operate simultaneously. This additional condition could be incorporated into system theory, as we discuss later.

Perhaps the easiest way to show how a system theory or process overlap theory can explain many of the phenomena cited by Kovacs and Conway is to actually simulate a system. Detterman, Petersen, and Frey (2001) did that with a simple system, and their findings corroborate and, to some extent, clarify what Kovacs and Conway discuss.

One simulated person from this system is shown in Figure 1. Each circle represents a single element. There are nine basic elements labeled A to I. Three of the elements (D, E, F) are central elements, and six are peripheral elements (A, B, C, G, H, I). Arrows indicate the direction of flow through the system. This system has nine routes, which are the three input elements (A, B, C) each crossed with the three output elements (G, H, I). This yields a system with nine routes (A→G, A→H, A→I, B→G, B→H, B→I, C→G, C→H, and C→I). Note that in this system, each route must include the central elements D, E, and F. D, E, and F are called central elements because no output occurs without their involvement. They are central to the system's operation. The score obtained from each route through the system could be regarded as equivalent to a score on a mental test or a test of cognitive processing. IQ would be equivalent to the average result from all of these nine mental tests or cognitive processes.

To simulate this system, we generated 2,500 cases of nine individual elements representing a single person similar to the person shown in Figure 1. For each person, a random normal deviate was assigned to each of the nine elements indicating how well that element worked. (This was done in SPSS but could have as easily been done with Microsoft Excel.) Because no route can work any better than its weakest element, the output of each route can be computed as the minimum value of the five processes. For example, route A→I would be computed as min(A, D, E, F, I) or, for the person in Figure 1, min(0.66, 0.31, 0.34, 2.30, 1.14) = 0.31.

Each route now has a value indicating its efficiency. Each route's value corresponds to what can be considered a score on a single cognitive or mental test. An IQ is estimated as the average efficiency of all the routes in the system by averaging the values of all nine routes for each simulated person. In the case of the simulated person in Figure 1, the average of each of the nine routes is −0.43. (Note that this IQ has not been standardized to a mean of 100 and a standard deviation of 15, but this could have easily been done. This transformation would have no effect on the following results.)

Examining Figure 1, it is clear that the elements that will most affect the outcome of each process are those that are central elements (D, E, F), because a low minimum value for any one of these elements will assure that every route receives a score as low or lower than this minimum value. Also note that the values for each element are entirely independent and randomly assigned. Although the current model is a very simple system, almost cartoonish in character, compared to the processes that are probably involved in intelligence, it is sufficient to demonstrate some of the major findings about much more complicated systems like human intelligence. The major findings demonstrated by this model should be pertinent to any similar system, but numerical values would almost certainly change according to the complexity of the model. In the following analyses, data exactly as just described are used without modification unless otherwise stated.

Positive Manifold

Positive manifold means that all mental tests will be positively correlated with each other. Remember that all elements were assigned a random normal deviate. If the nine individual elements were correlated with each other, the average correlation would be zero. However, the main interest is in the output of
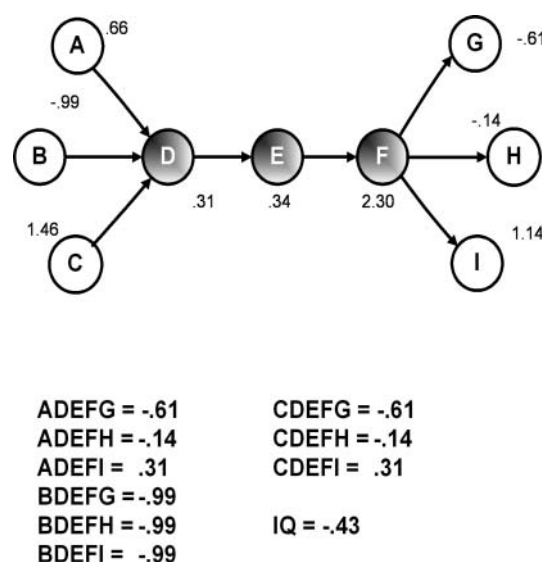
ADEFG = -.61   CDEFG = -.61
ADEFH = -.14   CDEFH = -.14
ADEFI = .31   CDEFI = .31
BDEFG = -.99
BDEFH = -.99   IQ = -.43
BDEFI = -.99

Figure 1. A model simulating a single person's mental architecture. *Note.* Letters indicated individual elements in the system, and arrows indicate direction of flow through the system. The pathways through the system are indicated by the sets of letters below the flow diagram, for example, ADEFG. Each route represents a cognitive test or mental test. Each route is assigned the minimum value in the route. IQ is the average of the route values.

each route because those are assumed equivalent to the scores obtained from cognitive tasks or mental tests.

To determine if positive manifold results from the system constructed here, a correlation matrix for each of the scores for the nine routes through the system for all persons simulated was calculated. The average correlation of the variables in this matrix was then calculated. The average correlation can be easily calculated from the eigenvalue of the first principal component using the following formula: $r_{ave} = (\lambda - 1) / (n - 1)$ where λ is the eigenvalue and n is the number of routes (Kaiser, 1968).

The average correlation among the routes is $r = .65$. There is substantial positive manifold. This value might be larger than would be obtained from a real battery of mental tests, but that is to be expected given the simplicity of the model simulated. It is also clear from the size of the correlation that there is a very large first principal component; that is, the general intelligence factor, g, is very large in this simulation.

We find it amazing that the correlation among tests arises from the simple fact that for simulated people, the central elements work better for some than for others. Because of this, some people with one or more poor central elements generally perform worse on all processes or routes than others with better central processes. Across persons, these differences produce positive manifold.

Differentiation

Differentiation refers to the fact that positive manifold is not uniform across IQ level, age, or other variables (e.g., Detterman & Daniel, 1989). (This is sometimes called Spearman's Law of Diminishing Returns, because Spearman pointed out that correlations change across ability level and referred to this as the law of diminishing returns. Because higher ability levels do not provide diminished returns for individuals, the name seems inappropriate, and we prefer differentiation.) Although the

methodology of demonstrating this effect in actual human data is subject to many potential methodological problems (Molenar, Dolan, Wicherts, & van der Maas, 2010), those issues are ignored for now and are probably less likely in a simulation. In much of the work that has been done, the counterintuitive finding is that correlations between tests are higher among lower IQ persons than higher IQ persons. Because the correlations are higher in lower IQ persons, it means that general intelligence (g) accounts for a larger portion of the total variance for lower IQ person than for higher IQ persons. If general intelligence (g) is what makes people smart, it seems contradictory that low-IQ persons would have more of it than high-IQ persons.

To see if there were differentiation effects in the simulated data, the distribution of IQs was divided at the median. (Although the distribution of IQ scores was roughly normal, the raw mean was −1.15 and the standard deviation was 0.56. These scores were not restandardized into the typical IQ distribution with a mean of 100 and a standard deviation, though they could have been as raw scores are when an IQ test is standardized.) A correlation matrix of the nine routes was then computed for the 1,250 cases below the median and the 1,250 cases above the median. The correlation for the low-IQ cases (below the median) was $r = .51$, whereas the correlation for high-IQ cases (above the median) was $r = .27$.

It seems clear why this happens. As previously discussed, when very low values affect central elements, they have a much more generalized, masking effect on the entire system. That means that when low values do not have a severe effect on central elements, this masking effect does not occur, and so higher IQ cases will be more variable.

Test Complexity

Why do complex tests have a higher correlation with each other than less complex tests (e.g., Vernon & Jensen, 1984)? Although complexity, in general, has no agreed-upon scientific definition when it comes to tests, most researchers agree that people can tell when one test is more complex than another. There is also general agreement that the more complex the test, the higher it correlates with other complex tests. Complexity is specifically defined here as the number of central elements involved in a test. The more central elements a test includes, the more complex the test. Figure 2 shows three tests varying in complexity.

To simulate results from these tests varying in complexity, the same process as before was used for each test. Each process was assigned random deviates, and routes were calculated assigning the minimum value to each route. Each of the tests still consists of nine routes, with only the number of central processes changing. The correlation among the routes was as follows: most complex test, $r = .65$; less complex test, $r = .58$; and least complex test, $r = .44$.

The explanation is much the same as for previous effects. With fewer central elements there is less of a chance that the minimum value occurs in a central element. That means that route scores are more variable in the less complex tests within cases, and therefore correlations are lower for less complex tests.
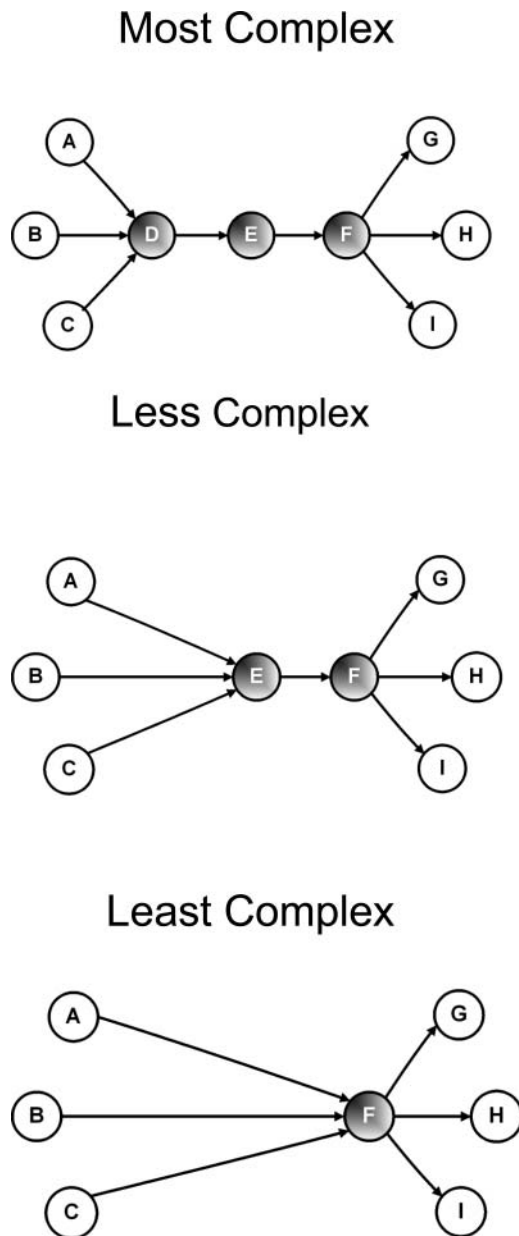
Most Complex

Less Complex

Least Complex

Figure 2. Three models differing in complexity.

Less Complex Tests Can Be Combined to Predict More Complex Tests

Detterman et al. (1992) attempted to use more basic cognitive tasks to predict intelligence tests. How good can such prediction be? To reiterate, in the model shown in Figure 1, each route through the system was scored as the minimum of the elements included in the route. IQ was defined as the average score for all routes through the system. To test if less complex tests can predict more complex tests, the more complex system can be decomposed into simple systems that include only a single central element (D, E, or F) producing three systems. These systems would be equivalent to the least complex test in Figure 2, but there would be three of them, one for each central process. In addition, we use exactly the same random deviates as were used for each process for each case in the complex model. Each route in each of these systems is then scored as for

the more complex system. For each least complex system, there would still be nine scores.

These scores were then used in various combinations using multiple linear regression to predict IQ from the most complex system having three central processes. First, one simple system at a time was used to predict the IQ score from the more complex system. The simple system containing D, E, or F was entered into a multiple regression separately, and the multiple Rs were averaged. The result was an average $R = .62$. Next, the simple systems were entered two at a time so that systems containing, D and E, D and F, or E and F were entered. The multiple Rs were again averaged. The result was an average $R = .75$. Finally, all three systems were entered in the same multiple regression including D, E, and F as separate systems. The result was $R = 0.81$. It should be noted that degrees of freedom were increasing for each complication but that the only new information was the central elements, as all other elements were the same across the simple systems.

Not surprisingly, we found that the more information about the complex system that was contained in the multiple regression, the better the prediction was. However, prediction was never perfect. This indicates that important information about the relationship of the parts of the system is not conveyed by the simple systems assembled from the more complex system. To perfectly predict the functioning of the complex system, each part of the system has to be measured independent of other parts of the system. When that is done, prediction becomes perfect. This suggests that it will be difficult to perfectly predict more complex tests from simpler tests unless it is possible to directly measure each element of the system.

Why Is It That Complex Tests Cannot Be Used to Diagnose Specific Deficits?

Since the first IQ tests, clinicians have attempted to find diagnostic clues among the test results. They thought these clues might be in subtest scatter or in other differences across subtests. Unfortunately, none of these clinical signs have been supported by research. If a child has an IQ of 55, an IQ test gives little information about why. For known genetic defects, a diagnosis can be made on the basis of other phenotypic anomalies associated with the disorder and ultimately by genetic screening. Even more basic cognitive tasks provide very little information about the source of cognitive deficits. For a large portion of intellectually disabled children, no specific cause can be identified for their low IQ. They are often thought to represent the low end of the normal distribution, though this is not a very satisfying or explicit identification of the reason for their low IQ.

The previous simulation suggests why clues from subtests may not be diagnostic. Central elements are confounded with each other and, even when measured by less complex tasks, cannot be completely isolated and do not result in perfect prediction of system outcomes. To demonstrate this fact, the data from the complex system that has been used in the first example are used with one modification. The value assigned to D was changed to −3.00 for half of the cases, and for the other half of the cases, the value assigned to F was changed to −3.00. This creates a large sample with equal deficits in different

central elements. For one group, the deficit is in element D, and for the other group it is in element F. The question to be answered is how these groups can be diagnosed.

The first attempt to separate the groups used information available from the system output from the nine routes through the system. These are the scores that, when averaged, represent the IQ score. To differentiate the groups, we used discriminate analysis. This is a method that uses the data provided to form the most discriminating function between groups in which each case's membership is identified. As expected, this analysis was not able to discriminate group membership beyond a chance level, meaning that it did no better than flipping a coin.

A second analysis entered the values originally assigned to each of the nine elements in the system. This was nearly perfectly discriminating, assigning 91.7 and 92.8 of cases to their respective groups, $\chi^2(9) = 2740.9$, $p < .000$. The point here is much the same as in the previous simulation: Perfect prediction requires direct measurement of the basic, independent elements. The lesson from this simulation is that it will be important to isolate individual elements if we are to understand how cognitive processes work. This suggests that basic brain processes involved in cognitive processes will have to be understood. Cognitive tests, no matter how basic, are unlikely to be perfectly diagnostic at isolating the individual elements of the system.

Simulation Assumptions

These simulations have made as few assumptions as possible. This allowed a focus on the ancillary effects even a very simple system would produce. A first assumption was that abilities are somehow assigned randomly to each element of the system. This was done to be sure that positive manifold could not be attributed to a priori correlations among the abilities. Everything we know about human intelligence suggests that human abilities are not random across individuals. At the very least, assortative mating would likely produce some correlations among abilities. Because assortative mating is high for intelligence, it should be expected that the basic elements of ability are, to some extent, correlated from conception.

The assumption that the weakest process in a serial chain of processes would determine the outcome seems like a reasonable one. It is hard to see how it would be otherwise. However, future simulations could complicate this assumption by making each process probabilistically variable. This would be much closer to human performance, which is never as static as the outcomes in these simulations.

The next assumption is what is meant by an element. That has never been specified. Each process could be defined molecularly or modularly. It would seem that as long as some processes were more important than others and had greater effects on the outcome, what each process represents has little effect on the results found here. The effects demonstrated here should hold up at any level of analysis. It would, therefore, seem that overlap theory and system theory are very nearly identical. For example, the central elements in the model shown in Figure 1 could be thought of as nodes that reflect lower level elements. D might reflect the summation of three lower level elements, which could be called D_1, D_2, and D_3. D, then, would reflect some combination of these more basic elements. The value

assigned to D could be the lowest of the three more elemental processes, or it could be some more complicated combination of the three more basic elements, D_1, D_2, and D_3.

Besides the assumptions made in these simulations, we have ignored some important issues for simplification of the model. Chief among these is the fact that the simulated model is static. It represents a system at a particular moment in time, but intelligence is not static but dynamic. Elements in a complex system change across time and are more or less efficient depending on circumstances. The system simulated here could be made dynamic by including a second parameter for each element, a measure of variability. The random deviate assigned to each element of the system could then be supplemented by a measure of variability. Each time a path was traced through the system, the measure of central tendency could be altered by randomly selecting a value within the range of the element's variability, which could be added to the central tendency measure. Using this method, it would be possible to generate actual data for multiple trials for each simulated person. We have not done this, because a static model is much easier to conceptualize and explains many phenomena associated with intelligence without further complications.

The worst performance rule requires a dynamic model for explanation. Although we have not yet modeled a dynamic system, it is possible to speculate about why the worst performance rule works well at predicting overall performance. When a person is performing at his or her worst, it is reasonable to expect that it is because important elements within the system are at their lowest levels. In other words, it provides the lower bound for performance and indicates how badly central elements can perform.

Conclusions

The existential nature of g has been a question for many years (Detterman, 1982). These simulations indicate that general intelligence and its properties are actually a by-product of the operation of a complex system, as Kovacs and Conway (this issue) suggest. Central elements have an overwhelming effect on the characteristics of the system. In general, the most important thing these simulations demonstrate is the possibility that there is no such thing as general intelligence or g. Rather, what we call general intelligence is a by-product of the operation of a complex system. General intelligence appears to be an index of the efficiency of the brain in carrying out cognitive processing

In addition, a number of other properties attributed to general intelligence can also be derived simply from the simple system structure simulated here. General intelligence and many of the phenomena associated with it may be epiphenomena of the way our brains are organized and how this organization is genetically programmed. In our opinion, Kovacs and Conway are on the right path.

References

Detterman, D. K. (1982). Does "g" exist. *Intelligence*, 6, 99–108.

Detterman, D. K. (1986). Human intelligence is a complex system of separate processes. In R. J. Sternberg & D. K. Detterman (Eds.), *What is intelligence? Contemporary viewpoints on its nature and definition* (pp. 57–61). Norwood, NJ: Ablex.

Detterman, D. K. (1987). Theoretical notions of intelligence and mental retardation. *American Journal of Mental Deficiency*, 92, 2–11.

Detterman, D. K. (1994). Theoretical possibilities: The relation of human intelligence to basic cognitive abilities. In *Current topics in human intelligence volume 4: Theories of intelligence* (pp. 85–115). Norwood, NJ: Ablex.

Detterman, D. K. (2000). General intelligence and the definition of phenotypes. In G. R. Bock, J. A. Goode, & K. Webb (Eds.). *The nature of intelligence, Novartis Foundation Symposium 233* (pp. 136–148). Chichester, UK: Wiley.

Detterman, D. K., & Daniel, M. H. (1989). Correlations of mental tests with each other and with cognitive variables are highest for low IQ groups. *Intelligence, 13,* 349–359.

Detterman, D. K., Mayer, J. D., Caruso, D. R., Legree, P. J., Conners, F., & Taylor, R. (1992). Assessment of basic cognitive abilities in relation to cognitive deficits. *American Journal of Mental Retardation, 97,* 251–286.

Detterman, D. K., Petersen, E., & Frey, M. C. (2001, December). *Simulation of a system theory of intelligence: Critical issues.* Paper presented at the 2nd annual meeting of the International Society for Intelligence Research, Cleveland, OH.

Kaiser, H. F. (1968). A measure of the average intercorrelation. *Educational and Psychological Measurement, 28,* 245–257.

Molenaar, D., Dolan, C. V., Wicherts, J. M., & van der Maas, H. L. J. (2010). Modeling differentiation of cognitive abilities within the higher-order factor model using moderated factor analysis. *Intelligence, 38,* 611–624.

Vernon, P. A., & Jensen, A. R. (1984). Individual and group differences in intelligence and speed of processing. *Personality and Individual Differences, 5,* 411–423.

No Population Is Frozen in Time: The Sociology of Intelligence

James R. Flynn

Submit your article to this journal ⧉

View related articles ⧉

View Crossmark data ⧉

Routledge
Taylor & Francis Group

No Population Is Frozen in Time: The Sociology of Intelligence

James R. Flynn

Department of Psychology, University of Otago, Dunedin, New Zealand

The target article by Kristof Kovacs and Andrew Conway (this issue) is of great importance. Of particular interest is its division of the common factor in IQ test performance into three elements: induction, working memory capacity, and executive functions. I divide my commentary as follows: the theory's superiority to Jensen's theory on the within-population (individual differences) area of intelligence, how the three elements listed illuminate what occurs in the between-population (sociological) area of intelligence, and how social trends over time generate novel hypotheses on both the individual differences and brain physiology areas.

Process Overlap Theory Versus *g*-Theory

Many deductions from Jensen's theory have now been falsified. All of them rest on the proposition that *g* is psychological process (with its own seat in the brain) tapped by all intelligence tests to obtain the positive manifold. I call this the "irreplaceable fuel" theory, which has four parts: (a) On the neural level, there is a mental energy (having to do perhaps with neural speed). (b) On the psychological level, this engenders a problem-solving capacity tapped by special problem-solving abilities (say those measured by different Wechsler subtests): (c) the degree to which it is tapped varies by subtest insofar as they pose problems of cognitive complexity and this is measured by their *g*-loadings, and (d) score gains over time that do not tally with a subtest hierarchy according to *g*-loadings are not true *g*-gains or intelligence gains.

They are gains of lesser consequence caused by social change (more familiarity with test content). They are not true intelligence gains because these come only with the upgrading of brains, thereby upgrading the quality of *g*. Improved brains are caused by factors that impinge on brains directly. For example, natural selection, hybrid vigor, better nutrition, either during pregnancy or in early childhood (breast-feeding), elimination of childhood diseases that damage brain development, and so forth. Needless to say, only a fragment of IQ gains over time can be explained by biological factors, and thus only this fragment is accorded the status of cognitive progress. For an analysis of Jensen's theory versus my own, see my recent book (Flynn, 2016, Part 2).[1]

I elaborate the criterion Jensen (1998) offered to determine whether score differences over time tallied with *g*. Take IQ gains from one generation to the next: You rank the 10 Wechsler subtests in order of the magnitude of the gains on each sub-test, and then you rank the same subtests in order of the size of their *g*-loadings. The *g*-loading tell you the extent to which a particular subtest measured *g*, in the sense of what subtest was most predictive of the positive manifold, that is, the tendency of a good subtest performance to be sustained over all 10 subtests. Unless you find a robust positive correlation between the two hierarchies (biggest gain = highest *g* loading, etc.), the score gains do not constitute a *g* difference. IQ gains over time generally flunked this criterion and were therefore "hollow." Whatever fuelled them was not improved *g*-fuel.

We now know that whether IQ gains are *g* gains does not rob them of real-world significance. Coyle and Pillow (2008) showed that the cognitive skills measured by the Scholastic Aptitude Test predict university grades even after *g* has been removed. Woodley (2012a) showed that education in particular cultivates specialized patterns of cognitive abilities and that these improve independently of whether they correlate with *g*. Ritchie et al. (2015) were quite explicit: The association of education with improved cognitive performance is not mediated by *g*; education directly affects specific IQ subtests. Woodley (2012b) showed that the historical trend of IQ gains (which of course are not correlated with *g*) both parallels and predicts the growth in GDP per capita experienced by Western nations over the last 10 decades or so ($r = .930$). Meisenberg (2014) argued that over time we are accumulating "cognitive human capital" that is interdependent with economic growth.

There is an inference here I want to defend: Schooling promotes a variety of cognitive skills (*g* aside), and these promote economic progress. Note that the causal arrows could go in the opposite direction: x causes us to get richer, and we spend more on schools and get "smarter." My inference is more probable when we look at "lagged correlations" or what happens when the dimension of time is included. Ireland enhanced education, its tests scores rose, and its per capita gross domestic product rose above that of England—*in that order*. Finland enhanced education of its poorest students and duplicated Ireland's trend (Nisbet, in press).

Fox and Mitchum (2013) showed that IQ gains on Raven's reflect the kind of problems we can solve, despite the fact that they are not correlated with *g* and are not factor invariant. Using the Advanced Progressive Matrices test, Fox and Mitchum allow us to analyze what has altered in people's minds when one generation scores higher on Raven's than the last.

The following analysis is in my language (reproduced from Flynn, 2012a, pp. 284–286). However, we met at the University of Richmond, and they confirmed that my interpretation is compatible with their analysis.

Some 115 years ago, people just beginning to enjoy modernity were still focused on the concrete objects of the real world. They wanted to manipulate the real world to their advantage, and therefore the representational images of objects was primary. If you are hunting you do not want to shoot a cow rather than a deer; if a bird is camouflaged in a bush, you flush it out so its shape can be clearly seen. Raven's poses a problem that is quite alien to your "habits of mind": You must divine relations that emerge only if you "take liberties" with the images presented. It is really a matter of perceiving analogies hidden behind distracters. I present a series of analogies (the first three are my own) to illustrate the point.

1. Dogs are to domestic cats as wolves are to (wild cats). Presented with these representational images people a century ago would have no difficulty.
2. ■ is to ♦ as ↑ is to (→) where the choices are ↑, →, ↖, and ↗. Here you must ignore everything about an image except its shape and position. Just as the square has been rotated a half turn, so has the arrow.
3. □ is to / as O is to (|) where the choices are ∅ - Θ -| - ⊗. Here you must ignore everything but the number of dimensions: The analogy compares two-dimensional shapes to one-dimensional shapes and all else is irrelevant. Representational images are of course three-dimensional, so such a contrast requires being well removed from them.
4. &#B is to B&# as T&T is to ##_ (enter what symbol fits). This is an item from Fox and Mitchum that illustrates the kind of analogical thinking you must do on the Advanced Raven's Progressive Matrices.

Note that the right answer in the fourth item has been left blank. Because no alternatives were presented to choose from, you had to deduce that "&" is the correct answer. I got it right, which was reassuring given that I was then 78 years old, by reasoning as follows. In the first half of the analogy, all that has altered is the sequence of symbols: labeling them 1, 2, 3, they have become 3, 1, 2. Applying that to the second half of the analogy, T&T changes to TT&. Clearly you are supposed to ignore the fact that the doubled letter (TT) has changed to a doubled symbol (##), so the right answer is ##&. This would really discriminate between the generations. We have moved far away from the "habit of mind" of taking pictorial images at face value; indeed, we are interested only in their sequence and treat images themselves as interchangeable if the logic of the sequence demands it.

The key is this: Anyone fixated on the literal appearance of the image "T," as a utilitarian mind would tend to be, would simply see no logical pattern. Contrast this with Wechsler Adult Intelligence Scale Vocabulary (here gains are large as distinct from Wechsler Intelligence Scale for Children Vocabulary). The etiology of enhanced scores over time would be quite different. People over time, thanks to the bonus of more education, simply accumulated a larger store of core vocabulary and got no bonus from the shift from utilitarian toward "scientific" thinking. Except of course for words that labeled abstractions (like species), which now appeared in the new subjects taught.

Fox and Mitchum (2013) classified Raven's items in ascending order of "relational abstraction," more specifically: "for analogical mapping when relations between objects are unrelated to objects themselves." Once again, in Example 4, the relationship can be derived only if one sees that a "T" does not have to retain its identity as a "T." Their core assumption was that "analogical mapping of *dissimilar* objects is more difficult than mapping *similar* objects" (italics mine). I certainly found this to be true. The fact "TT&" had to be translated into "##&" rendered the item harder to solve. And if I were my father (born in 1885), and wedded to taking images at face value for reasons of utility, I suspect I would have found it insuperable.

They analyze the performance of two samples of young adults tested in 1961 and circa 2006, respectively. They found that as the degree of deviation toward the abstract increased, certain items became less predictive of performance within the two generations than between the two generations.

We now know why Raven's scores are so sensitive to environmental change over time. Like our ancestors, we can still use logic to analyze the concrete world. But we have entered a whole new world that allows us to use logic on symbols far removed from the concrete world. We organize the concrete world using abstract concepts that are not represented there.

Premodern people see fish as having nothing in common with crows. You can eat one and not the other; one swims, the other flies. We use DNA analysis to divide living creatures into categories that are nonobservable but offer understanding, and this language has become that of every person who has been exposed to several years of formal schooling. We know that bacteria differ from one-celled animals, that whales are more akin to land animals than fish, and that the tiny hyrax is more akin to the huge elephant than to the rodents it resembles. We know that stars are different from planets (they look the same in the sky), and indeed, our whole picture of the universe (and even our approach to explaining human behavior) is based on logic and abstractions. We are exposed to the symbolism of algebra. No one has ever observed an "x."

In other words, using logic on symbols detached from concrete reality has become a habit of mind in no way alien to us. These skills are not merely useful in mathematics and science and computer programming (programmers do very well on Raven's). They help us to create (and comprehend) a nonrepresentational map of the London underground, or an organizational map that functionally relates the tasks a complex business organization performs. We are more ready to engage with Raven's because the rise of modernity altered our perspective. And the rise of modernity has occurred over only a few generations. Only a test that is sensitive to the new minds that modernity has put into our heads could measure something so malleable. Raven's, more than any other test, *is a barometer of the stages of modernity* and thus continues to play a crucial role in the study of intelligence.

Fox and Mitchum (2014) extended their analysis to Letter Series and Word Series and showed that the fact that the present generation has developed new habits of mind is the very reason gains are not factor invariant. Woodley, Figueredo, Ross, and Brown (2013) concluded that autonomous mental skills allow people to cognitively adapt to modernity and thus score higher on personality indexes. Flynn (2012a) showed that the fact that American adults with some tertiary education went from 12% to 52% between 1953

and 2007 registered as huge gains on the WAIS Vocabulary subtest. These were the equivalent of 17 IQ points (over 1 standard deviation). Irrespective of whether the overall pattern of American subtest gains correlated with g, this had real-world consequences: They could carry on different conversations and read a wider range of books. Flynn (2013) suggested how cognitive progress independent of g has enhanced moral maturity (but not political maturity).

There always was something odd about that the notion that performance gains must tally with complexity for the gains to have real-world consequences. Two basketball teams are evenly matched. The coach of one decides to drill his players on the fundamentals, layups, and foul shots, simple tasks that are less "basketball-g" loaded. Therefore, the performance gains they make do not correlate with a hierarchy of basketball-skill g loadings (no gains on complex tasks like fade-away jump shots). Yet there are real-world consequences: His team beats their rivals by 10 points.

Flynn, te Nijenhuis, and Metzen (2014) put a nail in g's coffin. They compared the Wechsler subtests scores of typical subjects with those who suffered from iodine deficiency, prenatal cocaine exposure, fetal alcohol syndrome, and traumatic brain injury. The typical subjects were higher on every subtest. However, the magnitude of their advantages by subtest had zero correlation with the size of the subtest g loadings. It is difficult to deny that the typical subjects had a significant real-world cognitive advantage over the four comparison groups. This is not to say that their advantage was analogous to that of one generation over another. The latter was influenced by the new habits of mind that evolved over the 20th century.

Now process overlap theory puts a second nail in the coffin. It strips g of its central role in Jensen's theory of intelligence: that of a psychological process tapped by all IQ tests to obtain the positive manifold. It shows that g emerges *because* of the positive manifold rather than *explaining* it. The proffered explanation of the positive manifold involves three elements: induction, working memory capacity, and executive function. These overlap, and the combination is always to some degree involved in performing cognitive tasks. Better still, it adds specificity by identifying the central role played by *induction*. No one actually solves g problems (whatever they might be), that is, it is not a functional mental ability. But people do induction, and it is clear why Raven's is the best test of "g": It is a test of induction beyond all other tests (John C. Raven called it eduction).

There is nothing odd about why the three elements cohere. Working memory capacity is clearly a prerequisite of induction: The greater your capacity to hold abstract concepts in mind, the more you can look for relevant similarities and differences. Executive function in this context is the ability to exclude both cognitive and emotional interference with the inductive task at hand. It is clearly a prerequisite for both induction and high working memory—and indeed, the solution of any other cognitive task.

One thing troubles me: the Wechsler subtests scores of typical subjects and those who suffered from iodine deficiency, prenatal cocaine exposure, fetal alcohol syndrome, and traumatic brain injury. Although typical subjects were higher on every subtest, the magnitude of their advantages by subtest had zero correlation with the size of the subtest g loadings. If we substitute for g the three-factor concept of induction, working memory capacity, and executive function, should there not be a correlation between the extent to which this package is relevant to the subtest and the score difference between normal and damaged subjects? Unless these maladies collectively (and indeed virtually singly) damage the prefrontal lobes in a way that somehow cancels out their differential contribution to the cognitive task set by the different subtests, perhaps by reducing its contribution in all cases to a minimum. This does not seem very plausible, and the authors may wish to comment.

Sociology and the Three Elements

What goes on in people's minds as they solve cognitive problems is a product of the kind of person they are in a particular social setting. Kovacs and Conway (this issue) confine themselves to a within-generation analysis (the common factor weaker at high levels of ability) with only a nod at between-generation analysis (the common factor weakens as generations produce more people of high ability). There is one exception: They imagine the difference between a normal and a color-blind population when they try to solve the colored version of Raven's Progressive Matrices, with the former population largely defeated by the test.

Different societies and different stages of society on the path to modernity alter the hierarchy of problems that are considered important and the habits of mind of the people who try to solve them. They produce radically different populations not unlike the difference between those who are color-blind and those who are not. Moreover, going from one population to another affects the balance between the three elements of induction, working memory capacity, and executive function.

