Learning and Individual Differences 29 (2014) 41–49

Contents lists available at ScienceDirect

Learning and Individual Differences

journal homepage: www.elsevier.com/locate/lindif

The rule-dependence model explains the commonalities between the Flynn effect and IQ gains via retesting

Elijah L. Armstrong [a,*,1], Michael A. Woodley [b,c,1]

[a] 34 Acacia Road, Fairfax, CA, 94930, United States
[b] Umeå University, Department of Psychology, Umeå, Sweden
[c] Center Leo Apostel for Interdisciplinary Research, Vrije Universiteit Brussel, Belgium

ARTICLE INFO

Article history:
Received 26 April 2013
Received in revised form 23 August 2013
Accepted 12 October 2013
Available online xxxx

Keywords:
IQ
Flynn effect
Retesting
Inductive ability
Life history

ABSTRACT

We present a new model of the Flynn effect. It is proposed that Flynn effect gains are partly a function of the degree to which a test is dependent on rules or heuristics. This means that testees can become better at solving 'rule-dependent' problems over time in response to changing environments, which lead to the improvement of lower-order cognitive processes (such as implicit learning and aspects of working memory). These in turn lead to apparent IQ gains that are partially independent of general intelligence. We argue that the Flynn effect is directly analogous to IQ gains via retesting, noting that Raven's Progressive Matrices is particularly sensitive to both the effects of retesting and the Flynn effect. After an extensive review of the relevant supporting literature, we test our thesis by developing a rule-dependence typology and then correlate the vector of a test's position in the typology with the vector of the Flynn effect that it yields. We find a significant vector correlation of $r \sim .60$ ($N = 14$). Finally, we make a number of novel and testable predictions based on our model.

1. Introduction

The Flynn effect describes the tendency for IQ scores to rise across the board at a rate of approximately .30 points per year, or three points per decade (Flynn, 2009). Amongst developed countries the effect had its origins in the early decades of the 20th century (Lynn, 2013), but seems to have been most pronounced in the period immediately following the conclusion of World War II, where especially pronounced gains in Europe and Japan were recorded (Flynn, 2009). Large gains have also been detected in South Korea following the cessation of hostilities in the post-Korean war period (te Nijenhuis, Cho, Murphy, & Lee, 2012). The Flynn effect has recently been detected in a number of developing countries, including Dominica (Meisenberg, Lawless, Lambert, & Newton, 2005), Saudi Arabia (Batterjee, Khaleefa, Ali, & Lynn, 2013), South Africa (te Nijenhuis, Murphy, & van Eeden, 2012), Turkey (Kagitcibasi & Biricik, 2011; Rindermann, Schott, & Baumeister, 2013), Brazil (Colom, Flores-Mendoza, & Abad, 2007), Kenya (Daley, Whaley, Sigman, Espinosa, & Neuman, 2003) and Sudan (Khaleefa, Sulman, & Lynn, 2009).

There are three related and significant issues concerning the Flynn effect: 1) To what extent does the effect constitute a 'real' gain in IQ, as opposed to a simple change in test-taking habits, such as an increasing reliance guessing the answers to multiple choice format items (e.g., Brand, 1996)? 2) To what extent does the Flynn effect concern changes in the level of g, the common factor among many different cognitive ability measures, rather than more narrow sources of ability variance (e.g., Jensen, 1998a)? 3) What has caused the Flynn effect?

With respect to the first issue, the presence of apparent real world corollaries involving the Flynn effect, such as historical increases in GDP (Purchasing Power Parity adjusted) per-capita paralleling the historical trends in the effect (Woodley, 2012a), an increase in precociousness in intellectual games such as Chess, Bridge and Go, teacher ratings indicating that students are becoming increasingly practically 'intelligent' (Howard, 1999, 2001), and neurological evidence indicating that the effect may be directly related to both increasing brain size (Lynn, 1989) and to enhanced right hippocampal functioning (Baxendale & Smith, 2012) suggests that the effect is associated with actual increases in certain abilities, and is therefore not solely an artifact of changing attitudes towards test-taking. Recent research, however, indicates that changing test-taking attitudes, especially the tendency towards the increased use of guessing on harder items, may nevertheless account for a portion of the Flynn effect (Must & Must, in press).

With respect to the second issue, two complementary lines of evidence indicate that the Flynn effect is not occurring on g.

The first line of evidence concerns the use of the method of correlated vectors, where the g loading of an association between IQ and another variable is calculated by correlating the vector of the magnitude of the effect with the vector of the g loadings of different tests (Jensen, 1998a). Generally, the relationships between IQ and biological or part-biological sources of individual and group differences, such as subtest heritabilities, inbreeding depression scores (Rushton, 1999;

* Corresponding author. Tel.: + 1 415 456 8777.
E-mail addresses: elijahlarmstrong@gmail.com (E.L. Armstrong), Michael.Woodley@psy.umu.se (M.A. Woodley).
[1] Both authors are to be considered joint first authors.

1041-6080/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.lindif.2013.10.009

Rushton & Jensen, 2010; van Bloois, Geutjes, te Nijenhuis, & de Pater, 2009), reaction time measures (Jensen, 1998a), brain size (Rushton & Ankney, 2009), fluctuating asymmetry (Prokosch, Yeo, & Miller, 2005), and dysgenic fertility (Woodley & Meisenberg, 2013a) are g loaded. Collectively, such effects are termed "Jensen effects" (Rushton, 1998). Conversely, culturally driven effects, such as the IQ gains accrued via the retesting effect and IQ gains in adopted children, are generally anti-Jensen effects in that they are significantly more pronounced on the least g loaded subtests (Jensen, 1998b; te Nijenhuis, van Vianen, & van der Flier, 2007). Given the presence of this apparent biological vs. cultural division, where does the Flynn effect fall? In other words, is it closer to being a purely biological or cultural effect?

The preponderance of studies indicate that the effect is either uncorrelated or mildly negatively correlated with subtest g loadings (Jensen, 1998a; Must, Must, & Raudik, 2003a, 2003b; Rushton, 1999; te Nijenhuis, 2013, te Nijenhuis & van der Flier, 2007; Woodley & Meisenberg, 2013b). A recent meta-analytic study of over 17,000 individuals revealed that the Flynn effect is in fact a statistically significant anti-Jensen effect ($rho = -.38$; te Nijenhuis & van der Flier, in press), indicating that it is likely to be substantially environmental in origin, given the monotonic positive relationship between g loadings and subtest heritabilities (Rushton & Jensen, 2010; van Bloois et al., 2009).

The second line of evidence demonstrating the Flynn effect's lack of g loading comes from the study of Wicherts et al. (2004), who utilized multi-group confirmatory factor analysis to examine factorial invariance across a number of cohorts exhibiting the Flynn effect. If the effect occurs on g, it would be expected that the factor structure of g will be preserved across time between cohorts, i.e., will be invariant. The study found that lack of factorial invariance was characteristic of the Flynn effect, which indicates that the effect is associated with heterogeneous gains on specific tests rather than a gain at the level of latent variables (such as g). This finding was replicated subsequently in Estonian cohorts employing the National Intelligence Test (Must, te Nijenhuis, Must, & van Vianen, 2009), and using a different method, at the item level in the Raven's Progressive Matrices (Fox & Mitchum, 2013). Despite the finding of no factorial or measurement invariance in the Flynn effect, te Nijenhuis and van der Flier's (in press) finding of a modest (rather than monotonic) anti-Jensen effect suggests that some small portion of the Flynn effect may still occur on g. One possible explanation for this discrepancy is that secular gains resulting purely from changing test-taking habits may mask the size of the anti-Jensen effect on the remainder of the Flynn effect, especially in so much as gains through guessing concern the use of increased guessing on *harder* items, which are generally more g loaded (Must & Must, in press), hence may *mimic* the Jensen effect. This is a sound theoretical reason for suspecting that the *guessing-controlled* Flynn effect is more strongly negatively associated with g loadings than the data currently indicate.

