Intelligence 92 (2022) 101643

Contents lists available at ScienceDirect

Intelligence

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

The ups and downs of intelligence: The co-occurrence model and its associated research program

Jonathan Egeland

University of Stavanger, Department of Social Studies, 4021 Stavanger, Norway

A R T I C L E I N F O

Keywords:
Co-occurrence model
Attenuation model
Flynn effect
Woodley effect
g
IQ
Selection

A B S T R A C T

This paper discusses the co-occurrence model and its associated research program, and it argues that the model provides the best supported theory of secular changes in cognitive ability. The co-occurrence model offers a better solution to Cattell's paradox (relative to the alternatives in the literature), and it is able to accommodate Flynn's four major paradoxes also. A review of empirical work conducted in order to test the model's predictions demonstrates that many populations in which selection favors lower intelligence have experienced a decline in *g* or some cognitive ability variable that correlates with *g*, at the same time that average phenotypic IQ has increased. Moreover, since the co-occurrence model makes predictions about variables that are not directly concerned with cognitive ability testing, its research program can be extended to other domains of research.

1. Introduction

Are humans becoming more or less intelligent? This question has occupied intellectuals for over a century and a half (Galton, 1869), and it has been systematically studied at least since the 1930s. Runquist (1936) was the first to observe a secular increase in IQ scores, shortly followed by Johnson (1937), Roesell (1937), Smith (1942) and Wheeler (1942). Their findings suggested that people were generally becoming more intelligent, as indicated by their increasing average IQ scores. Despite being a significant discovery, it remained relatively unremarked upon until almost 50 years later when the effect was rediscovered by Lynn (1982) and Flynn (1984)—the latter of whom famously found evidence for a 13.8 point increase in IQ scores among Americans on successive versions of the Wechsler and Stanford-Binet IQ tests between the years 1932 and 1978, amounting to a decadal increase of approximately 3 points.[1] And only 3 years later, he replicated the finding with respect to 13 additional nations (Flynn, 1987). After this, the apparently ubiquitous secular increase in IQ scores gained widespread attention, and it eventually became known as the "Flynn effect"—a term coined by Herrnstein and Murray (1994, p. 307).[2]

The Flynn effect certainly provides prima facie evidence for an affirmative answer to the question with which this article began, especially given the fact that the IQ tests on which the Flynn effect have been observed are both (typically) highly *g* loaded and highly correlated with real world outcomes that are antecedently known to draw upon important intellectual abilities, such as job performance and educational attainment (Deary, Strand, Smith, & Fernandes, 2007; Jensen, 1998; Schmidt & Hunter, 1999). However, despite massive accumulating data demonstrating a Flynn effect of approximately 3 IQ points per decade in countries from all around the world (Pietschnig & Voracek, 2015; Trahan, Stuebing, Fletcher, & Hiscock, 2014), intelligence researchers have often been skeptical of the claim that the increase in IQ scores during the 20th century reflects a genuine improvement in intellectual ability. Indeed, skepticism is expressed by Flynn in his seminal (1987) article, and it is now being echoed by commentators demonstrating that IQ tests do not have the property of measurement invariance when different generational cohorts are compared, meaning that they do not have the same underlying ability factor structure across time (Fox & Mitchum, 2014; Must, te Nijenhuis, Must, & van Vianen, 2009; Wicherts et al., 2004).

E-mail address: jonathan.egeland@uis.no.

[1] Schaie and Strother (1968) also deserve recognition, as they were the first to attribute test score changes to more than measurement/design artifacts.

[2] Lynn (2013) has suggested that the trend should be called the "Runquist effect" after its discoverer. Others, like Rushton (1999) and Voracek (2006), prefer to call it the "Lynn-Flynn effect", and some even call it the "FLynn effect" (Rindermann & Becker, 2018), in order also to acknowledge Lynn's contributions. It should be further noted that Furnham (2008) has proposed that the term "Lynn effect" be reserved specifically for instances where the Flynn effect has gone into reverse (i.e., in the case of so-called "anti", or "negative" Flynn effects).

https://doi.org/10.1016/j.intell.2022.101643
Received 10 December 2021; Received in revised form 12 March 2022; Accepted 12 March 2022
Available online 23 March 2022

An alternative view of the matter is provided by theories operating within what may be called the *postindustrial selection research program* (PSRP).[3] The theories that constitute the PSRP all argue that there is a sense in which people generally are becoming *less* intelligent, and that the correct etiological explanation of this fact will have to include selection as an important causal factor.[4] The latest and most sophisticated of these theories is the co-occurrence model, first articulated by Woodley (2012a) and later tested and developed in much greater detail by Woodley and Figueredo (2013). This article will offer a statement of the theory, an evaluation of its explanatory power and predictive accuracy, and, by attempting to derive novel and risky predictions from it, the article will extend the PSRP to other fields of scientific research. The aim of the article is thus twofold: it aims to provide a coherent statement of the theory that is the co-occurrence model, by focusing on its explanatory power and predictive accuracy, as well as arguing that the theory plausibly has the capacity to explain and predict phenomena outside the realm of cognitive ability testing—in which case the PSRP should be extended to other fields of research.

2. The co-occurrence model and Cattell's paradox

During the first half of the 20th century, many intellectuals were interested in measuring and understanding the effects of selection pressures favoring lower intelligence. Following Galton (1869), interest in the effects of such selection pressures in the early 1900s was primarily spurred by the discovery that there was a negative association between fertility, as measured by numbers of siblings or numbers of offspring, and intelligence, as measured by IQ tests or IQ proxies, such as educational attainment or social status (Bradford, 1925; Burks & Jones, 1935; Cattell, 1936, 1937; Lentz Jr, 1927; Lynn, 1996; Reeve, Heeney, & Woodley of Menie, 2018; Skirbekk, 2008). As a heritable trait, it was predicted that IQ should be declining due to the action of this fertility differential, with early predictions placing the anticipated decline at around one IQ point per decade (Cattell, 1937).

However, contrary to the expectation of IQ decline based on differential fertility rates, early attempts at measuring the effects of selection that favors lower cognitive ability were not just unsuccessful, but their results even appeared paradoxical as in several studies it was found that younger cohort groups were outperforming older ones on the same IQ battery (Burt, 1948; Cattell, 1950; Tuddenham, 1948). This quandary became known as "Cattell's paradox" (Higgins, Reed, & Reed, 1962), which concerns the apparent inconsistency in the following set of claims, each of which is individually plausible and supported by the evidence:

1. Differences in reproductive success favor lower intelligence.
2. Intelligence is a highly heritable trait.
3. There has been a secular increase in IQ scores in Western populations of approximately 3 points per decade (i.e., the Flynn effect).[5]

What arguably is the most prominent potential solution to Cattell's paradox is now known as the *attenuation model*, and its core claim is that selection against IQ cannot be measured at the phenotypic level, but rather has to be inferred since it is being masked by the comparatively larger Flynn effect (Woodley & Figueredo, 2013). In other words, the attenuation model suggests that the Flynn effect—which is likely caused by environmental or environmentally-mediated factors, such as greater access to education (Baker, Eslinger, Benavides, Peters, & Dieckmann, & León, J., 2015; Flynn, 1984) and other socioenvironmental improvements, such as increased nutritional quality (Lynn, 1990, 2009) and a slowing of life history speed (Pietschnig & Voracek, 2015; Woodley, 2012b)—has attenuated the relatively smaller losses in "genotypic IQ".[6] Moreover, the model also claims that attenuation of reduced cognitive ability due to the Flynn effect happens both at the level of performance on IQ tests and at the level of social trends, such as economic and technological development. Indeed, just as the decline in "genotypic IQ" is attenuated by the comparatively larger Flynn effect, there is a similar decline in aptitude relevant to economic and technological improvement that is being attenuated by the comparatively larger beneficial effects that the environmental factors responsible for the Flynn effect have—at least according to the model in question.

The attenuation model was first proposed by Burt (1948), and it has later been taken up by Lynn (1996), who argues that the Flynn effect is revealing an improvement in "phenotypic IQ", whereas the effects of selection (for the time being) only occur at the level of "genotypic IQ" and therefore have to be inferred, due to the attenuating effects of the comparatively larger phenotypic gains due to environmental improvements. More sophisticated and recent theoretical restatements of the attenuation model invoke the action of gene-by-environment interactions and cultural transmission pathways as the mechanism allowing for the translation of qualitative environmental improvements into increased expressivity of genetic factors associated with educational attainment and related phenotypes (e.g., Sauce & Matzel, 2018; Uchiyama, Spicer, & Muthukrishna, 2021), despite an apparent decline in the latter due to negative associations between these and fertility patterns (Beauchamp, 2016a, 2016b; Courtiol, Tropf, & Mills, 2016).

The attenuation model is therefore one of the theories that figures in the PSRP mentioned previously, since it agrees with the claims that there is a sense in which people generally are becoming less intelligent, and that the correct etiological explanation of this fact will have to include selection as an important causal factor. More specifically, according to the model, people are becoming less intelligent in the sense that their "genotypic IQ" is decreasing due to the effects of lower fertility among higher IQ individuals, but this reduction is (for the time being) unobservable at the level of test-performance or social trends, which are only influenced by "phenotypic IQ". An apt illustration of the model in question is thus offered by Loehlin's (1997) analogy of rising tides and leaky boats, where the decline in "genotypic IQ" is represented by leaky boats, and the Flynn effect is represented by the (comparatively much greater influence of) rising tides.