Contrary to Jensen, I make these assumptions: (a) The brain is like a muscle and is modified by exercise; (b) Societies (and generations) have very different hierarchies as to what problems are most important; (c) Practice at solving these problems create different "habits of mind" suited to solving problems in order of importance; and (d) These habits of mind alter how induction, working memory capacity, and executive functions interact. To elaborate, people in 1900 did not need to confront everyday problems that required these habits of mind: taking the hypothetical seriously, using abstract concepts to classify, using logic to analyze relation between such concepts. Therefore, when confronting the inductive tasks of Raven's they were like the color-blind confronting the colored matrices, except worse: Not only were their minds unprepared for the inductive tasks, but also they could not see the point of them, which would undermine their executive capacity to ignore distractors. As to whether they had lower working memory capacity than we do, who knows? I cannot estimate whether we need to hold more things in mind to analyze the relationship between abstractions than to analyze the relationships of coping with everyday life.

This is an example drawn from our own society as it progresses toward modernity, but other preindustrial societies also rank the importance of cognitive tasks differently than we do. Australian Aboriginal society put a high premium on "map reading," that is, noting signals of the presence of water and game on the horizon and calculating the distance that must be traversed. Thus, they would put map reading at the top of an importance of cognitive skills hierarchy and inductive analysis

of abstractions would hardly count. I do not say there would be no input from induction—you use induction to some degree in everyday life—but I suspect the input would be limited. Even in our own society the balance between map reading and induction has probably altered over time. When people began to pilot autos, they got more practice in map reading. When only the rich could own cars, those only who had developed their inductive capacity by more formal education would drive, and this would inflate the correlation between induction and mapping. Then the poor got cars, which would lower it. Now cars have road-trip planners or an automatic guidance system that should put map reading problems much lower on our scale of priorities and return us to the pre-car state.

The best illustration of how executive functions correlate with induction arises from an analysis of a consequence of cognitive progress often not perceived, namely, its role in promoting moral progress. Remember that the modern mind broke its ties with the concrete world, the dominant theme as late as 1900, and asked us to take the hypothetical seriously and use logic to analyze abstract concepts. How did these habits of mind take moral reasoning away from the Stone Age of simply accepting the bias and cruelty of the past?

First, there is taking the hypothetical seriously. When combating racism, taking the hypothetical seriously is the foundation of mature moral argument. In 1955, when Martin Luther King began the Montgomery bus boycott, young men of my acquaintance, home from college at 21, had dialogues with their parents or grandparents. Question: "What if you woke up tomorrow and had turned Black?" Reply: "That is the dumbest thing you have ever said, who do you know that turned Black overnight?" My father believed that problems had to be grounded in the real world to take them seriously and had no room for hypothetical problems.

As for nationalism, my *Beyond Patriotism* (Flynn, 2012b) tries to diagnose the retreat from patriotism by some of the American public between World War II and today. Try this question: "What if your home was hit by a drone because someone nearby was sheltering a Taliban?" Or better: "If a war killed so many foreigners to save 3,000 Americans, where would you fall off the boat: at 10,000 or 100,000 or one million?" The answer tends to divide the youth from the aged (the latter: "Their government protects them and our government protects us"). Voltaire said that all man's reason flies before a drum. Well, it depends on how much reason and how loud the drum.

Today we use logic to analyze abstract concepts. This is a powerful weapon against local norms that incorporate the cruelty of the past as a residue. An Islamic father (guided by local norms, not the Koran) shocks the world when he kills a daughter because she has been raped. We would ask: "What if you had been knocked unconscious and sodomized?' He is unmoved. He sees moral maxims as concrete things, no more subject to logic than any other concrete thing like a stone or a tree. He does not see them as universals to be generalized by logic. Today the tendency is to express moral maxims as generalizations and try to make them logically consistent with one another. Question for one of my students: You say we should never judge the customs of another culture, yet you are also an advocate of women's rights. What do you say about the practice of female circumcision? Whatever the conclusion, this is a far cry from primitive moral reasoning.

In other words, the new habits of mind did not merely help us to adapt to modernity. They also taught us how to modify the modern world thanks to more mature moral reasoning. They taught us to stride toward freedom with Martin Luther King and take seriously the "collateral damage" of killing foreigners in Vietnam and Iraq and Afghanistan. No general today would talk about "bombing the Vietnamese back to the Stone Age."

This makes it seem as if the evolution of society toward modernity has made the use of induction on moral problems merely a matter of developing new habits of mind, ones that are friendlier to logical analysis. However, the social setting has a profound influence on the role of executive functions. An affluent resident of an area in the Middle East or Africa may have had formal education, and thus modern habits of mind, but also come from a family dominated by an inherited sexist morality. The stress placed on his executive functions to banish emotions irrelevant to the application of logic to a moral question may be extraordinary compare to our own: The raped girl just seems somehow tainted. The same is true of someone who comes from a family dominated by racial prejudice: Real-world Blacks just seem alien in a way that impedes analysis based on the traits of hypothetical Blacks.

Intelligence and the Three Areas

In the area of explaining intelligence gains over time, causal explanation involves several levels: (a) Ultimate causes are the industrial revolution and the resulting trend toward modernity; (b) Intermediate causes are the effects of industrialization on society, more education, emancipation of women, smaller families (with a better adult to child ratio), more cognitively demanding jobs, more cognitively demanding leisure, a new pictorial and symbolic world from television and the Internet, better nutrition, and medical advance; and (c) Proximate causes have to do with how people's minds altered, so that in the test room they could do better when taking IQ tests (e.g., new habits of mind).

The Dickens/Flynn (Dickens & Flynn, 2001) model predicts that the size of the IQ advantage between generations will vary depending on the age at which we compare a later cohort (say those born in 1936) with an earlier cohort (say those born in 1921). Both of these groups live their own lives. During those lives the causal factors that differentiate the later from the earlier cohort vary greatly. This means that the IQ gap that separates the two will vary in magnitude with age according to the potency of the differential factors that kick in at each age. This prediction remained only a prediction until a recent study. As Staff, Hogan, and Whalley (2014) say, their study is the first to compare two cohorts at two different ages.

The Lothian Birth Cohorts were born in 1921 and 1936, respectively. They included almost every child born in Scotland in those years (and still attending school there at the age of 11). Both were tested on Raven's Progressive Matrices: The later cohort outscored the earlier by 3.7 IQ points at age 11 and by 16.5 IQ points at the age of 77. The difference is huge: The rates of gain differ at 0.247 points and 1.100 points per year over a period of 15 years. If anything the gain in old age is an underestimate: The earlier cohort lost more people by death (earlier death is negatively correlated with IQ) than the later. The differing gains must reflect the relative

potency of the causal factors that separated the cohorts at those two ages. What might these be?

When you test two cohorts at the age of 11, they both have approximately the same number of years of formal schooling and this serves as a leveler: The small IQ gap (and vocabulary gap) would reflect only the fact that the later cohort came from homes a bit higher in socioeconomic status and any progress made in the quality of schooling. The IQ gap doubles at the age of 21 and, indeed, the Vocabulary gap quadruples: This is thanks to more students going on to tertiary education; the later cohort would have more years of formal education. By age 35, the influence of more schooling would have faded in favor of the later cohort working at more cognitively demanding jobs. No data reveal whether this would confer a greater or lesser advantage than was present in the university years.

At the age of 70, one might anticipate a lessening of the gap, as both cohorts would have retired from work—*except* that the later cohort would be far more healthy and alert. Modern medicine has alleviated the many of the illnesses of old age, and older people today have a better diet and do more exercise (I still run at 82, and my father took no exercise after age 14). Elderly people also have leisure activities that are more cognitively demanding. At age 77, we have real data. We know that the three factors named produce a huge gap (16.50 points for two cohorts only 15 years apart), a gap unlikely to be matched at any earlier age.

I have often rejected the hypothesis that generational IQ gains reflect gains in health and nutrition, at least in advanced nations since 1950. This was because we were looking for them in the wrong place: We thought they would weigh in at the beginning of life (they do not); rather they weigh in at the end. At any rate, we now know that Raven's is not merely sensitive to the global environment enriched by modernity. It is also sensitive to each and every one of the particular factors that have triggered IQ gains over time. This has implications not only on between-generation IQ differences but also on individual differences within a cohort. If one person gets more formal schooling than another, or a more demanding job than another, or better diet and medical care in old age, they will at the appropriate stage of life have an IQ advantage.

By their very nature, theories of brain physiology would ideally accommodate both individual differences and the evolution of cognitive abilities over time. First, we want to map the areas/networks that are activated when people perform various cognitive skills; then we will want to observe differences in those areas/networks that rank people's performance for each cognitive skill. In principle, brain physiology should also illuminate cognitive trends from one generation to another. It is a plausible hypothesis that as people began to drive motorcars, more mapping exercise enlarged the hippocampus between 1900 and today, and that the introduction of automatic guidance systems will erode the size of the hippocampus in the future. We must wait for data about the future but could project back into the past by studying drivers versus nondrivers—or ethnic groups that do not drive cars (the Amish).
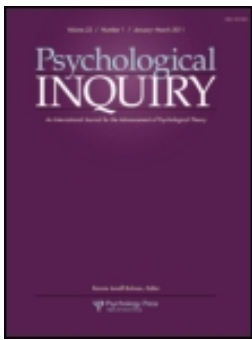
Integration of All Areas

We want an adequate theory of intelligence in the area of brain physiology. However much we may succeed, we will have to resist the temptation of reductionism. Physiology cannot replace psychology and sociology in the sense that we will still need causal explanations in all three areas of the study of human intelligence. Physiology may be able to predict exactly who will be the best basketball player, but we still need to know why someone is doing something as trivial as running around a court to try to throw a ball through a hoop, and why basketball became more popular after World War II, so that greater participation rose and triggered a huge rise in standards of performance.

Kovacs and Conway's primary contribution is in the area of individual differences. At times, they say that the emergence of a positive manifold is a function of the psychological processes of individuals solving problems. I see no reason to assume that this implies that scholars can neglect the fact that individuals are the product of different social circumstances, and that this affects how they solve problems. Which is to say it does not assume we can neglect the sociology of intelligence. Nonetheless, I want to emphasize that a comprehensive understand of intelligence must integrate all three areas. What we think we know about individual differences will always be qualified by what is true about both the brain and society.

References

Coyle, T. R., & Pillow, D. R. (2008). SAT and ACT predict college GPA after removing "g." *Intelligence*, 36, 719–729.

Dickens, W. T., & Flynn, J. R. (2001). Heritability estimates versus large environmental effects: the IQ paradox resolved. *Psychological Review*, 108, 346–369.

Flynn, J. R. (2012a). *Are getting smarter? Rising IQ in the twenty-first century*. New York, NY: Cambridge University Press.

Flynn, J. R. (2012b). *Beyond patriotism: From Truman to Obama*. Exeter, UK: Imprint Academic.

Flynn, J. R. (2013). *Intelligence and human progress: The story of what was hidden in our genes*. London, UK: Elsevier.

Flynn, J. R. (2016). *Does your family make you smarter? Nature, nurture and human autonomy*. Cambridge, UK: Cambridge University Press.

Flynn, J. R., te Nijenhuis, J., & Metzen, D. (2014). The *g* beyond Spearman's g: Flynn's paradoxes resolved using four exploratory meta-analyses. *Intelligence*, 44, 1–10.

Fox, M. C., & Mitchum, A. L. (2013). A knowledge-based theory of rising scores on "culture-free" tests. *Journal of Experimental Psychology: General*, 142, 979–1000.

Fox, M. C., & Mitchum, A. L. (2014). Confirming the cognition of rising scores: Fox and Mitchum (2013) predicts violations of measurement invariance in series completion between age-matched cohorts. *PLoS One*, 9(5), e95780.

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

Meisenberg, G. (2014). What are the causes of cognitive evolution? A critique and extension of psychogenetic theory. *Mankind Quarterly*, 54, 326–328.

Nisbett, R. E. (2015). *Mindware: Tools for smart thinking*. New York, NY: Farrar, Straus and Giroux.

Ritchie, S. J., Bates, T. C., & Deary, I. J. (2015). Is education associated with improvements in general cognitive ability, or in specific skills? *Developmental Psychology*, 51, 573–582.

Staff, R. T., Hogan, M. F., & Whalley, L. J. (2014). Ageing trajectories of fluid intelligence in late life: The influence of age, practice and childhood IQ on Raven's Progressive Matrices. *Intelligence*, 47, 194–201.

Woodley, M. A. (2012a). A life history model of the Lynn-Flynn effect. *Personality and Individual Differences*, 53, 152–6.

Woodley, M. A. (2012b). The social and scientific temporal correlates of genotypic intelligence and the Flynn effect. *Intelligence*, 40, 189–204.

Woodley, M. A., Figueredo, A. J., Ross, K. C., & Brown, S. D. (2013). Four successful tests of the cognitive differentiation integration effort hypothesis. *Intelligence*, 41, 832–842.

A g Theorist on Why Kovacs and Conway's Process Overlap Theory Amplifies, Not Opposes, g Theory

Linda S. Gottfredson

Published online: 02 Aug 2016.

Submit your article to this journal ⌇

View related articles ⌇

View Crossmark data ⌇

A *g* Theorist on Why Kovacs and Conway's Process Overlap Theory Amplifies, Not Opposes, *g* Theory

Linda S. Gottfredson

School of Education, University of Delaware, Newark, Delaware

Kristof Kovacs and Andrew Conway (this issue) raise a question that *g* theorists have sought to answer since Spearman (1904) statistically demonstrated the existence of the general intelligence factor, *g*, more than a century ago. Namely, what in the particulars of brains and biology could generate such a domain-general (content-independent) cognitive tool in everyday life? Like the great pioneers in *g* theory—Charles Spearman (1863–1945), Hans Eysenck (1916–1997), and Arthur Jensen (1923–2012)—Kovacs and Conway seek to understand the underpinnings of *g*'s domain generality by looking into the more elemental processes by which brains process information. I am pleased that these talented cognitive scientists are joining the effort.

They argue that their theory is superior to prior explanations, with special attention to *g* theory. They dispute a series of crucial claims that they associate with *g* theory, showing how the evidence is more consistent with their own theory. But the *g* theory they portray is not the one to which *g* theorists actually subscribe. The good news is that process overlap theory amplifies *g* theory exactly where its pioneers searched hardest for answers—in how the mind and brain process information to learn and solve problems.

My commentary explains why their contributions to understanding intelligence are concordant with, not contrary to, *g* theory. I do so by summarizing the contrasts they draw between their overlap theory and *g* theory, and how the seeming discordance is resolved by distinguishing between different levels of analysis in the full body of evidence on what *g* is and is not. I also suggest a strategy for simultaneously advancing the two theories, specifically, by exploiting a key "trait" of tests and tasks—their relative complexity—that activates the domain-general processes and abilities of keen interest to both. I draw on my work as a *g* theorist (Gottfredson, 1985, 1986, 1997a, 1997c, 2002a, 2002b, 2004, 2007, 2011). Although trained in sociology, my inquiries into the roots of social inequality and job aptitude demands led me inexorably to *g* (Gottfredson, in press).

Process Overlap Theory Offers an Alternative to *g* Theory for Explaining Psychometric *g*

The authors propose a new explanation—process overlap theory—for the "most replicated result in the field of intelligence" (p. 151). As Spearman discovered long ago, all cognitive tests correlate positively with each other, regardless of their manifest content (verbal, figural, etc.) or format (written, aural, individually or group administered, etc.). In technical jargon, mental tests exhibit positive manifold. In practical terms, individuals who perform well on one mental test tend to perform well on all others. In theoretical terms, *g* represents the most generic mental capacity possible: an all-purpose cognitive tool that enhances performance on all tasks requiring any mental manipulation of information. Spearman developed a statistical technique, factor analysis, to quantify the shared overlap (covariation) among mental tests, extract their common factor (*g*) for study as a phenomenon in itself, determine how well each test measures it (the test's *g* loading), and calculate test takers' relative standing (*g* scores) on this latent trait. He did so not to develop tests of intelligence but to understand this most astonishing phenomenon.

Kovacs and Conway, however, depart sharply from this conception of *g* because they do not regard *g* as a phenomenon in its own right. In their view, the general factor exists only as "a necessary algebraic consequence" of the positive manifold among tests. Under process overlap theory, "what is discarded is '*psychological g*': the interpretation of psychometric *g* as a psychological construct" (p. 241). In other words, the *g* factor is not an indicator of "general intelligence," as *g* theory holds, but merely a description of the positive manifold among tests' scores when quantified by factor analysis. The authors' aim, therefore, is to explain the positive manifold, not the algebraic representation of it as a unitary general factor.

To do so, they propose that many discrete cognitive and neural processes interleave—"overlap"—for individuals to answer test items correctly. Only mental processes that are globally useful (domain general) will contribute consistently to the positive manifold observed among tests of diverse content. Their overlap theory thus draws on information-processing constructs of this sort from cognitive psychology (working memory, executive function, attention, inhibition), cognitive neuroscience (the connectome, small world networks), and intelligence research (fluid *g*, reasoning). Conversely, tests of domain general constructs exhibit what Spearman (1927, p. 197–198) called "indifference of the indicator," meaning they line up individuals in basically the same order regardless of the tests' intent or appearance. To illustrate, tests of verbal ability and mathematical reasoning are for many purposes functionally equivalent because both measure mostly differences in *g*. That is why both are almost as good in predicting performance
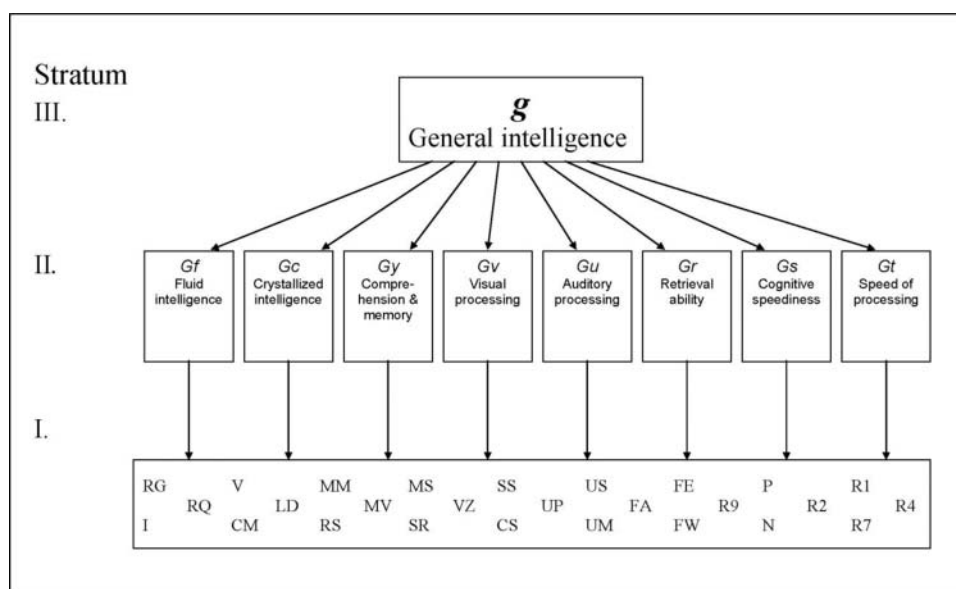
Figure 1. Carroll's (1993) three-stratum model of human cognitive abilities.

in the other's content domain as their own. As Kovacs and Conway (this issue) point out, what "a test a *purports* to measure" is not necessarily what it actually does measure (p. 165).

They argue the superiority of their theory by contrasting it with other explanations of this functional overlap among mental tests. They briefly describe several theories that likewise eschew a general factor, the best known being Thomson's (1916) sampling theory. They focus their contrasts, however, on the theory that gives the *g* factor a starring role in intelligence—*g* theory. To explain their departure from *g* theory more clearly, they refer to Carroll's (1993) three-stratum model, which organizes humans' many cognitive abilities according to their relatedness and scope of application. Figure 1 illustrates how his hierarchical model arrays cognitive abilities from the most general (Stratum III) to relatively narrow (Stratum I) based on his massive reanalysis of prior factor analytic studies.

Psychometric *g* sits alone at the apex, Stratum III, of Carroll's (1993, p. 627) empirically derived model. In Stratum II are arrayed eight factors that are less general but still quite broad in scope, including General Memory and Learning, Broad Visual Perception, and Processing Speed. In Stratum I are many specific abilities of relatively narrow scope, such as Reading Decoding, Free Recall Memory, and Ideational Fluency. This pattern of overlap or relatedness of distinct abilities, from broad to narrow, can be said to represent "intelligence" (cf. Carroll, 1993, p. 627). When referring specifically to the general factor atop the hierarchy, many of us refer to *g* as "general intelligence."

The broad abilities in Stratum II reflect patterns of covariation among the many specific abilities populating Stratum I. The pattern is that Stratum I abilities correlate more strongly when in the same content domain (verbal, quantitative, spatial, etc.).This indicates that the tests in a cluster measure something in common, in addition to *g*, which is content related (domain specific). When factor analyzed, they yield the broad but domain-specific abilities at Stratum II. These broad abilities also covary, but more tightly than do those at Stratum I. The most general, Stratum III abilities are extracted from the positive manifold (correlations among test

results) at Stratum II. Carroll found evidence for only one highly general ability, *g*. He also showed how models that stopped short of extracting a Stratum III *g*, such as Cattell's (1971) model of crystallized and fluid intelligence, could be integrated into his three-stratum model. Carroll determined that fluid *g* and crystallized *g* are Stratum II factors, so Carroll's model is now commonly referred to as the Carroll–Horn–Cattell model.

Fluid *g* is often found to be isomorphic with *g*, and Jensen (1998) considered them to be "one and the same" (p. 106). This makes theoretical sense because both manifest as a domain general capacity for reasoning and solving novel problems. It also accords with Spearman's earlier conceptualization of *g* as a facility for the "education of relations and correlates"—in effect, fluid *g*. Crystallized *g* represents broad cultural knowledge and skills (e.g., language) acquired from "investing" fluid *g*. Individual differences in crystallized *g* track changes in fluid *g* until the two trajectories diverge in early middle-age. Crystallized *g* begins to level off, but fluid *g* tends to decline in tandem with the aging of body and brain. As the two trajectories increasingly diverge, crystallized *g* becomes an increasingly misleading indicator of the individual's capacity for learning and reasoning effectively (fluid *g*). For these reasons I conceptualize *g* in terms of fluid *g* when speaking of Stratum III's general factor, *g*.

Kovacs and Conway also report that Stratum III's *g* and Stratum II's fluid *g* "correlate perfectly or almost perfectly" but argue that they "are conceptually different": "Gf represents individual differences in fluid reasoning while *g* does not represent any psychological process" (p. 166). They accept the existence and validity of trait constructs only at Strata I and II in Carroll's hierarchical model. "Therefore, for the stratum (or strata) below *g*, process overlap theory is compatible with a standard oblique model" (p. 161).[1] They then describe why they like Cattell's oblique model, which does not extract a higher order *g*. "A particular appeal of the Gf/Gc model is that the group factors are relatively easy to interpret as within

---

[1]Presumably oblique models like Thurstone's Primary Abilities and Cattell's Gf-Gc theory.

individual abilities [i.e., processes], which can account for correlations at lower levels of the hierarchy. For example, Gf is interpreted as fluid reasoning, a thoroughly studied cognitive ability, the neural correlates of which are also identified" (p. 155). Moreover, "the reason why tests of fluid intelligence are particularly successful at measuring the processes responsible for the across-domain correlations between mental tests is that they are more or less free from particular domains" (p. 166).

Now, group (Stratum II) factors might seem more interpretable on their face because their scope is defined by particular content domains (verbal reasoning, mathematical reasoning, etc.), yet the high *g* loadings of all group factors indicate that they all tap mostly general processes (reasoning) that cross all domains (reasoning with language, reasoning about mathematical operations), hence the tests' positive manifold. To illustrate the greater interpretability of group factors than *g*, they single out fluid *g*, which they interpret as fluid reasoning, that is, reasoning "more or less free from particular [content] domains" (p. 166). This sounds to me just like Stratum III *g*—domain-independent reasoning—which *g* theorists like Jensen and me have concluded is "one and the same" as fluid *g* and which, as the authors report, are "perfectly or almost perfectly correlated" (p. 166). Another similarity is that tests of fluid *g* produce the same type of scores as does any *g* factor derived from a battery of tests: They reflect only between-individual differences in cognitive ability, not "within-individual processes" (cf. Jensen, as quoted approvingly by the authors on p. 153). The authors nonetheless reject *g* but accept fluid *g* as a valid psychological construct.

Contrasting Understandings of *g* Theory

In fact, Kovacs and Conway reject *g* theory's most foundational conclusions, namely, that Stratum III *g* is a trait (a real dimension of individual differences), that it is a unitary trait (neither an amalgam of disparate abilities nor a "single" process; pp. 158 on evidence that "fractionates" *g*), and that it generates (causes) individual differences in performance on cognitive tests intended to tap more specific abilities (verbal ability, mathematical reasoning, spatial rotation, short-term memory, working memory, processing speed, etc.).

g theory refers to intelligence research in the Galtonian tradition. It was distinctive (and controversial) throughout the 20th century for positing that intelligence has a biological basis and that a general intelligence factor dominates in the pantheon of mental abilities. The tradition is also distinctive for its leaders' sophistication in conceptualizing and measuring human traits, as well as their acumen in formulating and testing hypotheses. Eysenck (1979), for instance, was well versed in both psychometrics and the philosophy of science, and Jensen (1998) was especially adept at making novel predictions and designing incisive experiments that could falsify a favored hypothesis, his or others'.

Kovacs and Conway correctly associate *g* theory with its key developers—Spearman, Eysenck, and Jensen. But *g* theorists would be puzzled by their characterization of *g* theory and its pioneers. For example, the authors argue the superior merits of their theory over *g* theory by sometimes disputing claims attributed to *g* theory that *g* theorists themselves reject. For instance, Kovacs and Conway protest that "There is no psychological process that corresponds to

psychometric *g*" (p. 171) and "it appears as if there is simply no place in the brain for general intelligence" (p. 187). But no *g* theorist has ever made that claim, to my knowledge. Even Spearman (1927, Chaps. 15, 16) spoke of multiple cognitive processes involved in *g*, including attention, memory, and mental span. Cognitive psychologist Hunt (2011, pp. 176, 190) concisely echoed the *g* theorists' stance when he wrote that "The brain functions as a system…. There is no single hot spot in the brain associated with all aspects of cognition."

At other times the authors propose views that *g* theorists are said to reject but have actually promulgated for decades. For instance, Kovacs and Conway's process overlap theory "proposes that *g* is characterized as an emergent property, a result of how processes overlap to produce cognitive activity required by mental tests" (p. 171). Yet, far from rejecting this view, Eysenck (1998) argued that *g* is an emergent property of a highly complex system:

> The brain acts like a unit, but this unit is made up of 10 billion cells, interacting in complex ways through numerous structures, hormones, neurotransmitters, neurological structures and physiology mechanisms; supplied with glucose, oxygen and other necessary foods that provide the energy to keep the engine going. … What the IQ really measures is the total effectiveness of the brain. (p. 79)

Jensen likewise referred to *g* as a property of the brain, not an ability per se.

The Seeming Contradictions Explained

How can this be, that the authors and the *g* theorists whom they dispute actually agree on the very issues that Kovacs and Conway say most distinguish them? To explain, I first provide an overview of the full nomological network for *g*, which ranges across the seven levels of analysis sketched in Figure 2. I use it to illustrate how confusion can arise from conflating constructs and evidence at different levels of analysis, in this case (a) test takers' behavioral responses to cognitive tests (Intelligence), and (b) the cognitive processing system by which their brains manipulate information to generate a response (Brain). Figure 2 also highlights the central importance of the external stimuli that activate the cognitive abilities and processes we wish to observe, in particular the complexity of the tasks to be performed. Knowing the overall complexity of tasks also allows us to predict *g*'s gradients of effect in everyday settings.

Different Levels of Analysis in Explaining Intelligence

Any theory of intelligence has to take account of replicated findings at all levels of analysis. Figure 2 depicts the major seven levels for *g*, ranging from the most molecular (genes) to most macro (evolution). Psychometric *g* ("Intelligence" in Figure 2) sits at the junction of the biological and social manifestations of *g*. Jensen referred to these, respectively, as the vertical and horizontal aspects of *g*.

Kovacs and Conway (this issue) integrate evidence primarily at two of the seven levels of analysis: people's brains (Brain in Figure 2) and their responses to cognitive tests and tasks (Intelligence). Their aim is to explain the positive manifold among test scores and hence *g* at the latter level of analysis (Intelligence). They do so by providing evidence of process overlap at
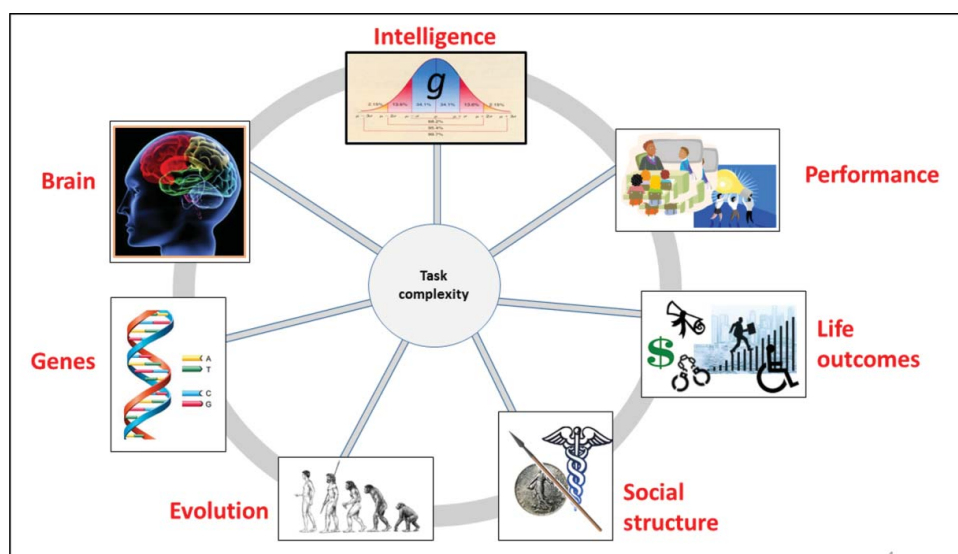
Figure 2. Networks of evidence on *g* spanning different levels of analysis, 2016. *Source:* Figure 4 in Gottfredson (in press). © Elsevier. Reproduced by permission of Elsevier. Permission to reuse must be obtained from the rightsholder.

both levels of analysis. Tests of working memory and other major constructs in cognitive psychology do not measure brain processes directly but provide psychometric "analogs" of them (Hunt, 2011). The authors provide considerable evidence of "process overlap" at this level of analysis (Intelligence). They also call upon research at the Brain level of analysis to support their process overlap theory, including imaging studies of neural networks responding to particular experimental tasks.

Considering evidence at different levels of analysis, as they do, is essential in building theory and testing hypotheses, but levels of analyses must be distinguished, which they do not. A theory is strengthened when data and conclusions are consistent and mesh across levels of analysis, but theoretical coherence does not entail identical conclusions at the different levels. For instance, *g* need not be unitary in the brain if it is unitary at the psychometric level. This, however, is what the authors imply when they criticize unspecified *g* theorists for concluding that *g* exists as a unitary process in the brain, presumably because *g* theorists claim that *g* is *psychometrically* unitary. Only by conflating the two levels of analysis can the *g* theorists' claim that *g* is unitary at the psychometric level be taken simultaneously as a claim that *g* is a unitary process in the brain as well.

Conflating levels of analysis creates a related confusion. It concerns the authors' discussion of whether *g* is a cause rather than an emergent result of the overlap observed among tests and processes in the brain. As Kovacs and Conway repeatedly and correctly stress, psychometric *g* is an emergent property of interacting brain systems, so *g* is their singular *result.* *g* theorists agree, of course, but the authors attribute the opposite belief to them: that *g causes* the overlap in brain processes. As described earlier, *g* theorists believe that psychometric *g* is an emergent property of the brain but also that, as the brain's unitary product, *g* generates a cascade of effects in the real world.

Ambiguities in the following passage illustrate how the confusion arises. I illustrate the authors' inadvertent conflation of two levels of analysis in the following statement by adding bracketed text to distinguish the two levels, tests and physical brains.

The most important difference, then, from *g*-oriented accounts of the positive manifold is that whereas reflective general factor theories propose a causal influence of a latent variable, *g*, on the positive manifold [among psychometric tests and life outcomes], according to process overlap theory *the positive manifold [among tests] is an emergent property [of the brain]*, the result of the specific patterns in which item response processes [i.e., information processing systems in the brain] overlap. (p. 162)

With these insertions, the "important difference" disappears. An emergent *g* produced by the brain can, in fact, cause the positive correlations among responses to psychometric tests and experimental tasks in information processing. These patterns of overlap in scores can then be used, in bootstrap fashion, to infer how the brain does and does not go about its work (e.g., working memory) in a way that produces a unitary *g*, which, in turn, produces its own cascade of effects as people go about their lives.

The authors rightly conclude that *g* is not a unitary or single process in the brain. Imaging research has demonstrated that the processes and structures associated with higher intelligence are widely distributed across the brain, whereas verbal and other broad abilities call upon particular brain modules as well. Domain-general processes are concentrated in the prefrontal lobes (e.g., executive function), as would be expected given their remarkable expansion during human evolution. At the Gene level, molecular genetic research is finding that intelligence is radically polygenic and that individual alleles, or single nucleotide polymorphisms, account for only minuscule proportions of variance in intelligence.