With respect to the third issue, the apparent independence of the Flynn effect from g permits us to better discriminate amongst potential causes (te Nijenhuis, 2013). Narrow and more 'hollow' sources of variance in cognitive abilities are sometimes substantially less heritable than g itself (Carroll, 1993; Rushton & Jensen, 2010; van Bloois et al., 2009). Hence, these sources are far more amenable to environmental manipulation of a sort that could give rise to relatively large gains in measured IQ over a relatively short time frame (Woodley, 2011a, 2012b). This presents a plausible solution to the so-called 'IQ paradox', or the observation that measured IQ has risen despite IQ exhibiting a high additive heritability (Dickens & Flynn, 2001). Proposed causes of the Flynn effect such as heterosis (Mingroni, 2004, 2007) can thus be ruled out (Flynn, 2009; Woodley, 2011a), as such gains are biological and associated with the Jensen effect (Nagoshi & Johnson, 1986). Sources of IQ gains that are associated with significant environmental and social improvements, such as decreased neurotoxic pollution (Nevin, 2000) and the expansion of the education system (Husén & Tuijnman, 1991; Teasdale & Owen, 1989; Tuddenham, 1948), are more plausible causes of the Flynn effect by comparison, since neither education nor neurotoxins seem to impact g (e.g., Christian, Bachnan, & Morrison, 2001; Lezak, 1983). However, there is substantial debate about which of the many proposed causes are predominantly involved in the effect, with different studies frequently indicating simultaneous contributions from multiple causes (e.g., Neisser, 1997; Williams, in press).

In this manuscript, we present a link between the Flynn effect and the retesting effect, i.e., the gain in IQ that accrues from retesting individuals on certain IQ tests. This link is the degree to which tests are reliant upon the repeated reapplication of solution rules, where rules are defined as specific procedures or pieces of information that can be consistently relied upon to locate solutions to specific problems. In essence, the more reliant a particular test is on the identification and repeated use of specific rule-sets, the bigger the Flynn and retesting effects. We consider the results of studies that have examined and formalized the use of rules in the solving of the Raven's Progressive Matrices (Carpenter, Just, & Shell, 1990), and others which have found that performance improvements via familiarity with the RPM are related to increases in the efficiency with which individuals can successfully sample rules on this test (Verguts, Boeck, & Maris, 1999; Verguts & De Boeck, 2002). We connect these findings with the observation that this battery is especially sensitive to both Flynn and retesting effects, despite its high g loading. We also use this rule-dependence model to propose two subsidiary models concerning the ways in which specific sources of environmental improvement can translate into massive IQ gains. In conducting a test of the model, we infer the existence of a four-level typology into which any IQ test can be assigned based on the degree to which it is dependent upon the reapplication of specific rule-sets to problem solving — from least dependent (*Level I*) to most (*Level IV*). It is hypothesized that an IQ test's position in this typology should be both positively and significantly correlated with the actual recorded size of the Flynn effects on those batteries. This is tested using real data on secular gains from 14 scales. Finally in the discussion we consider the broader implications of this model in terms of testable predictions and the debate surrounding the meaningfulness of both the Flynn and retesting effects.

1.1. Rules and IQ tests

What has been lacking in most previous attempts to find causes of the Flynn effect is an appreciation of the core cognitive processes needed to yield secular gains in the first place (Flynn, 2013).

The biggest Flynn effect gains have been recorded on tests of fluid intelligence, in particular the Raven's Progressive Matrices (five-seven points per decade; Flynn, 1987, 2007, 2009). The smallest gains typically occur on measures of crystallized knowledge (Jensen, 1998a). It must be noted that not all measures of IQ show long-term gains. For example, purely psychophysical elementary cognitive tasks (ECTs), such as inspection and reaction times, show no Flynn effect (Nettlebeck & Wilson, 2004; Silverman, 2010). The latter, in fact, indicate a long-term decline in functioning since the last decades of the 19th century (Silverman, 2010; Woodley, te Nijenhuis, & Murphy, in press). Psychophysical ECTs (in particular, reaction times) are considered by some to be extremely pure measures of speed of information processing, as they relate to physiological properties of the central nervous system that are believed to be fundamental to g (Jensen, 2006). Reflecting this, the common variance amongst psychophysical mental chronometric tests is sometimes termed 'chronometric g' (Jensen, 1998a, 2006). Given the presence of Jensen effects on both reaction times and dysgenic fertility magnitudes (Jensen, 1998a, 2006; Woodley & Meisenberg, 2013a), the finding of diminishing performance on reaction times has been interpreted as evidence that the effect of dysgenic fertility throughout the 20th century (Lynn, 2011) has been concentrated on speed of information processing, whereas other cognitive abilities have become less dependent on this faculty over time as a function of environmental improvements and the increased demand for cognitive specialization (Woodley & Meisenberg, 2013b; Woodley, te Nijenhuis, et al., in press).

Nonetheless, the Flynn effect's high occurrence on tests of fluid intelligence has been taken as evidence that the effect might be occurring at the level of *g*, owing to the notion that fluid intelligence and *g* are identical (e.g., Colom, Juan-Espinosa, & García, 2001). This claim is usually backed up (e.g., Jensen, 1998a) with references to studies indicating that the Raven's Progressive Matrices (RPM) test, a test of fluid intelligence, is very highly *g* loaded relative to other psychometric tests. Thus high gains on the Raven's seem to be an exception to the aforementioned finding that the Flynn effect is not associated with *g*. One potential resolution to this 'Raven's paradox' comes from the possibility that the RPM's *g* loading would be lower amongst an extremely diverse set of tests. Consistent with this is the observation that when pooled with a very broad array of test batteries and loaded onto a common 'super' *g* factor, the RPM seems to lose some of its *g* loading — especially relative to crystallized knowledge subtests, such as vocabulary (Raven's = .67 vs. three vocabulary subtests = .82, .81 and .81; Johnson, Bouchard, Krueger, McGue, & Gottesman, 2004). Nevertheless, these authors indicate that their results must be interpreted cautiously, since the RPM employed in their study was modified.

Another complementary solution is the aforementioned observation that RPM items lack measurement invariance across cohorts; hence, the IQ gains on this battery may have come *solely* as a consequence of these items having changed their *meaning* with time (Fox & Mitchum, 2013). In other words, even though the RPM is highly *g* loaded when scores are compared *within* a cohort, this has no bearing on the stability of items in terms of their ability to consistently tap *g* *between* cohorts separated in time.

A third complimentary solution is the theory that performance on the RPM and RPM-like batteries might relate to individual differences in a cognitive domain that is discriminable from pure *g*, termed *fluid cognition* (Blair, 2006). Blair conceptualizes this domain as being related to executive functioning, specifically components of working memory which deal with the capacity to store solution-sets for given tasks in the form of mental representations (Welsh & Pennington, 1988), which are in turn described as being distinct from *g*-as-information-processing speed (see also; Lewandowsky & Oberauer, 2009; Saito & Miyake, 2004 for a discussion of the distinction between speeded and non-speeded working memory components).

These three possibilities may assist in solving the 'Raven's paradox': firstly, the RPM might be less *g* loaded than is commonly assumed, although this is uncertain; secondly, RPM items lose their ability to measure *g* over time, which suggests a large role for other sources of variance in performance; and thirdly, performance on the RPM may be heavily contingent specifically upon the use of cognitive processes relating to working memory that are capable of operating independently of *g*.

1.2. The anatomy of secular IQ gains

Granting the aforementioned solutions to the 'Raven's paradox', a question that arises is how fluid cognition (Blair, 2006), as distinct from *g* as speed of information processing, e.g., 'chronometric *g*' (Jensen, 1998a, 2006), might be used in independently solving more 'domain specific' IQ-type problems (i.e., problems that tap test-specific rather than general sources of variance). One possibility is that tests like the RPM contain a kind of 'cognitive scaffolding' that functions to guide the use of non-*g* abilities in solving problems.