However, a problem with the attenuation model is that there exists evidence that appears to be incompatible with it. One such piece of evidence is provided by Murray (2003) who demonstrated historiometrically that the per capita number of eminent figures in both the arts and sciences has been declining since the late 1800s. This finding was echoed by Simonton (2013) who argued that the phenomenon of genius was essentially "extinct" after Einstein. These trends might indicate that selection favoring lower cognitive ability may have had real, unattenuated social effects. Another piece of relevant evidence is provided by Bloom, Jones, Van Reenen, and Webb (2020) and Boeing and Hunermund (2020), who show that there have been large declines in research productivity. More evidence is also offered by Huebner (2005), who demonstrates that there has been a significant worldwide decline in the per capita number of major innovations since the 1870s, again

[3] For more on the nature of scientific research programs, see Lakatos (1978a).

[4] The term "postindustrial selection research program" is fitting since its contemporary proponents generally agree that the selection pressures favoring lower intelligence arose in the middle of the 19th century, after the advent of the Industrial Revolution (Lynn, 1996; Woodley, Younuskunju, Balan, & Piffer, 2017).

[5] Implicit in the paradox is an assumption that these co-occurring correlational patterns have causal effects—more specifically, that differences in reproductive fitness have causal effects that are expressed phenotypically in the subsequent generation (Cattell, 1937). This is an assumption on which the co-occurrence model also relies.

[6] An individual's "genotypic IQ" should be conceptualized in terms of their genetic potential for a certain IQ level, irrespective of whether they attain it or not (Lynn, 1996).

indicating that there are certain negative social developments that the increase in "phenotypic IQ" might not have been able to offset. Now although the evidence provided by these authors appears to falsify the attenuation model, it does not for that reason undermine the PSRP as a whole. Indeed, it is in light of this evidence that the more sophisticated co-occurrence model—which also belongs to that same research program—was developed as an alternative potential solution to Cattell's paradox.

The co-occurrence model argues that selection pressures favoring lower means of cognitive ability and the Flynn effect have affected different variance components of IQ. For purposes of illustrating this model the IQ "atom", which first was "split" by Spearman (1904), can be (crudely) separated into four different variance components: heritable general cognitive ability (*g.h*), environmental general cognitive ability (*g.e*), heritable specialized abilities (*s.h*), and environmental specialized abilities (*s.e*) (Woodley & Figueredo, 2013). Since the Flynn effect is concentrated on the less *g* loaded aspects of intelligence (te Nijenhuis & van der Flier, 2013), which also tend to be much less additively heritable (Panizzon et al., 2014), secular gains are for the most part due to an improvement in environmentally influenced specialized abilities. Selection, on the other hand, is concentrated on the most heritable and most *g* loaded aspects of intelligence (Meisenberg, 2010; Peach, Lyerly, & Reeve, 2014; Woodley & Meisenberg, 2013a; Woodley of Menie et al., 2017), meaning that we should expect a decline in *g* and in other abilities and trends associated with it.

The central claim of the co-occurrence model is thus that there have been co-occurring gains and losses in intelligence, but that these diametrically opposed trends are concentrated on different cognitive variance components that can vary independently of each other. This presents a neat solution to Cattell's paradox since it eliminates the apparent inconsistency associated with the three claims mentioned above.[7] Not only can all of them be true at the same time, but the effects of selection should also be measurable at the phenotypic level since the predicted diminution in cognitive ability occurs on an ability component with respect to which the Flynn effect is disparate. Moreover, in contrast to the attenuation model, the co-occurrence model successfully predicts the declining rates of major innovation and eminent innovators, as a society's capacity to produce such innovations is likely to be highly conditional upon there being sufficient numbers of individuals within a population possessing very high levels of *g* (Woodley, 2012b; Woodley & Figueredo, 2013). Continuing, the next section will summarize and examine more closely the most important studies that have tested the predictions made by the co-occurrence model.

3. The 'Woodley effect'

The co-occurrence model posits that there has been a decline in *g* since the middle of the 19th century when fertility became negatively associated with intelligence (and proxies, such as educational attainment) in many populations,[8] and, moreover, it predicts that the decline is likely to be measurable on any cognitive ability variable that satisfies one, or both, of two criteria:

1) *High heritability:* If *g* is the primary source of heritability among subtests (Panizzon et al., 2014), and assuming that gene-by-environment interactions do not substantially modify the heritability of *g* in adults (as is indicated by some of the available data; Mingroni, 2007; Sesardic, 2005), then a highly heritable measure of cognitive ability is likely not only to be a strong indicator of *g*, but is also likely to be resistant to the sorts of environmental improvements that cause the Flynn effect. This means that such a measure might also exhibit secular trend characteristics that are caused (in part) by selection (example: vocabulary knowledge within cultures).
2) *Measurement invariance:* This describes the statistical property of a given scale or set of scales (such as an IQ test battery) to reliably measure the same parameter across different measurement occasions (i.e., across different cohorts separated by time in the case of secular trends). The Flynn effect is known to violate measurement invariance, as different parameters are being measured in more recently assessed cohorts relative to older ones (Fox & Mitchum, 2014; Must et al., 2009; Wicherts et al., 2004). This is consistent with the Flynn effect being primarily associated with *s.e* (i.e., a performance increase in one, or a small number of specialized abilities, or test specificities, rather than in *g*). A measure might be able to reliably track *g* over time even if it is not strongly correlated with it at the individual differences level, by virtue of being proximate for a more biologically or cognitively fundamental process (such as an endophenotype) that can potentially be more directly measured across cohorts (examples: simple reaction time, working memory).

On this basis, we would expect to be able to measure secular declines on any prospectively highly heritable and/or invariant measure of *g*, in any population where fertility is negatively associated with intelligence. Since the development of the co-occurrence model by Woodley and Figueredo (2013), a number of studies have been conducted in order to test the model's core prediction of declining *g*. The upshot of these studies is an impressive amount of coherent evidence, most of which is consistent with the model's predictions by virtue of showing either that there has been a decline in *g* among Western populations, or that there has been a decline in some manifestation of cognitive ability that functions as a proxy measure of *g* by virtue of significantly correlating with it.

The evidence so far presented in the literature can be broken down into the following classes:

3.1. Direct observations of IQ decline

Decreasing IQ scores (a phenomenon sometimes called the negative/anti-Flynn effect) have, in terms of magnitude, been found to positively associate with the *g* loadings of subtests in two countries, the Netherlands (Woodley & Meisenberg, 2013b), and France (Woodley of

[7] Another potential solution to Cattell's paradox is that what causes the co-occurring gains in phenotypic IQ and losses in *g* is increasing cognitive ability differentiation, perhaps due to more schooling and education that encourages the use of specialized abilities rather than *g* (Pietschnig, Deimann, Hirschmann, & Kastner-Koller, 2021; Pietschnig, Voracek, & Gittler, 2019). However, by way of objection, increased cognitive ability differentiation does not imply a lower level of *g*—and there is ample evidence for such a diminution (cf. section 3). It is possible to imagine a situation in which the breadth of abilities encompassed by the positive manifold of *g* might shrink, owing to the Flynn effect "leveraging" gains among these narrower skills and competencies. However, under such conditions, a population's *level* of *g* might still remain constant (this just being the integral of their performance level with respect to whatever abilities are captured by *g* at that instant). In theory, *g* could increase or decrease entirely independently of the degree of ability differentiation. Moreover, in so far as there is any kind of an individual-differences level association between the level of *g* and the strength of its positive manifold, Spearman's Law of Diminishing Returns suggests that higher levels of *g* should be accompanied by *greater* ability differentiation (Blum & Holling, 2017). Furthermore, another reason as to why the co-occurrence model is more plausible, is that it predicts that certain heritable dimensions of brain quality, which we already know have implications for individuals' level of *g*, will undergo decline as a result of selection that favors lower intelligence (cf. Sections 3 and 6). This means that the co-occurrence model can explain the declining levels of *g* on the basis of variability in physiologically relevant variables.

[8] Before that, individuals possessing characteristics that are suggestive of higher cognitive ability (such as higher social status and wealth) appear to have enjoyed a fitness advantage as they generally had more surviving offspring compared to those with lower levels of these characteristics (Clark, 2007; Skirbekk, 2008).

Menie & Dunkel, 2015). In the latter case, where the declines were identified using the Wechsler Adult Intelligence Scale (WAIS), it was also found that the magnitude of the association between ability scores and fertility across subtests, simple visual reaction time-subtest correlations, along with heritabilities (derived from US WAIS samples) also clustered with subtest *g* loadings and the secular IQ declines, suggesting a biological "nexus" among these effects. The interpretation of this finding as evidencing a biological causal nexus has recently been critiqued however, on the basis that "culture loadings", which are clearly non-biological, also correlate with subtest *g* loadings and the performance declines across subtests, suggesting that changes in French culture in the gap between measurement and remeasurement might more plausibly explain the declines (Gonthier, Grégoire, & Besançon, 2021). A recent study (Hegelund et al., 2021) has yielded some support for the involvement of selection pressures in these secular declines. It noted that a very small proportion of the variance (0.01%) in the negative Flynn effect among Danish conscripts, measured using the Danish Draft Board's intelligence test (Børge Priens Prøve), can be explained by the change in year-by-year strength of the negative correlation between IQ and family size.