In contrast, decades of research in psychometrics, personnel selection, and other behavioral sciences have established that *g* is a psychometrically unitary (indivisible) dimension of human competence. It is unitary at the level of test behavior (Intelligence) and in life outcomes, which are increasingly global and cumulative at higher levels of analysis: Performance in school and work, Life Outcomes like level of education, occupation, and income, and Social Structures such as education, employment practices, and the occupational hierarchy. Psychometric *g* is indivisible, not "fractionated," at these levels because the brain (and person) responds as a unit,

whether answering items on a test or calculating the tip for a meal in real life.

More important, evidence converges from various disciplines at these higher, "horizontal" levels of analysis to show that *g* is an especially powerful force in human affairs, shaping even culture itself, precisely because it is a unitary, domain general capacity for learning, reasoning, and problem solving in any life domain (for overviews, see Gottfredson, 1986, 1997b, 2011, in press; Lubinski, 2004). For instance, when broad batteries of ability and personality tests are used to predict individual differences in performance in school and work or in health and socioeconomic success, *g* always "carries the freight of prediction." Stratum II abilities add little or nothing beyond *g* to predicting who will perform best in school, jobs, guarding their health, avoiding premature death, and more. Moreover, general intelligence tends to be the single best predictor in the behavior scientist's toolkit of variables, including social disadvantage, for predicting the level of education, occupation, and income that adults attain. *g* is hardly the be-all and end-all of human performance, but it has unrivaled power when life presents individuals with the need to learn, connect the dots, and figure things out. No specific ability, personality trait, social advantage, or fund of experience has been identified that can compensate for mental powers too weak to lift a task's cognitive load.

How to Determine What g Is and Is Not

As Figure 2 illustrates, the nomological network for *g* has expanded greatly since Spearman set out to explain his discovery. It now reaches into all realms of human functioning, and thereby guides and constrains our theorizing about what *g* is and is not. Some of this hard-won knowledge is captured in the following description of general intelligence (Gottfredson, 1997b). All descriptors are content-free, domain-general manifestations of information processing that lay people also recognize as "intelligence."

Intelligence is a very general mental capability that, among other things, involves the ability to reason, plan, solve problems, think abstractly, comprehend complex ideas, learn quickly and learn from experience. It is not merely book learning, a narrow academic skill, or test-taking smarts. Rather, it reflects a broader and deeper capability for comprehending our surroundings—"catching on," "making sense" of things, or "figuring out" what to do.

Factor analysis does not explain the factors it yields, as Kovacs and Conway note. Nor did Spearman or any other any *g* theorist of the Galtonian tradition believe that it could. Indeed, when Hans Eysenck returned to the topic of intelligence in the late 1960s,[2] he argued (Eysenck, 1979) that factor analysis had nothing more to contribute to understanding *g*. He also complained that psychometrics had become focused on the technology of testing and showed scant interest in the constructs tests actually measures.

As the authors also illustrate, understanding intelligence is a long investigative process, with many iterations in collecting data and revising hypotheses. Intelligence is not "defined" but described by laboriously creating a portrait of the phenomenon

as embedded in broader networks of human functioning—a nomological network. *g*'s theoretical meaning is inferred from replicated patterns gleaned from multiple, ever-evolving bodies of evidence.

Eysenck approached intelligence as a biological phenomenon, so his laboratory began noninvasive studies of elemental processes in the brain. He used the only tool available at the time, the EEG, to watch the brain in real time responding to experimental stimuli. He also developed choice reaction time tasks (e.g., the odd-man-out task) that better instantiated Spearman's (1927, p. 410–411) theoretical description of highly *g*-loaded tests as requiring the "eduction of relations and correlates." EEG brain waves and reaction time on exceedingly simple tasks (e.g., touch a button when it lights up) were as close to the brain as he could get.

Arthur Jensen, another pioneer in understanding *g*, wrote often about the "g beyond factor analysis." His review (Jensen, 1998) of the many biological and sociological correlates of *g* helped demonstrate that *g* was no chimera of factor analysis, Gould (1981) notwithstanding. It was especially important to Jensen to determine whether *g* was a replicable phenomenon across human populations. He and others therefore investigated whether different populations and different test batteries produce different *g* factors, or whether they all converge on the same "true" *g*. Prominent psychologists such as Anne Anastasi (1970, 1983) had been arguing that different cultures create different abilities and, later, would argue that the *g* dimension of correlated individual differences is a product of Western education. However, all derived *g*s turned out to converge on the same "true" *g*, surely a biological fact in itself.

Kovacs and Conway (this issue) argue that "*g* is far from being a constant" (p. 155), but they mean something different. For them, it means that *g* (the positive manifold) does not account for the same proportion of variance in a test battery's scores in all groups of people or batteries of tests, though admittedly the lion's share in all. It is theoretically intriguing that *g* accounts for a smaller proportion of test score variance among high-*g* than low-*g* individuals, but the construct validity of a domain-general human capacity does not rest on its being equally dominant among cognitive abilities in all circumstances and populations.

The positive manifold that is *g* is similar in this respect to the heritability of intelligence, which is just the proportion of phenotypic variation in a population that can be attributed to genetic variation. The proportion of total variance accounted for by the "general factor" in question (genetic variation, variation in *g*) can differ depending on age, statistical artifacts (e.g., measurement unreliability, restriction in range in test scores), and conditions that allow versus block individuals from expressing their potentials and proclivities (e.g., relaxed vs. rigid rules for behavior; tests that are not too hard or too easy vs. those that are). Not being "constant" in this narrow sense does not contradict the universality of the *g* dimension in human populations. The validity of *g* as a human universal rests instead on whether the *g*s derived from different populations and test batteries exhibit the same properties, such as showing the same pattern of relations with other variables after correction for statistical artifacts. Stated another way, what matters is evidence that cognitive differences in all populations align

---

[2]Eysenck's (1939) first publication reconciled Spearman and Thurstone's dueling factor analytic models: Spearman posting only a general factor (*g*) and test specificity, and Thurstone positing a set of distinct primary abilities but no general factor.

themselves in the same relation to one another along the same underlying continuum, or "true" g.[3]

Kovacs and Conway draw on various sorts of evidence, including their own, to conclude that psychometric g is an emergent property of the brain and to rule out notions of it being a single process or place in the brain. So did Eysenck and Jensen. As noted earlier, Eysenck argued that the brain acts as a unit but its internal workings are exceedingly complex. He started his inquiries into the brain by focusing on speed of processing (e.g., latencies of particular brain waves in response to a sound) but soon concluded that speed of processing involved more than mere conduction speed. He and his research team speculated that physiological properties such as myelination of axons in the brain's white matter might explain differences in efficiency or error rates in neural transmission, which would also slow speed of processing. In his last book, Eysenck (1998) discussed the nascent body of research on brain-wide efficiency in information processing, including the first imaging study of normal intelligence (Haier et al., 1988), which found that brighter brains use less glucose when solving problems. He anticipated, but sadly did not live to see, the enormous advances in tracing neural networks that Kovacs and Conway (this issue) mention.

Jensen[4] (2006) was particularly interested in reaction time studies as a window into the brain, not because he thought speed alone explained intelligence but because units of time (e.g., milliseconds) provide ratio-level measurement of mental processes. Standard cognitive tests do not. He considered norm-referenced test scores (performance relative to some reference group's mean) a major barrier to progress in understanding general intelligence. I should note that norm-referenced measurement is far less a problem for understanding g's causal effects at the horizontal levels of analysis in intelligence. The reason is that social life operates as a comparative, competitive system of (being the more qualified job applicant, "getting ahead"), as does evolution itself.

How Variations in Task Complexity Help Expose What g Is and Does

Figure 2 places task complexity at the hub of all seven levels of g-related phenomena. In my view, it is the key to explaining g, from how it evolved to how it operates in the real world. Why? Because cognitive abilities and processes manifest themselves, become observable, and exert their causal power only when activated by some stimulus. In fact, abilities are named and classified by the range of tasks on which they enhance performance.

> As used to describe an attribute of individuals, *ability* refers to the possible variation over individuals in the … levels of task difficulty … at which, on any given occasion in which all conditions appear

favorable, individuals perform successfully on a defined class of tasks. (Carroll, 1993, p. 9)

The question, then, is what features of a task or stimulus evoke domain general processes and *only* domain general processes, ones not limited in scope by any content boundaries, which in turn generate the positive manifold among tests? The literatures in many domains of human performance, from ergonomics and academics to health and occupational advancement, point to how the cognitive complexity of work performed drives the magnitude of individual differences and effect sizes in performance (e.g., variances, correlations, mean differences). As sociologists documented in the 1970s, even the worldwide occupational prestige hierarchy orders occupations by overall complexity and thus cognitive demands and average IQ of incumbents. These literatures discuss task complexity at different levels of granularity: For example, a functional literacy item might require the individual to use two rather than one bit of information, and a job might routinely require workers to analyze information rather than just code it.

Psychometric tests are carefully contrived stimuli for evoking information-processing behavior at increasing levels of difficulty. Spearman and Jensen both sought to understand what made some items and tests more difficult and zeroed in on how complexity increases item difficulty, for instance, abstractness of the information to be processed. So have the developers of the U.S. Department of Education's adult literacy tests. They (Kirsch, Jungblut, Jenkins, & Kolstad, 2002) traced item difficulty on all their scales (Prose, Quantitative, Document) to the same "processing complexity": principally, abstractness of information, amount of information, and distracting information (the third requiring cognitive "inhibition" as described by Kovacs and Conway). Daily life is suffused with such cognitive complexity. The more novel and complicated a task, the more g loaded it will be. Patterns in the complexity (g loading) of tests and life tasks allow one to predict g's gradients of effect in any performance domain or life arena because they are so regular.

I was therefore delighted to see Kovacs and Conway (this issue) describe how experimental tasks in cognitive psychology that are more complex show larger effect sizes. Indeed, the authors highlight complexity as one of four important features of the positive manifold among tests that their theory explains (p. 155): "more complex tests load higher on g than less complex tests (Jensen, 1981)." They provide numerous examples when discussing research on working memory (pp. 156–158), which they repeatedly illustrate throughout their article. "Of course, the characteristics of the task determine the nature of the processes involved at arriving at a correct solution" (p. 164).

Yet they argue that this feature is theoretically uninformative: "However, 'complexity' is not an explanatory concept that can help our understanding of g" (p. 155). Their reasons are that experts do not agree about (a) "how complex a test is" or (b) "how complexity differs from difficulty (Mackintosh, 1998)" and because (c) "there are certainly different 'complexities' … that probably invoke rather different cognitive processes" (pp. 155). They suggest that understanding the g-complexity relation requires first understanding "the cognitive processes involved in more 'complex' tests" (p. 156). However, it would seem more useful to reverse the order and

---

[3]Jensen always cautioned that precision in measurement and conceptualization was essential for theoretical purposes. Degree of error must be taken into account to avoid misinterpreting research results, for example, by not realizing that mean differences or correlations have been artificially lowered by common statistical artifacts.

[4]Jensen began his career as what we would now call a cognitive psychologist, for instance, conducting experiments with the Stroop test to understand general principles in learning.

use the elements of a task's complexity to identify the processes they call forth.

If I understand their argument correctly,[5] their first and second rationales for rejecting a theoretical link between *g* and complexity—experts cannot agree on what complexity is or how it differs from difficulty—would (if valid) seem to apply to process overlap theory as well. However, consensus is not a criterion for demonstrating validity or utility, and Jensen (1998, p. 94) explained the difference between a test's complexity (*g* loading) and its difficulty (% passing), as well as how tasks can be difficult without being complex (memorize 100 telephone numbers in 10 min). Complexity is an attribute of cognitive tasks and refers to differences in the cognitive load they impose for successful performance (e.g., bits of information to integrate, inferences required, abstractness of concepts, irrelevancies to ignore). In contrast, difficulty refers to the proportion of test items that are failed in a specified population, meaning difficulty depends not only on the intensity of the test's cognitive demands but also on the ability level of the individuals tested. Less able populations pass fewer items, so the same test earns a higher difficulty rating when administered to lower-*g* than higher-*g* populations. Complexity is an attribute of tests that can be ascertained independent of whoever might take them, if anyone. In contrast, a test's difficulty and its *g*-loading are population dependent because they derive from the scores of people who took the test.

Their third reason ("there are certainly different complexities") is more to the point, but precisely because understanding what makes tasks more versus less cognitively complex is absolutely crucial for understanding the nature, origins, and consequences of human variation in a capacity that transcends the particulars of time, place, form, and content of information. If we better understood the various task attributes that call for additional sorts of information processing, we might be in a better position to understand the nature, number, and relations among the processes themselves.

Kovacs and Conway are correct that there is no consensus on the meaning of complexity, at any level of analysis, despite researchers' frequent appeal to the concept. However, the authors are ideally qualified to resolve that matter. As they say, "Of course, the characteristics of the task determine the nature of the processes involved at arriving at a correct solution" (p. 164). It would be an enormous contribution, both to research and theory on intelligence, for them to spell this out. I have searched in vain for a system that allows one to systematically identify and catalog the elements of a cognitive task that ratchet up its complexity. Such a system would have practical applications as well: for example, to chart and reduce the heavy cognitive demands in health self-care today that generate high rates of patient error and nonadherence to treatment, which mightily frustrate health care providers and endanger patients. When critical self-care tasks are too difficult for patients, the tasks can be restructured but patients' brains cannot.

---

[5] I cannot be sure because Kovacs and Conway (this issue) refer to complexity sometimes as an attribute of cognitive processes ("This implies that g is related to the complexity of cognitive activity," p. 155), sometimes as an attribute of experimental tasks that evoke them ("how complex a test is," p. 155), and at other times as the extent to which one particular class of processes is used in solving problems ("the overlap is caused by executive functions," p. 171).

How Task Complexity Links Experimental and Differential Research on Intelligence (Within- vs. Between-Individual Differences)

Systematic attention to the elements of task complexity would have another important benefit, namely, directly joining the experimental and differential approaches to intelligence. The authors refer to them, respectively, as the within-individual versus between-individual approaches because that is the partition of variance in mental performance that each tries to explain. Cronbach (1957) referred to them as the "two worlds of scientific psychology" because it was as if they inhabited different planets. Even today, they still speak different dialects, pursue different goals using different methods, convene separately, publish in different journals, and trace different lineages. It is no surprise that they sometimes misunderstand one another. I describe one such misunderstanding reflected in the authors' article so that I can better explain the second way they could exploit task complexity to great benefit.

Kovacs and Conway (this issue) offer a "critique of the interpretation of *g* as a within-individual construct" (p. 153). Their concern is that "the concept of general intelligence interprets *g* as a within-individual mental ability" (p. 153). Their concern is misplaced, however, if by "concept of general intelligence … interprets" they mean *g* theory, and if by "within-individual mental ability" they are referring to how brains typically process information rather than how some brains work better than others. They themselves (p. 153) quote Jensen (although to support a different point) clarifying how studies of individual differences in intelligence do not capture thought processes measurable only by studying what goes on within the minds of individuals. Once again, the apparent contradiction between process overlap theory and *g* theory dissolves into agreement.

All traits are by definition accounts of differences between people, and virtually all if not all measures of psychological traits report scores on a norm-referenced scale (distance from the average) such as IQ, z, T, and stanine scores, rather than on an absolute scale such as minutes, inches, pounds. Intelligence, extraversion, neuroticism, self-esteem, and such refer to continua along which individuals differ, but ones not anchored to any meaningful zero point (total absence). We scientists foster confusion among nonscientists by not prefacing trait names with "differences in" because nonscientists often wrongly assume we are referring to absolute measures like height and weight (e.g., "Casey is 40% smarter than Meredith"). That shorthand for traits is why *g* is sometimes mistaken as "a within-individual construct," to which Kovacs and Conway rightly object.

Although not directly illuminating how brains process information, differential studies are nonetheless valuable for generating and testing hypotheses about how they do so. Haier et al.'s finding of differential glucose uptake by intelligence level is an early example. A decade earlier, in 1973, cognitive psychologist Earl Hunt and his colleagues (Hunt, 2011, p. 143) published a series of studies on the information-processing correlates of verbal and mathematical reasoning. It stimulated a "blizzard" of such studies. As Kovacs and Conway's review of evidence illustrates, cognitive psychologists today often turn to differential studies to further their experimental work on information-processing constructs, such as working memory and executive

function. In like manner, brain imaging neuroscientists are supplementing their correlational studies of intelligence and brain action with experimental studies.

Both the experimental ("within") and differential ("between") approaches require the administration of cognitive tasks, and they often use the same or similar ones. Both approaches have discovered that domain processes ("within") and general abilities ("between") are activated by the domain-general demands of a task, referred to generically as its "complexity" as distinct from its content. The key difference is that first approach would compare two tasks performed by the same individual, whereas the second would compare two individuals performing the same task (within vs. between individuals variation). Either approach can provide clues for the other—how do minds operate, and how do minds differ?

Being able to characterize tasks according the attributes generating their complexity, and by how much, would provide a common metric for integrating results from the two types of research. For instance, if both administered three timed tasks of increasing complexity, an experimental study would look at how given increases in task complexity (ΔX) change individuals' successive responses (ΔY), perhaps by slowing them down as more cognitive processes are recruited to answer the more complex task correctly. A differentialist study would look at how much the same increments in task complexity (ΔX) expand the differences in how quickly individuals respond ($\Delta \sigma_y^2$) and tighten the correlation (Δr_{xy}) between response times and intelligence level. A metric for task complexity would also allow placing findings from both approaches into a common, quantitative frame of reference. In effect, to reunite the two partitions of variance.
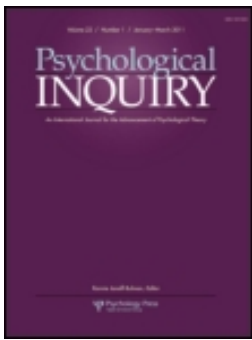
Conclusion

Kovacs and Conway have provided a critique of g theory to justify proposing a new theory, process overlap, for explaining an old but still remarkable discovery about human intelligence. I have explained various ways in which their critique is misplaced. But my main point is that the critique was unnecessary. Not because the two theories actually align, not collide, but because the authors' illumination of how cognitive processes themselves align stands on its own. They need no theory to fall for theirs to stand. More than that, I believe they could make major contributions in understanding how the confluence of domain-general reasoning processes is evoked by external demands and opportunities to solve problems effectively and efficiently. To that end, I encourage them to parse the complexity of the stimuli that instigate cognitive action. Success in quantifying the cognitive load of different experimental tasks would also help bridge the "two worlds" of intelligence research.

Acknowledgments

References

Anastasi, A. (1970). On the formation of psychological traits. *American Psychologist*, 25, 899–910.

Anastasi, A. (1983). Evolving trait concepts. *American Psychologist*, 38(2), 175–184.

Carroll, J. B. (1993). *Human cognitive abilities: A factor-analytic study*. Cambridge: Cambridge University Press.

Cattell, R. B. (1971). *Abilities: Their structure, growth, and action*. Boston, MA: Houghton Mifflin.

Cronbach, L. J. (1957). The two disciplines of scientific psychology. *American Psychologist*, 12, 671–684.

Eysenck, H. J. (1939). Review of Primary mental abilities by L. L. Thurstone. *British Journal of Educational Psychology*, 9, 270–275.

Eysenck, H. J. (1979). *The structure and measurement of intelligence*. Berlin, Germany: Springer-Verlag.

Eysenck, H. J. (1998). *Intelligence: A new look*. London, UK: Transaction Press.

Gottfredson, L. S. (1985). Education as a valid but fallible signal of worker quality: Reorienting an old debate about the functional basis of the occupational hierarchy. In A. C. Kerchoff (Ed.), *Research in sociology of education and socialization* (Vol. 5, pp. 119–165). Greenwich, CT: JAI Press.

Gottfredson, L. S. (Ed.). (1986). The g factor in employment [Special issue]. *Journal of Vocational Behavior*, 29(3).

Gottfredson, L. S. (Ed.). (1997a). Intelligence and social policy [Special issue]. *Intelligence*, 24(1).

Gottfredson, L. S. (1997b). Mainstream science on intelligence: An editorial with 52 signatories, history, and bibliography. *Intelligence*, 24(1), 13–23.

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

Gottfredson, L. S. (2002a). g: Highly general and highly practical. In R. J. Sternberg & E. L. Grigorenko (Eds.), *The general factor of intelligence: How general is it?* (pp. 331–380). Mahwah, NJ: Erlbaum.

Gottfredson, L. S. (2002b). Where and why g matters: Not a mystery. *Human Performance*, 15(1/2), 25–46.

Gottfredson, L. S. (2004). Intelligence: Is it the epidemiologists' elusive "fundamental cause" of social class inequalities in health? *Journal of Personality and Social Psychology*, 86(1), 174–199.

Gottfredson, L. S. (2007). Innovation, fatal accidents, and the evolution of general intelligence. In M. J. Roberts (Ed.), *Integrating the mind: Domain general versus domain specific processes in higher cognition* (pp. 387–425). Hove, UK: Psychology Press.

Gottfredson, L. S. (2011). Intelligence and social inequality: Why the biological link? In T. Chamorro-Premuzic, A. Furnham, & S. von Stumm (Eds.), *Handbook of individual differences* (pp. 538–575). New York, NY: Wiley-Blackwell.

Gottfredson, L. S. (in press). Hans Eysenck's theory of intelligence, and what it reveals about him. *Personality and Individual Differences*. doi:10.2016/j.paid.2016.04.036

Gould, S. J. (1981). *The mismeasure of man*. New York, NY: Norton.

Haier, R. J., Siegel, B. V. Jr., Nuechterlein, K. H., Hazlett, E., Wu, J. C., Paek, J., … Buchsbaum, M. S. (1988). Cortical glucose metabolic rate correlates of abstract reasoning and attention studied with positron emission tomography. *Intelligence*, 12, 199–217.

Hunt, E. (2011). *Human intelligence*. Cambridge, UK: Cambridge University Press.

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

Jensen, A. R. (2006). *Clocking the mind: Mental chronometry and individual differences*. Oxford, UK: Elsevier.

Kirsch, I. S., Jungblut, A., Jenkins, L. & Kolstad, A. (2002). *A first look at the findings of the national adult literacy survey* (3rd ed., Report NCES 1993–275). Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement.

Lubinski, D. (2004). 100 years after Spearman's 1904 ""General intelligence," objectively determined and measured [Special section]. *Journal of Personality and Social Psychology*, 86(1).

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.

Thomson, G. H. (1916). A hierarchy without a general factor. *British Journal of Psychology*, 15, 271–281.

The Psychometric Brain

Richard J. Haier & Rex E. Jung

Published online: 02 Aug 2016.

Submit your article to this journal

View related articles

View Crossmark data

Routledge
Taylor & Francis Group

The Psychometric Brain

Richard J. Haier[a] and Rex E. Jung[b]

[a]School of Medicine, University of California, Irvine, Irvine, California; [b]Department of Neurosurgery, University of New Mexico, Albuquerque, New Mexico

When neuroimaging first became available to study intelligence, one of the most exciting possibilities was the ability to test competing psychometric models against quantifiable brain characteristics (Haier, 1990). For example, Maxwell and colleagues had predicted inverse correlations between brain function and cognitive test scores based on IQ test factor loading differences in children between good and poor readers (Maxwell, Fenwick, Fenton, & Dollimor, 1974). Subsequently, we found inverse correlations between scores on the Raven's Advanced Progressive Matrices Test and glucose metabolic rate assessed with Positron Emission Tomography (Haier et al., 1988). At the time we were unaware of Maxwell's prediction, but we interpreted the finding as evidence that brain efficiency was related to intelligence: Higher scores were achieved at a lower energetic cost. Understanding the relationship between brain efficiency and intelligence factors is still a matter of empirical research interest (Haier, 2016, in press).

The thoughtful and wide-ranging model proposed by Kristof Kovacs and Andrew Conway (this issue) once again invigorates the concept of using neuroimaging to test psychometric models the way animal models were once used. For example, a study of systematic brain lesions after rats learned multiple tasks indicated that lesions in some areas adversely effected performance on certain specific tasks but lesions in other areas impaired performance on many tasks (Thompson, Crinella, & Yu, 1990). According to the authors, the former lesions were assumed to identify brain areas associated with "psychometric" intelligence (like the subtests of an IQ test), and the latter regions were related to "biological" intelligence necessary for survival. Early Positron Emission Tomography study results in humans did not map very strongly onto the rat brain areas enumerated for each kind of intelligence (Haier, Siegel, Crinella, & Buchsbaum, 1993), but the concept of matching brain areas to intelligence factors has evolved into human neuroimaging studies. Sophisticated neuroimaging can be applied to test the ideas and the hypotheses suggested by the process overlap theory. As Kovacs and Conway (this issue) note, recent imaging studies in human lesion patients are a major step in this direction (Barbey, Colom, Paul, & Grafman, 2014; Glascher et al., 2010; Glascher et al., 2009; Barbey et al., 2012).
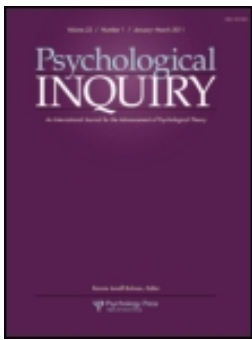
Overall, in our view, the neuroimaging research cited to support the process overlap theory is still simmering, and not quite ready to serve for drawing firm conclusions. One aspect of neuroimaging research, however, that seems to have a compelling weight of evidence is that the frontal lobes are not the sole "locus" of intelligence—fluid, g, or otherwise conceptualized. Our parieto-frontal integration theory (P-FIT) drew attention to the distributed nature of intelligence based on neuroimaging measures of brain structure, biochemistry, and/or function (Jung & Haier, 2007). When neuropsychological and lesion study results were added to imaging studies, it was evident that the distributed view was not an artifact of a bias toward measures of crystallized intelligence. Some of the earliest studies used the Raven's, a good estimate of fluid g, and found, for example, that lesions by missile wounds (frontal lobe or otherwise) result in no significant decline in scores (e.g., Newcombe, 1969). Other neuroimaging studies also used fluid measures like Cattell's Culture Fair Test (Duncan et al., 2000) but studiously avoided discussing significant activation regions that lay outside of the frontal lobes (Colom, Jung, & Haier, 2006). The P-FIT put the parietal lobe (and some other areas) back into the discussion. A number of subsequent studies have supported the P-FIT (Basten, Hilger, & Fiebach, 2015; Pineda-Pardo, Martinez, Roman, & Colom, 2016; Shehzad et al., 2014). Most researchers no longer hold a frontal lobe locus view for intelligence, whatever the role of frontal lobe areas may be for aspects of working memory related to intelligence. We now have compelling brain network parameters that are related to the psychometrics of this important human capacity, and we would remark upon the striking overlap between the P-FIT and the so-called cognitive (or "executive") control network (Niendam et al., 2012), which Kovacs and Conway (this issue) touch upon but do not fully engage.

It takes time to establish a compelling weight of evidence, and there are inconsistencies in the imaging literature, as shown in the commentary by Colom et al. in this issue (see also Haier, in press, for detailed analyses of current imaging/intelligence studies). Nonetheless, the process overlap theory is a thoughtful consideration of current g-related issues and a road map for neuroimaging studies that might succeed in testing the respective validities of competing psychometric models where purely statistical evaluations cannot. No one would have been more excited and enthusiastic about the predictions made by the process overlap theory than Arthur Jensen, and he surely would have congratulated Kovacs and Conway for their important contribution. So do we.

References

Barbey, A. K., Colom, R., Paul, E. J., & Grafman, J. (2014). Architecture of fluid intelligence and working memory revealed by lesion mapping. *Brain Structure & Function*, 219, 485–494.

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.

Basten, U., Hilger, K., & Fiebach, C. J. (2015). Where smart brains are different: A quantitative meta-analysis of functional and structural brain imaging studies on intelligence. *Intelligence*, 51, 10–27.

Colom, R., Jung, R. E., & Haier, R. J. (2006). Distributed brain sites for the g-factor of intelligence. *Neuroimage*, 31, 1359–1365.

Duncan, J., Seitz, R. J., Kolodny, J., Bor, D., Herzog, H., Ahmed, A., & Emslie, H. (2000). A neural basis for general intelligence. *Science*, 289, 457–460.

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, 4705–4709.

Glascher, J., Tranel, D., Paul, L. K., Rudrauf, D., Rorden, C., Hornaday, A., & Adolphs, R. (2009). Lesion mapping of cognitive abilities linked to intelligence. *Neuron*, 61, 681–691.

Haier, R. J. (1990). The end of intelligence research. *Intelligence*, 14, 371–374.

Haier, R. J. (in press). *The neuroscience of intelligence*. New York, NY: Cambridge University Press.

Haier, R. J., Siegel, J. R., B. V., Crinella, F. M., & Buchsbaum, M. S. (1993). Biological and psychometric intelligence: Testing an animal model in humans with positron emission tomography. In E. Detterman (Ed.), *Individual differences and cognition* (pp. 157–170). Norwood, NJ: Ablex.

Haier, R. J., Siegel, B. V., Nuechterlein, K. H., Hazlett, E., Wu, J. C., Paek, J., & Buchsbaum, M. S. (1988). Cortical Glucose metabolic-rate correlates of abstract reasoning and attention studied with positron emission tomography. *Intelligence*, 12, 199–217.

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

Maxwell, A. E., Fenwick, P. B. C., Fenton, G. W., & Dollimor, J. (1974). Reading ability and brain function: Simple statistical model. *Psychological Medicine*, 4, 274–280.

Newcombe, F. (1969). *Missile wounds of the brain: A study of psychological deficits*. London: Oxford University Press.

Niendam, T. A., Laird, A. R., Ray, K. L., Dean, Y. M., Glahn, D. C., & Carter, C. S. (2012). Meta-analytic evidence for a superordinate cognitive control network subserving diverse executive functions. *Cognitive, affective & behavioral neuroscience*, 12, 241–268.

Pineda-Pardo, J. A., Martinez, K., Roman, F. J., & Colom, R. (2016). Structural efficiency within a parieto-frontal network and cognitive differences. *Intelligence*, 54, 105–116.

Shehzad, Z., Kelly, C., Reiss, P. T., Cameron Craddock, R., Emerson, J. W., McMahon, K., & Milham, M. P. (2014). A multivariate distance-based analytic framework for connectome-wide association studies. *Neuroimage*, 93(Pt. 1), 74–94.

Thompson, R., Crinella, F. M., & Yu, J. (1990). *Brain mechanisms in problem solving and intelligence: A survey of the rat brain*. New York, NY: Plenum Press.

Process Overlap Theory: Strengths, Limitations, and Challenges

Kees-Jan Kan, Han L. J. van der Maas & Rogier A. Kievit

Published online: 02 Aug 2016.

Submit your article to this journal 🖉

View related articles 🔍

View Crossmark data ↗

Routledge
Taylor & Francis Group

Process Overlap Theory: Strengths, Limitations, and Challenges

Kees-Jan Kan[a], Han L. J. van der Maas[b], and Rogier A. Kievit[c]

[a]Department of Biological Psychology, VU University Amsterdam, Amsterdam, The Netherlands; [b]Department of Psychological Methods, University of Amsterdam, Amsterdam, The Netherlands; [c]Cognition and Brain Sciences Unit, Medical Research Council, Cambridge, United Kingdom

Kristof Kovacs and Andrew Conway (this issue) offer a new theory for the positive manifold of intelligence (PM) and thus for the presence of a statistical general factor of intelligence. This aim is highly ambitious and deserves praise, especially if the new theory—process overlap theory (POT)—turns out to be true. If so, Kovacs and Conway argue, the general factor of intelligence needs to be regarded as a summary (formally, a constructivist or formative variable) rather than a realistic underlying source of individual differences in cognitive performance (a reflective variable), even in cases where a reflective measurement model is statistically tenable. In this sense, POT contrasts strongly with mainstream theories of intelligence (e.g., Cattell, 1963; Jensen, 1998; Spearman, 1904, 1924) in which the general factor of intelligence is conceptualized as representing a hypothetical yet realistic variable, dubbed g. If g-theory would be true, meaning a realistic g indeed exists, then reflective modeling is not only possible but also appropriate.

Despite differences in interpretation of the statistical general factor of intelligence, there are also strong commonalities between POT and g-theory. For example, in both theories the subtests' (or items') factor loadings on a general factor of intelligence is a simple function of task complexity: The more complex a task, the higher its loading on the general factor, the better it indicates intelligence. Another example is that in both POT and g-theory the factors general and fluid intelligence are strongly related. Given such communalities, one may wonder if the interpretation of the general factor as being a realist or a constructivist variable is important, or if the reflective versus formative measurement approach matters; prediction of work success, health, and other important life outcomes (Gottfredson, 1997) will not change, for instance. In our view the distinction between formative and reflective perspectives does matter, and increasingly so given new insights from various fields.

Due to the influence of scientific reductionism, modern studies of intelligence focus increasingly on the neuronal or genetic "basis of intelligence." If the general factor of intelligence is nothing beyond a constructivist variable, the search for a simple neuronal instantiation of g ("neuro-g"; Haier et al., 2009) will not prove fruitful (e.g., Kievit et al., 2012). In addition, in the quest to detect "genes for general intelligence," lack of power will become an even bigger issue than it already is (e.g., van der Sluis, Kan, & Dolan, 2010). In other words, if

a constructivist conceptualization of the higher order factor is most appropriate, this informs and constrains our search for neural and genetic antecedents: The most fruitful path in such cases would be to focus on those lower order variables that do allow for a realist, causal interpretation.