Carpenter et al. (1990) were among the first to formalize the existence of such a 'scaffolding' in the RPM in the form of *rules*, or specific mental operations that could be reliably reapplied to solving a particular problem. Carpenter et al.'s model is based on the idea that integral to the RPM are five rules that can be used in some combination to solve all tasks. Carpenter et al. furthermore proposed that two sources of individual differences were necessary for solving Raven's-type problems: one dimension, akin to working memory, which is the ability to store the subresults of applying rule permutations to solving an item, and another which corresponds to the more basic capacity for rule induction.

Studies by Verguts et al. (1999) and Verguts and De Boeck (2002) have examined this second dimension more closely, and have found that in experiments, participants do repeatedly reapply a small number of rules when attempting to solve RPM-type items. The rule induction process seems to be characterized by the sequential sampling of rules, each of which have a certain probably of being resampled (Verguts et al., 1999). Furthermore, the speed with which individuals are able to sample and induce rules is strongly related to prior task exposure (Verguts & De Boeck, 2002).

These studies indicate that narrower abilities matter in solving RPM items, and that furthermore, reliance upon these narrow abilities increases with increasing test experience. This is consistent with the phenomenology of both retesting and Flynn effects, as not only are both effects anti-Jensen effects, but in both cases these performance gains seem to come at the expense of a test's *g* saturation (Juan-Espinosa, Cuevas, Escorial, & García, 2006; Kane, 2000; Kane & Oakland, 2000; Lynn & Cooper, 1993, 1994; Must et al., 2003a, 2003b, 2009; te Nijenhuis et al., 2007; Woodley & Madison, 2013).

The presence of specific rules in the RPM might therefore be guiding the use of the two cognitive mechanisms identified by Carpenter et al. (1990) in the case of both effects, thus accounting for the fact that these gains occur primarily on non-*g* sources of ability variance. This is consistent with the idea that whilst it is necessary to use features exclusive to the domain of *g*, such as information processing capacity, to establish the relevance of rules as opposed to other procedures, and to establish the boundary conditions for the use of those rules, in solving test items, other cognitive processes can come to principally determine performance once sufficient familiarity with the format of the test is achieved. This is because the ability to efficiently reuse acquired rules becomes easier over time as they become implicitly learned (Ackerman, 1986, 1987; Verguts & De Boeck, 2002).

1.3. Inductive reasoning increases, independently of g

The work of Carpenter et al. (1990) and Verguts and De Boeck (2002) yields important information as to the identities of these additional, more domain-specific cognitive mechanisms. To recap, they found that there were two cognitive dimensions that people came to increasingly rely upon in using rules to solve RPM items. One relates to the ability to effectively induce rules — a performance component that may draw on a cognitive dimension termed *implicit learning*, which corresponds to the ability to automatically detect simple or already-seen regularities in an otherwise noisy and complicated cognitive background. This construct has been found to modestly relate to *g* but not to working memory (Kaufman et al., 2010).

The other major dimension discussed by Verguts and De Boeck (2002) corresponds to the dimension of working memory, or the ability to store and manipulate mental representations of rule-sets. Its role in RPM performance was well characterized in earlier work by Carpenter et al. (1990). Extending this finding into the domain of the Flynn effect, the idea that *non-speeded* components of working memory (i.e., those that are independent of measures of processing speed) are foundational to the effect is consistent with Blair's (2006) contention that working memory as a source of executive functioning is integral to fluid cognition — the primary ability grouping on which the Flynn effect occurs. Reinforcing this, there is evidence that the Flynn effect operates directly on working memory-related measures (e.g., Baxendale, 2010; Resing & Tunteler, 2007; Rönnlund & Nilsson, 2008, 2009). Additionally consistent with

Blair's (2006) model is the finding that hippocampal (specifically right) pathology inhibits the Flynn effect in clinical samples (the hippocampus is associated with memory encoding; Baxendale & Smith, 2012). Furthermore working memory components are directly amenable to manipulation via training, in particular verbal and visuospatial working memory components, which based on meta-analysis, produce short term and long-term but narrow gains in response to training respectively (Melby-Lervåg & Hulme, 2013).

Both implicit learning as a source of individual differences and working memory share variance with *g* (Conway, Jarrold, Kane, Miyahe, & Towse, 2007; Kaufman et al., 2010), although as was mentioned previously, individual differences in these two constructs do not seem to correlate (Kaufman et al., 2010). Despite this, there appear to be interactions amongst aspects of both these sources of variance, which have direct relevance to performance gains on the RPM. This is revealed in the study of Bauernschmidt, Conway, and Pisoni (2008), who found transfer effects between working memory training on probabilistic sequences (i.e., learned sequences containing patterns) and implicit learning, as measured using tests of novel grammar acquisition. The same study also found that there were transfer effects from this training to RPM performance.

This is significant because both of these cognitive variables have also been found to be associated with the capacity for *inductive* or *categorizational reasoning* (e.g. Girelli, Semenza, & Delazer, 2004; Süss, Oberauer, Wittman, Wilhelm, & Schulze, 2002). This is the ability to explicitly induce solutions to problems via a process of *decontextualization* (i.e., determining what is relevant from what is not for the purposes of solving an item) and *generalization* (i.e., applying the relevant solutions to different items in the same test; Carroll, 1993; Jensen, 1998a; Flynn, 2009). There are respectable Flynn effects on measures that tap aspects of inductive reasoning, such as the Similarities subtest on the WAIS (Flynn, 2009). This observation suggests that it may be meaningful to talk about the interaction between working memory components and implicit learning in terms of an inductive reasoning ability complex — a source of *explicit* (i.e., effortful as opposed to purely automatic) cognitive processing variance that is both distinct from g and very important to the etiology of both the Flynn and retesting effects, though it may involve still other components, like hypothetical reasoning, which has also improved (e.g., Flynn, 2009). Note that inductive ability does not always have to relate to rule dependence. Some inductive tasks, like vocabulary (e.g., Jensen, 1998a, 1998b), which is inductive because it requires the inference of meaning from contexts (Jensen, 1998a, 1998b), may not extensively rely on the reapplication of rules. Vocabulary has not shown a drastic Flynn effect (Flynn, 2009), though vocabulary gains on the Wechsler are considerable among adults (Flynn, 2012). Hence, inductive reasoning has probably only increased as it relates to rule dependence.

1.4. Summary of the rule-dependence model

In summary, therefore, both the Flynn effect and also the retesting effect operate largely on the ability to induce and utilize rules, which we define as specific procedures or pieces of information that can be consistently relied upon in problem solving. The presence and also the number of rules are the principal determinants of the degree to which the relay between g-as-information-processing-speed and the inductive reasoning ability complex comprised of non-speeded components of working memory plus implicit learning ability can take place. Tests that are highly reliant upon a relatively small number of rules, such as the RPM, should be more sensitive to rule induction, and therefore IQ gains than those that do not employ rules (such as measures of information), or those that use relatively more complex and less predictable rule-sets (such as the Cattell Culture Fair).

As a consequence of the relay, the working memory and implicit learning components that partly define the inductive reasoning ability complex become increasingly independent of g-as-information-processing-speed, which results in those items increasingly coming to measure these two sources of individual differences rather than *g*, both as a function of retesting and the Flynn effect. Hence, with both effects, the RPM will fail to exhibit measurement invariance (e.g., Fox & Mitchum, 2013; te Nijenhuis et al., 2007). This mechanism explains the commonalities between IQ gains via retesting and the Flynn effect in terms of their independence from g, their tendency to come at the expense of a battery's g loading, and their greater effect on highly rule-dependent tests such as the Raven's.

This constitutes the theoretical core of the rule-dependence model of IQ gains.