In a synthetic review of negative Flynn effects spanning 87 years, 13 countries, and 66 observations, Woodley of Menie, Peñaherrera-Aguirre, Fernandes, and Figueredo (2018) found that the "*g*-ness" (i.e., the aggregate domain *g* loading assigned to different batteries) negatively predicted the magnitude of the decline, meaning that taken as a whole, the negative Flynn effect is not on *g* and may be driven in large part by negative cultural and environmental factors undoing the benefits of positive factors that may have driven the Flynn effect historically. However, a positive residual impact of demographic changes was found on the rates of IQ decline between countries, and this was also found to interact positively with domain *g*-ness, indicating that independently of the main non-*g* environmental/cultural drivers of the negative Flynn effect, some of the decline is also occurring at the level of *g* in response to demographic factors. Woodley of Menie, Peñaherrera-Aguirre, et al. (2018) note that this process could involve declines in the *g.e.* variance component.

3.2. Perceptual and psychomotor declines

Slowing of simple visual reaction times was first noted in Western populations by Silverman (2010), who compared the performance means of modern populations with respect to a large reference study conducted on the UK population in the 1880s by Francis Galton. In reanalyzing an expanded version of Silverman's dataset, Woodley, te Nijenhuis, and Murphy (2013) noted that the decline was robust to the use of random effects meta-regression, and that in terms of IQ equivalent change the slowing translated to a loss of slightly over one point per decade. The conclusions of these authors were however highly controversial, with it being argued instead that the apparent decline was likely an illusion caused by method variance associated with the use of different measurement systems for evaluating reaction time speed in different studies (Dodonova & Dodonov, 2013; Nettelbeck, 2014; Parker, 2014; Woods, Wyma, Yund, Herron, & Reed, 2015). Reanalysis of these data using corrections for method variance and tighter cultural homogenization (i.e., restricting the analysis to Anglosphere countries, or just to the UK) continued to yield indications of substantial decline (amounting to 20–40 ms per century; Woodley, te Nijenhuis, & Murphy, 2014; Woodley of Menie, te Nijenhuis, & Murphy, 2015a; Woodley of Menie, te Nijenhuis, & Murphy, 2015b). These findings have also been critiqued on the basis that more recently born cohorts appear to be able to solve problems on IQ tests at a faster rate than older ones when accounting for other factors (Must & Must, 2018). However, Must and Must (2018) only showed that test-taking speed increased, not simple reaction time, and it is not clear whether the former measure has the same relationship to *g*.

Analysis of cross-sectional Swedish data on simple auditory reaction time found strong indications of a decline that exceeded the expected age-related slowing of reaction time, such that older participants exhibited faster reaction times than younger ones (Madison, Woodley of Menie, & Sänger, 2016). In this study all participants were evaluated using the same testing platform, implying that the apparent decline was not likely to be an artefact of method variance. Finally, a systematic review of studies investigating both cross-sectional and longitudinal measurements of reaction time speed as a function of ageing, found evidence that, in the case of simple reaction time, cross-sectional slowing is generally greater than longitudinal slowing (ratio = 0.9, $N = 4078$, $K = 3$ studies), which is consistent with an overall secular decline in performance (Verhaeghen, 2014, see reanalysis in; Woodley of Menie, te Nijenhuis, & Murphy, 2015b). In another study, Woodley of Menie and Fernandes (2016b) found indications of a secular increase in the error rates associated with performance on a standard measure of color acuity across four normalization samples. The association between color acuity and cognitive performance was found to be strongly positively moderated by subtest *g* loadings.

3.3. Patterns in vocabulary usage

Several studies have found indications that presently "abstract", "difficult", or otherwise "hard to learn" words have been declining in usage across the English language corpus since the late 19th and early 20th centuries (e.g., Hills & Adelman, 2015; Roivainen, 2014). As vocabulary knowledge functions (within cultures) as both a potent and highly heritable measure of *g* (e.g., Kan, Wicherts, Dolan, & van der Maas, 2013), the co-occurrence model predicts that these cultural patterns might be reflective of selection acting to reduce *g*. This effect was first predicted in Woodley et al. (2014) and was directly tested in Woodley of Menie, Fernandes, Figueredo, and Meisenberg (2015). In this study target vocabulary words were selected from the WORDSUM test, along with item-level difficulties, and the degree to which the pass-rate on each item predicted fertility among those in completed or near-completed fertility (> 40 years old); these parameters were all computed using data from the US General Social Survey. The utilization frequencies of these words were then computed across the textual corpus of Google's Ngram Viewer between the years 1850 and 2000. In modelling the factors that influenced the change in frequency of word use for each word, it was found that the four high-difficulty words were all decreasing throughout this time period, whereas the six low-difficulty words were all increasing. The degree of decrease was predicted by the item's 'age' (e.g., how long it has existed in the English corpus), difficulty parameter, by the strength of the negative association between the item pass rate and fertility, and by the two-way interaction between item difficulty and the pass-rate-fertility association (items reduced in utilization faster given higher magnitude pass-rate-fertility association and difficulty). The major factors predicting the increase among items over time were low difficulty, rising written literacy rates (modelled as a time trend), and the interaction between these two (low difficulty items rose more rapidly in response to rising literacy).

3.4. Decreasing working memory and spatial reasoning

Decreasing working memory ability was first noted in Woodley of Menie and Fernandes (2015), who reanalyzed data first published by Gignac (2015) on secular trends in digit span across WAIS normalizations, starting in the 1930s. It was noted that there was a small, but significant decrease in the number of "bits" of information that subjects recalled across decades in the "digits backward" configuration, which was accompanied by a slightly larger magnitude gain in "bits" recalled in the "digits forward" configuration in the same set of subjects. It was also noted that working memory is both more *g* loaded and more heritable than short-term memory, which means this pattern of opposing secular trends is precisely in line with the predictions of the co-occurrence model. A much more thorough and expansive meta-analysis of the available literature on two recall tasks (Corsi blocks and digits) replicated both sets of findings as there were significant secular declines on both working memory and secular gains on both short-term memory configurations (Wongupparaj, Wongupparaj, Kumari, & Morris, 2017). These trends persisted even after controlling for a large number of covariates.

More recently, Graves et al. (2021) have found further evidence of secular declines with respect to the attention/working memory and learning trials when performance on the third edition of the California Verbal Learning Test (CVLT) was compared with the second edition of the CVLT. A meta-analysis by Pietschnig and Gittler (2015), which specifically set out to look for evidence of declining *g*, found indications of substantial declines in performance among German and Austrian samples on three-dimensional rotation ability.

3.5. Decreases in developmental stability

Evidence of long-term (i.e., across more than one century) decreases in potential proxies for developmental stability with respect to two indicators that correlate with *g* have been found. In the study of Woodley of Menie and Fernandes (2016a), a dataset on craniofacial fluctuating asymmetry was reanalyzed, yielding indications of a secular increase (starting in the early 19th century). In the study of Woodley of Menie, Fernandes, Kanazawa, and Dutton (2018) a regression discontinuity design was used to identify a multi-decadal secular increase in the proportion of individuals exhibiting sinistrality, which was found (in the same study) to be weakly negatively related to *g* in three datasets. As these indicators of developmental stability are only very weakly correlated with *g* at the individual-differences level (possibly via the action of mildly pleiotropic deleterious mutations that may be accumulating due to increased environmental mildness since the onset of the Industrial Revolution, e.g., Kondrashov, 2017), the estimated secular declines in *g* associated with these trends were noted to likely be extremely small in both cases. Although the estimated declines in *g* are small, they are nevertheless consistent with the co-occurrence model and its predictions, in so far as increased mutation load constitutes another source of genetic change that can potentially negatively influence levels of *g.h*.

3.6. Pre-Industrial gains in g

As noted in footnote 8, prior to the 1800s in many Western countries, signifiers of social status tended to be positively associated with larger numbers of surviving offspring (Clark, 2007; Skirbekk, 2008). On this basis, it might be expected that measures of *g* will have been increasing. Evidence consistent with this has been found in historical measures of per-capita rates of innovation and in the occurrence of eminent individuals (Huebner, 2005; Murray, 2003), both of which show signs of having increased from the 14th to 15th centuries until the mid-1800s, after which they decrease. In tracking the utilization frequencies of the four highest difficulty items from WORDSUM using Ngram viewer, Woodley of Menie, Figueredo, et al. (2017) found that the utilization frequency of the set of these words was increasing between 1600 and 1850, and that it was decreasing thereafter. Another converging line of evidence supporting potential historical increases in *g* is the finding of Woodley of Menie et al. (2017), who estimated various cognitive ability polygenic scores for a sample of ancient genomes (most of which were sourced from Bronze Age populations living in Eurasia). The levels of these scores were then compared with the equivalent scores among ancestrally matched genomes from the 1000 genomes dataset, where it was found that they were significantly lower. Even among the sample of ancient genomes, a positive correlation was noted between the frequencies of these variants and sample recency. This suggests that the variants comprising these polygenic scores might have been under positively directional selection throughout much of the intervening period. Further evidence for historical polygenic selection favoring higher *g* has been identified by Srinivasan et al. (2018), who found that more evolutionarily salient regions of the human genome were significantly enriched for variants associated with educational attainment and cognitive ability, suggesting that selection had acted to enrich these regions in period since human-Neanderthal divergence.