Comparing the plausibility and merit of scientific theories is a complex challenge, requiring balancing many desiderata including parsimony, explanatory power, internal consistency, falsifiability, and coherence across a range of settings. This is especially challenging in situations where multiple competing theories predict similar or even identical outcomes, like in the preceding examples, which has historically often been the case in the intelligence literature. We here focus on what we see as two possibly outstanding challenges of POT: first, internal consistency, and second, how we may go about testing (and therefore supporting or refuting) the model.

In examining the consistency of POT across representations of the theory, we follow the authors and make a distinction between the theory as stated verbally (POT-V) and the theory as stated more formally, first as a structural relations model of the interindividual variance–covariance structure among intelligence test scores (POT-Structural Model [POT-S]) and second as a test theoretical model (a multidimensional item response model) in the form of Kovacs and Conway's equation (POT-Item Response Theory [POT-I]). We maintain the following position: If POT is a valid theory, POT-V, POT-S, and POT-I should align and should all explain the PM, hence the existence of a statistical general factor, together as well as individually. In addition, inconsistencies or contradictions between POT-V, POT-S, and POT-I will provide a threat to the validity of POT as a whole, or at least require further investigation regarding what representation of POT should be considered the correct conceptualization.

We agree with the authors that a strong theory of intelligence should account for more major findings than simply the positive manifold. Kovacs and Conway (this issue) identify four such findings: (a) the fact that higher order general factor of intelligence and the factor fluid intelligence are strongly correlated (e.g., Detterman & Daniel, 1989; Gustafsson, 1984; Kan, Kievit, Dolan, & van der Maas, 2011; Kvist & Gustafsson, 2008); (b) the finding that the positive manifold is stronger at lower levels intelligence than at higher levels of intelligence (Detterman & Daniel, 1989; Molenaar, Dolan, Wicherts, & van

der Maas, 2010); (c) compared to noncomplex cognitive processing tests, complex cognitive processing tests load relatively highly on the general factor of intelligence (Jensen, 1998); and (d) variability in item performance in certain cognitive domains (e.g., reaction time) relates more strongly to general intelligence than mean item performance (Jensen, 1998; Larson & Alderton, 1990).

At least as important are findings that are thought to *differentiate* between theories of intelligence. Consider, for instance, the finding that the general factor is more heritable than specific factors, such that subtests' factor loadings on the general factor and heritability coefficients are positively correlated (Jensen, 1998). This correlation, dubbed the Jensen-effect for heritability (Rushton, 1998), or simply the Jensen-effect, is often taken as in support of *g*-theory (Rushton & Jensen, 2010), because the correlation would follow naturally if *g* would indeed be the most heritable variable that influences IQ. Conversely, this correlation does not naturally follow from theories in which general intelligence is merely a formative variable. However, recent work has shown how additional hypotheses allows formative accounts of intelligence that also account for the Jensen-effect (which has been accomplished sucessfully; see, e.g., Dickens, 2008; van der Maas et al., 2006; van der Maas, Kan, Hofman, & Raijmakers, 2014). On the other hand, a number of developmental effects, most notably the growth of cognitive performance, do not follow automatically from mainstream *g*-factor models (unless additional assumptions are made), whereas they follow naturally in reciprocal interaction models of intelligence. Ideally, a new theory of intelligence would account for both the Jensen-effect and developmental effects.

We welcome the approach taken by Kovacs and Conway in bringing together various strands of evidence, but we argue that certain aspects deserves critical examination. We end our comment by providing challenges and questions to be answered, in order to help integrating and converging insights from genetics, developmental psychology, and (cognitive) neuroscience. We propose some possible inroads for future extensions.

Pot as Stated Verbally (POT-V)

In a nutshell, Kovacs and Conway's POT-V can be regarded as a particular instance or concretization of Thomson's (1946) sampling theory of intelligence, which in turn was inspired on Thorndike's idea of positive associations between cognitive test score as a result of "overlapping bonds" (see Bartholomew, Deary, & Lawn, 2009; Jensen, 1998, for treatments). Although Thomson and Thorndike speculated about the nature of these bonds, this nature was never specified concretely within their models. This lack of specification is still present in recent variants of sampling theories, such as the model of Bartholomew et al. (2009). In the end the "bonds" in sampling theories must be regarded as no further defined as representing "the variables that underlie individual differences in cognitive performance." In mainstream theories of intelligence, which are inspired on (higher order) factor analytic models of intelligence, the hypothetical underlying variables are generally considered to be

limited in number and positively correlated due to their common dependence on *g*, whereas in sampling theory these underlying variables (*x*) are many (*n*) and considered statistically independent. These characteristics are crucial distinctions between the two theories.

In sampling theory in its simplest form (see Bartholomew et al., 2009, for an overview and more elaborated models), the score of individual *i* on subtest (or item) *j* can be expressed as:

$$y_{ij} = \sum_{k=1}^{n} b_{jk} x_{ik},$$

where b_{jk} is either 1 (x_k is being tapped by subtest *j*) or 0 (x_k is not being tapped by subtest *j*). As the intelligence subtests will draw from the same set of *n* variables and draws will thus show overlap, any two subtest scores will tend to correlate positively. Moreover, the more variables a subtest draws from the population of variables (i.e., the more complex a test is), the stronger the correlations between the subtests scores (if two subtests would both draw all variables, their correlation would be 1, after correction for measurement error).

As acknowledged by Kovacs and Conway (this issue), "process overlap theory can be considered a modern sampling theory" (p. 169). New in POT, and a big step forward, the nature of the cognitive variables (the bonds) is specified more concretely. Based on Baddeley's model of working memory (Baddeley, 1992, 2000; Baddeley & Hitch, 1974), which consists of multiple, functionally independent components, including the Central Executive, the Phonological Loop, and the Visuospatial Sketchpad, a distinction is made between (a) individual differences in capacities that limit domain general executive functioning and (b) capacities that limit domain specific (verbal and visuospatial) processing. In addition, it is hypothesized that during intelligence testing the demand on executive processing is relatively high as compared to the demand on domain specific processes, so that individual differences in cognitive performance reflect to a relatively large extent individual differences in the domain general capacities that limit executive functioning.

Ideally, a theory described verbally is accompanied by formal modeling, that is, as a system of mathematical equations. One may think of sampling models, such as described earlier in this commentary, but also of dynamical system models or traditional psychometric models, such as structural equation models or item response theoretical (IRT) models.

Structural Model (POT-S)

Rather than in mathematical equations, Kovacs and Conway's (this issue) structural model (POT-S) is only presented path diagrammatically (in their Figure 8). Unfortunately, this makes POT-S ambiguous in several key aspects. For instance, the diagram does not show unambiguously whether executive functioning capacities (the black dots) should be conceived of as overlapping (partly shared) *among* verbal, fluid reasoning, and visuospatial tasks. Yet they must do so, as in the absence of such overlap the verbal factor, fluid factor, and visuospatial factor would not correlate. This in turn would mean that POT-S leaves open the explanation of the positive manifold and thus

the existence of a general factor. We assume therefore that the black dots represent executive functioning capacities that are partly shared across subtests. However, we would recommend the structural model to be made explicit somehow in order to avoid ambiguity, because as we illustrate next, POT-S may be formalized such that the general factor has the status of a reflective variable.

A mathematical sampling model that would be in line with both POT-V and POT-S could be, for instance,

$$fluid_i = \sum_{k=1}^{ne} b_k E_{ik}$$

$$verbal_i = \sum_{l=1}^{nv} b_l V_{il} + \sum_{k=1}^{ne} c_k E_{ik}$$

$$visuospatial_i = \sum_{m=1}^{ns} b_m S_{im} + \sum_{k=1}^{ne} d_k E_{ik},$$

where ne is the number of capacities (E) that limit executive functioning, and nv and ns are the number of capacities (V and S) that limit verbal and visuospatial processing, respectively. The parameters b, c, and d are constants that take values of either 0 or 1. (Note: For reasons of simplicity, we sometimes drop the index for test in the equations, but they should be thought of as being present.) Subsequently, one can include the assumption that the variables E_k, V_l, and S_m are multivariate normally and independently distributed.

In this POT-sampling model, differences on intelligence test j would all indicate individual differences in the sum of executive functioning capacities. Verbal and visuospatial tests would both provide biased estimates, toward the sum of the phonological loop capacities and visuospatial sketchpad capacities respectively, whereas executive functioning tests (fluid tests)

would not show such a bias. It is for this reason that the three indices of cognitive functioning will not correlate perfectly with one another.

To verify that our formalization of the POT sampling model indeed results in a statistical model consistent with POT-S, we carried out a series of simulations (code available on http://sites.google.com/site/keesjankan/intelligence) and created performance scores on (three) fluid intelligence tests, (three) verbal tests, and (three) visuospatial tests. The number of capacities was set at 500 each (so 500 executive-functioning capacities, 500 verbal-processing capacities, 500 visuospatial-processing capacities). Individual values were drawn from a (1,500) multivariate standard normal distribution. The 1,500 variables were assumed all statistically independent. The sample size was set at 250, which is a typical sample size in intelligence research (not small, not large). Following POT-V, the probability that a test samples a capacity was set relatively low for domain-specific capacities ($p\_b_{l=1} = p\_b_{m=1} = .35$) and relatively high for executive-functioning capacities ($p\_c_{k=1} = p\_d_{k=1} = .50$; $p\_b_{k=1} = .60$).

The results of the simulation indeed provided support of the factor structure as presented by Kovacs and Conway. Figure 1 gives a typical outcome. In most cases a three (correlated) factors model (with the same fit as a hierarchical model) was tenable, although sometimes a bifactor model (Gignac & Watkins, 2013; Hood, 2008) fitted better (especially when sample size was increased). The correlation between the fluid intelligence factor and general intelligence (modeled as reflective) was generally very high, so much so that in the translation to a higher order model the relation between the two often needed to be fixed at 1 in order to avoid Heywood cases (negative residual variance in Gf).

Whereas Kovacs and Conway (this issue) claim that POT "challenges the idea that the across-domain correlations between diverse mental tests are caused by an underlying factor" and that according to this theory "the positive manifold is
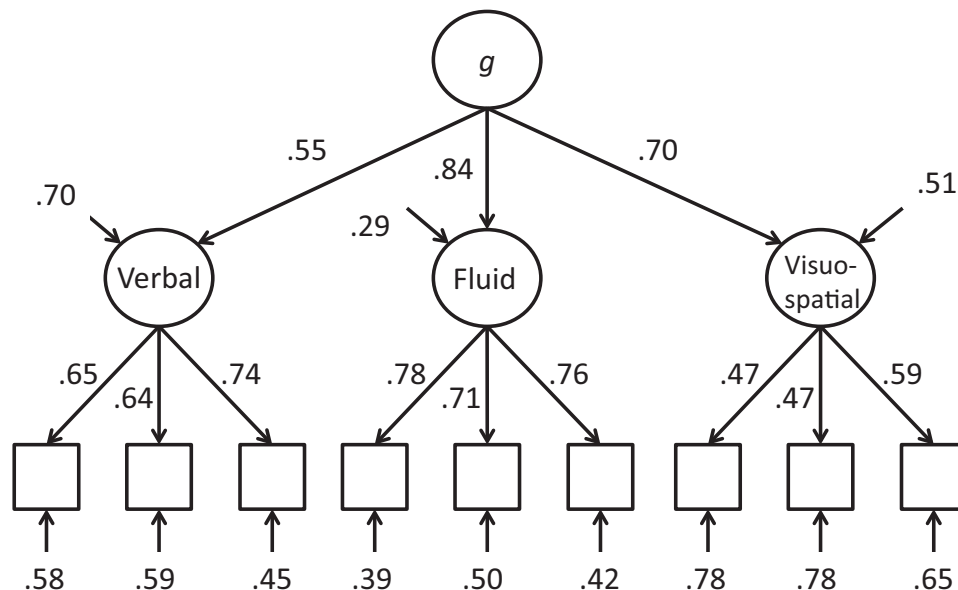

Figure 1. Showing a typical outcome of a simulation with a model that is in line with process overlap theory as stated verbally and as presented as structural model: The latent variable g (the sum of all fluid processing capacities) predicts the latent variables Verbal, Fluid, and Visuospatial Intelligence.

an emergent property" and "translates to a formative model with regard to the general factor" (p. 162), we argue that POT does not *necessarily* do so. From the simulations with the sampling model just cited, which is completely consistent with POV-V and POS-S, we can conclude that the general factor is not so much a variable *constructed out of* the verbal, visuospatial, and fluid factor but rather *is* the fluid factor, which Kovacs and Conway consider to be reflective. In the structural model as just depicted, the factors fluid and general intelligence both represent an (unbiased estimate of the) sum of the executive functioning capacities: any imperfect relation between the two is (literally) due to sampling error. The more complex items or subtests are, the more bonds will be sampled, the smaller the sampling error, the more evidence the fluid and general factor are one and the same variable (the total of the executive functioning capacities). Moreover, because the effects of the individual executive-functioning capacities are purely additive, the underlying factor that explains across-domain correlations between diverse mental tests can simply be interpreted as "total executive functioning capacity."

We conclude that a key element from POT, the *bottleneck*, (somehow) needs to be incorporated in POT-S, because according to path diagrammatical conventions, the performance on the task would be estimated as a weighted sum of the underlying variables. Other than viewing these variables from different levels of analyses, there would not be much difference between POT-S and *g*-theory. In the former, the analysis is on the level of the many individual capacities, which add up to a small number of total capacities, whereas in the former the analysis is on the level of the relatively small number of total capacities, which are all composed of a large number of smaller capacities. Yet one may then distinguish between different types of *g*-theories: In the one *g*-theory, *g* may indeed be a *sum* of multiple capacities that may act as whole or as a fraction thereof (like the force that a pound of marbles or a fraction of these marbles can exert), whereas in the other, *g* *consists* of these multiple capacities and always acts as a whole (like the force that a single marble weighing pound can exert).[1]

We acknowledge of course that POT can be formalized differently, but our contention is that POT needs to be precise and formalized in such a way that the key phenomena can be derived, for example, by simulation or analytical proof. Whatever form it will take, it should make the crucial distinction with *g*-theory (of the first kind). In our view, the most promising candidate of POV is therefore the proposed IRT model (POT-I). To be able to fully separate POT from *g*-theory, POT-I should show that the appropriate interpretation of the general factor is (a) a variable distinct from the fluid factor and (b) of the formative kind.

Multidimensional IRT Model (POT-I)

The interpretation of *g* as a summary variable stems from arguments given by proponents of sampling models. Following these arguments, not only *g* but also the verbal, visuospatial, and fluid factor should be regarded as summary variables (formative). However, as Kovacs and Conway (this issue) consider those latter three as reflective (see POT-S in their Figure 8), the reflective interpretation of *g* may also still be defensible, at least in certain specifications. Kovacs and Conway did not provide their readers with simulations that could further illustrate the claim that *g* must indeed be regarded as a formative in POT-I, or specify precisely how the general and fluid factor are different. Hence, we conclude that the authors still need to provide more formal evidence for this aspect of POT. Note that we do not mean to say the new theory is *invalid* but merely that certain assumptions of POT-I may have crucial consequences for the interpretation of the model. For instance, a novel feature of POT-I is the choice to let general executive-processing capacities be noncompensatory (multiplicative in the equation). According to the authors, this property leads to the crucial bottleneck feature of POT-I. Yet this leaves open the choice for the nature of the domain-specific processes capacities. Why are these, in contrast to the general executive-processing capacities, compensatory (additive in the equation)? In addition, although the choice of general processes capacities as being noncompensatory was based on empirical findings that are in favor of this choice, it is in principle possible to adduce evidence that argues for the idea that it is *general* processes capacities are compensatory. It may even be possible to argue for the opposite assumption, namely that the domain-specific capacities should be taken as multiplicative and domain-general processes as additive.

First note that POT-I pertains only to domain-specific tasks, in which both domain-specific and domain-general capacities are important, and not to purely fluid tasks, as in the latter domain-specific processes play no role. Second, domain-specific tasks are often crystallized tasks, meaning that they rely on acquired knowledge and abilities that are essential to solve the task. If one does not know certain facts (the capital of Spain) or certain words ("curriculum") when answering items of a knowledge or vocabulary test, this cannot be compensated with domain-general processes or other domain-specific processes (such as arithmetic knowledge). This also true for "real-life" crystalized tasks. We take chess as an example. If one does not know the rules of chess, one can simply not play chess. In addition, whereas differences in general intelligence explain some part of the variance in chess playing, more variance is explained by differences in chess expertise, such as differences in hours of serious practice (Grabner, Stern, & Neubauer, 2007).

Of course, without any working memory and other domain-general processes we probably are unable to do arithmetic, play chess, or take a vocabulary test. But this case is less realistic within the normal population. In addition, some experts are able to display amazing levels of performance in spite of lack of access to domain-general processes. To stay with chess, think of blitz chess, blindfolded chess, or more prosaic of a very drunk chess grandmaster who easily beats amateur chess players while

---

[1]This is an important distinction, because only the latter kind of *g*-theory would provide an explanation of the Jensen-effect (the relation between *g*-loading and heritability), for instance. In the former kind, the heritability of the observed scores is the average of the heritability of the sampled capacities. In principle, *g*-loading and heritability are then unrelated. The POT-sampling model as just formulated would fall within this category and will thus not provide an account for the Jensen-effect, unless perhaps additional assumptions are included.

discussing politics with the public as a double task in a crowded, noisy chess cafe.

We thus call for an investigation of the (possibly competing) properties and predictions of alternative POT-IRT models.

POT and Major Findings

According to Kovacs and Conway (this issue), the integrated theory explains several major findings, including ability differentiation and the law of worst performance (not evaluated here). However, it leaves open how other important findings that are considered to differentiate between theories of intelligence should be explained. Although it is not necessarily a criticism of their model that it cannot explain every empirical fact (to the best of our knowledge, no model can), it is still worth considering these findings in detail. Ultimately, they should be captured in a comprehensive model of cognitive abilities. Our discussion that follows can therefore be seen as much as a criticism of Kovacs and Conway as of virtually all other models, and as such is best seen as an appeal to expand POT (or any other theory) to accommodate outstanding challenges.

First and foremost, the notable omission of the subscript *t* in a model of intelligence means that at least three important phenomena cannot (yet) be accounted for: (a) Cognitive performance increases early on in life and declines in old age, and in different paces for different cognitive abilities (e.g., Baltes & Lindenberger, 1997; Horn & Cattell, 1967; Swagerman et al., 2016); (b) the (possibly related) effects called age differentiation and integration (for a review, see Tucker-Drob, 2009), which denote the varying proportion of variance explained by the general factor of intelligence across age (rather than across level of ability); and (c) the increase of the heritability of intelligence throughout development (Haworth et al., 2010; Trzaskowski, Yang, Visscher, & Plomin, 2014). In the literature, one can find hypotheses that can account for those effects. We propose these can be incorporated in POT.

POT-PLUS

POT already does an admirable job in bringing together various strands of evidence and is undoubtedly a considerable step forward in the challenge of developing an integrated model of general cognitive abilities. However, there are also several central outstanding questions that remain for POT or any successor. Inspired by POT, we next describe what we consider main remaining challenges for any comprehensive theory of intelligence. They may provide an initial outline toward how these may be tackled by (versions of) POT or new models.

Test Sampling

Kovacs and Conway (this issue) borrow Baddeley's architecture of a multicomponent working memory and the idea that these components are each limited by their own (total) capacity, thereby causing individual differences in cognitive-processing performance. We would agree with the idea that tests may sample from multiple of those capacities. That is, we believe in the possibility that any two tests or test items may tap from different cognitive processes. We denote this idea *test sampling*.

However, we also believe that psychometricians aim to construct psychometric tests such that the overlap is as small as possible. In the end, test sampling in additive models should reveal itself through the presence of cross-loadings in factor models of intelligence. A good psychometric instrument will minimize these cross-loadings, such that a correlated first-order factor model or hierarchical model is tenable. Because of the simplicity of a hierarchical factor, this model may be preferred over the bifactor model, in which it is nested by imposing proportionality constraints; in the realist interpretation of the hierarchical model this is due to mediating roles of the lower order factors (for discussion, see, e.g., Gignac & Watkins, 2013; Hood, 2008). However, the larger the sample size, the more power to detect imperfections, hence the more likely the hierarchical will be rejected and the bifactor is the preferred model, statistically speaking. A challenge for POT-I, as it is not an additive sampling model, is to show if or in what situations POT-I predicts good fit for the hierarchical model and in what situations for the bifactor model.

As POT explains the positive manifold and the factorial structure of intelligence as resulting from test sampling, it would follow naturally that changes in the positive manifold and factor structure would reflect changes in test sampling. However, due to the omission of subscript *t*, this actually remains an open question. Age integration, differentiation and de-differentiation effects (Deary et al., 1996, 2004; Juan-Espinosa et al., 2002; Tucker-Drob, 2009) are thus left unexplained. One might argue that the empirical evidence for such effect is mixed, and thus inconclusive or difficult to interpret (Tucker-Drob, 2009), yet the subject must be taken seriously, as they may relate to the Flynn-effect, for which Kovacs and Conway (this issue) do aim to provide an account in terms of differentiation. This account boils down to a second way of sampling (which also is not clear from POT-S). Apart from the idea that subtests or items sample, Kovacs and Conway implement the idea of individual differences in the sampling procedure, which we may denote as *individual sampling*.

In the additive POT-sampling model we just specified, one could implement the idea of individual sampling by introducing a subscript for the individual concerning the chances the underlying capacities are samples, so that the model would read

$$fluid_i = \sum_{k=1}^{ne} b_{ik} E_{ik}$$

$$verbal_i = \sum_{l=1}^{nv} b_{il} V_{il} + \sum_{k=1}^{ne} c_{ik} E_{ik}$$

$$visuospatial_i = \sum_{m=1}^{ns} b_{im} S_{im} + \sum_{k=1}^{ne} d_{ik} E_{ik}.$$

This different way of sampling may also be interesting in the light of research into the relation between fluid intelligence and working memory capacity. Strong relations between the two constructs have been found (e.g., Ackerman, Beier, & Boyle, 2005), but overall findings are mixed again and inconclusive in order to provide a definitive answer to the question of whether

the two constructs are the same. The work of Chuderski (2013), however, may provide a reason for these mixed results; when individuals are under pressuring circumstances, the two constructs become identical, while under less demanding circumstances they are not. As individual sampling suggests that individuals with low levels of intelligence have lower levels of any of the total capacities and need to recruit more of their capacities in order to solve a problem, one might hypothesize that, especially under time pressure, individuals with a low total central executive capacity need to recruit more of central executive capacities as compared to individuals with a high total central executive capacity; under less demanding circumstances these sampling differences may be smaller. Again in additive sampling models like the aforementioned, differences between the constructs can be explained relatively easily, namely, as the result of "sampling error": The variables both represent an estimate of total of executive functioning capacity, but relatively small samples of bonds yield relatively small overlap and thus lower correlations. A challenge for Kovacs and Conway would be to show if this identity also occurs in their IRT model.

Genetics

POT does not make any claims regarding the heritability of the cognitive abilities, their underlying capacities, hence general intelligence. One simple explanation is that as each of the underlying variables are to some extent heritable, their sum is also heritable. However, in itself this will not provide an account for the relation between factor loading and heritability, thus for the way the Jensen-effect arises. We encourage proponents of sampling theory to develop such hypotheses. We believe this should be possible, as the genetic literature also captures the idea of sampling, which is central to POT. One can distinguish again between theories that assume genetic determinants (genetic variants or genetic mutations) cognitive processing have general effects ("generalist genes"; Kovas & Plomin, 2006) and theories that assume what we may call *genetic sampling*, by which we mean that any two cognitive-processing capacities always share some of their genetic determinants but that there are no determinants that influence all cognitive processes (Anderson, 2001; Cannon & Keller, 2006; Penke, Denissen, & Miller, 2007). Both mechanisms will lead to genetic correlations between the underlying capacities, whereas in the original POT theory these are unrelated. The question becomes what implications such genetic correlations may have for POT. Does POT need to assume the absence of any shared genetic effects, that is, the absence of pleiotropy for which there is ample empirical evidence (Trzaskowski, Shakeshaft, & Plomin, 2013)?

Other behavioral genetic challenges for POT are to explain why heritability of intelligence is higher in adults than in children (Haworth et al, 2010), why genetic stability increases (Deary et al., 2012), why over development genetic variance can be described by a single latent factor (Deary et al., 2012), and why genetic correlations among the various abilities appear to increase (Hoekstra, Bartels, & Boomsma, 2007). Of these findings, the first may be the easiest to account for: In standard genetic models, genotype–environment correlation contributes to heritability, so increase in genotype–environment

correlation, as proposed by Scarr and McCartney (1983), will therefore result in an increase of estimated heritability. In the model proposed by Dickens (2008), such relation between genotype and environment will result in increasing genetic stability and genetic correlations among the different cognitive abilities. To disentangle such explanations, it would be crucial to determine whether POT assumes the absence of any shared genetic effects, as implied by the assumption that the underlying capacities are independent.

Development

There is increasing empirical evidence for the presence of mutual beneficial interactions between cognitive abilities during their development. One question needs to be answered: Are such interactions also present in POT's architecture, for instance, among the multiple components in Baddeley's working memory model? If such interactions exist, they will result in stronger correlations between measures of cognitive performance as compared to the correlations between their underlying limiting capacities (van der Maas et al., 2006). Similarly, cognitive abilities have mutual beneficial relationships with educational attainment. As educational institutions provide training in many cognitive skills simultaneously, educational attainment also increases positive correlations among these skills.

The missing role of education reveals other challenges for POT. POT, as well as many other theories of intelligence, explains individual differences in cognitive-processing capacities but not how these may lead to individual differences in their outcomes, namely, knowledge and skills (often denoted "crystallized intelligence"). Cattell's investment theory of fluid and crystallized intelligence might be considered an important exception, yet this theory clearly falls within the *g*-theoretical framework. In those theories, as well as POT, *g*-loadings of fluid tests are a function of complexity (the more complex a test, the more *g*-loaded). Yet crystallized knowledge tests, which are themselves noncomplex, demonstrate high *g*-loadings as well (and often the highest, e.g., Kan, Wicherts, Dolan, & van der Maas, 2013). The relation between complexity and *g*-loading is thus not one-to-one. The relation between *g*-loading and test content may be better characterized as being a function of cultural load (indicating the subtests' dependency on individual differences in prior knowledge). That is, the more individual differences in successful task completion depend on individual differences in cultural dependent knowledge, the higher the tasks' loading on the general factor of intelligence. The finding becomes even more puzzling because the larger the role of culturally dependent knowledge, the higher the heritability of individual differences in performance. Ideally, a new theory of intelligence, hence POT, should also account for this (rather paradoxical) finding.

Neuroscience

A final open question is how to reconcile converging insights from (cognitive) neuroscience with POT. In terms of existing evidence, it is clear that POT represents a considerable step forward in this regard compared to traditional *g* theories as a

single neural property or dimension is likely not fruitful. The empirical evidence is rapidly converging on the conclusion that intelligence is best seen as determined by the (weighted) sum of many neural properties, rather than as some underlying "neuro g" (Haier et al., 2009). This conclusion has been supported across multiple cohorts and neuroimaging metrics, showing how gray and white matter play complementary roles in supporting (fluid) intelligence (Kievit et al., 2014; Kievit et al., 2012) and how cortical, subcortical, and even different metrics of white matter determine fluid intelligence in old age (Ritchie, Booth, et al., 2015). This neuroimaging evidence further supports the hypothesis central to POT that g is best seen as a (formative) summary of lower levels, both cognitive and neural, rather than a single underlying entity. In short, POT naturally accommodates the emerging consensus in neuroimaging that higher cognition depends on a broad and partially complementary set of low-level determinants.

However, other findings may be more challenging to reconcile. First, emerging work suggests that the canonical role of the dorsolateral prefrontal cortex—(dl)PFC—that of actively maintaining representations by means of continuous (spiking) activity, is likely an oversimplification: Working memory representations can, in principle, be maintained even in the absence of continuous activity (Stokes, 2015). More worryingly, the canonical explanation of the role of the (dl)PFC is likely incomplete: A recent study of a nine patients with considerable (dl)PFC lesions (Mackey, Devinsky, Doyle, Meager, & Curtis, 2016) showed a surprising lack of cognitive sequelae, both in terms of spatial working memory and general cognitive function (both were largely preserved). Although neither of these are direct threats to POT, it does suggest that our ability to translate our psychometric, structural representations into precise underlying neural mechanisms is still limited. It seems likely that executive processes that are at the heart of POT comprise a complex set of cognitive processes, including but not limited to maintenance of interim representations, metacognition, inhibition, and set-shifting, all of which are likely operating partially simultaneously and dependent on overlapping neural systems.

To truly get at the heart of the neural processes underlying executive processes and their relation to general intelligence, we reiterate the importance of the subscript t; in both the short term (intraindividual task-related processing) and long term (developmental timescales). One of the strengths of the POT model compared to g-theory is that it simultaneously bears upon interindividual differences as well as intraindividual processes. In one way, POT can be seen as a process model for different contributions of executive and low-level abilities when performing a given task. It should be possible, in principle, to separate these contributions in time (response duration and activation across a trial) and space (across the cortex). By decomposing trial-level activity across the cortex, neuroimaging techniques offer the promise of testing process level, intraindividual theories of cognition. Recent work provides a proof of principle in terms of spatial activity, using an IRT showing how intraindividual processes differ even when conditioned on interindividual difference in fluid intelligence (Kievit, Scholte, Waldorp, & Borsboom, 2016), illustrating how neuroimaging can be used to go beyond well-fitting behavioral models.

Moreover, if POT is true, we would expect that it may be possible to selectively disrupt or even temporarily improve cognitive abilities that form POT. Initial evidence suggests this may be possible, with TMS-based disruption of prefrontal activity disrupting visual-spatial memory (Costa et al., 2013), whereas prefrontal stimulation (γ-tACS) shows task and frequency-specific improvement of fluid reasoning tasks (Santarnecchi et al., 2013). Although these findings are far from settled, they show how we may, in principle, be able to utilize neuroscience to test specific aspects of POT and related theories and separate the hypothesized interactions between executive, visuospatial, and verbal processes over time during task performance in such a way that it can be predicted or derived from the model.

Arguably the biggest challenge remaining for both behavior only and neuroscientific inquiry is developmental change. An influential study showed that cohort differences in cognitive abilities (low, middle, high IQ) were associated with distinct patterns of neural maturation or rates of change (Shaw et al., 2006), further illustrating the fact that one-slice cross-sectional samples likely omit the key features that underlie the phenomenon of interest. Most promising in this regard are longitudinal psychometric investigations of concurrent changes in cognition and brain structure. These allow one to investigate whether changes in cortical structure precede changes in cognitive ability (compatible with a causal view of brain structure), whether changes in neural structure are the consequence of improving cognition (a plasticity-based view), and whether both are dependent on some other (e.g., genetic) cause or uncorrelated. Recent work in older adults shows the promise of these approaches, with studies showing greater white matter health predicts less decline in processing speed in older adults (Ritchie, Bastin, et al., 2015), whereas another sample suggested greater baseline gray matter volumes were associated with greater gains in fluid intelligence (Persson et al., 2016).

Conclusion

POT represents an ambitious step forward in our understanding of, and thinking about, the structure of general cognitive abilities. Like all other theories of intelligence, key empirical phenomena cannot yet be captured. By further formalizing and extending POT, it may very well be possible to do so in the future. This endeavor is increasingly feasible with the advent of large, multimodal, publically available data sets. Ultimately, our hope is that the intelligence field moves toward the integration of formalized models of inter- and intraindividual differences, such as POT and the Q diffusion model (van der Maas, Molenaar, Maris, Kievit, & Borsboom, 2011), together with genetic and neuroimaging data over developmental timescales. Only then will we be able to tease apart the interplay between inter- and intraindividual processes and make further steps in unraveling "the well-aged puzzle of g."
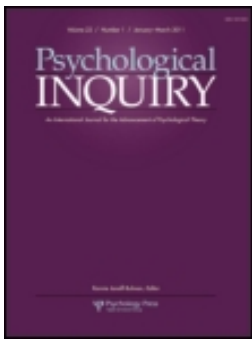
Funding

References

Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2005). Working memory and intelligence: The same or different constructs? *Psychological Bulletin*, *131*(1), 30–60.

Anderson, B. (2001). g as a consequence of shared genes. *Intelligence*, *29*, 367–371.

Baddeley, A. D. (1992). Working memory. *Science*, *255*, 556–559.

Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, *4*, 417–423.

Baddeley, A. D., & Hitch, G. (1974). Working memory. *The Psychology of Learning and Motivation*, *8*, 47–89.

Baltes, P. B., & Lindenberger, U. (1997). Emergence of a powerful connection between sensory and cognitive functions across the adult life span: A new window to the study of cognitive aging? *Psychology and Aging*, *12*, 12.

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.

Cannon, T. D., & Keller, M. C. (2006). Endophenotypes in the genetic analyses of mental disorders. *Annual Review of Clinical Psychology*, *2*, 267–290.

Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology*, *54*(1), 1–22.

Chuderski, A. (2013). When are fluid intelligence and working memory isomorphic and when are they not? *Intelligence*, *41*, 244–262.

Costa, A., Oliveri, M., Barban, F., Bonnì, S., Koch, G., Caltagirone, C., & Carlesimo, G. A. (2013). The right frontopolar cortex is involved in visual-spatial prospective memory. *Plos One*, *8*(2), e56039.

