PSYCHOLOGICAL INQUIRY
2016, VOL. 27, NO. 3, 220–228
http://dx.doi.org/10.1080/1047840X.2016.1182000

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

CONTACT Kees-Jan Kan k.j.kan@vu.nl Department of Biological Psychology, VU University Amsterdam, van der Boechorststraat 1, 1081 BT Amsterdam, The Netherlands.

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

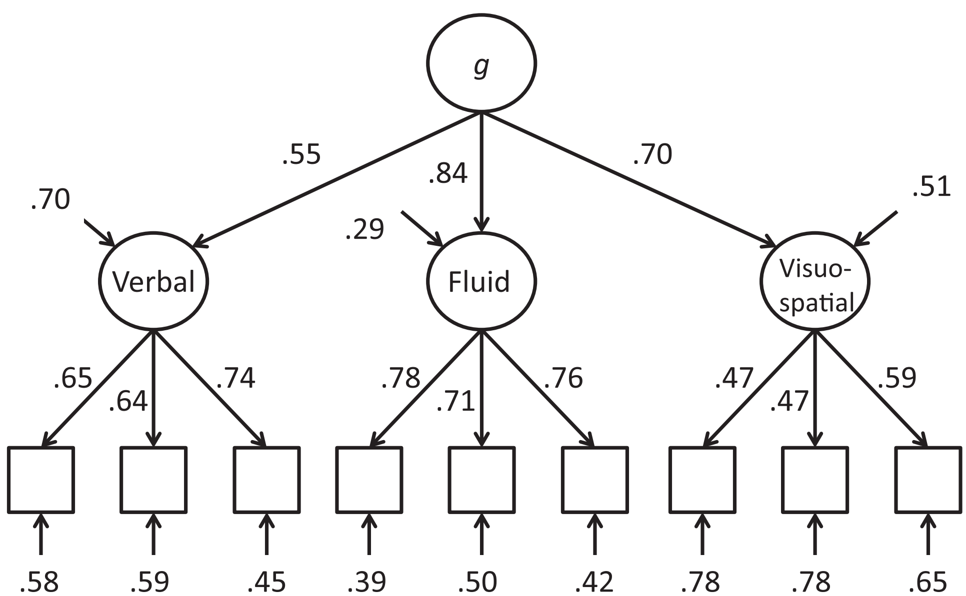

Figure 1. Showing a typical outcome of a simulation with a model that is in line with process overlap theory as stated verbally and as presented as structural model: The latent variable *g* (the sum of all fluid processing capacities) predicts the latent variables Verbal, Fluid, and Visuospatial Intelligence.

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.

[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.

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 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

RAK is supported by the Sir Henry Wellcome Trust (grant number 107392/Z/15/Z).

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.