What might be considered remarkable about some of these lines of evidence is that in some cases, the full set of predictions associated with the co-occurrence model (i.e., simultaneous increases in less *g* loaded, narrower, and more environmentally sensitive abilities should co-occur with respect to declines in more *g* loaded and more heritable abilities) have been evidenced (e.g., Wongupparaj et al., 2017). This is most strikingly evident in the case of Woodley of Menie, Fernandes, et al. (2015), where both selection (as a cause of declining *g*) and rising literacy rates (as a cause of narrow ability gains, for evidence of this see: Must, Must, & Raudik, 2003) were found to be significantly associated in the theoretically expected direction with the secular decrease in utilization frequencies among the high-difficulty vocabulary items and co-occurrent secular increase among the low-difficulty items respectively.

Another notable demonstration of the co-occurrence model can be found in Woodley of Menie, Figueredo, et al. (2017), where a number of diverse measures of *g* (simple visual reaction time, working memory, per capita major innovation rate, high-difficulty vocabulary utilization, and usage of altruism-connoting words) were found to covary strongly in time with one another, yielding a common ***g.h decline*** factor (with loadings ranging from 0.724 to 0.927). A similar factor tracking the co-occurrent rise in ***s.e*** was also estimated, employing the temporal correlations among short-term memory, GDP per capita (as a proxy for "microinnovation" rate, growth in which has been found to strongly correlate with the Flynn effect, Pietschnig & Voracek, 2015; Rindermann & Becker, 2018), and data on the utilization frequencies of three different sets of easy-to-learn words. These two factors correlated strongly and negatively with one another across time, consistent with expectations.[9]

In his work on the attenuation model, Lynn (2009) has estimated that there should be a decline in "genotypic IQ" of approximately 0.3 points per decade—an estimate that is consistent with that of Kong et al. (2017), who, in examining the associations between fertility and polygenic scores for educational attainment among a large, Icelandic cohort, similarly estimated an expected decline in the region of 0.3 IQ points per decade when correcting for missing heritability. This "traditional" estimate may however be slightly deflated. The reason being that Kong et al. (2017) relied on an implausibly low estimate of the additive heritability of IQ ($h^2 = 0.3$). When the estimated decline in Kong et al. (2017) was corrected using better empirically supported values for the additive heritability of *g* (Panizzon et al., 2014), the expected decline reached a value of almost 1 point per decade (Woodley of Menie, Figueredo, et al., 2017). Moreover, this corrected estimate is similar to the estimate in Woodley of Menie, te Nijenhuis, et al. (2015b, see: supplement S1) on the basis of highly homogenized reaction time data from the UK.

That said, it is not at the moment possible to provide a single value for the decadal decline in *g*, since the various studies calculating estimates make use of measures with very different *g* loadings, and different relationships to the development of *g* across the life course (for discussion of this in relation to working memory and processing efficiency, see: Demetriou & Spanoudis, 2017). There exists substantial heterogeneity in the strengths of selection across subpopulations also (Hugh-Jones & Abdellaoui, 2021; Reeve et al., 2018). However, the most accurate value

[9] Woodley of Menie, Fernandes, et al. (2017), Woodley of Menie, Figueredo, et al. (2017) also found evidence for the existence of a third temporal factor capturing various somatic modifications (specifically increased brain size, height, BMI, fluctuating asymmetry, and sinistrality). The three factors yield a higher order temporal factor which the authors termed the *Nexus*. This factor was found to be strongly positively predicted by indicators of increasing environmental mildness (such as increased mean global temperature) over 200 years.

to date is probably found in Woodley of Menie, Figueredo, et al. (2017), who note that the average decline estimated for three measures (working memory, high-difficulty vocabulary usage and simple visual reaction time) based on their *g.h* common-factor model, is 0.44 IQ points per decade. This value is highly consistent with both Beauchamp's (2016a) estimated decline in educational attainment (−0.3 months of attained education per generation) based on genetically informed analysis of the Health and Retirement Study, which rescales to a corresponding decadal IQ decline of around 0.4 points, and the similar estimate of declines in the Wisconsin Longitudinal Study based on polygenic scoring data, using the formula from Kong et al. (2017). The actual decline in *g* probably has a value of approximately half an IQ point per decade; higher estimates are likely to capture both *g* and non-*g* declines, or possibly some combination of *g.h* and *g.e* declines, in both cases capturing changes due to selection as well as adverse environmental factors. Consistent with this are the results of Woodley of Menie, Peñaherrera-Aguirre, et al. (2018), where it was noted that the negative Flynn effect was associated with decadal IQ declines ranging from −0.98 points (among trends in the spatial domain), to −1.7 points (in the fluid domain). These declines (−1.51 averaged IQ points across four domains) are far too large to have been caused by selection alone, and they are certainly consistent with a substantial impact stemming from environmental and cultural factors which might be acting to reverse the Flynn effect among these populations.[10]

However, it must be noted that there are some observations that might be taken as evidence against the co-occurrence model. For example, Flynn (2012) found that American adults experienced increasing IQ scores on the WAIS between 1995 and 2006 whose subtest magnitudes correlated positively with the *g* loadings of the subtests, with values between 0.540 and 0.621.[11] On the other hand, a meta-analysis of the available literature on these effects yielded a negative aggregate vector correlation, suggesting that these findings may have been an outlier (te Nijenhuis & van der Flier, 2013).

Another potential issue with the model is how to accommodate the fact that head and brain size—two endophenotypes that correlate with IQ (Nave, Jung, Karlsson Linnér, Kable, & Koellinger, 2019; Pietschnig, Penke, Wicherts, Zeiler, & Voracek, 2015)—have increased, seemingly in tandem with the Flynn effect (Lynn, 2009). Woodley of Menie, Peñaherrera-Aguirre, Fernandes, Becker, and Flynn (2016) demonstrated, via reanalysis of existing data, the presence of robust multidecadal secular gains in brain mass in both UK and German samples. However, the magnitude of the corresponding Flynn effect is extremely small owing to the low magnitude of the association between brain size and IQ (with the gains only being able to account for very small percentages in the observed secular gains in these two countries matched for time range). Woodley of Menie, te Nijenhuis, Fernandes, and Metzen (2016) considered a number of possible reasons for this apparent anomaly with respect to the co-occurrence model. The best supported of these is that brain components do not scale isometrically as a function of overall brain size, which means that most of the gain may be occurring in specific neuroanatomical regions with known associations with the Flynn effect (e.g., the right-hippocampal formation, Baxendale & Smith, 2012). This means that overall brain volume among individuals might contribute substantially to variation in components of ability without necessarily contributing substantially to *g* itself. This prediction was corroborated in a meta-analysis (Woodley of Menie, te Nijenhuis, et al., 2016), where it was found that subtest *g* loadings only very weakly positively moderated the association between brain volume and subtest score.

Secular trends toward reduced genomic autozygosity (a genetic index of inbreeding) have also been noted in genetically informed data representatively sampled from the US (Nalls et al., 2009). This trend might be expected to counteract the effects of selection acting against *g*, as reduced inbreeding pressures are known to raise IQ via heterosis at the level of *g* (Nagoshi & Johnson, 1986). This has in fact been proposed as a major potential, genetic contributor to the Flynn effect and related trends (such as increases in stature and brain mass) (see Mingroni, 2004, 2007, 2014; cf. Woodley, 2011). However, it has been observed that the IQ increase expected on the basis of the data in Nalls et al. (2009), 0.07 points per decade, is substantially lower than the approximately half an IQ point decadal loss expected on the basis of selection, which means that any gains due to heterosis are likely to be substantially attenuated by losses due to selection (Woodley of Menie, Figueredo, et al., 2017).[12]

Finally, not all efforts to find evidence for declining *g* have proved successful. One notable failure in this regard is Pietschnig and Gittler's (2017) attempt to find indications of declining *g* in a meta-analysis of secular trends among ability-based measures of emotional intelligence. With the exception of perceiving emotions, where they noted a small negative time trend, they found that there were no indications of any secular trend on the other measures between the years 2001 and 2015. This should be contrasted with the finding of longer-term (i.e., over two centuries) declines in a social-cognitive indicator of altruism noted in Woodley of Menie, Figueredo, et al. (2017).