Deary, I. J., Egan, V., Gibson, G. J., Austin, E. J., Brand, C. R., & Kellaghan, T. (1996). Intelligence and the differentiation hypothesis. *Intelligence*, *23*(2), 105–132.

Deary, I. J., Whiteman, M. C., Starr, J. M., Whalley, L. J., & Fox, H. C. (2004). The impact of childhood intelligence on later life: Following up the Scottish Mental Surveys of 1932 and 1947. *Journal of Personality and Social Psychology*, *86*(1), 130–147.

Deary, I. J., Yang, J., Davies, G., Harris, S. E., Tenesa, A., Liewald, D., ... & Redmond, P. (2012). Genetic contributions to stability and change in intelligence from childhood to old age. *Nature*, *482*(7384), 212–215.

Detterman, D. K., & Daniel, M. H. (1989). Correlations of mental tests with each other and with cognitive variables are highest for low IQ groups. *Intelligence*, *13*, 349–359.

Dickens, W. T. (2008). *What is g? Paper presented at the annual meeting of the Behavior Genetics Association*, Louisville, KY.

Gignac, G. E., & Watkins, M. W. (2013). Bifactor modeling and the estimation of model-based reliability in the WAIS-IV. *Multivariate Behavioral Research*, *48*, 639–662.

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

Grabner, R. H., Stern, E., & Neubauer, A. C. (2007). Individual differences in chess expertise: A psychometric investigation. *Acta Psychologica*, *124*, 398–420.

Gustafsson, J.-E. (1984). A unifying model for the structure of intellectual abilities. *Intelligence*, *8*, 179–203.

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*, 136–144.

Haworth, C., Wright, M., Luciano, M., Martin, N., de Geus, E., van Beijsterveldt, C., ... Davis, O. (2010). The heritability of general cognitive ability increases linearly from childhood to young adulthood. *Molecular Psychiatry*, *15*, 1112–1120.

Hoekstra, R. A., Bartels, M., & Boomsma, D. I. (2007). Longitudinal genetic study of verbal and nonverbal IQ from early childhood to young adulthood. *Learning and Individual Differences*, *17*, 97–114.

Hood, S. B. (2008). *Latent variable realism in psychometrics.* Unpublished doctoral dissertation, Indiana University.

Horn, J. L., & Cattell, R. B. (1967). Age differences in fluid and crystallized intelligence. *Acta Psychologica*, *26*, 107–129.

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

Juan-Espinosa, M., Garcıa, L. F., Escorial, S., Rebollo, I., Colom, R., & Abad, F. J. (2002). Age dedifferentiation hypothesis: Evidence from the WAIS III. *Intelligence*, *30*, 395–408.

Kan, K.-J., Kievit, R. A., Dolan, C., & van der Maas, H. (2011). On the interpretation of the CHC factor Gc. *Intelligence*, *39*, 292–302.

Kan, K. J., Wicherts, J. M., Dolan, C. V., & van der Maas, H. L. (2013). On the nature and nurture of intelligence and specific cognitive abilities the more heritable, the more culture dependent. *Psychological Science*, *24*, 2420–2428.

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*, Article 5658.

Kievit, R. A., Scholte, H. S., Waldorp, L. J., & Borsboom, D. (2016). Inter- and intra-individual differences in fluid reasoning show distinct cortical responses. *bioRxiv*, 039412. http://dx.doi.org/10.1101/039412

Kievit, R. A., van Rooijen, H., Wicherts, J. M., Waldorp, L. J., Kan, K.-J., Scholte, H. S., & Borsboom, D. (2012). Intelligence and the brain: A model-based approach. *Cognitive Neuroscience*, *3*, 89–97.

Kovas, Y., & Plomin, R. (2006). Generalist genes: Implications for the cognitive sciences. *Trends in Cognitive Sciences*, *10*, 198–203.

Kvist, A. V., & Gustafsson, J.-E. (2008). The relation between fluid intelligence and the general factor as a function of cultural background: A test of Cattell's investment theory. *Intelligence*, *36*, 422–436.

Larson, G. E., & Alderton, D. L. (1990). Reaction time variability and intelligence: A "worst performance" analysis of individual differences. *Intelligence*, *14*, 309–325.

Mackey, W. E., Devinsky, O., Doyle, W. K., Meager, M. R., & Curtis, C. E. (2016). Human dorsolateral prefrontal cortex is not necessary for spatial working memory. *The Journal of Neuroscience*, *36*, 2847–2856.

Molenaar, D., Dolan, C. V., Wicherts, J. M., & van der Maas, H. L. (2010). Modeling differentiation of cognitive abilities within the higher-order factor model using moderated factor analysis. *Intelligence*, *38*, 611–624.

Penke, L., Denissen, J. J., & Miller, G. F. (2007). The evolutionary genetics of personality. *European Journal of Personality*, *21*, 549–587.

Persson, N., Ghisletta, P., Dahle, C. L., Bender, A. R., Yang, Y., Yuan, P., ... Raz, N. (2016). Regional brain shrinkage and change in cognitive performance over two years: The bidirectional influences of the brain and cognitive reserve factors. *NeuroImage*, *126*, 15–26.

Ritchie, S. J., Bastin, M. E., Tucker-Drob, E. M., Maniega, S. M., Engelhardt, L. E., Cox, S. R., ... Pattie, A. (2015). Coupled changes in brain white matter microstructure and fluid intelligence in later life. *The Journal of Neuroscience*, *35*, 8672–8682.

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

Rushton, J. P. (1998). The "Jensen effect" and the "Spearman-Jensen hypothesis" of Black-White IQ differences. *Intelligence*, *26*, 217–225.

Rushton, J. P., & Jensen, A. R. (2010). The rise and fall of the Flynn effect as a reason to expect a narrowing of the Black–White IQ gap. *Intelligence*, *38*, 213–219.

Santarnecchi, E., Polizzotto, N. R., Godone, M., Giovannelli, F., Feurra, M., Matzen, L., ... Rossi, S. (2013). Frequency-dependent enhancement of fluid intelligence induced by transcranial oscillatory potentials. *Current Biology*, *23*, 1449–1453.

Scarr, S., & McCartney, K. (1983). How people make their own environments: A theory of genotype→ environment effects. *Child Development*, 424–435.

Shaw, P., Greenstein, D., Lerch, J., Clasen, L., Lenroot, R., Gogtay, N., ... Giedd, J. (2006). Intellectual ability and cortical development in children and adolescents. *Nature*, *440*, 676–679.

Spearman, C. (1904). "General Intelligence," objectively determined and measured. *The American Journal of Psychology*, *15*, 201–292.

Spearman, C. (1924). The nature of "intelligence" and the principles of cognition. *Journal of Philosophy, 21*(11), 294–301.

Stokes, M. G. (2015). "Activity-silent" working memory in prefrontal cortex: A dynamic coding framework. *Trends in Cognitive Sciences*, *19*, 394–405.

Swagerman, S. C., de Geus, E. J., Kan, K.-J., van Bergen, E., Nieuwboer, H. A., Koenis, M. M., ... Boomsma, D. I. (2016). The Computerized

Neurocognitive Battery: Validation, aging effects, and heritability across cognitive domains. *Neuropsychology*, *30*(1), 53–64.

Thomson, G. H. (1946). *The factorial analysis of human ability* (2nd rev). London: University of London Press.

Trzaskowski, M., Shakeshaft, N. G., & Plomin, R. (2013). Intelligence indexes generalist genes for cognitive abilities. *Intelligence*, *41*, 560–565.

Trzaskowski, M., Yang, J., Visscher, P. M., & Plomin, R. (2014). DNA evidence for strong genetic stability and increasing heritability of intelligence from age 7 to 12. *Molecular Psychiatry*, *19*, 380–384.

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

van der Maas, H. L. J., Dolan, C. V., Grasman, R. P., Wicherts, J. M., Huizenga, H. M., & Raijmakers, M. E. (2006). A dynamical model of general intelligence: The positive manifold of intelligence by mutualism. *Psychological Review*, *113*, 842–861.

van der Maas, H. L. J., Kan, K., Hofman, A., & Raijmakers, M. (2014). Dynamics of development: A complex systems approach. In P. C. M. Molenaar, R. M. Lerner, & K. M. Newell(Eds.), *Handbook of developmental systems theory and methodology* (pp. 270–286). New York, NY: Guilford Press.

van der Maas, H. L. J., Molenaar, D., Maris, G., Kievit, R. A., & Borsboom, D. (2011). Cognitive psychology meets psychometric theory: on the relation between process models for decision making and latent variable models for individual differences. *Psychological Review*, *118*(2), 339–356.

van der Sluis, S., Kan, K.-J., & Dolan, C. V. (2010). Consequences of a network view for genetic association studies. *Behavioral and Brain Sciences*, *33*, 173–174.

Commentary on Kovacs and Conway, Process Overlap Theory: A Unified Account of the General Factor of Intelligence

Scott Barry Kaufman

Published online: 02 Aug 2016.

Submit your article to this journal ⬀

View related articles ⬀

View Crossmark data ⬀

Routledge
Taylor & Francis Group

Commentary on Kovacs and Conway, Process Overlap Theory: A Unified Account of the General Factor of Intelligence

Scott Barry Kaufman

The Imagination Institute and Positive Psychology Center, University of Pennsylvania, Philadelphia, Pennsylvania

I appreciate the opportunity to respond to Kristof Kovacs and Andrew Conway's (this issue) target article. As they note, some aspects of their theory are not new. For instance, the strong connection between general cognitive ability and working memory has a lot of support in the literature and is well known by this point. Indeed both researchers have contributed significantly to this literature. However, I'd like to focus on a more novel aspect of their theory, and point out some practical implications.

As they point out, the positive manifold is a well-replicated finding. What still lacks consensus, however, is the explanation for this positive manifold. Their idea that g is an emergent property (not the cause) of multiple domain-general executive functions is a novel way of looking at the g factor. But to me, the most interesting puzzles they've helped to shed light on are (a) the law of diminishing returns, and (b) the finding that the worst performance on a cognitive test battery is a better predictor of the g factor than the best performance. The cause of these two findings has never been satisfactorily explained. Their solution is reasonable: Individual differences in executive processes can serve as a bottleneck for cognitive functioning across a wide range of tasks.

Practically speaking, this solution suggests that it may be more difficult for individuals with executive functioning deficits to showcase their intellectual capabilities. Chuderski (2013) reviewed 26 studies that administered a measure of working memory and a measure of fluid reasoning and found that the studies that increased the time pressure of the fluid reasoning task significantly increased the correlation between working memory and fluid reasoning. In a follow-up experiment, Chuderski found that when participants were required to complete a test of fluid reasoning in 20 min, working memory explained all of the variation in fluid reasoning, whereas when participants were given 60 min to complete a measure of fluid reasoning, working memory accounted for only 38% of the variation in fluid reasoning. This is a big difference! These findings are consistent with other research showing that the processes involved in fast and slow responses can be differentiated (Partchev & De Boeck, 2012). Future iterations of the process overlap theory should address the importance of changes in test administration (e.g., timing) on their theory.

Chuderski also found that a measure of relational learning—that assessed the ability to learn from prior relations to increase efficiency of future processing of relations—predicted variation

in fluid reasoning above and beyond the effects of working memory. Taken together, the implication is that tests that relax the demands on executive functioning may give those with executive functioning difficulties more of a chance to bring to bear other cognitive processes—such as relational learning or associative learning (see Kaufman, DeYoung, Gray, Brown, & Mackintosh, 2009)—that may allow them a chance to perform well on complex tests of cognitive ability.

This is a real issue in the learning disability literature. Various learning disabilities, such as dyslexia and attention deficit hyperactivity disorder (ADHD), are accompanied by deficits in executive functioning. For instance, people who exhibit ADHD-like symptoms tend to score lower on tests of working memory (see Kolger, Rapport, Bolden, Sarver, & Raiker, 2010). However, in one recent study, Fugate, Zentall, and Gentry (2013) studied a sample of academically advanced students who either scored in the 90th percentile or above on a standardized test or had a grade point average of 3.5 or greater in a specific academic domain. Students with ADHD characteristics such as "inattention" scored lower in working memory than the students who did not display ADHD characteristics, *even though the groups did not differ in fluid reasoning ability*. How would the process overlap model explain these findings? I think if the model is going to be comprehensive, it needs to explain how it is possible for those with executive functioning difficulties to still be highly intelligent.

The explanation has important implications for how we recognize intelligence in students with extreme scatter in their cognitive profiles. Due to their area of disability, students with learning disabilities tend to score much higher in one cognitive area compared to others. Various researchers are attempting to develop methods for eliminating the attenuating influences of cognitive-processing deficits on an estimate of a child's general cognitive ability (Flanagan, Ortiz, & Alfonso, 2013). It would be great if the process overlap model could help inform real selection decisions that influence the course of a child's future education.

There are also implications for students on the higher end of the IQ spectrum. Multiple studies support the idea that intellectually precocious youth show "multipotentiality"—they tend to show more extreme discrepancies in their cognitive profiles compared to students with average cognitive ability (Achter, Benbow, & Lubinski, 1997; Lohman, Gambrell, & Lakin, 2008). This result suggests that for those with high general cognitive
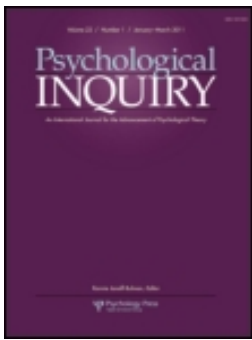
ability, their *g* factor scores may mask their particular specific cognitive strengths. I'd like to see future iterations of the process overlap theory further explain the meaning of the general cognitive ability factor among those on the highest end of the spectrum. What's the difference in the cognitive mechanisms that give rise to general cognitive ability among those with an IQ greater than 160 compared to an IQ of 130, for instance?

Finally, it would be great to see how the process overlap theory relates to creativity. Fugate et al. (2013) found that the lower the working memory scores among their population of high-achieving students, the *higher* their creativity. Clearly, creative cognition is a form of intelligence. But it's a form of intelligence that doesn't necessarily always benefit from domain-general executive functions. An interesting future line of research would be to investigate interactions between the executive control network and other networks in the brain. One recent study found that communication between the executive network and the default network contributed to idea generation (Beaty et al., 2015). However, the time course of the task also mattered. The executive network was much more important for later stage processing than early stage processing.

I look forward to seeing how process overlap theory develops and how it makes connections with other areas of psychology, such as educational psychology and creativity research. There is a lot of potential for integration.

References

Achter, J. A, Benbow, C. P., & Lubinski, D. (1997). Rethinking multipotentiality among the intellectually gifted: A critical review and recommendations. *Gifted Child Quarterly*, *41*, 5–15.

Beaty, R. E., Benedek, M., Kaufman, S. B., & Silvia, P. J. (2015). Default and executive network coupling supports creative idea production. *Scientific Reports*, *5*, 10964.

Chuderski, A. (2013). When are fluid intelligence and working memory isomorphic and when are they not? *Intelligence*, *41*, 244–262.

Flanagan, D. P., Ortiz, S. O., & Alfonso, V. C. (2013). Pattern of strengths and weaknesses analyzer (PSW-A v1.0). In D. P. Flanagan, S. O. Ortiz, & V. C. Alfonso (Eds.), *Essentials of cross-battery assessment* (3rd ed., CD). Hoboken, NJ: Wiley.

Fugate, C. M., Zentall, S. S., & Gentry, M. (2013). Creativity and working memory in gifted students with and without characteristics of attention deficit hyperactive disorder. *Gifted Child Quarterly*, *57*, 234–246.

Kaufman, S. B., DeYoung, C. G., Gray, J. R., Brown, J., & Mackintosh, N. (2009). Associative learning predicts intelligence above and beyond working memory and processing speed. *Intelligence*, *37*, 374–382.

Kolfer, M. J., Rapport, M. D., Bolden, J., Sarver, D. E., & Raiker, J. S. (2010). ADHD and working memory: The impact of central executive deficits and exceeding storage/rehearsal capacity on observed inattentive behavior. *Journal of Abnormal Child Psychology*, *38*, 149–161.

Lohman, D. F., Gambrell, J., & Lakin, J. (2008). The commonality of extreme discrepancies in the ability profiles of academically gifted children. *Psychology Science Quarterly*, *50*, 269–282.

Partchev, I., & Boeck, P. D. (2012). Can fast and slow intelligence be differentiated? *Intelligence*, *40*, 23–32.

Parameters, Not Processes, Explain General Intelligence

Klaus Oberauer

Published online: 02 Aug 2016.

Submit your article to this journal ⤤

View related articles ⤤

View Crossmark data ⤤

Routledge
Taylor & Francis Group

Parameters, Not Processes, Explain General Intelligence

Klaus Oberauer

Department of Psychology–Cognitive Psychology Unit, University of Zurich, Zurich, Switzerland

Kovacs and Conway make a compelling argument for process overlap theory as an explanation of the positive manifold among cognitive tests. They make a strong case for their theory to be considered as a viable alternative to the dominant view of *g*. The general idea of explaining *g* as an emerging phenomenon rather than a common cause of performance is probably on the right track. At the same time, I am skeptical about some of the more specific assumptions. I will discuss two concerns that, so far, let me hesitate to fully endorse process overlap theory, and finish with an alternative version of the idea of *g* as a formative factor.

Are There General Processes?

The core assumption of process overlap theory is that every response to a cognitive test reflects a sample from a set of cognitive processes. Performance in two cognitive tests is positively correlated to the degree that their samples of processes overlap. Some processes are more general than others, so that they are sampled by a larger and more heterogeneous set of tests; these processes are primarily responsible for positive correlations even among very different kinds of tests.

What is a process? Literally, it is the transition between two states. In the context of process overlap theory, it is better defined as the skill or ability to produce a particular kind of transition between two states reliably. For instance, a person who has mastered arithmetic can reliably produce the transition between a representation of "2 + 3 = ?" to "2 + 3 = 5." In production system architectures of the mind, this concept of a process is embodied in productions: representations of condition–action links that, when the conditions are given (i.e., perceived or held in working memory), reliably produce the action linked to them (Anderson & Lebiere, 1998).

For process overlap theory to work, processes must have at least some degree of generality so that they are used in more than one specific trial of a cognitive test. For some processes, the case for generality can be made. The most compelling examples come from language processing. For instance, a speaker who can understand "Peter loves Mary" is likely to also understand "Jane loves John" and "Peter hates Mary." Observations of this kind have been summarized as demonstrating the systematicity of language and interpreted as showing the operation of abstract rules that incorporate our linguistic knowledge (Fodor & Pylyshyn, 1988). In this case, we can assume that

competent speakers of English have at their disposal a process of parsing (at least) simple subject–verb–object constructions that generalizes across all sentences of that form.

However, outside the domain of language, examples for systematicity are much harder to come by. For instance, mastery of elementary arithmetic relies only to a small degree on rules that generalize across all numbers, and much more on knowledge of individual addition, subtraction, and multiplication facts (Ashcraft, 1992; Zbrodoff, 1995). A person who has acquired the fact "2 + 3 = 5" does not necessarily know that "5 + 7 = 12." Can we assume that the processes of generating the result to the first and to the second problem are the same? In both cases, a skilled person arrives at the correct result by retrieving the relevant arithmetic fact from long-term memory. Therefore, one could argue that the same process—retrieval of arithmetic facts—underlies solution of all elementary arithmetic problems. We need to recognize, however, that conceptualizing all these processes as "retrieval of arithmetic facts" reflects an arbitrary choice of abstraction on our side. The same processes could equally be conceptualized on a more concrete level as "retrieval of 2 + 3 = 5" and "retrieval of 5 + 7 = 12," which would make them look like different processes. Of course, we could also conceptualize them on an even more abstract level by regarding them as "fact retrieval," which would suggest that retrieval of "2 + 3 = 5" and of "Paris is the capital of France" are instances of the same process.

The brain, I presume, does not care about the level of abstraction we choose to conceptualize its processes. Process overlap theory, therefore, must be formulated in a way that does not depend on them. We need a way to decide—at least conceptually—whether two processes are the same, in the sense that is relevant for process overlap theory, such that the decision is independent of how we choose to classify them. A reasonable criterion would be to call two processes the same if (and only if) their chances of success are necessarily highly correlated, such that if a person is able to reliably carry out one process, that person is able to reliably carry out the other. By necessarily highly correlated I mean that the success probabilities are not correlated for merely accidental reasons, such as the fact that they tend to be practiced together. Rather, whenever one process is improved (e.g., by practice) or impaired (e.g., by brain damage), the other always follows suit. This is the empirical criterion for systematicity that is met by some aspects of language processing but is less prevalent in other domains. I

call it the correlation criterion of process identity. The correlation criterion is reasonable because it matches exactly what is needed in process overlap theory: If, and only if, the success probabilities of two processes, 1 and 2, are highly correlated, then it is the case that when solving Test Task A requires Process 1, and Test Task B requires Process 2, the solution chances of Tests A and B are positively correlated. Hence, by regarding Processes 1 and 2 "the same" according to the correlation criterion, the core assumption of process overlap theory holds: Tasks A and B correlate because they use the same process.

The critical question for process overlap theory therefore is, Is there a sufficient number of cognitive processes that are reasonably general according to the correlation criterion? This is where I have doubts. Studies of skill acquisition have shown again and again that skills acquired through practice have a very narrow scope of generalization. Even within a narrow domain such as elementary arithmetic, practice with individual arithmetic facts improves selectively retrieval of those facts (Zbrodoff, 1995). More generally, skill acquisition has been shown to depend strongly on the acquisition of highly specific episodes, facts, and procedures (Rosenbaum, Carlson, & Gilmore, 2001), and there is little transfer from one domain of expertise to another.

The stubbornly narrow scope of transfer of practice poses a challenge for the assumption in process overlap theory that there are domain-general processes—in particular, executive processes—that are enrolled in a multitude of tasks across different content domains. Kovacs and Conway (this issue) assume that these processes play an important part in generating the positive manifold because their use in many different cognitive activities is responsible for positive correlations among the success rates of these activities. What is the evidence that such domain-general processes exist, and that they are critical for performance?

If domain-general processes of major importance for performance exist, we should expect them to benefit from practice. This notion underlies the optimistic perspective on cognitive training (Jaeggi, Buschkuehl, Jonides, & Perrig, 2008): The idea is that we can improve general intelligence through training of transferable cognitive processes. Repeated waves of optimism regarding this possibility have been met regularly with a sobering lack of robust and reproducible transfer of training on measures of general intelligence (Au et al., 2015; Melby-Lervåg & Hulme, 2016; Spitz, 1986; von Bastian & Oberauer, 2014). If people improve massively through training on a task with strong demands on executive functions, should we not expect strong transfer of training effects to other tasks also making heavy demands on executive functions?

There are two possible explanations for the weakness of general training benefits within process overlap theory. One is that we all, by using our general executive processes day in and day out, have practiced them to a near-asymptotic level that leaves very little room for improvement. This assumption begs the question why there are still large individual differences in the effectiveness of these processes. Perhaps people with low IQ have lower asymptotes? That raises the question of why some individuals have lower asymptotic efficiency of executive processes than others—whatever is responsible for individual differences in these asymptotes would be the root cause of

individual differences in general intelligence, and hence, of g. Such an explanation, I believe, would undercut the thrust of process overlap theory. The other possible explanation is that executive processes cannot be improved by practice. The training gains on trained tasks might be due entirely to increased efficiency in the more specific processes involved in these tasks. I have no strong objection to this possibility, but I note that it introduces a qualitative difference between specific processes, which are highly malleable through practice, and general process, which for some reason are impervious to practice.

Domain-General Executive Processes?

Kovacs and Conway (this issue) identify the family of executive processes as the domain-general processes that are primarily responsible for the positive manifold. This assumption is not necessary for process overlap theory, and I am skeptical about it.

The term of executive functions is used with a variety of meanings, with frustratingly little agreement among researchers. Some define it as the set of all cognitive mechanisms and processes that enable goal-directed behavior (Diamond, 2013; Welsh, Pennington, & Groisser, 1991). This broad definition comes down to saying that executive functions are all those mechanisms and processes that enable intelligent behavior. Others offer a list of more specific cognitive functions as a definition, which begs the question of what they have in common (Baddeley, 1996). Some working-memory researchers define executive processes in contrast to storage, implying that every process that transforms representations in working memory is an executive process (Smith & Jonides, 1999). Perhaps the best approximation to a useful definition is to identify executive processes with the set of processes involved in cognitive control, that is, processes that control other processes. My understanding is that Kovacs and Conway use the term in this sense.

The lack of a clear definition of executive processes renders the assumption that they underlie the positive manifold virtually untestable. In principle, this assumption could be tested by Jensen's method of correlated vectors: The g-loading across many test tasks should be positively correlated with their degree of dependence on executive processes. But how can we determine to what degree a task depends on executive processes if the definition of executive processes is exceedingly vague? This vagueness invites circular reasoning: We infer how strongly a task depends on executive processes from its correlation with g. I perceive instances of such circular reasoning when Kovacs and Conway (this issue) argue that all varieties of working-memory tasks that have shown high correlations with g place strong demands on executive functions. For instance, in visual change-detection tasks, participants briefly see an array of simple visual stimuli (e.g., colored squares) and about 1 s later see another array, and decide whether one element has been changed. I cannot imagine why this task should place particularly high demands on executive functions, whereas a digit span task does not.

One way to test the hypothesis of a close link between executive processes and g is to use task paradigms for which there is broad agreement that they involve executive processes and to

isolate the contribution of these processes to performance. There are many experimental paradigms that have been proposed for studying cognitive control, and many efforts have been made to isolate the contribution of cognitive control through difference scores between a condition with high and a condition with lower control demand. Positive correlations of such difference scores—such as the size of the Stroop effect or the flanker effect—with working memory capacity and fluid intelligence constitute the main evidence that Kovacs and Conway (this issue) cite in favor of the central role of executive processes. However, these correlations have mostly been very modest in size, and often zero, despite good reliability of the difference scores (Keye, Wilhelm, Oberauer, & van Ravenzwaaij, 2009; Oberauer, Süß, Wilhelm, & Sander, 2007; Wilhelm, Hildebrandt, & Oberauer, 2013).

Of the three executive-function factors that Miyake et al. (2000) identified, only one—updating of working memory—contributed unique variance in a regression model predicting fluid intelligence (Friedman et al., 2006). In this study (as in many others), updating of working memory was measured simply through the accuracy in a working-memory task involving updating. When the updating process is isolated from other factors contributing to success in a working-memory task, the updating component was found not to correlate at all with working-memory capacity (Ecker, Lewandowsky, Oberauer, & Chee, 2010). Kovacs and Conway (this issue) dismiss the idea that general mental speed explains *g* on the grounds that measures of speed account for only about 10% of the variance in *g*. My reading of the literature on the correlation between measures of executive functions and *g* (or working-memory capacity) is that they explain no more, and probably less, than 10% of the variance. Moreover, the shared variance between executive-process measures and intelligence is shared not with *g* or *Gf* but with lower level factors: A recent reanalysis of studies relating the three Miyake-Friedman factors of executive functions to measures of intelligence found that in most cases the executive-function factors could be identified with group-level factors of the Cattell-Horn-Carroll model, primarily the speed factor *Gs* (Jewsbury, Bowden, & Strauss, 2016).

If executive processes are to fill the explanatory role that Kovacs and Conway assign them, they need to be general, that is, they need to be enrolled by many tasks across different content domains. Why should we believe that executive processes are more general than other processes? In production-system architectures such as ACT-R, cognitive control is not exerted by a set of general productions for control. Rather, it emerges from the interaction of task-specific productions and task-specific declarative representations, in particular representations of goals (Anderson et al., 2004; Salvucci & Taatgen, 2008). In conflict-monitoring and control theory (Botvinick, Braver, Barch, Carter, & Cohen, 2001), an influential computational model of cognitive control, representations of the current task, or the currently relevant feature dimension to attend to, exert control by biasing the processing of stimuli such that task-relevant stimuli or stimulus features are processed more strongly than others. Arguably, these task representations are task specific. According to an alternative model aiming to explain the same phenomena, cognitive control arises from Hebbian learning of connections between stimuli, responses,

and task representations. These connections are specific not only to a task but even to individual stimuli and responses within that task (Verguts & Notebaert, 2008).

No computational model of cognition, or of cognitive control, includes general executive processes. This is probably not an accident: It is very hard to write a job description for a general executive process, because what an executive control process needs to do to ensure goal-oriented behavior differs for each task and situation. Consider, for instance, different versions of the Stroop task. In the classic version the person needs to suppress processing of the word meaning while processing the print color instead. In the numerical Stroop task, people need to suppress processing the digit's meaning and process their number instead. We could conceptualize the control processes in both cases as "inhibition of irrelevant information" or even more specifically "inhibition of processing the symbol's meaning," but that does not make them the same process. In fact, measures of Stroop interference from different versions of the Stroop task have often found to correlate weakly with each other (Hull, Martin, Beier, Lane, & Hamilton, 2008; Salthouse & Meinz, 1995; Shilling, Chetwynd, & Rabbitt, 2002), casting doubt on the assumption that they measure the same domain-general control process.

An Alternative: Parameter Overlap Theory

I find it plausible that *g* is a formative rather than a reflective factor, but I don't think that its constituents are best described as processes. As an alternative I want to propose that the formative constituents are parameters of the cognitive system. This hypothesis is motivated by efforts to build a computational model of working memory (Anderson, Reder, & Lebiere, 1996; Oberauer & Kliegl, 2006; Oberauer, Lewandowsky, Farrell, Jarrold, & Greaves, 2012; Oberauer, Souza, Druey, & Gade, 2013). In every computational model there are parameters that influence performance measures, such as reaction time or accuracy, and that can reasonably be assumed to vary between individuals. These parameters describe general features of the mechanisms that the model assumes to produce behavior. Candidate parameters include the amount of source activation in a spreading-activation network (Anderson et al., 1996; Lovett, Reder, & Lebiere, 1999), distinctiveness of representations (Farrell, 2012; Oberauer et al., 2012), the noisiness of evidence accumulation in response selection (Healey & Kahana, 2015; Schmiedek, Oberauer, Wilhelm, Süß, & Wittmann, 2007), and the efficiency of removing irrelevant information from working memory (Oberauer et al., 2012).

Computational modeling of working memory is in its infancy, and there is no converging evidence yet favoring one or a set of parameters as explanations of individual differences in (fluid) intelligence. I suspect that the set of relevant parameters will include more than one. If that is the case, then the positive manifold could be explained in a way very similar to process overlap theory: The performance of the cognitive system depends on the values of a set of N parameters. These parameters might be uncorrelated across individuals. Performance on each cognitive task j depends on a large subset n_j of those N parameters. Because the subsets are large, they overlap substantially across tasks. The correlation between two tasks, j

and k, depends on the proportion of shared parameters in their subsets n_j and n_k. This overlap is rarely, if ever, zero so that most pairs of tasks are positively correlated.

The difference between process overlap and parameter overlap is that parameters, by definition, are variables characterizing the general mechanisms of the cognitive system. In neural terms, they correspond to features of the entire brain—or at least of large networks—such as the number of neurons (which arguably influences the precision, and thereby the distinctiveness, of neural population codes); the degree of myelination of neuronal connections, which influences the noisiness of information processing (Miller, 1994); or the gain parameter of neurons, which also influences the signal-to-noise ratio (Cohen & Servan-Schreiber, 1992).

The distinction between processes and parameters helps to explain why training with working-memory tasks is very effective in improving performance on the trained tasks but yields little, if any, transfer to fluid intelligence: Training makes the processes involved in the trained task more efficient, but it does not change the parameters of the system.

To conclude, I think that the general idea of explaining g as a formative factor holds much promise. The positive manifold could be explained as emerging from the fact that performance on all cognitive tasks is dependent on sets of overlapping units. Processes, however, are unlikely to be these units because they are not general enough. Parameters that characterize features of the cognitive system as a whole are better candidates.

Funding

References

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review*, 111, 1036–1060.

Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought*. Mahwah, NJ: Erlbaum.

Anderson, J. R., Reder, L. M., & Lebiere, C. (1996). Working memory: Activation limits on retrieval. *Cognitive Psychology*, 30, 221–256.

Ashcraft, M. H. (1992). Cognitive arithmetic: A review of data and theory. *Cognition*, 44, 75–106.

Au, J., Sheehan, E., Tsai, N., Duncan, G. J., Buschkuehl, M., & Jaeggi, S. M. (2015). Improving fluid intelligence with training on working memory: a meta-analysis. *Psychonomic Bulletin & Review*, 22, 366–377. http://dx.doi.org/10.3758/s13423-014-0699-x

Baddeley, A. D. (1996). Exploring the central executive. *Quarterly Journal of Experimental Psychology*, 49A, 5–28.

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, 108, 624–652.

Cohen, J. D., & Servan-Schreiber, D. (1992). Context, cortex, and dopamine: A connectionist approach to behavior and biology in Schizophrenia. *Psychological Review*, 99, 45–77.

Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, 64, 135–168. http://dx.doi.org/10.1146/annurev-psych-113011-143750

Ecker, U. K. H., Lewandowsky, S., Oberauer, K., & Chee, A. E. H. (2010). The components of working memory updating: An experimental decomposition and individual differences. *Journal of Experimental Psychology-Learning Memory and Cognition*, 36, 170–189.

Farrell, S. (2012). Temporal clustering and sequencing in short-term memory and episodic memory. *Psychological Review*, 119, 223–271.

Fodor, J. A., & Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, 28, 3–71.