1.5. The rule-dependence model and proposed causes of the Flynn effect

How might the rule-dependence model relate to the Flynn effect at the level of environmental causes? The retesting effect obviously relates to repeated exposure to specific test batteries under controlled conditions. Implicit in the previous discussion is the notion that increased battery exposure may also be at the root of the Flynn effect. However, it is somewhat less clear how environmental improvements might have translated into Flynn effect gains via this mechanism.

Two models provide plausible and complimentary mechanisms.

1.5.1. Increasing rule exposure

In a broad sense, education may be eliciting from whole populations the equivalent of a mass retesting effect. Education is known to boost performance on IQ tests (e.g., Ceci, 1991), and moreover educational gains show low transfer effects between disparate performance domains, consistent with the idea that they are concentrated on narrow abilities (Christian et al., 2001). Perhaps this relates to certain components of modern educational styles, such as the increasing use of heuristics or ‘rules-of-thumb’ (an example might include the use of mnemonic devices such as the colors of the rainbow or element series and periods in the periodic table) in conveying and memorizing complex concepts; or, more directly, increased ‘teaching to the test’, i.e., explicitly teaching students the solution rules to tests (Jensen, 1998a). For example, Lynn (1990, 1998) has noted that mathematics education is highly heuristic-based, and that the increase in RPM performance may be a consequence of the transferability of these heuristics within narrow domains tapped by both Similarly, Kaufman (2010) has noted that some popular puzzle books and games require the induction of patters, along the lines of the RPM (on cultural exposure to patterns, cf. Greenfield, 1998; Pinker, 2011). Finally, the RPM may be highly sensitive to increased analogical reasoning ability (Fox & Mitchum, 2013) because of its few rules.

1.5.2. Increasing rule sensitivity

Populations may be becoming more sensitive to rules as a consequence of changes in brain functioning. One possible source of this may be changes in the pattern of life history in response to the demographic transition. Life history (*K*) describes the ways in which organisms allocate bio-energetic resources into components of fitness given the level of environmental stability (Figueredo, Cabeza de Baca, & Woodley, 2013). In an unstable environment, organisms opt for a high mating effort strategy (low-*K*), characterized by the channeling of effort into offspring production. High offspring yields can compensate for the high death rates in unstable environments. In a stable environment, organisms can afford to produce relatively fewer offspring, into whom greater parenting effort can be allocated per capita. Furthermore, the organisms themselves can afford to develop more slowly and live longer, healthier lives (somatic effort). Organisms that opt for this strategy also devote resources into building complex, inclusive fitness-enhancing pro-social networks (nepotistic effort).

People have become increasingly high-*K* over the past (Clark, 2007; Figueredo, 2009), especially over the last century due to the demographic transition (Mace, 2000). K encompasses most of behavior and personality (Figueredo et al., 2013), and it exhibits modest environmentality (e^2 ~ .35; Figueredo & Rushton, 2009; Figueredo, Vásquez, Brumbach, & Schneider, 2004). If mortality sources like diseases and malnutrition are eliminated, and stable environmental contexts with opportunities to acquire somatic capital (e.g., education) are provided, people will develop slower life histories. That is, they will reallocate resources into parenting effort longevity, complex social organization, and the ability to acquire greater amounts of somatic capital (Mace, 2000; Woodley, 2012b). These tradeoffs are very well documented in humans and other animals (Ellis, Figueredo, Brumbach, & Schlomer, 2009).

A consequence of increasing K might be increasing cognitive differentiation as a means of socio-cultural specialization — a useful predisposition given the high population densities characteristic of K-selected human ecologies (Woodley, 2011b). The data indicate that higher-K individuals are indeed more cognitively specialized, and like the Flynn effect, this tendency is both independent of the *level* of g (i.e., it manifests independently of a $g \times K$ correlation) and lacks correlation with more g loaded abilities (the effect of K on cognitive ability structure is not a Jensen effect; Woodley, Figueredo, Brown, & Ross, in press). It has been theorized that higher-K individuals need to be more sensitive to rules, as they must navigate complex socio-cultural ecologies in which the capacity to identify socially salient rules translates into opportunities for ecological specialism, and therefore fitness (Woodley, 2011b, 2012b; Woodley, Figueredo, et al., in press). Hence the Flynn effect and, by extension, modernity with all of its concomitant complexity and cognitive diversity may constitute an example of *evoked culture* (Barkow, Cosmides, & Tooby, 1992). That is to say, the Flynn effect and modern life may simply be a by-product of the elicitation of an evolutionarily prepared response to radically altered and improved environmental conditions (Mace, 2000; Woodley, 2012b).

This model is also consistent with both the theory and the observation that diminishing family size is causally involved in the Flynn effect (Rönnlund & Nilsson, 2008, 2009; Zajonc & Mullally, 1997), as this indicates an increase in inter-generational transfers of parental effort. Furthermore, the model is concordant with the involvement of hippocampal functioning in the Flynn effect (Baxendale & Smith, 2012), as cognitive modules relevant to behavioral manifestations of high K-strategies are believed to relate to the hippocampus (Figueredo, Hammond, & McKiernan, 2006; Figueredo et al., 2006).

Whilst g and K do not correlate at the individual differences level (Woodley, 2011b), measures of executive functioning show modest positive correlations with K (Wenner, Bianchi, Figueredo, Rushton, & Jacobs, 2013), suggesting that increases in the former could have been driven by increases in the latter.

Finally, education may increase the ability to induce rules — that is to say, in addition to exposing cohorts to the specific rules on IQ tests, it may increase inductive or categorizational ability in general. One related form of increased rule sensitivity has been proposed by Fox and Mitchum (2013), who argue that performance on culture-fair tests (such as the RPM) is associated with more recent cohorts being better able to utilize 'initial representations' in inferring abstract analogies amongst items. They associate this with Flynn's (2009) concept of 'scientific spectacles', i.e., modern scientific 'habits of thought' adopted by Western or Westernizing populations. These heuristics and 'habits of thought', which could have been acquired either actively (i.e., within an educational context) or passively (i.e., via exposure to the broader culture) may have increasingly prepared modern populations for enhanced performance on highly rule-dependent test batteries like the RPM, such that its 'cognitive scaffolding' is simply more readily apparent to modern generations than it was to previous generations.

It is unknown whether increasing rule exposure, i.e., acquaintance with highly specific rule-sets has any impact on the Flynn effect above and beyond that which is due to increasing rule sensitivity, i.e., increased ability to induce and re-apply rules. However, we believe that it is likely. Cohorts have been exposed to the specific rules on the RPM through education and cultural background, and even minimal acquaintance with the specific rules of the RPM is likely to lead to an increase in scores (this is because it is relatively easy to reapply learned rules when there are relatively few of them).

Neither of the above mechanisms are mutually exclusive to other proposed environmental causes of the Flynn effect. The idea that increasing K may be behind the Flynn effect is especially inclusive of a variety of potential determinants of the effect, as anything that decreases mortality and environmental instability and increases biological condition (i.e., overall phenotypic quality) will elicit the development of higher population levels of K, and therefore potentially higher levels of rule sensitivity. If better quality environments are evoking higher K phenotypes and the Flynn effect at an *epigenetic* level (Greiffenstein, 2011; Storfer, 1999), then people might even be expected to show Flynn effects via heightened rule sensitivity from birth, thus potentially accounting for the apparent finding of Flynn effects in infants (Lynn, 2009).

2. A preliminary test of the rule-dependence model

2.1. Introducing the rule-dependence typology

It is obvious that not all ability batteries contain rules that are equally central to performance. Furthermore, some tests do not rely upon rules at all for their solutions (such as certain crystallized intelligence tests which require that vocabulary items be recalled; cf. Lehrer (2012) for a discussion of similar distinctions in one test of fluency). In order to operationalize this distinction amongst IQ tests and batteries, we propose a qualitative four-level typology into which any ability battery can be placed based on the degree to which the induction of rules is integral to performance. The typology is structured as follows:

Level IV: Batteries in this level possess relatively few rules, but these rules are needed to solve items in the majority of cases. Moreover, items are arranged by the relevant rules, so that the test emphasizes its rule-dependent nature. *Example*: Raven's Standard Progressive Matrices.