Another finding that may appear problematic is that there are some Western populations were the data indicate that IQ is declining even in cases where it is apparently not being selected against, such as Norway (Bratsberg & Rogeberg, 2018; cf. Hegelund et al., 2021; Sundet, 2014; Sundet, Borren, & Tambs, 2008). However, there are three points that must be considered in this regard, and that illustrate why this finding is not necessarily problematic for the co-occurrence model. First, the fertility data from Norway is primarily sourced from male conscripts. Since it is well established that the strength of the negative association between intelligence and fertility is significantly greater among females (Reeve et al., 2018), more data are required before it can be determined whether there might indeed be an overall positive fertility-intelligence association in Norway. Second (and this is the most important point), assuming that there is no negative directional selection for intelligence in Norway, Norwegian data cannot be used in order to evaluate the co-occurrence model, since its predictions are conditional (in modern populations) on the existence of a negative association between fertility and intelligence. The existence of a negative Flynn effect and no *g* decline is therefore perfectly compatible with the co-occurrence model, as long as intelligence is not being selected against. Third, the paper by Bratsberg and Rogeberg (2018) *does* illustrate a limitation of the co-occurrence model, in that there is at least one population (i.e., that of Norway) with respect to which the model is mute, in the sense that it does not offer any relevant prediction.[13] After all, the model says that there is likely to be observable declines in *g*, *if* the population in question is characterized by an overall negative relationship between fertility and

[10] For further discussion on these sorts of cultural factors, especially in relation to the very large decrease in IQ estimated on the basis of Piagetian staging noted among school children in the UK, see Flynn and Shayer (2018).

[11] Interestingly, he also found that American children during the same period experienced increasing IQ scores on the Wechsler Intelligence Scale for Children, but that the subtest magnitudes had negative correlations with the *g* loadings of the subtests, with values between −0.302 and − 0.409.

[12] Although as Woodley of Menie, Fernandes, et al. (2017), Woodley of Menie, Figueredo, et al. (2017) note, there is a sense in which these opposing processes might be expected to visibly co-occur at the molecular level, as selection against *g* primarily acts on the additive genetic variance associated with IQ, whereas changes in directional dominance caused by heterosis will occur on non-additive genetic variance. Therefore, we might expect to be able to track, via genotyping, these two, seemingly opposing secular genetic trends.

[13] The data from Norway may be suggestive of an alternative model. Indeed, it may be the case that there are grounds for developing a reversed co-occurrence model claiming that in Norway (and possibly other nations too) *g* is stagnating, or possibly even increasing (depending on the sign of its correlation with fertility) at the same time as phenotypic IQ is decreasing.

cognitive ability. However, it should be noted that Norway is likely an outlier, since cognitive ability apparently is being selected against in the majority of nations around the world (Reeve et al., 2018; Skirbekk, 2008). That said, the case of Norway, and possibly other nations too, shows that the co-occurrence model has somewhat limited generalizability, in that it does not offer relevant predictions for all populations. Moreover, another limitation is that the model does not predict anything about how long high cognitive ability can be selected against before phenotypic IQ gains stagnate or reverse.

As the findings listed above on balance suggests that there is a decrease in *g* or some cognitive ability associated with *g*, and as intelligence researcher Michael A. Woodley of Menie was the first to explicitly theorize this, these effects have been referred to by the term "Woodley effect". Sarraf (2017) was the first to provide a definition of the term:

> [Woodley effects are] defined as secular trends that plausibly result, in part or in whole, from population-level [declines in] genetic factors that underlie *g*
>
> (Sarraf, 2017, p. 239)

Given that there now is ample evidence for Woodley effects on many different prospective indicators of *g* (and that they covary among themselves across time to a great extent), it is reasonable to assert that the co-occurrence model on balance is supported by the preponderance of the available empirical evidence.[14] The empirical work conducted in order to test the model does indeed suggest that co-occurring with the Flynn effect there is also a Woodley effect, in the sense that various phenotypic indicators of *g* (and especially *g.h*) have been declining for more than a century, together with a host of other *g* loaded cognitive abilities and phenotypic proxies.

4. Solving Flynn's paradoxes

In addition to making novel predictions that the majority of the empirical tests conducted thus far have corroborated, the co-occurrence model can also be used to provide plausible solutions to several psychometric paradoxes (in addition to Cattell's paradox) that have plagued the discipline of intelligence research for quite some time. In his book *What is intelligence?* Flynn presents four paradoxes that the Flynn effect appears to engender. The paradoxes are as follows:

> *The factor analysis paradox:* how can intelligence be both one and many at the same time or how can IQ gains be so contemptuous of *g* loadings? How can people get more intelligent and have no larger vocabularies, no larger stores of general information, no greater ability to solve arithmetical problems?
> *The intelligence paradox:* if huge IQ gains are intelligence gains, why are we not struck by the extraordinary subtlety of our children's conversation? Why do we not have to make allowances for the limitations of our parents? A difference of some 18 points in Full Scale IQ over two generations ought to be highly visible.
> *The mental retardation paradox:* if we project IQ gains back to 1900, the average IQ scored against current norms was somewhere between 50 and 70. If IQ gains are in any sense real, we are driven to the absurd conclusion that a majority of our ancestors were mentally retarded […]
> *The identical twins paradox:* there is no doubt that twins separated at birth, and raised apart, have very similar IQs, presumably because of their identical genes. Indeed, a wide range of studies show that genes dominate individual differences in IQ and that environment is feeble. And yet, IQ gains are so great as to signal the existence of environmental factors of enormous potency. How can environment be both so feeble and so potent?
>
> (Flynn, 2007, pp. 9–10)

The co-occurrence model can be used to offer plausible solutions to all of these paradoxes. Consider, by addressing the paradoxes in reverse chronological order, the identical twins paradox first. How can IQ be both highly heritable and highly malleable at the same time? On the one hand, studies consistently show that IQ is one of the most highly heritable psychological traits, with narrow-sense heritability often exceeding 70% in adult samples—which means that >70% of the variance in IQ is explained by additive genetic effects (Bouchard, 2013; when *g* is estimated directly this value increases to >80%; Panizzon et al., 2014).[15] On the other hand, as evidenced by the Flynn effect, IQ scores are also highly malleable, as they have increased approximately 30 points on average during the 20th century (Flynn, 1984, 1987). But how—the paradox asks—can these two facts be reconciled? Does not one of them (high heritability/malleability) entail the negation of the other?

According to the co-occurrence model, the solution to the identical twins paradox involves recognition of the fact that the psychometric "atom" that is IQ actually can be split into different components and, furthermore, that the less heritable intelligence components also are more malleable in response to changes in environmental conditions. It is in fact precisely because individual differences in *s.e* are unconstrained by one's genes that environmental factors apparently can have such huge effects on IQ. However, from this fact, it does not follow that the heritability of IQ cannot be as high as >70%. The reason is simply that *s.e* can be extremely sensitive to changes in the environment without other, comparatively larger variance components (such as *g*) having to be so as well. Indeed, as evidenced by a subset of the studies demonstrating the Woodley effect, the secular increase in certain specialized abilities seems to have directly co-occurred with a secular decrease in *g* in the same sample of individuals (e.g., Wongupparaj et al., 2017).

Consider next the so-called mental retardation paradox. How can it be that by projecting the Flynn effect back to 1900, the average IQ at that time would be in the range of 50 to 70 points? The idea that the average person in the year 1900 had such a low IQ that they would be considered mentally disabled by modern standards is of course absurd. Moreover, it is also contradicted by the evidence indicating that the "smart faction"—i.e., the proportion of a population above a certain

[14] One reviewer expressed some concern over the fact that the evidence supporting the co-occurrence model has to do with certain patterns of covariation, and that it occasionally relies on the method of correlated vectors (MCV), which has been critiqued (the reviewer specifically cites Wicherts, 2017, in relation to this). However, it should be noted that almost all of the research on secular changes in psychological traits is evaluated with respect to evidence that is of a covariational nature. The author of this article does not see why this should be considered problematic. Moreover, Wicherts (2017) only presents problems for certain forms of item level MCV, not subtest or test level MCV. Aston and Lee (2005) do however note that a positively signed vector correlation using subtest data cannot be taken to conclusively demonstrate the claim that it is *g* that is moderating the relevant associations. Such a moderation pattern could result instead from the action of another latent variable, such as a lower-order general verbal or quantitative factor, depending on the composition of the battery. Only confirmatory methods can conclusively differentiate between latent variables. In so far as much of the relevant research that has been conducted on the co-occurrence model to date is of an exploratory, rather than confirmatory nature, this critique can be reframed in terms of there now being a need for confirmatory models to test the robustness of claims that it is *g* (rather than some other lower-order latent variable) that is the primary focal point of fertility differentials.

[15] It should be noted that the heritability of intelligence is much lower at a younger age, but that it increases well beyond 50% into adulthood—a phenomenon known as the Wilson effect (Bouchard, 2013).

threshold of intelligence[16]—in fact was larger around 1900 than it is today, as evidenced by Murray's (2003) work on the decreasing per capita number of eminent individuals during the 20th century, and Huebner's (2005) work on the decreasing per capita number of major innovations.

According to the co-occurrence model, the paradox in question is resolved by the proposition that the average person in 1900 had a higher, rather than lower level of *g* compared to the average person today. Moreover, since *g* is correlated to a much larger extent than *s.e* with other socially valued traits and real-world outcomes (Ackerman, Kanfer, & Calderwood, 2013; Jensen, 1998; Kell, Lubinski, Benbow, & Steiger, 2013; Zaboski, Kranzler, & Gage, 2018), it follows that the people living in 1900 cannot meaningfully be said to be intellectually disabled compared to contemporary populations. In fact, given what we now know about the prevalence and magnitude of the Woodley effect (based on both genetic and phenotypic indicators, losses in the range of half an IQ point per decade can be expected), a more puzzling question is why contemporary populations still are so scientifically and technologically advanced, despite the drop in *g*.