Friedman, N. P., Miyake, A., Corley, R. P., Young, S. E., DeFries, J. C., & Hewitt, J. K. (2006). Not all executive functions are related to intelligence. *Psychological Science*, 17, 172–179.

Healey, M. K., & Kahana, M. J. (2015). A four-component model of age-related memory change. *Psychological Review*. http://dx.doi.org/10.1037/rev0000015

Hull, R., Martin, R. C., Beier, M. E., Lane, D., & Hamilton, A. C. (2008). Executive function in older adults: A structural equation modeling approach. *Neuropsychology*, 22, 508–522. http://dx.doi.org/10.1037/0894-4105.22.4.508

Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences*, 105, 6829–6833. http://dx.doi.org/10.1073/pnas.0801268105

Jewsbury, P. A., Bowden, S. C., & Strauss, M. E. (2016). Integrating the switching, inhibition, and updating model of executive function with the Cattell–Horn–Carroll model. *Journal of Experimental Psychology: General*, 145, 220–245.

Keye, D., Wilhelm, O., Oberauer, K., & van Ravenzwaaij, D. (2009). Individual differences in conflict-monitoring: testing means and covariance hypothesis about the Simon and the Eriksen Flanker task. *Psychological Research-Psychologische Forschung*, 73, 762–776.

Lovett, M. C., Reder, L. M., & Lebiere, C. (1999). Modeling working memory in a unified architecture: An ACT-R perspective. In A. Miyake & P. Shah (Eds.), *Models of working memory. Mechanisms of active maintenance and executive control* (pp. 135–182). Cambridge, UK: Cambridge University Press.

Melby-Lervåg, M., & Hulme, C. (2016). There is no convincing evidence that working memory training is effective: A reply to Au et al. (2014) and Karbach and Verhaeghen (2014). *Psychonomic Bulletin & Review*, 23, 324–330. http://dx.doi.org/10.3758/s13423-015-0862-z

Miller, E. M. (1994). Intelligence and brain myelination: A hypothesis. *Personality and individual Differences*, 17, 803–832.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology*, 41, 49–100.

Oberauer, K., & Kliegl, R. (2006). A formal model of capacity limits in working memory. *Journal of Memory and Language*, 55, 601–626.

Oberauer, K., Lewandowsky, S., Farrell, S., Jarrold, C., & Greaves, M. (2012). Modeling working memory: An interference model of complex span. *Psychonomic Bulletin & Review*, 19, 779–819.

Oberauer, K., Souza, A. S., Druey, M., & Gade, M. (2013). Analogous mechanisms of selection and updating in declarative and procedural working memory: Experiments and a computational model. *Cognitive Psychology*, 66, 157–211.

Oberauer, K., Süß, H.-M., Wilhelm, O., & Sander, N. (2007). Individual differences in working memory capacity and reasoning ability. In A. R. A. Conway, C. Jarrold, M. J. Kane, A. Miyake, & J. N. Towse (Eds.), *Variation in working memory* (pp. 49–75). New York, NY: Oxford University Press.

Rosenbaum, D. A., Carlson, R. A., & Gilmore, R. O. (2001). Acquisition of intellectual and perceptual-motor skills. *Annual Review of Psychology*, 52, 45–470.

Salthouse, T. A., & Meinz, E. J. (1995). Aging, inhibition, working memory, and speed. *Journal of Gerontology: Psychological Sciences*, 50B, 297–306.

Salvucci, D. D., & Taatgen, N. A. (2008). Threaded cognition: an integrated theory of concurrent multitasking. *Psychological Review*, 115, 101–130.

Schmiedek, F., Oberauer, K., Wilhelm, O., Süß, H. M., & Wittmann, W. W. (2007). Individual differences in components of reaction time distributions and their relations to working memory and

intelligence. *Journal of Experimental Psychology: General*, *136*, 414–429.

Shilling, V. M., Chetwynd, A., & Rabbitt, P. M. A. (2002). Individual inconsistency across measures of inhibition: an investigation of the construct validity of inhibition in older adults. *Neuropsychologia*, *40*, 605–619.

Smith, E. E., & Jonides, J. (1999). Storage and executive processes in the frontal lobes. *Science*, *283*, 1657–1661.

Spitz, H. H. (1986). *The raising of intelligence*. Hillsdale, NJ: Erlbaum.

Verguts, T., & Notebaert, W. (2008). Hebbian learning of cognitive control: dealing with specific and nonspecific adaptation. *Psychological Review*, *115*, 518–525.

von Bastian, C. C., & Oberauer, K. (2014). Effects and mechanisms of working memory training: a review. *Psychological Research*, *78*, 803–820. http://dx.doi.org/10.1007/s00426-013-0524-6

Welsh, M. C., Pennington, B. F., & Groisser, D. B. (1991). A normative-developmental study of executive function: a window on prefrontal function in children. *Developmental Neuropsychology*, *7*, 131–149.

Wilhelm, O., Hildebrandt, A., & Oberauer, K. (2013). What is working memory capacity, and how can we measure it? *Frontiers in Psychology*, *4*, 433. http://dx.doi.org/10.3389/fpsyg.2013.00433

Zbrodoff, N. J. (1995). Why is $9 + 7$ harder than $2 + 3$? Strength and interference as explanations of the problem-size effect. *Memory & Cognition*, *23*, 689–700.
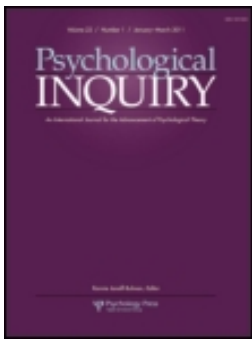
Psychological Inquiry

An International Journal for the Advancement of Psychological Theory

Groundhog Day: Is the Field of Human Intelligence Caught in a Time Warp? A Comment on Kovacs and Conway

Robert J. Sternberg

Published online: 02 Aug 2016.

Submit your article to this journal ⬈

View related articles ⬈

View Crossmark data ⬈

Routledge
Taylor & Francis Group

Groundhog Day: Is the Field of Human Intelligence Caught in a Time Warp? A Comment on Kovacs and Conway

Robert J. Sternberg

Department of Human Development, Cornell University, Ithaca, New York

Kovacs and Conway (this issue) present a well thought through "unified account of the general factor of intelligence." The account integrates cognitive, psychometric, and neuropsychological evidence. The account is plausible and well integrates the various sources of evidence. It is a compelling article. It is one of the most well-developed and thoughtful models of *g* currently to be found.

If many of my comments on the article in this essay are in the form of critique, it is because the goal of an interchange such as this one is largely to stimulate exchange of different ideas rather than simple to agree with what is said. I will argue in this article that at times, the field of intelligence seems to be stuck in a time warp, but first let's consider some other issues.

Problems of Scholarship

Let's start with a minor but nevertheless vexing issue—quality of scholarship. I need to start with this issue because it will be central to the main thesis of this article regarding the field of intelligence being stuck in a time warp. In their article, Kristof Kovacs and Andrew R. A. Conway (this issue) say they "propose a novel solution to this well-aged puzzle which we refer to as process overlap theory" (p. 152). The puzzle to which they refer is the nature of *g*. But is their solution to the puzzle "novel"?

The authors understandably if not quite correctly cite Godfrey Thomson (1916) as the founding father of what they call "process overlap" theory. Thomson's theory of bonds is arguably the original basis for the idea that *g* is not just one thing. Equally arguably, at least from an information-processing standpoint, the idea goes even further back to Edward Thorndike (1911), who is not cited. Thorndike believed that associations and their overlaps underlay intelligence. Thorndike, an avid experimentalist, later became heavily involved in the mental-testing movement during World War I and was instrumental in creating the Army Alpha (verbal) and Beta (performance) intelligence tests.

That said, the real father of the overlapping-process theory is not really either Thomson or Thorndike, but rather Charles Spearman, who formulated the theory of *g* (Spearman, 1904, 1927). Spearman's *truly novel* theory of overlapping processes was not presented in his papers and books on *g*, but rather in a separate work (Spearman, 1923), which is not cited by Kovacs

and Conway (this issue). Spearman believed that apprehension of experience (what I later called "encoding"; R. J. Sternberg, 1977), eduction of relations (what I later called "inference"), and eduction of correlates as used in analogical reasoning (what I later called "application") are the basic overlapping information processes of intelligence. My later work on analogy (R. J. Sternberg, 1977) was merely an elaboration on Spearman, as is the work of Kovacs and Conway. According to Spearman (1923), apprehension of experience, eduction of relations, and eduction of correlates occur in overlapping fashion in many diverse intellectual tasks, epitomized by the analogy. It is an interesting historical note that two of the great psychometricians of all time, Spearman (1904) and Carroll (1993), were also astute cognitive psychologists (Carroll, 1976; Spearman, 1923).

Spearman (1923) was not the only early famed psychometrician who constructed a cognitive theory of intelligence. Thurstone (1924), in a little remembered work that preceded his psychometric contributions, argued that intelligence is the ability to inhibit an instinctive response. This is one of the key components of working memory, as noted in Kovacs and Conway and elsewhere (Miyake et al., 2000). Thurstone's cognitive contribution, like Spearman's, was prescient, and is not cited by Kovacs and Conway.

With regard to more recent literature, Kovacs and Conway do cite Detterman (1994), whose ideas about process overlap are quite similar to theirs. Inexplicably, however, they do not cite three investigators whose earlier research makes any claim for this theory being a "novel solution" seem a bit stretched. The investigators are John B. Carroll, Earl Hunt, and me, and the relevant citations are Carroll (1976); Hunt, Frost, and Lunneborg (1973); Hunt, Lunneborg, and Lewis (1975); and R. J. Sternberg (1977, 1980, 1984b, 1985). Oddly, Kovacs and Conway also do not cite Howard Gardner (e.g., Gardner, 1983), who, like me, has viewed *g* as a much narrower construct than have most scholars. Gardner has amassed evidence that he, at least, interprets as suggesting that intelligence is modular rather than unified. Authors obviously do not have to cite every publication ever written on a topic, but the Kovacs and Conway's failure to cite key scholarly works relevant to their own is puzzling, to say the least.

Carroll, Hunt, and I proposed ideas quite similar to those in the Kovacs and Conway article. Hunt suggested that the

information processes found in traditional cognitive tasks, such as the S. Sternberg (1966) high-speed memory-scanning task, underlie general intelligence. Carroll and I proposed as well that overlapping information processes are the place to look, but we instead suggested looking at the processes involved in tasks similar or identical to those found on tests of intelligence. In my own case, I proposed that metacomponents, or executive processes, and to some extent performance components (e.g., inference, application) and knowledge-acquisition components (e.g., selective encoding) overlap among virtually all cognitive tasks and that they underlie what we have come to call g (R. J. Sternberg, 1977, 1984b; R. J. Sternberg & Gardner, 1982).

These theoretical accounts are not identical to those of Kovacs and Conway. Kovacs and Conway, writing more recently, elaborate on the role of working memory and of the brain in the theory of process overlap. Their account is up-to-date, elegant, and well researched. But it is somewhat astonishing that none of this earlier work was cited as antecedent to the authors' own, as the work collectively formed much of the basis for the cognitive-correlates (Hunt) and cognitive-components (Sternberg) work that later formed the basis for one of several dominant paradigms in the study of intelligence (see R. J. Sternberg, 1990).

Is the Theory Explanatory or Descriptive?

In their article, Kovacs and Conway state that "similar formative latent variables are socioeconomic status and general health, which each tap common variance between measures but do not explain it; according to process overlap theory, g is no different" (p. 162). I agree. If one were to do a factor analysis of tests of health or tests of wealth, one likely would end up with a g factor for each, but the g factor would not explain socioeconomic status (SES) or general health.

For SES, one could do a factor analysis on parental education, family income, number of books in the house, amount of money invested, and so on. The g for SES would reflect overlapping processes in getting a good education, buying books for the household, investing money wisely, and so forth. If you have enough of these factors at high-enough values, you would score well on the SES general factor.

General health, measuring levels of good cholesterol, bad cholesterol (weighted negatively), blood pressure (weighted negatively), blood sugar (weighted negatively), and similar indices, also could yield a g factor. Likely there are overlapping processes responsible for these, some of which are inherited (as with intellectual g) but others over which the individual has some control, such as diet, refraining from dangerous drugs, and exercise. People who have good general health are people who have a number of the good-health factors weighing in their favor. The g for general health is not one thing but a set of overlapping processes.

Through our overlapping-process theories of the gs of SES and good health, we have created … what? Do we have an explanatory theory of what *causes* high SES or good general health? Are lots of books in the house, high levels of education, and enough money to live comfortably causes of high SES, or are they in turn secondary variables that reflect deeper underlying causes? What might some of the deeper underlying causes

be? One would be familial inheritance—the individual inherited his or her SES. Happens all the time. Another deeper cause could be a high level of intrinsic motivation—someone who is extremely motivated to do whatever it takes to raise her or his SES. But then, these deeper causes could have still deeper causes, for example, that the parent from whom the child inherited her or his SES was extremely highly motivated to succeed.

Of course, the same logic applies to good health. Blood pressure and cholesterol numbers are certain overlapping in terms of being responsible for good health, but they in turn can be traced back to earlier antecedents, such as lucky inheritance, good diet, good exercise habits, or whatever. Those elements in turn could be traced further back, for example, to early parental or peer pressure to eat well or exercise regularly.

We end up in the same place, I believe, with the process-overlap theory of intelligence. Our models of SES and general-health g, like Kovacs and Conway's model of intellectual g, is largely descriptive rather than deeply explanatory.

Lest it sound as if I am quick to dish out criticism to Kovacs and Conway, I should add that the same critique applies to my own theory of intelligence. When I have spoken of the overlapping processes that are responsible for the general factor (R. J. Sternberg & Gardner, 1982—36 years before the Kovacs & Conway article), I realized that the model was basically descriptive. What I call "metacomponents"—recognizing the existence of a problem, defining what the problem is, choosing component processes to solve the problem, creating or selecting a strategy for problem solving, constructing a mental representation upon which the strategy will act, monitoring problem solving, and evaluating problem solving—are descriptive and can be further broken down, and then the chances are that whatever they are broken down into can be broken down still further.

Some theorists, including Kovacs and Conway, try to finesse this problem in part by identifying neuropsychological correlates of overlapping processes. But that is exactly what they are—correlates. As Satel and Lilienfeld (2015) and many others have pointed out, most of the neuropsychological models we have today are at best correlational and sometimes the methodology used in identifying parts of the brain involved in cognitive processing is suspect. When theorists try to give their work the cloak of causal modeling by evoking the brain, they really are doing something not entirely different from what factor analysts have done for many years—taking descriptive entities (as Vernon, 1950, realized with regard to factors) and viewing them as causal. Van der Maas et al. (2006) argued that it is not worth looking for g in the brain, because there is no entity of g, only a set of mutually beneficial interactions among cognitive processes (which, yes, constitutes yet another overlapping-processes model). With this introduction, we now can turn to the main point of the present critique.

Groundhog Day

In the movie *Groundhog Day* (Ramis, 1993), a weatherman named Phil (Bill Murray) gets stuck in some kind of time warp and keeps living the same day over and over again. Try as he

might, he can't get past that day (until, of course, the romantic end of the movie). Research on *g* seems to have gotten stuck in a similar time warp. Spearman (1904) probably had no idea of how his theory could get so stuck. The construct of *g* was a major finding, but it happened well over a century ago. Why haven't we, as a field, moved on?

One could argue that we haven't moved on because we are still trying to understand *g*. But what's there to understand? As early as 1923, Spearman proposed overlapping processes underlying *g*, and it seems we are still at it, with the theory of the day being presented as "novel" exactly one century after the publication of Thomson's (1916) overlapping bonds theory and almost a century after the publication of Spearman's (1923) overlapping-process theory. The language has gotten more sophisticated, and there is a larger experimental literature to draw on, but theorists keep dressing the same mannequin in new clothes.

Hunt (Hunt, Frost, & Lunneborg, 1973) and I (R. J. Sternberg, 1977) tried to go beyond redressing the same mannequin and reliving the same Groundhog Day by creating theories of intelligence that were not based on individual differences. Theories based on individual differences have been the major custom since Spearman's (1904) article, with the early exception of developmental researchers such as Piaget (1952, 1972), Luria (1976), and Vygotsky (1978). Lee Cronbach (1957) was prescient in arguing for an integration of psychometric and experimental methods. Even his comments were adumbrated by another great psychologist, Louis Thurstone (1938). Thurstone suggested that differential methods eventually would be supplemented or replaced by experimental ones. Hunt and I thought ours was a step forward—we were not basing our theories on individual differences—but apparently the step was tentative.

In their article, Kovacs and Conway state that

> latent variables are *differential* constructs that do not directly translate to within-individual processes or mechanisms … and without variation in mental abilities there would be no latent variables—the last survivor of a meteor collision with Earth would still have cognitive abilities and mental limitations but would not have *g*. (p. 153)

If so, it is unclear why almost the whole Kovacs and Conway article is devoted to a within-subjects account of overlapping processes within an individual of a between-subjects construct, *g* (as per the title of their article). Why, after more than a century, do investigators persist in starting with a construct based on individual differences—despite the work of Spearman (1923); Thurstone (1938); Hunt, Frost, and Lunneborg (1973); R. J. Sternberg (1977); and the developmental theorists. Perhaps those who do not learn from history are doomed to repeat it, which may explain why this article is reminiscent of the movie *Groundhog Day*. Even more reminiscent of *Groundhog Day* are the investigators who persist in finding more and more and more correlates of *g*, over a century after Spearman (1904) pointed out that *g* correlates with pretty much everything that has a mental-ability component.

Or does it? As Kovacs and Conway point out, correlations are dependent on the context of the measurement situation from which they are calculated. If there were no individual differences, there would be no *g*, and if there were only weak

individual differences, there would be only a weak *g*. But the ways in which mental abilities develop all depend on the interaction of genetics with environment, and *g* as we know it assumes in large part an environment in which the mental abilities measured by conventional ability tests are mutually developed and rewarded (R. J. Sternberg, 1984a, 2004).

Instead, imagine an environment in which this mutual development and reward is not the case. Then will *g* look different, given that it its existence depends on patterns of individual differences? Kovacs and Conway acknowledge this possibility when they state that "the average correlation between diverse tests and thus the domain-generality of the positive manifold varies across time and ability level, and *g* is only informative of the extent of domain-general variance in a given population at a given time" (p. 155). Thus, as they say, "the positive manifold is an emergent property" (p. 162).

Some evidence suggests that *g* is indeed much more fickle than some investigators have thought, and that Kovacs and Conway should have added "place" to "time" as a constraint on how *g* manifests itself.

My colleagues and I found that *g*, as we traditionally think of it, has only a minor role in some cultures' conceptions of intelligence (see R. J. Sternberg, 2004). But does *g* also have a different role in actual adaptive value in different cultures (R. J. Sternberg, 2014a)? Our work in Kenya suggests it does. A group of us (Grigorenko et al., 2001) found that *g*, as we usually think of it, plays only a minor role in rural Kenyans' conceptions of intelligence. But is their conception of intelligence divorced from their reality, or does it faithfully reflect the demands of their environment? Our research suggests the latter.

Prince et al. (2001) investigated adaptive behavior in rural Kenya, seeking out what it actually meant to be intelligent in their environment. These investigators found that a key element of adaptive intelligence is the learning of natural herbal medicines to combat parasitic illnesses, such as malaria, hookworm, whipworm, schistosomiasis, and related illnesses. R. J. Sternberg et al. (2001), in turn, measured village children's knowledge and of and ability to use these natural herbal medicines. The kind of test we employed would yield only chance scores among Western children, as none of them (or any of the adults) would have any knowledge of these medicines. When we correlated scores on our adaptive-functioning tests with scores on tests of intelligence, the correlations were negative, that is, children whose academic knowledge and skills were better did worse on the real-world adaptive tests. Why?

In the rural Kenyan villages we studied, schooling and its concomitant acquisition of academic knowledge and skills were viewed as a proposition for losers. Schooling did not lead to a job. What led to a job was an apprenticeship, and master craftsmen were interested only in taking on children who showed ability for their trades. Thus, in terms of the village conception of intelligence, the "intelligent" children left school early to learn a trade; the not-so-intelligent ones stayed in school because no one wanted to teach them a trade. Learning academic knowledge, beyond a certain point, was viewed as a sign of failure. This view is not limited to Kenya.

In much of Silicon Valley today, the college and business-school dropouts (like my son Seth!) manage the PhDs, not the other way around. Funders generally are not interested in

funding potential entrepreneurs while they are in still in school, so the students drop out. The top rewards go to the dropouts, not to the most educated, among whom are some of those highest in *g* (e.g., the software engineers).

In a more recent and as-yet-unpublished study, my wife, Karin Sternberg, and I found that a test of scientific reasoning administered to Cornell students correlated zero to negatively with tests of fluid intelligence. In earlier work on practical intelligence, we found only very weak positive correlations between tests of *g* and tests of practical intelligence, although both kinds of tests independently predicted job performance (R. J. Sternberg et al., 2000). Of course, the correlations in all instances were among a restricted range of cognitive abilities. But that is the point. How *g* functions will depend on when, where, and with whom the testing is done.

A Future for Research

Is there a future for research in this area that enables researchers to escape the Groundhog Day phenomenon whereby researchers keep doing the same things over and over again? I believe there is. But to escape from Groundhog Day, researchers have to want to escape, and the biggest problem so far in the field is not inability to escape but lack of desire to stop doing the same things over and over again—changing the clothing, changing the makeup, but leaving the mannequins the same. What else could researchers in the 21st century do besides what researchers did at the turn of the 20th century—constructing theories and finding correlates of *g*, over and over and over again? Here are just a few examples.

What Are the Roles of Various Cognitive and Noncognitive Processes in Intelligence?

We know that working memory is important to intelligence; and attention is important to intelligence; and inductive reasoning is important to intelligence. And we know that working memory, attention, and inductive reasoning are all related to each other. So, are working memory, attention, and inductive reasoning separate processes, or are there more basic processes underlying these and other labeled processes related to intelligence? If there are more basic (orthogonal) processes, what are they? And how do noncognitive processes play a role in intelligence, for example, motivation? Several theorists (Dweck, 2007; Hayes, 1962) have suggested that motivational differences are key to individual differences in intelligent performance.

Enculturation and Socialization

The authors allude to the effects of enculturation and socialization, but rather obliquely. They never deal with any of the effects and do not cite any of the relevant literature. They are right—how *g* manifests itself depends on people, time, and also place. Investigators such as Flynn (2012) have shown how levels of *g* (and of its correlate, IQ) change over time, and other investigators (e.g., Rindermann, 2007) have shown how they differ across space. So that's the easy part—showing the differences.

We need to understand why the differences occurs, what they mean, and how they can be mitigated.

Value of g-Related Skills

We tend to assume that *g*-related skills are important or even crucial to have. But are they? Our previously mentioned research among Kenyans, as well as research we have done among Yup'ik Eskimos (Grigorenko et al., 2004), suggests that in some sociocultural settings *g* counts for a lot and in others such settings it is not so important. For example, among the Eskimos, ice-fishing and hunting skills were far more important—and are not well predicted by *g*. Kanazawa (2012) suggested that people with higher levels of *g*-based skills often perform more poorly, on average, on evolutionarily significant tasks than do people with lower *g*-based skills. I have suggested (R. J. Sternberg, 1988) something related—that children high in *g*-based skills are so rewarded for these skills in school and at home that they often fail to develop other skills that will be as important or more important in life, such as creative and practical skills. They may even be lower in wisdom (R. J. Sternberg, 2002) because they fail to realize that their high level of *g* does not protect them from acting foolishly. In any case, when and where is *g* helpful, and where and when is it indifferent or even harmful?

Intelligence As Trait and State

We tend to view intelligence largely as a trait. But intelligence, like anxiety, has statelike properties as well as traitlike ones (R. J. Sternberg, 2014b). Intelligence can rise and fall with various variables, such as level of anxiety, drug intake, and level of wakefulness. How much of our intelligence as manifested in everyday life is more traitlike and how much is more statelike?

These problems are just a pitifully small sample of the problems intelligence researchers might address. I intend them only to illustrate that we don't have to be locked forever into more and more studies of the bases and correlates of *g*. But inertia is a powerful force, and perhaps 10, 20, or 30 years from now, the field still will not have broken out of Groundhog Day. Then, maybe it will have. Only time will tell.

References

Carroll, J. B. (1976). Psychometric tests as cognitive tasks: A new "structure of intellect." In L. B. Resnick (Ed.), *The nature of intelligence* (pp. 27–56). Hillsdale, NJ: Erlbaum.

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

Cronbach, L. J. (1957). The two disciplines of scientific psychology. *American Psychologist, 12*, 671–684.

Detterman, D. K. (1994). Theoretical possibilities: The relation of human intelligence to basic cognitive abilities. In D. K. Detterman (Ed.), *Current topics in human intelligence (Volume 4): Theories of intelligence* (pp. 85–115). Norwood, NJ: Ablex.

Dweck, C. (2007). *Mindset: The new psychology of success*. New York, NY: Ballantine.

Flynn, J. R. (2012). *Are we getting smarter? Rising IQ in the 21st century*. Cambridge, UK: Cambridge University Press.

Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences*. New York, NY: Basic Books.

Grigorenko, E. L., Geissler, P. W., Prince, R., Okatcha, F., Nokes, C., Kenny, D. A., … Sternberg, R. J. (2001). The organization of Luo conceptions of intelligence: A study of implicit theories in a Kenyan village. *International Journal of Behavioral Development*, 25(4), 367–378.

Grigorenko, E. L., Meier, E., Lipka, J., Mohatt, G., Yanez, E., & Sternberg, R. J. (2004). Academic and practical intelligence: A case study of the Yup'ik in Alaska. *Learning and Individual Differences*, 14, 183–207.

Hayes, K. J. (1962). Genes, drives, and intellect. *Psychological Reports*, 10, 299–342. Monograph Supplement 2-V10.

Hunt, E., Frost, N. & Lunneborg, C. (1973). Individual differences in cognition: A new approach to intelligence. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 6, pp. 87–122). New York, NY: Academic Press.

Hunt, E. B., Lunneberg, C., & Lewis, J. (1975). What does it mean to be high verbal? *Cognitive Psychology*, 7, 194–227.

Kanazawa, S. (2012). *The intelligence paradox: Why the intelligent choice isn't always the smart one*. New York, NY: Wiley.

Luria, A. R. (1976). *Cognitive development: Its cultural and social foundations*. Cambridge, MA: Harvard University Press.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "Frontal Lobe" tasks: A latent variable analysis. *Cognitive Psychology*, 41, 49–100.

Piaget, J. (1952). *The origins of intelligence in children*. New York, NY: International Universities Press.

Piaget, J. (1972). *The psychology of intelligence*. Totowa, NJ: Littlefield, Adams.

Prince, R. J., Geissler, P. W., Nokes, K., Maende, J. O., Okatcha, F., Grigorenko, E. L., & Sternberg, R. J. (2001). Knowledge of herbal and pharmaceutical medicines among Luo children in western Kenya. *Anthropology & Medicine*, 8(2/3), 211–235.

Ramis, H. (1993). *Groundhog day* [Motion picture]. United States: Columbia Pictures.

Rindermann, H. (2007). The *g*-factor of international cognitive ability comparisons: The homogeneity of results in PISA, TIMSS, PIRLS and IQ-tests across nations. *European Journal of Personality*, 21(6), 667–706.

Satel, S., & Lilienfeld, S. (2015). *Brainwashed: The seductive appeal of mindless neuroscience*. New York, NY: Basic.

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

Spearman, C. (1923). *The nature of intelligence and the principles of cognition*. London, UK: Macmillan.

Spearman, C. (1927). *The abilities of man*. New York, NY: Macmillan.

Sternberg, R. J. (1977). *Intelligence, information processing, and analogical reasoning: The componential analysis of human abilities*. Hillsdale, NJ: Erlbaum.

Sternberg, R. J. (1980). Sketch of a componential subtheory of human intelligence. *Behavioral and Brain Sciences*, 3, 573–584.

Sternberg, R. J. (1984a). A contextualist view of the nature of intelligence. *International Journal of Psychology*, 19, 307–334.

Sternberg, R. J. (1984b). Toward a triarchic theory of human intelligence. *Behavioral and Brain Sciences*, 7, 269–287.

Sternberg, R. J. (1985). *Beyond IQ: A triarchic theory of human intelligence*. New York, NY: Cambridge University Press.

Sternberg, R. J. (1988). *The triarchic mind: A new theory of human intelligence*. New York, NY: Viking.

Sternberg, R. J. (1990). *Metaphors of mind: Conceptions of the nature of intelligence*. New York, NY: Cambridge University Press.

Sternberg, R. J. (2002). Smart people are not stupid, but they sure can be foolish: The imbalance theory of foolishness. In R. J. Sternberg (Ed.), *Why smart people can be so stupid* (pp. 232–242). New Haven, CT: Yale University Press.

Sternberg, R. J. (2004). Culture and intelligence. *American Psychologist*, 59(5), 325–338.

Sternberg, R. J. (2014a). The development of adaptive competence. *Developmental Review*, 34, 208–224.

Sternberg, R. J. (2014b). Intelligence as trait—and state? *Journal of Intelligence*, 2, 4–5.

Sternberg, R. J., Forsythe, G. B., Hedlund, J., Horvath, J., Snook, S., Williams, W. M., … Grigorenko, E. L. (2000). *Practical intelligence in everyday life*. New York, NY: Cambridge University Press.

Sternberg, R. J., & Gardner, M. K. (1982). A componential interpretation of the general factor in human intelligence. In H. J. Eysenck (Ed.), *A model for intelligence* (pp. 231–254). Berlin, Germany: Springer–Verlag.

Sternberg, R. J., Nokes, K., Geissler, P. W., Prince, R., Okatcha, F., Bundy, D. A., & Grigorenko, E. L. (2001). The relationship between academic and practical intelligence: A case study in Kenya. *Intelligence*, 29, 401–418.

Sternberg, S. (1966). High-speed memory scanning in human memory. *Science*, 153, 652–654.

Thomson, G. H. (1916). A hierarchy without a general factor. *British Journal of Psychology, 1904–1920*, 8(3), 271–281.

Thorndike, E. L. (1911). *Animal intelligence: Experimental studies*. New York, NY: Macmillan.

Thurstone, L. L. (1924). *The nature of intelligence*. London, UK: Kegan Paul, Trench Trubner & Co.

Thurstone, L. L. (1938). *Primary mental abilities*. Chicago, IL: University of Chicago Press.

van der Maas, H. L. J., Dolan, C. V., Grasman, R. P. P. P., Wicherts, J. M., Huizenga, H. M., & Raijmakers, M. E. J. (2006). A dynamical model of general intelligence: The positive manifold of intelligence by mutualism. *Psychological Review*, 113(4), 842–861.

Vernon, P. E. (1950). *The structure of human abilities*. London, UK: Methuen.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Has g Gone to POT?

Kristof Kovacs & Andrew R. A. Conway

Published online: 02 Aug 2016.

Submit your article to this journal ↗

View related articles ↗

View Crossmark data ↗

REPLY

Has *g* Gone to POT?

Kristof Kovacs[a,b] and Andrew R. A. Conway[c]

[a]Center for Research and Development, Eszterházy Károly University, Eger, Hungary; [b]Department of Psychological Methods, University of Amsterdam, Amsterdam, The Netherlands; [c]Division of Behavioral and Organizational Sciences, Claremont Graduate University, Claremont, California

Introduction

There is no way to express how amazed and impressed we are by the commentaries received to our target article. We are particularly humbled by the simulations and other studies conducted to test predictions or clarify proposals of the theory. Such a diverse range of topics and commentaries like the ones we received makes it impossible to adequately respond to every issue raised, and with regret we had to select the most important and/or recurring points to address. First we deal with issues related to the novelty of the theory and the corresponding model. This is followed by a discussion of evidence for the theory. We turn to its predictions next and respond to each simulation or other study that tests them. Finally, we elaborate on the concept of a formative general factor and the value of prediction in evaluating the relevance of such a concept.

Old Wine in New Bottles?

"Discovery consists of seeing what everybody has seen, and thinking what nobody has thought," said physiologist Albert Szent-Györgyi, winner of the Nobel Prize in 1937 (quoted in Good, 1962). As it is likely the most replicated result in psychology, virtually everybody has seen the positive manifold. As it is a necessary algebraic consequence of the positive manifold, everybody has seen the general factor, too. But has anybody thought of the same explanation as the one provided in the target article?

Several commentators have accused us of selling old wine in new bottles. At the same time, some are certain that we are rebottling an Australian Chardonnay, whereas others clearly recognize that excellent 2008 Pinot Noir from Chile. That is, it has been argued by different commentators that our theory only slightly differs from what Thomson, Spearman, Carroll, Hunt, Sternberg, Eysenck, Jensen, or Anderson said and that we shamelessly borrow ideas from uncited pioneers. This is most pronounced in Sternberg's (this issue) commentary: "It is somewhat astonishing that none of this earlier work was cited as antecedent to the authors' own" (p. 237).

As we did explicitly admit, process overlap theory (POT) is "not the first cognitive approach to human intelligence" (p. 161). Yet it is the first cognitive theory that also provides a latent variable model and an item response model. POT is a whole with these three constituents, as highlighted by Kan,

Kievit, and van der Maas (this issue). As well, it draws from the study of neural networks and the neural correlates of cognitive abilities. In fact, as a fourth leg of the theory, it provides a testable account of the neural mechanisms underlying the proposed overlap of psychological processes.