Level III: Some rules are present; however, a basic rule-set cannot be relied upon to solve items in all instances. Hence these tests are more diverse in that new rules must sometimes be induced at different stages of progression through the tests. A test such as this is *always* dependent upon explicit categorization, generally figural. *Examples*: Jenkins Non-Verbal Test, Cattell Culture Fair.

Level II: Many rules may be required to solve problems in this level. However, there is no basic rule-set that can be used in all instances. Furthermore, there is a high need to frequently deduce new rules at different stages of progression through the tests. Rules may be more akin to problem-solving strategies or heuristics than to specific computational operations. It is very difficult to give a discrete number of rules for a *Level II* test. This is a broad category, which might include the majority of IQ tests and batteries. *Example*: Block Design.

Level I: No 'cognitive scaffolding' present. Solutions entirely dependent upon recalled knowledge or the use of spontaneous and unstructured problem-solving. *Examples*: Draw-a-Man.

Note that the number of a test does not correspond to its number of rules — i.e., a Level IV test *does not necessarily have 4 rules*. Each level in the typology simply indicates how dependent upon the *centrality of rules* a particular test may be.

2.2. Method and result

There exists a substantial amount of literature on the size of Flynn effects for a wide variety of cognitive ability measures and batteries, so one test of the rule-dependence model is to simply correlate the vector of the position of the test in the rule dependence typology with the vector of the size of the Flynn effect on that test. A positive correlation would indicate that tests that were more dependent upon rules were yielding the larger Flynn effects, consistent with the model.

To test this we have collected together data for Flynn effect gains on 14 different IQ tests, each of which we have assigned to a position in our rule-dependence typology. In terms of inclusion rules, we only consider tests that fall within the pen-and-paper tradition of psychometric testing initiated by Binet. To that end we exclude creativity measures, developmental measures of ability, such as Piagetian staging, and psychophysical measures such as ECTs. We also do not consider anti-Flynn effects on pen-and-paper tests, as recent research has indicated that these are modest Jensen effects (Woodley & Meisenberg, 2013b), and therefore may be etiologically distinct from the Flynn effect (they are probably in large part dysgenic effects rather than environmentally induced reversals of the Flynn effect). The results of this analysis are presented in Table 1.

Table 1 indicates the presence of a substantial and significant vector correlation between the location of a test in our rule-dependence typology and the actual magnitude of the Flynn effect on that test, consistent with predictions.

Fig. 1 presents a scatter plot illustrating the relationship between Flynn effect gains and test position in the rule-dependence typology.

A reasonable point of criticism concerns the subjectivity associated with assigning particular tests to a particular level in the typology. Subsequent research aiming to go beyond this test should employ multiple raters in assigning tests to the various levels. Our test is also somewhat cursory; future tests should be more thorough.

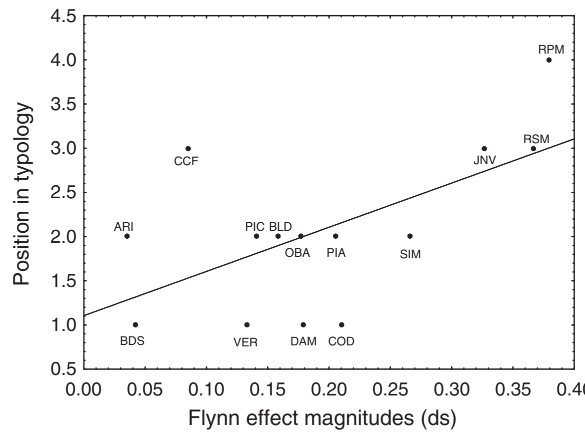

Fig. 1. A scatter plot of Flynn effect magnitudes in standard deviation unit gains per decade (ds) vs. test position in the rule dependence typology. $r = .597$, $P_{(\text{bi-directional})} < .05$, $N = 14$ tests.

3. Discussion

It is proposed that the rule-dependence model can account for the etiological commonalities between both IQ gains via the Flynn effect and those accrued via retesting, in terms of the apparent partial or total independence of these effects from *g*. This is because the tests that yield the biggest gains possess the most 'cognitive scaffolding' in that the reapplication of combinations of relatively simple rules can be used to reliably solve the majority of items. We argue that the mechanism through which these gains accrue involves a cognitive relay or partnership between different abilities. One of these abilities is *g*-as-speed-of-information-processing (or 'chronometric *g*').

Other dimensions – those that have improved – are working memory, which permits the testee to sample and reapply rules once learned; implicit learning, which permits the effortless induction of very simple rules and the quick reapplication of those rules once learned; inductive

Table 1
Flynn effect gains per decade (in standard deviation units per decade, or *ds* and IQ points) along with the position of the test in the rule-dependence typology as indicated by level number. N = 14 tests.

Test (abbreviation)	Gains per decade *ds* (IQ gain [d*15])	Test position in rule-dependence typology: level number	References
Verbal Tests (VER)	.133 (2)	1	From te Nijenhuis (2013)
Similarities (SIM)	.266 (3.995)	2	From Flynn (2009, 2012; averaged from the gains of adults and children)
Backward Digit Span (BDS)	.042 (.625)	1	From Flynn (2009, 2012)
Coding (COD)	.21 (3.15)	1	Flynn (2009, 2012);
Arithmetic (ARI)	.035 (.53)	2	Cf. Backwards Digit Span
Object Assembly (OBA)	.177 (2.65)	2	Cf. Backwards Digit Span
Picture Arrangement (PIA)	.205 (3.075)	2	Cf. Backwards Digit Span
Picture Completion (PIC)	.141 (2.11)	2	Cf. Backwards Digit Span
Block Design (BLD)	.158 (2.365)	2	Mean of Rönnlund and Nilsson (2008; simple regression-derived estimate of gains) and Flynn (2009, 2012).
Jenkins Non-Verbal Test (JNV)	.327 (4.9)	3	From Flynn (1987)
Raven's Standard Progressive Matrices (RPM)	.379 (5.69)	4	From Jensen (1998a); the type of Raven's test is not stated in Jensen's figure, but it is assumed to be primarily SPM data, because APM data is unrepresentative, and there is also less of it (Flynn, 1987; Brouwers, Van de Vijver, & Van Hemert, 2009).
Raven's Progressive Matrices Modified (RSM)	.367 (5.506)	3	This modification refers to an altered variant of the RPM administered in Norway and one other country (Flynn, 1987), which undermined the 5-groups-of-12 schema, thus making the rules harder to apply.
Cattell Culture Fair Test (CCF)	.085 (1.28)	3	Mean of Cattell (1950; given in Flynn, 1987) and Lynn, Hampson and Mullineux (1987).
Draw-a-Man (DAM)	.179 (2.68)	1	Mean of Colom, Flores-Mendoza and Abad (2007, urban data) and Lynn (2006).
Vector correlation (r)	.597*		

* $P < .05$ (non-directional, $N = 14$).

ability, which permits the induction of rules; and the learning of specific heuristics or learned rules, which are linked to the *specific* rules on a test.

This explains why the biggest Flynn effect gains occur on tests like the RPM, which exhibit few rules, combinations of which can be relied upon to solve all items (Carpenter et al., 1990). Our theory thus provides a potential solution to the "Raven's paradox", as it is proposed that tests like the RPM are only highly *g* loaded when encountered initially — even basic familiarity with the rules and heuristics on a test, or improved rule based reasoning, has the potential to radically diminish the *g* loading of this test over time, both under controlled conditions (such as in a retesting scenario) and over larger societal time scales (i.e., across generations in the case of the Flynn effect).