The intelligence paradox asks why the drastic secular increase in IQ scores is not noticeable when conversing or otherwise engaging with people of different generations in ordinary everyday contexts. Indeed, why is it that children today don't appear any more gifted than those of previous generations, or that the elderly don't appear especially slow or dull (beyond what one would expect given the effects of normal ageing on cognitive ability)?

The reason why children today don't appear any smarter, let alone gifted, compared to children of previous generations is simply that they aren't any smarter with respect to *g*, according to the co-occurrence model. Although they may have greater technological skill or a greater understanding of the society in which they live (due to better access to information), they do not have a greater level of *g*. Moreover, concerning the elderly, the reason they don't appear especially slow or dull is that they might be expected to have a higher level of *g*, after controlling for the effects of normal ageing of course. In brief, the reason why "a difference of some 18 points in Full Scale IQ over two generations" is not all that noticeable in everyday contexts[17] is that there has been a co-occurring decline in *g* that plausibly acts in the opposite direction when it comes to traits and trends that draw upon intellectual ability.

Lastly, the factor analysis paradox asks why the Flynn effect is not accompanied by improvements in cognition pertaining to one's vocabulary, general information storage, or quantitative reasoning ability. And the answer, which is also suggested in Flynn's formulation of the paradox above, is that the secular increase in IQ scores is not concentrated on *g*, but rather on *s.e* that have little or no impact upon other cognitive abilities. Indeed, the core claim of the co-occurrence model is that the different variance-components of IQ are free to vary independently of each other and, moreover, that co-occurring with the Flynn effect there has also been a Woodley effect in the sense that *g* has decreased at the same time as *s.e* has increased, as evidenced by e.g., decreasing use across texts of words with high item-difficulty.

In sum, the co-occurrence model not only offers a number of predictions that have been confirmed by subsequent empirical tests, but it also potentially solves all of Flynn's paradoxes, and Cattell's paradox too. Its ability to solve these paradoxes is a credit to its explanatory power.

5. Scientific progressiveness

By virtue of the accumulating merits of the co-occurrence model, the research program to which it belongs can be shown to be of a progressive nature. According to the philosopher of science Imre Lakatos, scientific theories are not (or at the very least should not be) evaluated in isolation. Rather, they are evaluated as parts of a larger research program to which they belong. A research program is constituted by a sequence of theories, and its quality is evaluated with respect to certain special properties that the sequence of theories may or may not have. If a research program has the relevant properties, then it is considered *progressive*. On the other hand, if it does not have the relevant properties, then it is considered *degenerating*.

According to Lakatos, there are two conditions that any good research program invariably satisfies. First, the research program is *theoretically progressive*, in the sense that "each new theory has some excess empirical content over its predecessor, that is, […] it predicts some novel hitherto unexpected fact." (Lakatos, 1978a, p. 33.) This means that later theories in the sequence that constitutes a research program have to provide novel predictions not contained in or derivable from earlier theories belonging to the same research program. Second, the research program is *empirically progressive*, in the sense that "some of this excess empirical content is also corroborated, that is, […] each new theory leads to the discovery of some *new fact*." (Lakatos, 1978a, p. 34.) This means that at least some of the novel predictions made on the basis of the theoretical developments within a research program have to be confirmed by the empirical evidence. Thus, a research program is progressive if it satisfies both of these conditions, and it is degenerating if it does not (Lakatos, 1978a, p. 34).

The PSRP is a paradigm case of a progressive research program. Consider for example the transition that occurred from Cattell and his inability to detect the effects of selection favoring individuals with lower intelligence, to Burt and Lynn's adoption of the attenuation model. The attenuation model first of all had a lot of new content not included in previous theory, some of which provided the basis for novel empirical predictions. One such prediction was that the effects of selection are undetectable at the phenotypic level, due to the masking or attenuating effects of the comparatively larger Flynn effect. Another prediction was that despite the increase in "phenotypic IQ", there should nevertheless be a decrease in "genotypic IQ". Moreover, this latter prediction has been confirmed several times over by empirical work showing both direct negative associations between polygenic scores predictive of cognitive ability (and proxies such as educational attainment) and fertility (Beauchamp, 2016a; Conley et al., 2016; Hugh-Jones & Abdellaoui, 2021; Kong et al., 2017; Woodley of Menie, Rindermann, Pallesen, & Sarraf, 2019; Woodley of Menie, Schwartz, & Beaver, 2016), and also a reduction in the frequency of alleles associated with cognitive ability over a span of multiple decades (Abdellaoui et al., 2019; Kong et al., 2017).

The same sort of progressive theory development within the PSRP also occurs with the transition from the attenuation model to the later co-occurrence model. The co-occurrence model is not only consistent with the genetic data evidencing a decline in "genotypic IQ". It also makes a whole host novel and risky predictions, since it follows from the theory that there is likely to be a Woodley effect on any trait or trend, in any population with a negative association between intelligence and fertility, that draws upon the same genetic factors that underlie *g*—at least if the trait or trend in question is not positively influenced by comparatively larger upward changes in *s.e* or other environmental factors. One development of note here is the finding of Woodley of Menie, Sarraf, Peñaherrera-Aguirre, Fernandes, and Becker (2018) in regard to testing the neurotoxin hypothesis, first proposed by Demeneix (2014, 2017), which proposes that endocrine disrupting chemical

[16] According to smart fraction theory, the size of this fraction of the population is more important for scientific and technological progress than the mean intelligence of the population (Rindermann, Sailer, & Thompson, 2009).

[17] Although non-*g* IQ gains are not noticeable, in the sense that people today don't seem especially smart compared to those of previous generations, it does not follow that the gains are "hollow" and without real-world consequences (Flynn, 2012; Flynn, 2013). Indeed, it may very well be that the Flynn effect has played a crucial role in enabling people to flourish in today's fast-paced and highly technological societies (Rindermann & Becker, 2018; Woodley, 2012a).

pollutants (rather than selection effects) are the primary cause of declining *g* in Western populations. However, Woodley of Menie, Sarraf, et al. (2018) found that cross-sectional declines in polygenic scores (which partially capture both selection and survival bias effects), sourced from the populations of Iceland (Kong et al., 2017) and the US (from the Health and Retirement Study), predict declining trends among the measures comprising the *g.h* factor, first introduced in Woodley of Menie, Figueredo, et al. (2017), whereas an "industrialization factor" comprised of changing environmental levels of four endocrine disrupting neurotoxins (lead, mercury, dioxins, and alcohol) doesn't. In this study the observed genotypic declines noted in previous work were shown to predict actual phenotypic declines among *g.h* components, thus linking the genetic to the phenotypic declines, evidencing the co-occurrence model.

We have already seen above in Section 3 that various predictions made by the co-occurrence model have been supported (albeit not wholly without controversy, e.g., Baker et al., 2015; Demeneix, 2017, p. 87; Dodonova & Dodonov, 2013; Gonthier et al., 2021) with respect to the negative Flynn effect (in two countries), simple visual and auditory reaction times, color acuity, utilization frequencies of high-difficulty vocabulary items sampled across texts, working memory capacity, three-dimensional rotation ability, and phenotypic indicators of declining developmental stability. On the basis of the Lakatosian standard, whereby the hallmark of scientific progress is that it "leads to the discovery of hitherto unknown novel facts" (Lakatos, 1978b, p. 5), it is indeed the case that the PSRP, with the co-occurrence model as its latest and most sophisticated theoretical development, has to be considered progressive.[18]

However, it should be noted that although the PSRP is progressive, it does not follow that its core theses are true, or that it will continue to be progressive in the future. Just because a research program has been fruitful in the past, that does not guarantee that it will be so in the future. This means that if the PSRP (including its constituent theories) is to retain its current epistemic status, then it must be exposed to empirical tests that put the predictions of its leading theory (i.e., the co-occurrence model) at risk of falsification—preferably by a greater number of researchers and/or research teams than has been the case thus far.

6. Extending the PSRP

In this section it will be argued that the PSRP can be extended to other fields of scientific research, by virtue of the fact that the co-occurrence model makes predictions with implications extending beyond the realm of cognitive ability testing. The co-occurrence model predicts that there is likely to be a Woodley effect on *any* trait or trend, in any population with a negative association between intelligence and fertility (or other biological change), that draws upon the same genetic factors that underlie *g*, if the trait or trend in question is not positively influenced by comparatively larger upward changes in *s.e* stemming from improvements in environmental factors. It should also be noted that some of the predictions to emerge from this research program have yet to be tested. For example, Woodley et al. (2014, p. 145) predict that the white matter density in older preserved brains should be higher than in more recent ones, consistent with reduced myelination and slowing reaction times. That being said, three novel predictions that are derivable from the co-occurrence model will now be presented:

6.1. The inefficient brains prediction

The brains of populations with fertility rates favoring individuals with lower intelligence are on average likely to be less efficient than the brains of previous generations, in the sense that the rate of cerebral glucose metabolism during cognitive activation will have increased and the variability of the rate of cerebral glucose metabolism will have decreased.