The target article was already testing the length limits of the journal, as well as the patience of the reader. Hence it was not possible to adequately discuss all the great previous research that employed a cognitive approach to intelligence. We chose to cite only those whose work is inherently linked to the main argument or whose solution to the same problem we dealt with was judged similar enough to justify a direct comparison. We could have, of course, simply provided a long list of references after our sentence that we just quoted, but we decided against it. In retrospect, and especially reading Sternberg's commentary, this might have been a mistake, but in no way did we mean to be disrespectful to the giants on whose shoulders we are standing. We agree with Sternberg that this list would have included Spearman's seminal book on the laws of cognition—and many others he did not cite. We also admire the work carried out under the *cognitive correlates* (Hunt) and *cognitive components* (Sternberg) approach; in particular, Sternberg's work on metacomponents has greatly influenced how we think about human intelligence in general.

Gottfredson (this issue) also suggests that we fail to acknowledge the tradition to which our work belongs, but she adopts a very different perspective than Sternberg. To our utmost surprise, Gottfredson judges that POT is perfectly compatible with *g*-theory and argues that we are following the same tradition as Eysenck and Jensen in our efforts to explain the positive manifold. Yet, according to Gottfredson, we misinterpret *g*-theory, and this results in our false impression that we are contradicting it. In particular, Gottfredon argues that we confuse different levels of explanation—the psychological (intelligence) and the biological (the brain)—but if we overcome this confusion, then we are saying the same things that *g*-theorists have always said: General intelligence (*g*) *is* a unitary trait at the psychological level but is an emergent property of the brain.

To make this point, Gottfredson (this issue) quotes a relevant section of the target article and adds bracketed text to highlight where we purportedly fail to distinguish between the psychological and the biological. We repeat this section, as we find it very indicative of our differences:

I illustrate the authors' inadvertent conflation of two levels of analysis in the following statement by adding bracketed text to distinguish the two levels, tests and physical brains.

> The most important difference, then, from g-oriented accounts of the positive manifold is that whereas reflective general factor theories propose a causal influence of a latent variable, *g*, on the positive manifold [among psychometric tests and life outcomes], according to process overlap theory *the positive manifold [among tests] is an emergent property [of the brain]*, the result of the specific patterns in which item response processes [i.e., information processing systems in the brain] overlap. (p. 162)

With these insertions, the "important difference" disappears. (Gottfredson, this issue, p. 213)

Indeed, this extended statement does render POT to be a version of g-theory, and we would have to reconsider our criticism, if only we agreed with the additions, which we do not.

Gottfredson (this issue) also interprets POT as follows:

> As Kovacs and Conway repeatedly and correctly stress, psychometric *g* is an emergent property of interacting brain systems, so *g* is their singular result. *g* theorists agree, of course, but the authors attribute the opposite belief to them: that *g* causes the overlap in brain processes. (p. 213)

This interpretation and extension is not in accordance with our point. We never said that, according to g-theorists, *g* causes an overlap in brain processes. Nor did we propose that *g* is the "singular result" of "interacting brain systems." Contrary to Gottfredson's assertions, POT not only says that "psychometric *g* is an emergent property of the brain" and not only does it "rule out notions of it being a single process or place in the brain," it also proposes that *g* is an emergent property at the psychological level, that is, at the level of tests. Proximally, *g* emerges because of an overlap of psychological processes tapped by cognitive tests. Ultimately, this indeed can be traced to an overlap of brain activity behind such processes—as we did argue in the target article. But POT definitely does not propose that *g* is an emergent property of the brain that results in a unitary (reflective, causal) trait or ability.

Therefore, POT is in complete disagreement with how Gottfredson (this issue) conceptualizes *g* according to the tenets of g-theory: In short "domain-independent reasoning," a "unitary trait," that "generates (causes) individual differences in performance on cognitive tests intended to tap more specific abilities" (p. 210). Also, "in theoretical terms, *g* represents the most generic mental capacity possible: an all-purpose cognitive tool that enhances performance on all tasks requiring any mental manipulation of information" (p. 210). Finally, Gottfredson repeats the definition of intelligence that originally appeared in the manifesto published in the *Wall Street Journal*: "Intelligence is a very general mental capability that, among other things, involves the ability to reason, plan, solve problems, think abstractly, comprehend complex ideas, learn quickly and learn from experience" (p. 214). POT is incompatible with this definition.

Gottfredson somehow interprets POT as a theory that claims that *g*, a unitary trait, is an emergent property of the brain and that we only thought that it contradicts g-theory because we incorrectly assumed that g-theorists insist on finding a single source of *g* in the brain. Therefore, in her view POT

is an extension or elaboration of g-theory rather than a new theory that contradicts it, as we claim. Why did she get the impression that when we propose an overlap of processes that result in the emergence of *g* we in fact restrict this explanation to the level of the brain, leaving a unitary, singular psychological *g* intact?

In our opinion, the answer is in the approach to so-called elementary cognitive tasks (and therefore to "elementary" psychological processes) in the tradition established primarily by Eysenck and Jensen and to which Gottfredson explicitly subscribes. This approach assumes that such tasks and processes are somehow more "biological" than mental test scores; that is, they reflect brain functioning more directly. This line of reasoning is exemplified in Gottfredson's (this issue) statement on working memory: "Tests of working memory and other major constructs in cognitive psychology do not measure brain processes directly but provide psychometric 'analogs' of them" (p. 213). As Gottfredson acknowledges, this approach is historically intertwined with the methodology of the day: "EEG brain waves and reaction time on exceedingly simple tasks (e.g., touch a button when it lights up) were as close to the brain as [Eysenck] could get" (p. 213).

Yet, even though most intelligence tests are complex and thus probably tap a large number of cognitive processes, we disagree with the assumption that cognitive tasks that purportedly tap more "elementary" processes are more biological. In our view, they are just as psychological as mental tests. In general, the distinction should be made between (a) functions (elementary and otherwise) and (b) their anatomical substrate or physiology, rather than assuming that a psychological response is in itself a—more or less—*direct* measure of the underlying neural mechanisms. There is evidence, for instance, that performance on simple reaction time tasks depends on much more than processing speed (e.g., Conway, Kane, & Engle, 1999). As well, with regard to measures of executive functions, the concept of *dysexecutive syndrome* was invented exactly to separate anatomy and function and to differentiate certain symptoms from the anatomy-based diagnosis of *prefrontal syndrome* (Baddeley & Wilson, 1988).

A broader discussion of levels of analysis must eventually lead to the larger issue of reductionism in psychology: whether psychological phenomena in general can be explained by biological phenomena. However, regardless of one's position, "process measures" and "ability measures" are, in our view, on the same level of explanation. That is, either both the Raven's matrices and operation span are "biological" or both of them are "psychological."

Let us turn now to the influences we did cite and to whose work we did compare to POT. We explicitly claimed that POT is a modern sampling theory. Deary, Cox, and Ritchie (this issue) correctly point out that our influence has been Thomson's early criticism of Spearman's *g* rather than his late theorizing. They write that "we should remind readers that [Thomson] 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" (p. 193). This is exactly what we argued for: *g* (or, to be more precise, reflective *g* or psychological *g*) is a sufficient but not necessary cause of the positive manifold—and, following Thomson's pioneering idea, it

can also be explained with an overlap of processes. Provided that *g* does not seem to have a within-individual interpretation, one should reasonably look for sampling as a viable alternative option. Admittedly, we have much less been influenced by Thomson's (and Thorndike's) bonds theory, for reasons already outlined in the target article. Hence we adopted Thomson's general line of explanation relying on a functional overlap rather than his actual theory of what overlaps and how.

Yet POT differs even from (early) Thomson's general idea of sampling/overlap, and because there are commentators who seem to have misunderstood our model, it is imperative to explain how. Oberauer (this issue), for instance, summarizes the main idea of POT as "Performance in two cognitive tests is positively correlated to the degree that their samples of processes overlap" (p. 231). This is a correct description of Thomson's ideas, as expressed by his dice-throwing experiment, but not of POT. As we wrote in the target article,

> Process overlap theory proposes a nonadditive overlap of psychological processes. In particular, the executive/attentional processes that typically overlap with domain-specific ones function as a bottleneck: Failure to pass the executive demands of a test renders individual differences in specific processes unimportant for overall performance. As a consequence, *the correlation between tests is not simply the function of the sheer number of overlapping processes in relation to the total number of activated processes, as in Thomson's account* [emphasis added]. (p. 170)

So, unlike Thomson, we do not propose a linear relationship between the extent of process overlap and the size of correlations, and this is a crucial difference. This aspect of POT is expressed in the M-IRT model, and we return to this point later when discussing that model.

Parts of POT have also been interpreted by Deary et al. (this issue, p. 192) as "a re-statement of Anderson's (1992) theory of intelligence differences." We disagree. Anderson's theory of minimal cognitive architecture is practically a marriage of the mental speed hypothesis of *g* and the theory of massive modularity of cognitive architecture (originally invented by Fodor, 1993, but heavily criticized by Fodor, 2000, himself later on). Unfortunately, we do not subscribe to either. The evidence leans against informational "encapsulatedness," which is a primary characteristic of Fodor (1993) and thus Anderson's modules. Instead of massive modularity, a cognitive architecture of component processes incorporating both modules and central systems (see Moscovitch, 1992, as an example for memory) seems more plausible—and admittedly more in line with our theorizing of individual differences in cognition. Obviously, contrary to Anderson, we do not subscribe to the mental speed theory of the general factor either. So in short: We believe that POT and minimal cognitive architecture provide drastically different cognitive explanations to the problem of the positive manifold.

Even though the systems theory by Detterman is one that we clearly referred to in the target article as a major influence, and consequently we have not been accused of not citing it, we grab the opportunity to mention it again here. In fact, even though POT has been labeled as a modern sampling theory, it could have also been reasonably labeled as a centralized system theory. Therefore, we read the simulation in the commentary by Detterman, Petersen, and Frey (this issue) with particular

interest. We are delighted that their results are in agreement with POT. Even more so because—as a natural extension of the target article—we have also started working on our own simulations. The main difference between their simulation and ours is that theirs has a different take on the bottleneck effect: They simulate a number of process scores and the worst score functions as the total limit for all the others. Our simulation, on the other hand, is based on our M-IRT model and thus employs a probabilistic rather than deterministic approach, which is arguably more appropriate. Also, to simulate the neural basis of overlapping processes, we employ a network-based approach to selecting the components that interact to create the total probability of passing an item.

At this point it is worth noting that, to our surprise, the M-IRT model is an aspect of the target article that remained largely unnoticed by the commentators. In fact, only Kan et al. (this issue) seem to address the model directly. This is unfortunate because the model is strongly linked to the main theme of this section, whether POT is a new idea or not. In our opinion, the really novel aspects of the theory are probably easiest to grab in the M-IRT model, which provides a mathematical formulation not only of the nature of functional overlap between different cognitive domains but also of the bottleneck effect created by the limitations of executive functions that is central to the theory. As noted before, this aspect of the theory is the most pronounced difference when compared to (early) Thomson's ideas on sampling.

Moreover, we believe that, besides the *actual* model, linking item response theory *in general* with a cognitive theory is unusual in the field. Of course, this on its own does not make it valuable. But we believe that there is reason to claim that item-level modeling is the right approach to formulate theories of mental test performance. In other words, it is at the level of items that cognitive processes should be modeled. There have been simulation-based studies of intelligence, but they usually use total scores. An item-level approach is more appropriate because, in fact, the behavior we are trying to explain is a person providing either a correct or an incorrect answer to a test item, and the processes we are trying to explain are the ones responsible for this behavior. An IRT-based approach also makes it possible to model differences between items appearing in the same test, which increases its scope of explanation compared to total score based models, and allows different processes to contribute to success to a different extent on items with different requirements and/or difficulty within the same test.

Finally, we would like to comment on the mutualism model once more. Even though it is not a direct precursor to POT, it also questions the existence of psychological *g*, and is compatible with an emergent approach and formative model of the general factor of intelligence. Yet it is also very different, because it assumes developmental interactions instead of a functional overlap (which is mathematically formulated as a nonlinear growth model rather than a multidimensional item response model). Yet, despite this difference, the two can be almost indistinguishable, because if one includes executive processes in the mutualism model to have very high interaction weights (M) with all other processes, the end result will be almost identical to the one in POT.

So why not simply use this solution instead of a new theory and a new model? The answer lies in the ample evidence about the very limited extent to which transfer occurs—as cited by Oberauer (this issue) in relation to POT, to which we return in the section discussing the predictions of POT. That is, in our opinion mutualism is more easily interpreted as the causal factor behind the development of specific abilities than of the large across-domain variance. Also, regardless of whether developmental evidence backs up mutualism or not, functional overlap between processes, as well as neural overlap between specific networks, does seem to take place when people solve a mental test item.

Empirical Evidence in Support of Process Overlap Theory

Several commentators—in particular, Ackerman, Colom, Chuderski, and Santarnecchi, and Oberauer—question the strength of one or more empirical findings that we cited as evidence in support of POT. In many instances we disagree with the conclusions these authors have drawn based on existing data. In this section we therefore address three lines of empirical evidence that were called into question by at least one of the commentators: (a) the relationship between executive functions, g, and fluid intelligence (Gf); (b) the distinction between simple span and complex span tasks; and (c) ability differentiation.

Executive Function, g, and Gf

POT doesn't make a *specific* prediction about the relationship between executive function abilities and g, or Gf, other than that they are all multicomponent systems with overlapping processes. However, given the proposed central role of executive attention in test performance and fluid reasoning, there should be moderate to strong correlations between executive functions and both g and Gf. According to Colom et al. (this issue) and Oberauer (this issue), the evidence for this relationship is weak. Colom et al., for example, claim that "studies addressing the relationship between executive tasks and Gf are consistent with the conclusion that it is weak and unstable" (p. 182). Oberauer agrees: "My reading of the literature on the correlation between measures of executive functions and g (or working-memory capacity) is that they explain no more, and probably less, than 10 percent of the variance" (p. 233).

Colom et al. and Oberauer base these arguments on two studies by Miyake, Friedman and colleagues (Friedman et al., 2006; Miyake et al., 2000) and a series of studies by Oberauer and colleagues (Keye, Wilhelm, Oberauer, & Sturmer, 2013; Keye, Wilhelm, Oberauer, & van Ravenzwaaij, 2009; Oberauer, Süß, Wilhelm, & Wittmann, 2008; Wilhelm, Hildebrandt, & Oberauer, 2013). In fact, only two of these studies (Friedman et al., 2006 and Oberauer et al., 2008) directly examined the relationship between executive function and intelligence. Miyake et al. (2000) is a latent variable study on the relationship between three types of executive function: inhibition, updating, and shifting. The Oberauer studies examined the relationship between executive function and working memory capacity, not

intelligence. So why are Colom and Oberauer so pessimistic about the relationship between executive functions and intelligence? First, Friedman et al. (2006), which is based on Miyake et al., found that only updating was related to Gf ($r = .64$). The relationship between inhibition and Gf was not significant ($r = .29$), nor was the relationship between shifting and Gf ($r = .13$). We agree that these latter two correlations are inconsistent with POT. As well, Oberauer and colleagues consistently fail to find significant correlations between working memory capacity and executive function as measured by conflict sensitive measures derived from the Simon and Eriksen flanker paradigms. These results, too, are difficult to reconcile with POT.

However, these results stand in stark contrast to a number of positive findings with respect to the relationship between executive function and intelligence, as well as working memory capacity. Colom et al. and Oberauer fail to cite several recent studies that show a moderate to strong relationship between attention control and fluid intelligence (Shipstead et al., 2014; Unsworth et al., 2014; Unsworth & Spillers, 2010; Unsworth, Spillers, & Brewer, 2010). In each of these studies multiple tasks per construct were administered, and in each study the data clearly demonstrate a strong link between latent variables for attention control and Gf. In order of magnitude: Unsworth and Spillers (2010), $r = .45$; Shipstead et al. (2014), $r = .69$; Unsworth et al. (2010), $r = .70$; Unsworth et al. (2014), $r = .77$. As well, Unsworth et al. (2009) found strong and significant correlations between Gf and various executive functions: fluency ($r = .58$), response inhibition ($r = .76$), and vigilance ($r = .52$). Finally, a latent variable study specifically designed to investigate the relationship between executive function and working memory capacity found the two constructs to be nearly identical ($r = .97$; McCabe et al., 2010).

In addition, Miyake et al. (2001) found strong correlations between executive function and various spatial reasoning tasks that consistently reveal high g loadings in the intelligence literature. Moreover, they interpret their findings as consistent with the importance of attentional factors for g.

Why do some studies find such weak evidence for the relationship between executive function and intelligence/working memory capacity and others find a strong relationship? We don't claim to have the definitive answer here, but it seems that one important factor pertains to the measurement model adopted by different research teams. One approach, adopted by Oberauer and colleagues, is to link specific executive functions to individual tasks. The other approach is to administer multiple tasks per executive function and derive a latent variable from the covariance observed across task paradigms (for examples, see McCabe et al., 2010; Unsworth et al., 2014; but for an exception, see Friedman et al., 2006).

It is also possible that executive functions are more strongly related to one another, and to fluid intelligence/working memory capacity, at lower levels of ability, and especially in clinical populations:

> Although the current data do not speak directly to the relations between EFs and intelligence in these populations, one possibility suggested by Rabbitt et al. (2001) is that when frontal lobe functioning is generally compromised, multiple EFs may be affected, leading to higher inter-EF correlations. These higher correlations could

then result in generally higher EF-intelligence correlations. Indeed, Salthouse et al. (2003), examining an aging sample, found substantially higher inhibiting-updating (.71), inhibiting-Gf (.73), and updating-Gf correlations (.93) than those found here. (Friedman et al., 2006, p. 178)

This particular result is in complete agreement with the bottleneck effect proposed by POT. As well, the fact that latent variable approaches are generally more effective in establishing such correlations than individual executive tasks is also in agreement with how intelligence and executive functions are conceptualized by POT as multicomponent systems.

Overall, our reading of the evidence is definitely *not* that there is a lack of relationship between executive function and intelligence. There seem to be contradictory results, with the majority of studies demonstrating a substantial proportion of shared variance, whereas others report null or minimal effects. A more complete understanding of the nature of executive functions along with the development of more reliable tools to measure them should lead to a clarification of what moderates the relationship between executive function and intelligence. Provided that "the central executive in the original working memory model is little more than a homunculus, a little man who takes all the decisions that are beyond the capacity of the slave systems" (Baddeley, 1998, p. 524), we have come a long way to understanding this component of working memory and the control processes of the mind in general. Undeniably, there is still a long way ahead. But in our opinion, our current understanding of executive functions and their relation to intelligence does provide the necessary foundation for POT.

Gottfredson (this issue) also discusses the relationship between *g* and Gf, independently of executive functions. Following Jensen, she proposes that *g* and Gf are "one and the same," extending their statistical (almost)-unity to conceptual identity: "This makes theoretical sense because both manifest as a domain general capacity for reasoning and solving novel problems" (p. 211). In our view, this is contradicted, among other things, by exactly the distinction that Gottfredson makes: "Crystallized *g* begins to level off but fluid *g* tends to decline in tandem with the aging of body and brain" (p. 211). She argues, "For these reasons I conceptualize *g* in terms of fluid *g* when speaking of Stratum III's general factor, *g*" (p. 211).

In our opinion, such a fractionation of *g* based on aging or other reasons is a basis of not conceptualizing it as a singular, unitary construct, rather than identifying it with one of its components. Contrary to Gottfredson, POT conceptualizes Gf as a trait reflecting novel problem solving but *g* as a formative trait emerging from the interaction of overlapping component processes. Their identity is thus restricted to the statistical level—a phenomenon that POT does explain.

The Distinction Between Simple and Complex Span Tasks

In our target article we make the distinction between simple span and complex span tasks. In a simple span task, also known as immediate serial recall, the subject is presented with a list of stimuli and is required to recall the list in serial order. The number of items per list varies, typically from two to nine, and the stimuli are presented fairly rapidly, for example, one per second. The digit span task is a well-known example. In contrast, complex span tasks involve both processing and storage of information. For example, in the counting span task, the subject is presented with an array of colored shapes (e.g., blue and red circles and squares) and is instructed to count the instances of a target (e.g., blue circles) and remember the total for later recall. After a series of displays, the subject is required to recall all the counts in serial order. It is important that the storage and recall demands of simple and complex span tasks are identical (i.e., serial recall of digits), but complex span has an additional processing component (i.e., count the blue circles).

Despite more than 30 years of research on the important distinction between simple and complex span tasks (for a review, see Unsworth & Engle, 2007), Colom et al. (this issue) present their own review of the literature and conclude that "complex span tasks cannot be clearly distinguished from simple span tasks" (p. 183). This conclusion is difficult to reconcile with several findings, including (a) different serial position curves: the recency effect is small in simple span and is exaggerated in complex span (Unsworth & Engle, 2007); (b) different error patterns: transposition errors are most common in simple span whereas omissions are most common in complex span (Unsworth & Engle, 2007); (c) different effects of phonological similarity: there is a phonological similarity decrement in simple span and a phonological similarity facilitation effect in complex span (Chow, Macnamara, & Conway, 2016; Macnamara, Moore, & Conway, 2011); (d) different neural correlates: both simple and complex span tasks exhibit activity in lateral prefrontal, anterior cingulate, and parietal cortices, but complex span exhibits greater activity in the medial temporal lobes during recall (Chein, Moore, & Conway, 2011); (e) different correlations with tests of educational achievement (in fact, this was the original reason for the interest in complex span, because contrary to simple span they were able to predict Scholastic Aptitude Test scores; see Daneman & Carpenter, 1980; Turner & Engle, 1989); and (f) markedly different correlations with tests of fluid intelligence—but less so for crystallized intelligence or clerical speed; see Figure 6 in the target article.

We agree with Colom et al. that the distinction between simple and complex span is not always observed but instead depends on how simple span tasks are administered and scored. When active maintenance, especially via articulatory rehearsal, is encouraged, for example, by using short lists and a relatively slow presentation time, then simple span tasks are not as predictive of higher order cognition as complex span tasks (for a review, see Conway, Getz, Macnamara, & Engel de Abreau, 2011). However, when rehearsal is prevented, and/or when longer lists are used, and/or when performance from longer lists is weighted more strongly by the scoring procedure, then simple span tasks are just as predictive of higher order cognition as complex span tasks (Colom et al., 2006; Unsworth & Engle, 2007).

The bottom line, to us, with respect to simple span and complex span tasks is that the distinction is undeniable (on the face of it, they are clearly different types of tasks), but both types of tasks tap many of the same cognitive processes. On this point we agree with Unsworth and Engle, who concluded that

"simple and complex span largely measure the same basic processes (e.g., rehearsal, maintenance, updating, controlled search) but differ in the extent to which these processes operate in a particular task" (Unsworth & Engle, 2007, p. 1039). Similarly, as already noted in the target article, in light of the available experimental evidence, we subscribe to the view that complex span tests "reflect primarily general executive processes and secondarily, domain-specific rehearsal and storage processes," whereas simple span tests "reflect domain-specific storage and rehearsal skills and strategies primarily and executive attention processes only secondarily" (M. J. Kane, Conway, Miura, & Colflesh, 2007, p. 24).

Ability Differentiation

Ability differentiation refers to the finding that the correlations among various tests tend to be larger at lower levels of ability than at higher levels. It was originally discovered by Spearman (1927), who referred to it as the "Law of Diminishing Returns," and has been replicated many times (Detterman & Daniel, 1989; H. D. Kane, Oakland, & Brand, 2006; Molenaar, Dolan, Wicherts, & van der Maas, 2010; Tucker-Drobb, 2009).

Despite this evidence, Colom et al. (this issue) question the strength of the case for ability differentiation. Colom et al. cite seven articles on differentiation. Of these, one is not a direct test of ability differentiation (Gignac & Weiss, 2015), four find evidence for ability differentiation (Abad et al., 2003; Deary et al., 1996; Detterman & Daniel, 1989; Lynn, 1992), and two do not (Fogarty & Stankov, 1995; Kroczek, Ociepka, & Chuderski, in press). Given that four of the six studies *that they cite* find evidence for ability differentiation, we admit to being perplexed as to why this is being called into question. As well, one study that did not find evidence for differentiation (Fogarty & Stankov, 1995) used an extreme-group design, comparing a high-ability group ($n = 25$) to a low-ability group ($n = 20$). This approach has been shown to be inappropriate for testing for differentiation (Molenaar et al., 2010; Tucker-Drobb, 2009), and the sample size is clearly insufficient ($N = 45$).

A more optimal approach to studying this phenomenon is to apply moderated factor analysis and either look for changes in the factor loadings as a function of g or use a proxy for g or Gf and look for external moderation. With this method, we have found evidence for differentiation in intelligence (Molenaar, Wicherts, & Kovacs, 2016) and working memory (Kovacs, Molenaar, & Conway, 2016). As well, Molenaar et al. (2010) and Tucker-Drobb (2009), using moderated factory analysis, found evidence for differentiation in intelligence. Overall, the available evidence clearly demonstrates that ability differentiation does exist.

Finally, not only the existence but also the importance of phenomena that POT heavily draws upon has been called into question. According to Deary et al. (this issue), differentiation and the worst performance rule are "relatively small-scale phenomena that are not particularly important for a general theory of intelligence to explain" (p. 193). We disagree, and we are happy to see that so does Kaufman (this issue), as well as Detterman et al. (this issue).

But their importance is not even the real issue. Differentiation highlights that g is a population-dependent phenomenon,

which is a very important result. With regard to the worst performance rule, we agree with Detterman et al. (this issue):

> When a person is performing at their worst, it is reasonable to expect that it is because important elements within the system are at their lowest levels. In other words, it provides the lower bound for performance and indicates how badly central elements can perform. (p. 203)

Differentiation and the worst performance rule are not only interesting phenomena in their own right (even though, as Kaufman, this issue, points out, "the cause of these two findings has never been satisfactorily explained" [p. 229]). Their real significance, in our opinion, lies in canalizing the line of explanation a theory of the positive manifold has to take. That is, such a theory has to explain why the positive manifold is stronger in *certain populations* rather than others and why a general factor extracted from the positive manifold correlates *with a certain level of performance* more than with others. Kaufman correctly summarizes that, according to POT, both phenomena occur because "individual differences in executive processes can serve as a bottleneck for cognitive functioning across a wide range of tasks" (p. 229). This might not be their only explanation, but still both phenomena serve as guidelines for an appropriate theory of the positive manifold.

Predictions, Consequences, and Implications of POT

Having discussed the foundations of POT, let us move to the predictions or implications of the theory that several authors have commented on. In the previous section we did our best to sensibly group topics, but because we wish to adequately respond to the honoring effort by many commentators who have performed simulations and other analyses, or cited challenging evidence, here we have to deal with issues commentary by commentary. At the same time, we still simply cannot answer every issue mentioned, so we focus on new analyses, new simulations, and the most central topics. We leave out the simulation by Detterman et al. because we have already addressed that in the previous section.

So we start with the simulation by Kan et al. (this issue). This seems to be very much in line with our conception and related assumptions. We are excited that their results are also in agreement with the factor structure predicted by POT. Yet, they also claim that (a) the general factor is not so much a variable *constructed out of* the verbal, visuospatial, and fluid factor but rather *is* the fluid factor, and (b) the results are also compatible with a bifactor model, as well as with the oblique model.

Indeed, in such an ideal case as the one depicted by the simulation, g and Gf might be completely identical, both reflecting pure executive functioning. Real test batteries are probably less balanced in terms of specific processes canceling out one another, and the empirical results are less ideal. This is probably the answer to the second issue as well: In such an ideal case, bifactor models with only Gf (instead of g) might fit, but such a factor solution will be problematic with real-life—and especially Gc-biased—batteries like the Wechsler scales.

Amazingly, Deary et al. (this issue) provide three small-scale studies to test POT. In the first, they test whether elderly adults with large frontal atrophy have more across-domain variance

in test results than people with low atrophy. They found the predicted difference (52.7% vs. 29.1%), even though it did not reach significance. We concur that it is in agreement with POT.

In the second study they compared the neural correlates of g and Gf within the same sample and found that the same areas were associated with g and Gf, where "the magnitude of associations for all sub-regions for g and Gf were near-identical (vector correlation for surface area: $r = .98$, for volume: $r = .99$)" (p. 197). They conclude that "these data provide clear evidence that g and Gf are virtually identical in terms of bivariate associations, and also with respect to their cortical correlates" (p. 197).

First of all, we must say that this result is quite unique and is definitely not in accordance with the bulk of the evidence from imaging studies, reviewed in the target article—even though, as Deary et al. correctly point out, correlates of g and Gf were not investigated in the same studies. Second, and related to this, it is informative to take a look on how they measured the constructs:

> To construct g, we used WAIS-III Digit-Symbol Substitution, a test of Choice Reaction Time, WMS-III Verbal Paired Associates, the National Adult Reading Test, and Verbal Fluency. … To construct Gf, we used Matrix Reasoning, Block Design, Letter-Number Sequencing (from the WAIS-III) and Spatial Span (from the WMS-III). (Deary et al., this issue, p. 198)

Apparently, their g factor is a mixture of Gc (with three measures: WMS-III Verbal Paired Associates, the National Adult Reading Test, and Verbal Fluency—with verbal fluency probably tapping fluid abilities as well) and speed (Gs, with two measures: WAIS-III Digit-Symbol Substitution and a test of Choice Reaction Time). Gf, on the other hand, is measured with a typical test of Gf (Matrix Reasoning), along with a test of visuospatial ability (Gv, Block Design), a spatial working memory task (Spatial Span), and a transformational working memory task (Letter-Number Sequencing).

In our opinion, their conception of g and Gf is somewhat unorthodox, which has important consequences for testing POT's predictions. It certainly does not falsify POT that what Gc and Gs have in common, on one hand, and what Gf, Gv, and working memory have in common, on the other, actually have a lot in common—even at the level of neural correlates. We might even say that this is actually predicted by POT.

Their third analysis, as Deary et al. (this issue) admit, "does not directly test a prediction of process overlap theory" (p. 198). In this fascinating study they fitted a model in which formative biological elements produce a reflective g factor. This approach to modeling g is an alternative to the structural model offered for POT (POT-S, by Kan et al., this issue) and is more in line with traditional, g-oriented approaches to brain functioning (see Gottfredson's commentary and our earlier reply). But because Deary et al.'s actual model does not include a fluid reasoning factor, it is not possible to directly compare the two.

Indeed, should further research on the brain basis of g provide a different picture than it does today, and should it be reasonable for a psychological g to congruently and ubiquitously reflect the same independent biological sources (which it does not do, according to our current knowledge), this family of models could be superior to POT-S.

Apart from performing new analyses to test POT, a number of commentators pinpointed empirical phenomena that seem to contradict what one could expect from POT. An objection raised by Kan et al. (this issue) is that POT cannot explain the Jensen-effect, "the finding that the general factor is more heritable than specific factors, such that subtests' factor loadings on the general factor and heritability coefficients are positively correlated" (p. 221). In particular, they point out that

> POT does not make any claims regarding the heritability of the cognitive abilities, their underlying capacities, hence general intelligence. One simple explanation is that as each of the underlying variables are to some extent heritable, their sum is also heritable. However, in itself this will not provide an account for the relation between factor loading and heritability, thus for the way the Jensen-effect arises. (p. 224)

Kan et al. are correct that POT does not make direct, specific predictions about heritability. Yet it does make the indirect prediction that the heritability of executive functions is central with respect to the heritability of g. Moreover, it is exactly because of this indirect prediction that we disagree with their claim that total heritability as the sum of the heritability of component processes will not provide an account of the Jensen-effect. There is a very simple way to reconcile this kind of additive heritability (which we in fact did not explicitly propose in the target article) and the Jensen-effect: *if the heritability of executive functions is substantially higher* than the heritability of domain-specific abilities or processes. This would result, according to POT, in a correlation between heritability and g-loading even in the case of summed heritabilities of processes.

Actually, there is evidence that this happens to be the case. A study compared the heritability of three brain networks involved in different aspects of attention: an "orienting network" (responsible for orienting to sensory events), an "alerting network" ("developing and maintaining the alert state"), and a "conflict network" ("executive control used in resolving conflict between stimuli and responses"). Clearly, of these three networks, the last is the one strongly related to the executive functions that, according to POT, are central to mental test performance. The study found that the heritability of the orienting, alerting, and conflict networks is .18, .00, and .71, respectively (Fan, Wu, Fossella, & Posner, 2001).

Another study that directly measured the heritability of executive functions found it to be so high that it concluded that variation in executive functions is almost entirely genetic in origin. In the same study they found the heritability of overall IQ to be smaller, with approximately 30% environmental effect (Friedman et al., 2008). A third study (McClearn, 1997) found that the heritability of specific abilities was generally lower than that of a general cognitive ability index (with genetic variance being 55% for verbal ability, 32% for spatial ability, 62% for speed of processing, 52% for memory, and 62% for general cognitive ability). Overall, our reading of the available evidence is that the Jensen-effect can be reconciled with POT, because executive functions do seem to have higher heritability than other aspects of cognition.