The increasing capacity of societies to detect and explicitly utilize rules as a function of the Flynn effect may be related to increasing rule exposure via mass education and to 'ways of thinking' endemic to cognitive modernity (Flynn, 2009). It may also be caused by increased rule sensitivity as a consequence of changes in brain functioning that might relate to the prevalence of slower life history strategies in the modern era. These in turn might be in part a change in life history strategy due to dramatic improvements in environmental quality.

A small test of the rule-dependence model revealed a .6 correlation between the vector of an IQ test's position in a four-level rule-dependence typology and the vector of the magnitude of the decadal Flynn effect gains for 14 tests. The correlation was statistically significant for an *N* of 14.

3.1. Predictions

The rule dependence model makes a variety of testable predictions, which could form the basis of future research. These include:

i) Measures of non-speeded working memory and implicit learning ability should become less correlated with chronometric *g* due to the retesting effect. This could be investigated using ECTs as a measure of stable or invariant 'chronometric *g*' in a sample being retested on the RPM in addition to measures of non-speeded working memory and implicit learning. *g*-Residualized measures of inductive capacity, working memory and implicit learning should in turn account for much of the variance in performance on tests on which IQ gains have been elicited via the retesting effect.

ii) Implicit learning ability should exhibit a Flynn effect, consistent with those detected on working memory and related measures.

iii) Reanalysis of data on IQ gains accrued via education (e.g., Ceci, 1991) should indicate that these are not occurring on *g* (i.e., that they are substantially hollow), and they should be correlated with rule-dependence numbers, consistent with the hypothesis that schooling is a driver of the Flynn effect.

iv) Individual differences in *K* measured using instruments like the Arizona Life History Battery (Figueredo, 2007) or the Mini-K (Figueredo et al., 2006) may predict the speed with which the *g* loadedness of tests like the RPM decrease as a function of retesting. That is to say, amongst those with high *K*, test *g* loadedness may decrease more rapidly, owing to their higher rule sensitivity.

v) Measures of the speed of the demographic transition coupled with the rate of educational expansion in developing countries should predict the rates of the Flynn effect in these countries. The first of these measures is related to rule sensitivity through changing life history; the second, to rule exposure and rule sensitivity through education.

3.2. Can Flynn effects still be important?

It could be claimed that in asserting that retesting and Flynn effects relate primarily to the rule-dependence of tests, we are simply relegating the effects to psychological insignificance. It has been argued, for example, that *g* is the only source of criterion validity in tests of cognitive ability, and that its removal in turn removes its power to predict real-life outcomes (Jensen, 1998a). This model is however increasingly at odds with the data: more recent research indicates that highly *g* loaded tests, such as the SAT I and ACT, still retain the power to predict real-life outcomes specific to the domain of the test (such as college GPA) even when completely residualized for *g* variance (Coyle & Pillow, 2008). This indicates that non-*g* sources of ability variance matter to specific domains of real-world attainment and success. With regard to our theory, the increases in categorizational ability have real-world consequences as pointed out in Flynn (2009) and Pinker (2011). We hypothesize a possible increase in inductive ability, which may improve certain domains of scientific achievement. After all, as Hume pointed out, induction is the *sine qua non* of science.

Another way in which non-*g* variance might enhance real world performance is at the level of cognitive differentiation. There are many sources of non-*g* variance unique to a variety of cognitive ability batteries and tasks. Differentiation with respect to these sources of variance should lead to enhanced cognitive niche filling and concomitant social and economic diversification (Woodley, 2011b, 2012a; Woodley & Madison, 2013). With enhanced diversity comes increased aggregate efficiency, which translates into enhanced wealth as an inevitable consequence of the operation of Ricardo's Law of Comparative Advantage (Woodley, 2012a; Woodley, Figueredo, et al., in press). Evidence for this comes from the aforementioned observation that the historical trend in the Flynn effect both parallels and predicts the growth in GDP (PPP) per capita experienced by Western nations over the last 10 decades or so (Woodley, 2012a). The Flynn effect may therefore have functioned as a powerful counter to the co-occurring dysgenic trend on 'chronometric *g*' (Woodley, te Nijenhuis, et al., in press), which appears to have inhibited per capita innovation rates in science and technology since the 1870's (Woodley, 2012a; Woodley & Figueredo, 2013). This is because the success of modernity seems to be more contingent upon a population's capacity to specialize with respect to the ability to acquire and utilize numerous sets of simple rules in solving a diversity of complex problems than on its capacity to generate substantive innovation (Woodley, 2012a; Woodley & Figueredo, 2013). A direct consequence of this finding is that increasing rule-dependence has attenuated the *g* loading of modernity such that contemporaneous 'everyday life' is likely to be a less *g* loaded intelligence test than has been previously claimed (Gordon, 1997; Gottfredson, 1997).

Acknowledgments

We would like to thank Roberto Colom, Janet Davidson, Douglas Detterman, Kelly Glover, Mia Laurence, Gerhard Meisenberg and Ted Hermann for extensive help with the manuscript. We are also deeply grateful to James Flynn and two anonymous reviewers.

References

Ackerman, P. L. (1986). Individual differences in information processing: An investigation of intellectual abilities. *Intelligence*, *10*, 101–139.

Ackerman, P. L. (1987). Individual differences in skill learning: An integration of psychometric and information processing skills. *Psychological Bulletin*, *102*, 2–27.

Barkow, J. H., Cosmides, L., & Tooby, J. (Eds.). (1992). *The adapted mind: Evolutionary psychology and the generation of culture*. New York: Oxford University Press.

Batterjee, A. A., Khaleefa, O., Ali, K., & Lynn, R. (2013). An increase of intelligence in Saudi Arabia, 1977–2010. *Intelligence*, *41*, 91–93.

Bauernschmidt, A., Conway, C., & Pisoni, D. (2008). Working memory training and implicit learning. *Research on spoken language processing progress report no. 29*. Bloomington, IN: Speech Research Laboratory, Indiana University.

Baxendale, S. (2010). The Flynn effect and memory function. *Journal of Clinical and Experimental Neuropsychology*, *32*, 699–703.

Baxendale, S., & Smith, N. (2012). Right hippocampal pathology inhibits the Flynn effect in temporal lobe epilepsy. *Journal of Clinical and Experimental Neuropsychology*, *34*, 1033–1040.

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. *Behavioral and Brain Science*, *29*, 109–160.

Brand, C. R. (1996). *The g factor — General intelligence and its implications*. Chichester: John Wiley & Sons, Inc.

Brouwers, S. A., Van de Vijver, F. J. R., & Van Hemert, D. A. (2009). Variation in raven's progressive matrices scores across time and place. *Learning and Individual Differences*, *19*, 330–338.

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

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

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.

Christian, K., Bachnan, H. J., & Morrison, F. J. (2001). Schooling and cognitive development. In R. J. Sternberg, & E. L. Grigorenko (Eds.), *Environmental effects on cognitive abilities* (pp. 287–335). Mahwah: Erlbaum.

Clark, G. (2007). *A farewell to alms: A brief economic history of the world*. Princeton: Princeton University Press.

Colom, R., Flores-Mendoza, C. E., & Abad, F. J. (2007). Generational changes on the Draw-a-Man test: A comparison of Brazilian urban and rural children tested in 1930, 2002 and 2004. *Journal of Biosocial Science*, *39*, 79–89.

Colom, R., Juan-Espinosa, M., & García, L. F. (2001). The secular increase in test scores is a 'Jensen effect'. *Personality and Individual Differences*, *30*, 553–559.

Conway, A.R. A., Jarrold, C., Kane, M. J., Miyahe, A., & Towse, J. N. (2007). *Variation in working memory*. New York: Oxford University Press.