Let's first take a closer look at the inefficient brains prediction. The brain's primary source of energy is a simple sugar called glucose. The rate of glucose uptake and subsequent metabolism by different brain regions during various kinds of activity can provide information on how much energy the different brain regions expend. The procedure starts with the injection of a positron emitting glucose source (such as F-18 deoxyglucose), after which the person will engage in some cognitively demanding tasks (such as an IQ test) during the uptake period when the glucose is being metabolized by the brain. The uptake period usually lasts for approximately 30 min. After that, a brain scanning technique called positron emission tomography (PET) can be used to measure neural activity that occurred during the uptake period, by detecting gamma radiation emitted by positron-electron annihilation in the brain.

Studies have used this technique in order to investigate individual cognitive ability differences and their physiological basis. One important finding is that cerebral glucose metabolic rate (GMR) during cognitive activation has a significant and highly negative correlation with IQ test performance. In a study by Haier (1993), participants took the Raven's Advanced Progressive Matrices (RAPM)—a highly *g* loaded, nonverbal test of cognitive ability—during the uptake period. He found that individual scores on the RAPM had significant negative correlations with GMR between −0.7 and − 0.8, and that there were negative correlations with respect to all regions of the cerebral cortex. Thus, during cognitive activation, the brains of highly intelligent subjects expended less energy than their less intelligent counterparts, as indicated by their comparatively lower GMR (cf. Haier, Siegel, Tang, Abel, & Buchsbaum, 1992).

Another interesting finding is that GMR appears to vary as a function of how much "mental effort" an individual is capable of expending on a task. In an interesting study by Larson, Haier, LaCasse, and Hazen (1995), groups of average and high scoring subjects (with respective mean IQ scores of 104 and 123) were compared on easy and difficult cognitive tasks, where task difficulty was determined with respect to the subjective ability standards of each group: tasks that elicited 90% correct responses within a certain group were considered "easy" for that particular group, whereas tasks that elicited only 75% of correct responses within a certain group were considered "difficult" for that particular group. Now assuming that difficult tasks require more mental effort, the study found a significant interaction between group IQ level and mental effort. Although the two groups showed little difference in GMR on the easy tasks, there was a difference favoring the high IQ group on the difficult tasks. This indicates that high IQ individuals are able to expend more mental effort, as measured by GMR, when needed, and that such individuals therefore will have higher GMR variability when attempting to solve a battery of cognitive test items that are of sufficiently varied difficulty.

Given that GMR and GMR variability during cognitive activation are both closely related to *g*, and they are endophenotypes that can potentially be directly measured invariantly across cohorts, the co-occurrence model predicts that there is likely to be Woodley effects on these variables—at least insofar as they are not influenced by comparatively larger upward changes in *s.e* or other environmental factors. Since a core claim of the co-occurrence model is that there has been a secular decline in *g*, the model predicts that GMR during cognitive activation is likely to have *increased*, as GMR during cognitive activation is negatively correlated with *g*, and it predicts that the GMR variability during cognitive activation that results from engaging with a battery of cognitive test items that are of sufficiently varied difficulty is likely to have *decreased*, since GMR variability is positively correlated with *g*.

Moreover, much of the empirical work on GMR and its relation to individual differences in cognitive ability is interpreted as supporting the *brain efficiency hypothesis*, according to which intelligence is a

[18] This is not the first time Lakatos's conceptual framework is used in order to evaluate psychological research. See, e.g., Ketelaar and Ellis (2000), Urbach (1974).

function of how efficiently the brain works, rather than how hard it works (Haier et al., 1992; Parks et al., 1989). Now assuming that the brain efficiency hypothesis is correct, it becomes justified to say that what the co-occurrence model predicts is that our brains have become less efficient—hence the name *the inefficient brains prediction*—and that the reduction in efficiency, as measured by GMR, may provide at least a partial physiological explanation for the Woodley effect on *g*, and may also account for the potential anomaly presented to the co-occurrence model by indications of secular increases in brain mass. This is because larger brains, while being potentially better able to accommodate specialized abilities (possibly through greater allocation of grey matter), may not necessarily be more efficient.

6.2. Declining attention

Executive functioning measures associated with attention should be evidencing a decline due to the mounting evidence indicating a substantial role for this not just in human individual differences *g* research (Conway, Kovacs, Hao, Rosales, & Snijder, 2021), but in individual differences research on *g* in non-human primates (Woodley of Menie, Fernandes, Te Nijenhuis, Peñaherrera-Aguirre, & Figueredo, 2017).

Now let's turn to the declining attention prediction. Recent research has found indications that executive functioning measures, and in particular those associated with attention, might play a central role in the psychometric etiology of *g* at the individual differences level. Conway et al. (2021) argue that executive attentional processes function as a "central bottleneck" that throttles task performance in such a way that yields apparent correlations among the performance of domain specific cognitive processes via inhibiting their development. The dependence of *g* on these is therefore reflected not just in terms of the level of this trait (i.e., better executive attention will permit for more efficient problem solving across domains), but in terms of the strength of the manifold among abilities (this is reflected in the degree to which these distinct processes can be made to overlap with one another). Fairly strong support for at least some of the predictions of this hypothesis has been found in comparative psychological studies of *g* using individual differences data on a variety of different primate species (including human infants) evaluated using a common psychometric-behavioral cognitive inventory, the Primate Cognition Test Battery (PCTB). One study compared human infants and chimpanzees with one another using the PCTB in order to determine whether or not a common source of *g* variance moderated the magnitude of the species difference in performance between them (Woodley of Menie, Fernandes, et al., 2017). Using vector correlation analysis, it was found that across the 11 (available) subtests of the PCTB, *g* loadings (estimated with reference to the humans, chimps, and the average of the two) did not correlate with the magnitude of species performance differences, contrary to the prediction that *g* should be the primary source of differences between species. When subtests were removed from the comparison on the basis of having low coefficients of variation (CVs) (suggesting that they impose floor or ceiling effects on performance by virtue of measuring more modularized abilities), it was found that the vector correlation among the pool of the remaining subtests increased. This process of subtraction was performed iteratively, reducing the numbers of subtests from 11 to 3 (for a total of nine iterations). The increase in the strength of the vector correlation between *g* and species *d* was extremely strongly correlated with the average CV of abilities across each iteration ($r > 0.9$), indicating that once more measures of more modularized cognitive processes are subtracted, *g* becomes the primary source of individual differences in performance between species. This finding was subsequently replicated using three lemur species with a total of three analyses, one for each pairwise comparison (Woodley of Menie & Peñaherrera-Aguirre, 2021). In both cases, it was found that PCTB subtests tapping executive attention were consistently among the most *g* loaded and most discriminative in terms of species performance differences in all species that have been compared thus far.

As was noted previously, it has been found that in comparing performance on the CVLT3 battery with the CVLT-II measures of attention/working memory and learning, there have been declines across measurement occasions, which is consistent with the prediction of secular attention decline. More broadly, significant secular increases in attention deficit/hyperactivity disorder diagnoses would seem to be broadly compatible with this prediction (Xu, Strathearn, Liu, Yang, & Bao, 2018). Consistent with the expectation that these may constitute a Woodley effect is the finding of robust inverse associations between attention deficit hyperactivity disorder and IQ (Frazier, Demaree, & Youngstrom, 2004), coupled with evidence that the correlation between the two stems from common (pleiotropic) genetic factors (Kuntsi et al., 2004). What remains to be established is whether the relationship is positively moderated by *g*.

A finding that might at first glance appear to contradict this prediction is a meta-analysis of performance on the "marshmallow task" (Protzko, 2020), a measure of delay discounting that likely draws on some aspects of executive functioning, which found that it was improving (as evidenced by lowering time preference) over time. However, it should be noted that this task is not very cognitively demanding, and these secular trends may reflect changes in aspects of personality, possibly related to slowing life history and the Flynn effect (this possibility is discussed by Protzko, 2020), independently of *g*. It is expected that a comprehensive meta-analysis of executive attentional performance should yield robust indications of a Woodley effect by contrast.

6.3. Solving the height paradox

Height, like IQ, is highly heritable (Mingroni, 2007), yet also exhibits substantial secular gains (Hatton & Bray, 2010). Therefore, there exists what has been termed a "height paradox" (Mingroni, 2007, p. 811) analogous to the one identified by Flynn in the case of IQ. It is predicted that, like IQ, height is decomposable into different phenotypic variance components with different heritabilities and environmentalities, and that, like IQ, it will be the more environmentally sensitive ones that show signs of having undergone positive secular change.

Finally, let's turn to the height heritability paradox. Studies have found that most of the secular trend in height can be accounted for specifically by lengthening of the bones of the leg (Cole, 2003). It is also notable that studies in which negative associations between identified variants predictive of cognitive ability and fertility (indicating negative directional selection) do not identify similar associations between height variants and fertility (e.g., Conley et al., 2016), which suggests that "genotypic height" might be under stabilizing selection. In terms of the co-occurrence model, these findings might be taken to indicate that variation with respect to bone length in the legs is associated with low heritability, and correspondingly high environmentality, whereas other "variance components" of the phenotype of height (such as neck and torso length) might be much more heritable by contrast (and assuming no substantially confounding gene-by-environment interactions) might not exhibit secular trends (assuming no significant associations between relevant variants and fertility outcomes).