Flynn (this issue) adds another item to the list of challenges for POT. He cites Flynn, te Nijenhuis, and Metzen (2014), who meta-analyzed "the Wechsler subtests scores of typical subjects and those who suffered from iodine deficiency, prenatal cocaine exposure, fetal alcohol syndrome, and traumatic brain injury"

and found that even though "typical subjects were higher on every subtest, the magnitude of their advantages by subtest had zero correlation with the size of the subtest g loadings" (p. 207). The challenge is:

> If we substitute for g the three-factor concept of induction, working memory capacity, and executive function, should there not be a correlation between the extent to which this package is relevant to the subtest and the score difference between normal and damaged subjects? Unless these maladies collectively (and indeed virtually singly) damage the prefrontal lobes in a way that somehow cancels out their differential contribution to the cognitive task set by the different subtests, perhaps by reducing its contribution in all cases to a minimum. This does not seem very plausible and the authors may wish to comment. (Flynn, this issue, p. 207)

Our response here is very similar to the one Flynn provided to Rushton to a similar challenge (Flynn, 1999). Rushton (1999) claimed that IQ gains are not related either to g or to inbreeding depression, whereas Black–White differences in IQ are related to both. As evidence, Rushton cited a principal components analysis, in which inbreeding depression, Black–White differences, and g loadings of the Wechsler scales form a cluster, whereas IQ gains appear on another cluster. This was supposed to demonstrate that whereas IQ gains are environmental, Black–White differences and g are genetic in origin.

Flynn replied:

> You get Rushton's clusters at all only because of a negative correlation between g-loadings and IQ gains over time. And that negative correlation is a product of two things: the WISC battery is biased towards crystallised g; and crystallised g is biased against IQ gains. (1999, p. 391–392)

Flynn also offered an alternative approach: To focus on Gf instead of Gc, one has to look at the Raven's correlations of each subscale of the WISC. This way IQ gains correlate positively with g-loadings and we get a more valid measure of the component of g that is the actual subject of IQ gain—which do manifest themselves more strongly on tests of Gf than Gc.

We completely agree with both of Flynn's claims. First, the Wechsler scales are indeed biased toward Gc. Because Flynn has clearly stated this in his writings, we are surprised as to why in this issue he suggests that "we substitute for [Wechsler!!] g the three-factor concept of induction, working memory capacity, and executive function" (p. 207). Wechsler g is not any more strongly linked to these constructs as to IQ gains, and probably for the same reason.

Second, we believe a correlation of each subtest with the Raven's matrices is much more optimal as the basis for comparison here, too—in his reply to Rushton with inbreeding depression and IQ gains, in our case with the magnitude of advantage of healthy subjects. We have not carried out this analysis, but if that would still find a zero (or negative) correlation, that would be indeed problematic for POT—under the condition that these four deficiencies univocally result in large prefrontal/posterior parietal damage relative to other parts of the brain.

Finally, Oberauer (this issue) cites a lack of transfer from training studies as evidence against POT. First, he challenges the existence of general processes based on training studies. Before we evaluate this argument, let us point out that evidence clearly points to the existence of general mechanisms of

cognition, and we need to emphasize the relevant results from neuroscience (see the section on overlapping brain networks) that we believe to be much more relevant than the issue of transfer.

Moreover, we are unfortunately not sure that we completely follow his reasoning. First, he says—claiming that it is an implicit assumption of training studies—that "if domain-general processes of major importance for performance exist, we should expect them to benefit from practice" (Oberauer, this issue, p. 232). We fail to comprehend this syllogism. Why would the existence of domain-general processes imply that they are malleable/trainable as a logical necessity?

But let us move on and assume that general processes just have to be trainable. Yet, according to the evidence, there is no transfer: "The stubbornly narrow scope of transfer of practice poses a challenge for the assumption in process overlap theory that there are domain-general processes—in particular executive processes—that are enrolled in a multitude of tasks across different content domains" (Oberauer, this issue, p. 232). As we pointed out before, we believe that a *general* lack of transfer is more problematic for the assumptions of the mutualism model than for POT.

There is a specific and relevant point, however, that Oberauer makes with regard to transfer: "If people improve massively through training on a task with strong demands on executive functions, should we not expect strong transfer of training effects to other tasks also making heavy demands on executive functions?" (p. 232). The short answer is: It depends. POT proposes that executive functions serve as bottlenecks for diverse cognitive performance. It also proposes that there are multiple executive functions involved in different tasks in an overlapping fashion.

So there are a number of conditions for the answer to Oberauer's question to be "yes." Task on which transfer should be experienced (a) has to tap the same executive functions, at least in part, and (b) has to have executive demands high enough so that the effect is substantial (remember: The M-IRT model for POT implicates an asymptote for the role of executive functioning in performance). Also, (c) the executive functioning in the trained population should be at a level where training improves the probability of a correct answer on the other task (remember: The same asymptote is relevant for the individual as well, because POT proposes an interaction between the executive demands of the task and the executive functioning of the individual).

In other words: Because the M-IRT model implies that the executive and the domain-specific component of a single task can be modeled as two subtasks in terms of the probability of success, executive training should be effective only if it meaningfully transfers to the other task (i.e., that other "executive sub-task" has not been "solved" already, and the training increases the chances of solving it).

Overall, a lack of trainability or a lack of transfer does not necessarily imply a lack of general processes. We believe that results of studies on transfer and training do not directly contradict POT but the results of particular training studies might pose a challenge. Alas, digging deep into the literature of cognitive training, one gets the impression that this whole matter poses a challenge for everyone.

Future Directions: Escape From Groundhog Day

In his commentary, Sternberg compared us to the protagonist of the movie *Groundhog Day*, who gets stuck in a time loop. The analogy is supposed to demonstrate that our target article is practically no different from all the myriad previous attempts to explain *g*. We are puzzled, because we thought that POT exactly breaks free from Groundhog Day. So far, the mainstream conception of psychometric *g* interpreted it as a psychological mechanism, resulting in a reflective model, in which a mysterious concept accounts for individual achievements on a number of different ability tests as well as the covariance between these tests. By explaining the positive manifold without a psychological *g*, POT proposes a whole new conception of intelligence.

It is unfortunate that, along with the M-IRT model (POT-I), the formative *g* concept (POT-S) remained largely unnoticed by many commentators, because it is probably the other really "unorthodox" aspect of the theory (even though mutualism also leads to the same kind of conceptualization for *g*; see van der Maas, Kan, & Borsboom, 2014). Reconceptualizing *g* as a formative construct is, in our opinion, exactly how we can break free from Groundhog Day. Therefore, when Sternberg (this issue) writes that "we don't have to be locked forever into more and more studies of the bases and correlates of *g*" (p. 239) and when he juxtaposes his results on the lack of importance of *g* in Kenya with POT, he breaks through two open doors at the same time.

First, contrary to what Sternberg asserts, POT's *g* is not simply another process-based explanation of reflective/psychological *g*. It is a process-based explanation of the positive manifold *without* proposing a causal *g* factor. Hence, POT does not explain reflective/psychological *g*, it explains reflective/psychological *g away*. Under this framework one can, then, as Kan et al. (this issue) point out, "focus on those lower order variables that do allow for a realist, causal interpretation" (p. 220).

Second, POT's (formative) *g* can be characterized as an index of cognitive abilities, as measured by mental ability tests developed in Western, industrialized societies. This *g* is a valuable construct only insofar as it can predict important real-life criteria from education to the workplace and beyond. If it does a valuable job in doing so, for which there is evidence (Gottfredson, 2007), then so far so good.

Gottfredson (this issue) argues that "the many biological and sociological correlates of *g* helped demonstrate that *g* was no chimera of factor analysis" (p. 214) and that "*g*-theorists believe that psychometric g is an emergent property of the brain but also that, as the brain's unitary product, g generates a cascade of effects in the real world" (p. 213). In our view, *g* is not a unitary product of the brain, but no chimera either. According to POT, the many sides of intelligence research depicted in Gottfredson's figure reflect completely different phenomena and not different aspects of the same thing. Yet, as a formative variable, not reflecting any real trait and not being causal to the positive manifold, *g* can still have many sociological correlates and can predict a cascade of phenomena in the real world.

Yet if, as Sternberg argues, *g* is an insufficient predictor of success in Kenya or for Australian aboriginals (see Flynn's commentary), then the measurement of other characteristics becomes necessary for that particular cultural context. In general, if it loses its value for prediction, formative *g* can be discarded as a construct, without this having any relevance for research on the causes of individual differences in cognition.

In fact, Gottfredson and Sternberg have been arguing about the importance of *g* "in real life" for decades. POT does not intervene in the content of this prolific debate; it only puts *g* in a different context as it conceptualizes it is a consequence, not a cause of the positive manifold. Regardless, as an index of mental abilities, it may, or may not, be enormously useful in real life, in the Western world and beyond. We leave it to researchers on the sociological side of intelligence to continue that debate.

A thorough elaboration of the concept of formative *g* is beyond the scope of the target article, let alone this commentary. Actually, besides the M-IRT-based simulation, this is the second line of extension of POT that is in progress at the writing of this response. *g* as an emergent property, characterized by a formative latent variable, is the scaffolding for a "new era for intelligence research" (Conway, 2014, p. 33). There are a number of characteristics of this new approach, and even though we cannot discuss all, two of them we must emphasize here: the focus on specific abilities instead of "general intelligence" and (b) a shift from explanation to prediction as a measure of validity for *g*.

Speaking of validity, we need to address the related criticism from Ackerman, who asks, "How can an adequate theory be so firmly disconnected from any considerations of external validity?" He also claims that "until such a time when process measures can even approach the validity of these intelligence tests, it is not reasonable to say that one has developed an adequate theory of the general intelligence factor" (p. 179).

First, we disagree about the importance of predictive validity as an indicator of the value of POT as a scientific theory. POT explains the positive manifold of cognitive abilities measured by mental tests—tests that have been used extensively in educational and occupational contexts. It does not purport to establish the value of the tests themselves. On the other hand, as we said, a formative *g* indeed has to have predictive validity, but it is not directly related to the evaluation of POT as an explanation of the positive manifold. (In other words: If one simply removes the formative part of POT-S, the explanatory part is still functional and translates to an oblique model without a higher-order general factor.)

Having said that, we do wish to respond to Ackerman's (this issue) remarks regarding validity, because we find them to be completely detached from the entire body of research on these constructs. Ackerman claims that process measures, like tasks tapping working memory, are detached from important real-life criteria. That is not the impression we get from reading the literature. Working memory capacity (WMC) has been demonstrated to predict performance on the SAT (Daneman & Carpenter, 1980; Turner & Engle, 1989), as well as educational aptitude and achievement in general (Alloway & Passolunghi, 2011; Blankenship, O'Neill, Ross, & Bell, 2015; Foroughi, Barragán, & Boehm-Davis, 2016). In particular, it has been demonstrated that WMC in kindergarten is a better predictor of school achievement than IQ (Alloway & Alloway, 2010), and better WMC in kindergarten reduces the risk of high school

dropout (Fitzpatrick, Archambault, Janosz, & Pagani, 2015). Working memory predicts vocabulary learning (Daneman & Green, 1986) as well as performance on a number of secondary or long-term memory tasks (Unsworth, 2010; Unsworth, Brewer, & Spillers, 2009). Moreover, tasks measuring WMC can predict diverse real-life phenomena ranging from early onset Alzheimer's (even before standardized psychometric tests, Rosen, Bergeson, Putnam, Harwell, & Sunderland, 2002), through simultaneous language interpreting performance (Macnamara & Conway, 2015) to skill in Texas Hold 'Em poker (Meinz et al., 2011).

Ackerman (this issue) also claims a lack of predictive validity for tests of Gf in general and the Raven's progressive matrices in particular: "Those who followed Spearman in attempting to measure *g* for application purposes were often substantially disappointed by the lack of validity shown by Raven's Progressive Matrices for real-world criteria, such as job performance (e.g., see Vernon & Parry, 1949)" (p. 179).

In 2016 we find a claim like that with a 1949 reference bizarre. An outstanding number of studies have demonstrated the predictive validity of the Raven's matrices, for both school and the workplace. A supplement of the test's manual has pages of references to individual validity studies (Raven & Court, 1989). A relatively recent publication reviews the evidence for the validity of the Raven's matrices with an emphasis on occupational performance (Raven, 2000). This publication also has numerous references to support the predictive validity of these tests for occupational (as well as educational) success. Most interesting, it cites the same study as Ackerman as *evidence for validity*:

> Vernon and Parry (1949) summarised the results of testing 90,000 British naval recruits with a short, non-cyclical, version of the SPM [Standard Progressive Matrices] during the Second World War. There were systematic differences in the mean scores of men from 12 general classes of occupation: clerical, electrical workers, precision workers, woodworkers, sheet metal workers, machine operators, retail tradesmen, building workers, "mates," drivers, farm workers, and labourers. (p. 62)

Overall, it appears that there *was* a correlation between Gf and the cognitive demands of different jobs, but according to Ackerman there was no correlation with job performance. This would mean that clerical or electrical workers had higher SPM scores than building or farm workers, but within each profession SPM did not predict success. In our opinion it hardly undermines the validity of the construct of fluid intelligence that having more of it did not necessarily make one a better farm worker in 1949.

And even if the 1949 study had indeed found a lack of validity and no studies with the opposite result had been carried out since, would that result still be relevant? Unlikely. Enormous changes in school curriculum—and in job demands—have taken place since. In fact, as Flynn argued on several occasions (e.g., Flynn, 2007), it is exactly such changes—that require more and more abstract ("scientific") thinking—that might be responsible for the Flynn-effect in general and for the finding in particular that IQ gains have been highest on tests of fluid reasoning, whereas gains on crystallized knowledge have been minimal.

Current discussions of the future of the job market also emphasize that a large proportion of today's schoolchildren will eventually be employed in jobs that do not exist yet. Rapid technological changes are transforming the world of work, and arguably fluid intelligence that enables one to adapt to ever-changing environments and solve problems in novel situations is more important in jobs today than 67 years ago, and it will be probably even more important 67 years from now. Probably the opposite trend takes place with regard to task-specific knowledge.

Besides a focus on prediction, another important feature of a formative model, like the upper part of Figure 8 in the target article, is the location of unexplained variance. In a reflective model (like the lower part of the same figure) the latent variable does not explain the entire variance in the measures; hence there is individual variance in each (denoted with ε). In a formative model, on the other hand, the latent variable is a common consequence rather than a common cause, and thus does not explain variance in any of the measures (or lower order factors). On the contrary, it is a consequence of variance in the measures or factors, but in such a model it is the latent variable that can have unexplained variance (denoted as ζ).

This, in less technical terms, means that according to POT-S process overlap is probably not the sole causal force behind across-domain variance, that is, for the positive manifold to emerge. Indeed, it seems that there might be other, independent sources, such as (a) white matter tract integrity (e.g., Penke et al., 2011), (b) mutualism (van der Maas et al., 2006), (c) associative learning (Kaufman et al., 2009), and (d) environmental effects (Dickens & Flynn, 2001). Having said that, in our opinion the functional overlap of processes is probably the primary reason, and it is capable of explaining most of the across-domain variance.

Approaching the limits of this reply, we wish to repeat our gratitude to the commentators who have provided extensions and new interpretations to the theory. We are particularly grateful to Kan et al. for introducing the concepts of POT-S, POT-V, and POT-I. As well, we are excited by their suggestion to add a time dimension (the "subscript t") to the model in order to range its explanatory scope.

Oberauer's suggestion that it is overlapping parameters rather than overlapping processes that are responsible for an emergent *g* is food for thought, especially because we have to confess that we struggled with what to call these *things* that overlap. At a given point of the development of the theory they were called components, later on component processes, but they ended up being just processes. Accordingly, we referred to earlier versions of the theory as component process account and multicomponent model, before settling for POT, which emphasizes the actual causal factor behind the positive manifold. The expression "parameter" at first reminded us of unitary source models (like in the mental speed theory), but what Oberauer advances is clearly different, as he proposes multiple, independent parameters for cognition. We have to digest his suggestion further, but we do not find this, at a first sight, to be unreconcilable with our conception.

We are grateful to Cowan, Detterman et al., and Kaufman, who paved the way for future directions of our theory that we had not even thought of. In particular, Detterman et al. and

Kaufman discuss the consequences of POT for mental diagnosis (Kaufman with a particular interest in twice exceptional children and those at the highest part of the IQ-distribution). Cowan suggests that new tests could be devised on the basis of POT. Kaufman suggests that POT might have important consequences for creativity research. These are all intriguing possibilities that need to be further investigated (even though we are a bit skeptical about creativity, as a recent study found that its correlation with WMC is $r = .04$; M. J. Kane et al., in press).

We also agree with many commentators in that it is further research on executive functions and the neuroscience of individual differences in cognition that will really extend the explanatory power of POT. Oberauer (this issue) judges that "the term of executive functions is used with a variety of meanings, with frustratingly little agreement among researchers" (p. 232). Gottfredson (this issue) points out that the concept of executive functions is not necessarily more consensual than the concept of complexity. Sadly we must agree with both of them. Indeed, identifying the individual processes responsible for the positive manifold, but also the underlying brain mechanisms (see commentaries by Detterman et al., Haier and Jung, and Kan et al.) would be the most valuable extension of the theory.

As a final remark, it is worth reciting the aims of the target article and examining what we have achieved—especially in the light of some commentaries, which, with the best intentions, request things of POT that it just might not be able to achieve. What we tried to explain was the positive manifold: most likely the most replicated result in psychology, of which, after more than a century of its discovery, there is still no adequate explanation. As we pointed out in the target article, from a cognitive perspective the problem of the positive manifold translates to this: *Why does the variation between people in test performance appear massively domain-general if the abilities they employ to solve such tests are largely domain-specific?*

On our journey to find an answer we had been strongly influenced by Borsboom's work on latent variables and their relation to within-individual constructs. In particular, his coauthored paper on the ontological status of latent variables (Borsboom et al., 2003) has been a major influence, and it is probably reasonable to say that what we tried to accomplish all along was meeting this challenge:

> A between-subjects latent variable must be parasitic on individual processes, because these must be the source of between-subjects variability. If it is shown that a given set of cognitive processes leads to a particular latent variable structure, we could therefore say that this set of processes realizes the latent variables in question. The relevant research question for scientists should then be, which processes generate which latent variable structures?" (Boorsboom et al., 2003, pp. 215–216)

According to Sternberg, POT is descriptive, not explanatory, because—to overly simplify his argument—it does not explain why some people are smarter than others. In our view, POT *is* an explanatory theory; explanatory of the positive manifold, an important and well-replicated phenomenon. Explaining individual differences in the actual abilities that are conceptualized as reflective by POT (fluid reasoning, spatial ability, etc.) is beyond the scope of the target article.

So does POT lock us forever in Groundhog Day, or does it help us break free? We believe the answer is the latter. Provided

that the mainstream explanation of this phenomenon, reflective/psychological *g* theory leads to a completely different research program, providing an alternative explanation that leads to a different take on intelligence research is an important addition to the field.

Our aim was not simply to explain existing phenomena but also to inspire new research—research that, as a necessity of science, will eventually lead to a new, better theory or at best an improved version of POT. In the light of the commentaries, we are very optimistic.

References

Abad, F. J., Colom, R., Juan-Espinosa, M., & García, L. F. (2003). Intelligence differentiation in adult samples. *Intelligence*, *31*(2), 157–166. http://doi.org/10.1016/S0160-2896(02)00141-1

Alloway, T. P., & Alloway, R. G. (2010). Investigating the predictive roles of working memory and IQ in academic attainment. *Journal of Experimental Child Psychology*, *106*(1), 20–29. http://doi.org/10.1016/j.jecp.2009.11.003

Alloway, T. P. & Passolunghi, M. C. (2011). The relations between working memory and arithmetical abilities: A comparison between Italian and British children. *Learning and Individual Differences*, *21*, 133–137.

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

Baddeley, A. (1998). The central executive: A concept and some misconceptions. *Journal of the International Neuropsychological Society*, *4*, 523–526.

Baddeley, A., & Wilson, B. (1988). Frontal amnesia and the dysexecutive syndrome. *Brain and Cognition*, *7*, 212–230.

Blankenship, T. L., O'Neill, M., Ross, A., & Bell, M. A. (2015). Working memory and recollection, but not familiarity, contribute to academic achievement. *Learning and Individual Differences*, *43*, 164–169.

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2003). The theoretical status of latent variables. *Psychological Review*, *110*, 203–219.

Chein, J. M., Moore, A. B., & Conway, A. R. A. (2011). Domain-general mechanisms of complex working memory span. *NeuroImage*, *54*(1), 550–559. doi:10.1016/j.neuroimage.2010.07.067

Chow, M., Macnamara, B. N., & Conway, A. R. A. (2016). Phonological similarity in working memory span tasks. *Memory and Cognition*. Advance online publication. http://doi.org/10.3758/s13421-016-0609-8

Colom, R., Rebollo, I., Abad, F. J., & Shih, P. C. (2006). Complex span tasks, simple span tasks, and cognitive abilities: A reanalysis of key studies. *Memory & Cognition*, *34*, 158–171. http://doi.org/10.3758/BF03193395

Conway, A. R. A. (2014). A new era of intelligence research. *Journal of Intelligence*, *2*(2), 33–35.

Conway, A. R. A., Kane, M. J., & Engle, R. W. (1999). Is Spearman's g determined by speed or working memory capacity? Book review of Jensen on Intelligence-g-Factor. *Psycoloquy*, *10*(074).

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*, 450–466. doi:10.1016/S0022-5371(80)90312-6

Daneman, M., & Green, I. (1986). Individual differences in comprehending in context and producing words. *Journal of Memory and Language*, *25*, 1–18. http://doi.org/10.1016/0749-596X(86)90018-5

Deary, I. J., Egan, V., Gibson, G. J., Austin, E. J., Brand, C. R., & Kellaghan, T., (1996). Intelligence and the differentiation hypothesis. *Intelligence*, *23*, 105–132.

Detterman, D. K., & Daniel, M. H. (1989). Correlations of mental tests with each other and with cognitive variables are highest for low IQ groups. *Intelligence*, *13*, 349–359. doi:10.1016/S0160-2896(89)80007-8

Dickens, W. T., & Flynn, J. R. (2001). Heritability estimates versus large environmental effects: The IQ paradox resolved. *Psychological Review*, *108*(2), 346–369. http://doi.org/10.1037/0033-295X.108.2.346

Fan, J., Wu, Y., Fossella, J. A., & Posner, M. I. (2001). Assessing the heritability of attentional networks. *BMC Neuroscience*, *2*(1), 1.

Fitzpatrick, C., Archambault, I., Janosz, M., & Pagani, L. S. (2015). Early childhood working memory forecasts high school dropout risk. *Intelligence*, 53, 160–165. http://doi.org/10.1016/j.intell.2015.10.002

Flynn, J. R. (1999). Reply to rushton: A gang of gs overpowers factor analysis. *Personality and Individual Differences*, 26(2), 391–393.

Flynn, J. R. (2007). *What is intelligence? Beyond the Flynn effect*. New York, NY: Cambridge University Press.

Flynn, J. R., te Nijenhuis, J., & Metzen, D. (2014). The g beyond Spearman's g: Flynn's paradoxes resolved using four exploratory meta-analyses. *Intelligence*, 44, 1–10.

Fodor, J. A. (1993). *Modularity of mind: An essay on faculty psychology*. Cambridge, MA: MIT Press.

Fodor, J. A. (2000). *The mind doesn't work that way: The scope and limits of computational psychology*. Cambridge, MA: MIT Press.

Fogarty, G. J. & Stankov, L. (1995). Challenging the "Law of diminishing returns." *Intelligence*, 21, 157–174.

Foroughi, C. K., Barragan, D., & Boehm-Davis, D. A. (2016). Interrupted reading and working memory capacity. *Journal of Applied Research in Memory and Cognition*. Advance online publication.

Friedman, N. P., Miyake, A., Corley, R. P., Young, S. E., DeFries, J. C., & Hewitt, J. K. (2006). Not all executive functions are related to intelligence. *Psychological Science*, 17, 172–179.

Friedman, N. P., Miyake, A., Young, S. E., DeFries, J. C., Corley, R. P., & Hewitt, J. K. (2008). Individual differences in executive functions are almost entirely genetic in origin. *Journal of Experimental Psychology: General*, 137, 201–225.

Gignac, G. E., & Weiss, L. G. (2015). Digit Span is (mostly) related linearly to general intelligence: Every extra bit of span counts. *Psychological Assessment*, 27, 1312–1323. http://doi.org/10.1037/pas0000105

Good, I. (1962). *The scientist speculates*. New York: Basic Books.

Gottfredson, L. S. (2007). Innovation, fatal accidents, and the evolution of general intelligence. In M. J. Roberts (Ed.), *Integrating the mind: Domain general versus domain specific processes in higher cognition* (pp. 387–425). Hove, UK: Psychology Press.

Kane, H. D., Oakland, T. D., & Brand, C. R. (2006). Differentiation at higher levels of cognitive ability: Evidence from the United States. *The Journal of Genetic Psychology*, 167, 327–341. doi:10.3200/GNTP.167.3.327-341

Kane, M. J., Conway, A. R. A., Miura, T. K., & Colflesh, G. J. H. (2007). Working memory, attention control, and the N-back task: A question of construct validity. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 33, 615–622. doi:10.1037/0278-7393.33.3.615

Kane, M. J., Meier, M. E., Smeekens, B. A., Gross, G. M., Chun, C. A., Silvia, P. J., & Kwapil, T. R. (in press). Individual differences in the executive control of attention, memory, and thought, and their associations with schizotypy. *Journal of Experimental Psychology: General*.

Kaufman, S. B., DeYoung, C. G., Gray, J. R., Brown, J., & Mackintosh, N. (2009). Associative learning predicts intelligence above and beyond working memory and processing speed. *Intelligence*, 37, 374–382. doi:10.1016/j.intell.2009.03.004

Keye, D., Wilhelm, O., Oberauer, K., & Stürmer, B. (2013). Individual differences in response conflict adaptations. *Frontiers in Psychology*, 4, 1–15. doi:10.3389/fpsyg.2013.00947

Keye, D., Wilhelm, O., Oberauer, K., & van Ravenzwaaij, D. (2009). Individual differences in conflict-monitoring: Testing means and covariance hypothesis about the Simon and the Eriksen flanker task. *Psychological Research*, 73, 762–776. doi:10.1007/s00426-008-0188-9

Kovacs, K., Molenaar, D., & Conway, A. R. A. (2016). *The domain-specificity of working memory: A matter of ability*. Manuscript in preparation.

Kroczek, B., Ociepka, M., & Chuderski, A. (in press). No Spearman's Law of Diminishing Returns for the working memory and intelligence relationship. *Polish Psychological Bulletin*.

Lynn, R. (1992). Does Spearman's g decline at high levels? Some evidence from Scotland. *Journal of Genetic Psychology*, 153, 229–230.

Macnamara, B. N., & Conway, A. R. A. (2015). Working memory capacity as a predictor of simultaneous language interpreting performance. *Journal of Applied Research in Memory and Cognition*. Advance online publication. http://doi.org/10.1016/j.jarmac.2015.12.001

Macnamara, B. N., Moore, A. B., & Conway, A. R. A. (2011). Phonological similarity effects in simple and complex span tasks. *Memory and Cognition*, 39, 1174–1186. http://doi.org/10.3758/s13421-011-0100-5

McCabe, D. P., Roediger, H. L., McDaniel, M. A., Balota, D. A., & Hambrick, D. Z. (2010). The relationship between working memory capacity and executive functioning: Evidence for a common executive attention construct. *Neuropsychology*, 24, 222–243. doi:10.1037/a0017619

McClearn, G. E., Johansson, B., Berg, S., Pedersen, N. L., Ahern, F., Petrill, S. A., & Plomin, R. (1997). Substantial genetic influence on cognitive abilities in twins 80 or more years old. *Science*, 276(5318), 1560–1563.

Meinz, E. J., Hambrick, D. Z., Hawkins, C. B., Gillings, A. K., Meyer, B. E., & Schneider, J. L. (2011). Roles of domain knowledge and working memory capacity in components of skill and Texas Hold 'Em poker. *Journal of Applied Research in Memory and Cognition*, 1, 34–40.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "Frontal Lobe" tasks: A latent variable analysis. *Cognitive Psychology*, 41, 49–100. doi:10.1006/cogp.1999.0734

Miyake, A., Friedman, N. P., Rettinger, D. A., Shah, P., & Hegarty, M. (2001). How are visuospatial working memory, executive functioning, and spatial abilities related? A latent-variable analysis. *Journal of Experimental Psychology: General*, 130, 621–640. doi:10.1037/0096-3445.130.4.621

Molenaar, D., Dolan, C. V., Wicherts, J. M., & van der Maas, H. L. J. (2010). Modeling differentiation of cognitive abilities within the higher-order factor model using moderated factor analysis. *Intelligence*, 38, 611–624. doi:10.1016/j.intell.2010.09.002

Molenaar, D., Wicherts, J. M., & Kovacs, K. (2016). *Testing ability and age differentiation in the Hungarian WAIS and WISC*. Manuscript in preparation.

Moscovitch, M. (1992). Memory and working with memory: A component process model based on modules and central systems. *Journal of Cognitive Neuroscience*, 4, 257–267.

Oberauer, K., Süβ, H. M., Wilhelm, O., & Wittmann, W. W. (2008). Which working memory functions predict intelligence? *Intelligence*, 36, 641–652. doi:10.1016/j.intell.2008.01.007

Raven, J. C. (2000). Psychometrics, cognitive ability, and occupational performance. *Review of Psychology*, 7(1–2), 51–74.

Raven, J. & Court, J. H. (1989). *Manual for Raven's Progressive Matrices and Vocabulary Scales*. Research Supplement No. 4. London: H. K. Lewis & Co.

Rosen, V. M., Bergeson, J. L., Putnam, K., Harwell, A., & Sunderland, T. (2002). Working memory and apolipoprotein E: What's the connection? *Neuropsychologia*, 40(13), 2226–2233.

Rushton, J. P. (1999). Secular gains in IQ not related to the g factor and inbreeding depression-unlike Black-White differences: A reply to Flynn. *Personality and Individual Differences*, 26, 381–390

Penke, L., Maniega, S. M., Bastin, M. E., Valdés Hernández, M. C., Murray, C., Royle, N. A., … Deary, I. J. (2012). Brain white matter tract integrity as a neural foundation for general intelligence. *Molecular Psychiatry*, 17, 1026–1030. doi:10.1038/mp.2012.66

Salthouse, T. A., Atkinson, T. M., & Berish, D. E. (2003). Executive functioning as a potential mediator of age-related cognitive decline in normal adults. *Journal of Experimental Psychology. General*, 132(4), 566–94. doi:10.1037/0096-3445.132.4.566

Shipstead, Z., Lindsey, D. R. B., Marshall, R. L., & Engle, R. W. (2014). The mechanisms of working memory capacity: Primary memory, secondary memory, and attention control. *Journal of Memory and Language*, 72(1), 116–141. doi:10.1016/j.jml.2014.01.004

Spearman, C. (1927). *The abilities of man: Their nature and measurement*. New York, NY: Macmillan.

Tucker-Drob, E. M. (2009). Differentiation of cognitive abilities across the life span. *Developmental Psychology*, 45(4), 1097–118. http://doi.org/10.1037/a0015864

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language*, 28(2), 127–154. doi:10.1016/0749-596X(89)90040-5

Unsworth, N. (2010). On the division of working memory and long-term memory and their relation to intelligence: A latent variable approach. *Acta Psychologica*, 134(1), 16–28. http://doi.org/10.1016/j.actpsy.2009.11.010

Unsworth, N., & Engle, R. W. (2007). On the division of short-term and working memory: An examination of simple and complex span and their relation to higher order abilities. *Psychological Bulletin*, 133(6), 1038–1066. doi:10.1037/0033-2909.133.6.1038

Unsworth, N., Fukuda, K., Awh, E., & Vogel, E. K. (2014). Working memory and fluid intelligence: Capacity, attention control, and secondary memory retrieval. *Cognitive Psychology*, *71*, 1–26. doi:10.1016/j.cogpsych.2014.01.003

Unsworth, N., Miller, J. D., Lakey, C. E., Young, D. L., Meeks, J. T., Campbell, W. K., & Goodie, A. S. (2009). Exploring the relations among executive functions, fluid intelligence, and personality. *Journal of Individual Differences*, *30*(4), 194–200. http://doi.org/10.1027/1614-0001.30.4.194

Unsworth, N., & Spillers, G. J. (2010). Working memory capacity: Attention control, secondary memory, or both? A direct test of the dual-component model. *Journal of Memory and Language*, *62*(4), 392–406. doi:10.1016/j.jml.2010.02.001

Unsworth, N., Spillers, G. J., & Brewer, G. A. (2010). The contributions of primary and secondary memory to working memory capacity: An individual differences analysis of immediate free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 240–247. doi:10.1037/a0017739

van der Maas, H. L. J., Dolan, C. V., Grasman, R. P. P. P., Wicherts, J. M., Huizenga, H. M., & Raijmakers, M. E. J. (2006). A dynamical model of general intelligence: The positive manifold of intelligence by mutualism. *Psychological Review*, *113*, 842–861. doi:10.1037/0033-295X.113.4.842

van der Maas, H. L. J., Kan, K.-J., & Borsboom, D. (2014). Intelligence is what the intelligence test measures. Seriously. *Journal of Intelligence, 2*(1), 12–15.

Vernon, P. E., & Parry, J. B. (1949). *Personnel selection in the British forces*. London, UK: University of London Press.

Wilhelm, O., Hildebrandt, A., & Oberauer, K. (2013). What is working memory capacity, and how can we measure it? *Frontiers in Psychology*, *4*, 1–22. doi:10.3389/fpsyg.2013.00433