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

Daley, T. C., Whaley, S. E., Sigman, M.D., Espinosa, M. P., & Neuman, C. (2003). IQ on the rise: The Flynn effect in rural Kenyan children. *Psychological Science*, *14*, 215–219.

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

Ellis, B. J., Figueredo, A. J., Brumbach, B. H., & Schlomer, G. L. (2009). Fundamental dimensions of environmental risk: The impact of harsh versus unpredictable environments on the evolution and development of life history strategies. *Human Nature*, *20*, 204–298.

Figueredo, A. J. (2007). The Arizona life history battery [electronic version]. http://www.u.arizona.edu/~ajf/alhb.html

Figueredo, A. J. (2009). Human capital, economic development, and evolution: A review and critical comparison of Lynn & Vanhanen (2006) and Clark (2007). *Human Ethology Bulletin*, *24*, 5–8.

Figueredo, A. J., Cabeza de Baca, T., & Woodley, M.A. (2013). The measurement of human life history strategy. *Personality and Individual Differences*, *55*, 251–255.

Figueredo, A. J., Hammond, K. R., & McKiernan, E. C. (2006). A Brunswikian evolutionary developmental theory of preparedness and plasticity. *Intelligence*, *34*, 211–227.

Figueredo, A. J., & Rushton, J. P. (2009). Evidence for shared genetic dominance between the general factor of personality, mental and physical health, and life history traits. *Twin Research & Human Genetics*, *12*, 555–563.

Figueredo, A. J., Vásquez, G., Brumbach, B. H., & Schneider, S. M. R. (2004). The heritability of life history strategy: The K-factor, covitality, and personality. *Social Biology*, *51*, 121–143.

Figueredo, A. J., Vásquez, G., Brumbach, B. H., Schneider, S. M. R., Sefcek, J. A., Tal, I. R., et al. (2006). Consilience and life history theory: From genes to brain to reproductive strategy. *Developmental Review*, *26*, 243–275.

Flynn, J. R. (1987). Massive IQ gains in 14 nations: What IQ tests really measure. *Psychological Bulletin*, *101*, 171–191.

Flynn, J. R. (2007). *What is intelligence?* New York: Cambridge University Press.

Flynn, J. R. (2009). *What is intelligence?* (2nd ed.)New York: Cambridge University Press.

Flynn, J. R. (2012). *Are we getting smarter?* New York: Cambridge University Press.

Flynn, J. R. (2013). The "Flynn effect" and the Flynn paradox. *Intelligence*. http://dx.doi.org/10.1016/j.intell.2013.06.014.

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.

Girelli, L., Semenza, C., & Delazer, M. (2004). Inductive reasoning and implicit memory: Evidence from intact and impaired memory systems. *Neuropsychologia*, *42*, 926–938.

Gordon, R. A. (1997). Everyday life as an intelligence test: Effects of intelligence and intelligence context. *Intelligence*, *24*, 203–320.

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

Greenfield, P. M. (1998). The cultural evolution of IQ. In U. Neisser (Ed.), *The rising curve: Long-term gains in IQ and related measures* (pp. 207–215). Washington DC: American Psychological Association.

Greiffenstein, M. F. (2011). Secular IQ increases by epigenesis? The hypothesis of cognitive genotype optimization. *Psychological Reports*, *109*, 353–366.

Howard, R. W. (1999). Preliminary real-world evidence that average human intelligence really is rising. *Intelligence*, *27*, 235–250.

Howard, R. W. (2001). Searching the real world for signs of rising population intelligence. *Personality and Individual Differences*, *30*, 1039–1058.

Husén, T., & Tuijnman, A. (1991). The contribution of formal schooling to the increase in intellectual capital. *Educational Researcher*, *20*, 17–25.

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

Jensen, A.R. (1998b). Adoption data and two *g*-related hypotheses. *Intelligence*, *25*, 1–6.

Jensen, A.R. (2006). *Clocking the mind: Mental chronometry and individual differences*. Oxford: Elsevier.

Johnson, W., Bouchard, T. J., Jr., Krueger, R. F., McGue, M., & Gottesman, I. I. (2004). Just one *g*: Consistent results from three test batteries. *Intelligence*, *32*, 95–107.

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

Kagitcibasi, C., & Biricik, D. (2011). Generational gains on the Draw-a-Person IQ scores: A three-decade comparison from Turkey. *Intelligence*, *39*, 351–356.

Kane, H. D. (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. (2000). Secular declines in Spearman's *g*: Some evidence from the United States. *The Journal of Genetic Psychology*, *161*, 337–345.

Kaufman, A. S. (2010). "In what way are apples and oranges alike?" A critique of Flynn's interpretation of the Flynn effect. *Journal of Psychoeducational Assessment*, *28*, 328–398.

Kaufman, S. B., De Young, C. G., Gray, J. R., Jiménez, L., Brown, J., & Mackintosh, N. (2010). Implicit learning as an ability. *Cognition*, *116*, 321–340.

Khaleefa, O., Sulman, A., & Lynn, R. (2009). An increase of intelligence in Sudan, 1987–2007. *Journal of Biosocial Science*, *41*, 279–283.

Lehrer, J. (2012). *Imagine: How creativity works*. Boston: Houghton Mifflin.

Lewandowsky, S., & Oberauer, K. (2009). No evidence for temporal decay in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*, 1545–1551.

Lezak, M.D. (1983). *Neuropsychological assessment*. New York: Oxford University Press.

Lynn, R. (1989). A nutrition theory of the secular increases in intelligence — Positive correlations between height, head size and IQ. *British Journal of Educational Psychology*, *59*, 372–377.

Lynn, R. (1990). The role of nutrition in secular increases in intelligence. *Personality and Individual Differences*, *11*, 273–285.

Lynn, R. (1998). In support of the nutrition theory. In U. Neisser (Ed.), *The rising curve: Long-term gains in IQ and related measures* (pp. 207–215). Washington DC: American Psychological Association.

Lynn, R. (2006). *Race differences in intelligence: An evolutionary analysis*. Augusta: Washington Summit Publishers.

Lynn, R. (2009). What has caused the Flynn effect? Secular increases in the development quotients of infants. *Intelligence*, *37*, 16–24.

Lynn, R. (2011). *Dysgenics: Genetic deterioration in modern populations* (2nd revised ed.) London: Ulster Institute for Social Research.

Lynn, R. (2013). Who discovered the Flynn effect? A review of early studies of the secular increase of intelligence. *Intelligence*. http://dx.doi.org/10.1016/j.intell.2013.03.008.

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

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

Mace, R. (2000). Evolutionary ecology of human life history. *Animal Behavior*, *59*, 1–10.

Meisenberg, G., Lawless, E., Lambert, E., & Newton, A. (2005). The Flynn effect in the Caribbean: Generational change of cognitive test performance in Dominica. *Mankind Quarterly*, *46*, 29–69.

Melby-Lervåg, M., & Hulme, C. (2013). Is working memory training effective? A meta-analytic review. *Developmental Psychology*, *49*, 270–291.

Mingroni, M.A. (2004). The secular rise in IQ: Giving heterosis a closer look. *Intelligence*, *32*, 65–83.

Mingroni, M.A. (2007). Resolving the IQ paradox: Heterosis as a cause of the Flynn effect and other trends. *Psychological Review*, *114*, 806–829.

Must, O., & Must, A. (2013). Changes in test taking patterns over time. *Intelligence*. http://dx.doi.org/10.1016/j.intell.2013.04.005 (in press).

Must, O., Must, A., & Raudik, V. (2003a). The secular rise in IQs: In Estonia the Flynn effect is not a Jensen effect. *Intelligence*, *167*, 1–11.

Must, O., Must, A., & Raudik, V. (2003b). The Flynn effect for gains in literacy found in Estonia is not a Jensen effect. *Personality and Individual Differences*, *34*, 1287–1292.

Must, O., te Nijenhuis, J., Must, A., & van Vianen, A. E. M. (2009). Comparability of IQ scores over time. *Intelligence*, *37*, 25–33.

Nagoshi, G. T., & Johnson, R. C. (1986). The ubiquity of *g*. *Personality and Individual Differences*, *7*, 201–207.

Neisser, U. (Ed.). (1997). *The rising curve. Long-term gains in IQ and related measures*. Washington DC: American Psychological Association.

Nettlebeck, T., & Wilson, C. (2004). The Flynn effect: Smarter not faster. *Intelligence*, *32*, 85–93.

Nevin, R. (2000). How lead exposure relates to temporal changes in IQ, violent crime, and unwed pregnancy. *Environmental Research*, *83*, 1–22.

Pinker, S. (2011). *The better angels of our nature: Why violence has declined*. NY: Viking.

Prokosch, M., Yeo, R., & Miller, G. (2005). Intelligence tests with higher *g* loadings show higher correlations with body symmetry: Evidence for a general fitness factor mediated by developmental stability. *Intelligence*, *33*, 203–213.

Resing, W. C., & Tunteler, E. (2007). Children becoming more intelligent: Can the Flynn effect be generalized to other child intelligence tests? *International Journal of Testing*, *7*, 191–208.

Rindermann, H., Schott, T., & Baumeister, A. E. E. (2013). Flynn effect in Turkey: A comment on Kagitcibasi and Biricik (2011). *Intelligence*, *41*, 178–180.

Rönnlund, M., & Nilsson, L. G. (2008). The magnitude, generality, and determinants of Flynn effects on forms of declarative memory and visuospatial ability: Time-sequential analysis of data from a Swedish cohort study. *Intelligence*, *36*, 192–209.

Rönnlund, M., & Nilsson, L. G. (2009). Flynn effects on sub-factors of episodic and semantic memory: Parallel gains over time and the same set of determining factors. *Neuropsychologia*, *47*, 2174–2180.

Rushton, J. P. (1998). The "Jensen effect" and the "Spearman–Jensen hypothesis" of Black–White IQ differences. *Intelligence*, *26*, 217–225.

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

Rushton, J. P., & Ankney, C. D. (2009). Whole brain size and general mental ability: A review. *International Journal of Neuroscience*, *119*, 691–731.

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.

Saito, S., & Miyake, A. (2004). On the nature of forgetting and the processing–storage relationship in reading span performance. *Journal of Memory and Language*, *50*, 425–443.

Silverman, I. W. (2010). Simple reaction time: It is not what it used to be. *American Journal of Psychology*, *123*, 39–50.

Storfer, M. (1999). Myopia, intelligence, and the expanding human neo-cortex: Behavioral influences and evolutionary implications. *International Journal of Neuroscience*, *98*, 153–276.

Süss, H. -M., Oberauer, K., Wittman, W. W., Wilhelm, O., & Schulze, R. (2002). Working-memory capacity explains reasoning ability — And a little bit more. *Intelligence*, *30*, 261–288.

te Nijenhuis, J. (2013). The Flynn effect, group differences and *g* loadings. *Personality and Individual Differences*, *55*, 224–228.

te Nijenhuis, J., Cho, S. H., Murphy, R., & Lee, K. H. (2012). The Flynn effect in Korea: Large gains. *Personality and Individual Differences*, *53*, 147–151.

te Nijenhuis, J., Murphy, R., & van Eeden, R. (2012). The Flynn effect in South Africa. *Intelligence*, *39*, 465–467.

te Nijenhuis, J., & van der Flier, H. (2007). The secular rise in IQs in the Netherlands: Is the Flynn effect on *g*? *Personality and Individual Differences*, *43*, 1259–1265.

te Nijenhuis, J., & van der Flier, H. (2013). Is the Flynn effect on g?: A meta-analysis. *Intelligence*. http://dx.doi.org/10.1016/j.intell.2013.03.001 (in press).

te Nijenhuis, J., van Vianen, A. E. M., & van der Flier, H. (2007). Score gains on *g* loaded tests: No *g*. *Intelligence*, *35*, 283–300.

Teasdale, T. W., & Owen, D. R. (1989). Continuing secular increases in intelligence and a stable prevalence of high intelligence levels. *Intelligence*, *13*, 255–262.

Tuddenham, R. D. (1948). Soldier intelligence in World Wars I and II. *American Psychologist*, *3*, 54–56.

van Bloois, R. M., Geutjes, L. L., te Nijenhuis, J., & de Pater, I. E. (2009). *g* loadings and their true score correlations with heritability coefficients, giftedness, and mental retardation: Three psychometric meta-analysesPaper presented at the 10th Annual Meeting of the International Society for Intelligence Research, Madrid, Spain.

Verguts, T., De Boeck, P, & Maris, E. (1999). Generation speed in Raven's Progressive Matrices Test. *Intelligence*, *27*, 329–345.

Verguts, T., & De Boeck, P. (2002). The induction of solution rules in Raven's Progressive Matrices Test. *European Journal of Cognitive Psychology*, *14*, 521–547.

Welsh, M. C., & Pennington, B. F. (1988). Assessing frontal lobe functioning in children: Views from developmental psychology. *Developmental Neuropsychology*, *4*, 199–230.

Wenner, C. J., Bianchi, J., Figueredo, A. J., Rushton, J. P., & Jacobs, W. J. (2013). Life history theory and social deviance: The mediating role of executive function. *Intelligence*, *41*, 102–113.

Wicherts, J. M., Dolan, C. V., Oosterveld, Hessen D. J., van Baal, P., G.C.M. Boomsma, D. I., et al. (2004). Are intelligence tests measurement invariant over time? Investigating the nature of the Flynn effect. *Intelligence*, *32*(2004), 509–537.

Williams, R. L. (2013). Overview of the Flynn effect. *Intelligence*. http://dx.doi.org/10.1016/j.intell.2013.04.010 (in press).

Woodley, M.A. (2011a). Heterosis doesn't cause the Flynn effect: A critical examination of Mingroni (2007). *Psychological Review*, *118*, 689–693.

Woodley, M.A. (2011b). The cognitive differentiation–integration effort hypothesis: A synthesis between the fitness indicator and life history models of human intelligence. *Review of General Psychology*, *15*, 228–245.

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

Woodley, M.A. (2012b). A life history model of the Lynn–Flynn effect. *Personality and Individual Differences*, *53*, 152–156.

Woodley, M.A., & Figueredo, A. J. (2013). *Historical variability in heritable general intelligence: Its evolutionary origins and socio-cultural consequences.* Buckingham: The University of Buckingham Press.

Woodley, M.A., Figueredo, A. J., Brown, S. D., & Ross, K. C. (2013). Four successful tests of the cognitive differentiation–integration effort hypothesis. *Intelligence*. http://dx.doi.org/10.1016/j.intell.2013.02.002 (in press).

Woodley, M.A., & Madison, G. (2013). Establishing an association between the Flynn effect and ability differentiation. *Personality and Individual Differences*, *55*, 387–390.

Woodley, M.A., & Meisenberg, G. (2013a). A Jensen effect on dysgenic fertility: An analysis involving the National Longitudinal Survey of Youth. *Personality and Individual Differences*, *55*, 279–282.

Woodley, M.A., & Meisenberg, G. (2013b). In the Netherlands the anti-Flynn effect is a Jensen effect. *Personality and Individual Differences*, *54*, 871–876.

Woodley, M.A., te Nijenhuis, J., & Murphy, R. (2013). Were the Victorians cleverer than us? The decline in general intelligence estimated from a meta-analysis of the slowing of simple reaction times. *Intelligence*. http://dx.doi.org/10.1016/2013.04.006 (in press).

Zajonc, R. B., & Mullally, P. R. (1997). Birth order: Reconciling conflicting effects. *American Psychologist*, *52*, 685–699.