7. Conclusion

This article has presented and discussed the co-occurrence model as the latest and most sophisticated theoretical development within the PSRP. More specifically, it has argued (1) that the model provides a sound solution to Cattell's paradox; (2) that the model is quite well supported based on the empirical work conducted in order to test its predictions; (3) that the model is able to provide plausible solutions to Flynn's paradoxes; (4) that the PSRP in which it figures clearly is progressive, rather than degenerative; and (5) that the PSRP can be extended to other domains of research, by virtue of the fact that the co-occurrence model makes predictions with implications extending beyond the realm of cognitive ability testing.

The article started by asking whether we humans are becoming more or less intelligent. It concludes that the question can be answered either way, depending on how one chooses to operationalize intelligence. If one thinks that IQ is the best operationalization of the intelligence construct, then we are becoming more intelligent as evidenced by the Flynn effect. If, however, *g* provides the best operationalization of intelligence, then the question should plausibly be answered in the negative, at least with respect to populations in which selection is favoring lower intelligence, due to all the accumulating evidence for a Woodley effect on *g* in said populations. As it is believed that *g* provides the best scientific representation of intelligence—since *g* is responsible for most of the predictive and explanatory power of IQ (Ackerman et al., 2013; Kell et al., 2013; Ree & Carretta, 2022; Zaboski et al., 2018)—it is concluded that populations in which lower cognitive ability has been selected for likely have become less intelligent since the onset of the shift in fertility patterns favoring lower intelligence.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of competing interest

The author declares that there are no conflicts of interest.

References

Abdellaoui, A., Hugh-Jones, D., Yengo, L., Kemper, K. E., Nivard, M. G., Veul, L., … Visscher, P. M. (2019). Genetic correlates of social stratification in Great Britain. *Nature Human Behavior, 3*, 1332–1342.

Ackerman, P. L., Kanfer, R., & Calderwood, C. (2013). High school Advanced Placement and student performance in college: STEM majors, non-STEM majors, and gender differences. *Teachers College Record, 114*, 1–43.

Aston, M. C., & Lee, K. (2005). Problems with the method of correlated vectors. *Intelligence, 33*, 431–444.

Baker, D. P., Eslinger, P., Benavides, M., Peters, E., Dieckmann, & León, J. (2015). The cognitive impact of the educational revolution: A possible cause of the Flynn Effect on population IQ. *Intelligence, 49*, 144–158.

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.

Beauchamp, J. P. (2016a). Genetic evidence for natural selection in humans in the contemporary United States. *Proceedings of the National Academy of Sciences, USA, 113*, 7774–7779.

Beauchamp, J. P. (2016b). Reply to Woodley of Menie: Natural selection, educational attainment, and cognitive variance components. *Proceedings of the National Academy of Sciences USA, 113*, E5782.

Bloom, N., Jones, C. I., Van Reenen, J., & Webb, M. (2020). Are ideas getting harder to find? *American Economic Review, 110*, 1104–1144.

Blum, D., & Holling, H. (2017). Spearman's law of diminishing returns. A meta-analysis. *Intelligence, 65*, 60–66.

Boeing, P., & Hunermund, P. (2020). A global decline in research productivity? Evidence from China and Germany. *Economics Letters, 197*, Article 109646.

Bouchard, T. J. (2013). The Wilson effect: The increase in heritability of IQ with age. *Twin Research and Human Genetics, 16*, 923–930.

Bradford, E. J. G. (1925). Can present scholastic standards be maintained? *Forum of Education, 3*, 186–194.

Bratsberg, B., & Rogeberg, O. (2018). Flynn effect and its reversal are both environmentally caused. *Proceedings of the National Academy of Sciences, USA, 115*, 6674–6678.

Burks, B. S., & Jones, H. E. (1935). A study of differentia fertility in two Californian cities. *Human Biology, 7*, 539–554.

Burt, C. (1948). *Intelligence and fertility: The effect of the differential birthrate on inborn mental characteristics*. Eugenics Society.

Cattell, R. B. (1936). Is national intelligence declining? *The Eugenics Review, 28*, 181–203.

Cattell, R. B. (1937). *The fight for our National Intelligence*. King & Son.

Cattell, R. B. (1950). The fate of national intelligence: Test of a thirteen-year prediction. *The Eugenics Review, 42*, 136–148.

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

Cole, T. J. (2003). The secular trend in human physical growth: A biological view. *Economics and Human Biology, 1*, 161–168.

Conley, D., Laidley, T., Belsky, D. W., Fletcher, J. M., Boardman, J. D., & Domingue, B. W. (2016). Assortative mating and differential fertility by phenotype and genotype across the 20th century. *Proceedings of the National Academy of Sciences, USA, 113*, 6647–6652.

Conway, A. R. A., Kovacs, K., Hao, H., Rosales, K. P., & Snijder, J. P. (2021). Individual differences in attention and intelligence: A united cognitive/psychometric approach. *Journal of Intelligence, 9*, 3.

Courtiol, A., Tropf, F. C., & Mills, M. C. (2016). When genes and environment disagree: Making sense of trends in recent human evolution. *Proceedings of the National Academy of Sciences, USA, 113*, 7693–7695.

Deary, I. J., Strand, S., Smith, P., & Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence, 35*, 13–21.

Demeneix, B. (2014). *Losing our minds: How environmental pollution impairs human intelligence and mental health*. Oxford University Press.

Demeneix, B. (2017). *Toxic cocktail: How chemical pollution is poisoning our brains*. Oxford University Press.

Demetriou, A., & Spanoudis, G. (2017). Mind and intelligence: Integrating developmental, psychometric, and cognitive theories of human mind. In M. Rosen (Ed.), *Challenges in educational measurement— Contents and methods* (pp. 39–60). New York: Springer.

Dodonova, Y. A., & Dodonov, Y. S. (2013). Is there any evidence of historical slowing of reaction time? No, unless we compare apples and oranges. *Intelligence, 41*, 674–687.

Flynn, J. R. (1984). The mean IQ of Americans: Massive gains 1932 to 1978. *Psychological Bulletin, 95*, 29–51.

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?* Cambridge University Press.

Flynn, J. R. (2012). *Intelligence and human progress: The story of what was hidden in our genes*. Elsevier.

Flynn, J. R. (2013). The "Flynn effect" and Flynn's paradox. *Intelligence, 41*, 851–857.

Flynn, J. R., & Shayer, M. (2018). IQ decline and Piaget: Does the rot start at the top? *Intelligence, 66*, 112–121.

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

Frazier, T. W., Demaree, H. A., & Youngstrom, E. A. (2004). Meta-analysis of intellectual and neuropsychological test performance in attention-deficit/hyperactivity disorder. *Neuropsychology, 18*, 543–555.

Furnham, A. (2008). *Psychology ideas you really need to know* (Vol. 50). Quercus editions Ltd.

Galton, F. (1869). *Hereditary genius*. MacMillan.

Gignac, G. E. (2015). The magical numbers 7 and 4 are resistant to the Flynn effect: No evidence for increases in forward or backward recall across 85 years of data. *Intelligence, 48*, 85–95.

Gonthier, C., Grégoire, J., & Besançon, M. (2021). No negative Flynn effect in France: Why variations of intelligence should not be assessed using tests based on cultural knowledge. *Intelligence, 84*, Article 101512.

Graves, L. V., Drozdick, L., Courville, T., Farrer, T. J., Gilbert, P. E., & Delis, D. C. (2021). Cohort differences on the CVLT-II and CVLT3: Evidence of a negative Flynn effect on the attention/working memory and learning trials. *The Clinical Neuropsychologist, 35*, 615–632.

Haier, R. J. (1993). Cerebral glucose metabolism and intelligence. In P. A. Vernon (Ed.), *Biological approaches to the study of human intelligence* (pp. 317–373). Ablex.

Haier, R. J., Siegel, B. V., Tang, C. Q., Abel, L., & Buchsbaum, M. S. (1992). Intelligence and changes in regional cerebral glucose metabolic rate following learning. *Intelligence, 16*, 415–426.

Hatton, T. J., & Bray, B. E. (2010). Long run trends in the heights of European men, 19th-20th centuries. *Economics and Human Biology, 8*, 405–413.

Hegelund, E. R., Teasdale, T. W., Okholm, G. T., Osler, M., Sørensen, T. I. A., Christensen, K., & Mortensen, E. L. (2021). The secular trend of intelligence test scores: The Danish experience for young men born between 1940 and 2000. *PLoS One, 16*, Article e0261117.

Herrnstein, R. J., & Murray, C. (1994). *The bell curve: Intelligence and class structure in American life*. Free Press.

Higgins, J. V., Reed, E. W., & Reed, S. C. (1962). Intelligence and family size: A paradox resolved. *Eugenics Quarterly, 9*, 84–90.

Hills, T. T., & Adelman, J. S. (2015). Recent evolution of learnability in American English from 1800 to 2000. *Cognition, 143*, 87–92.

Huebner, J. (2005). A possible declining trend for worldwide innovation. *Technological Forecasting and Social Change, 72*, 980–986.

Hugh-Jones, D. A., & Abdellaoui, A. (2021). Human capital mediates natural selection in contemporary humans. In *University of East Anglia School of Economics Working Paper Series 2021–22*.

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

Johnson, G. R. (1937). *High school survey* (vol. 35, pp. 2–34). Public School Messenger.

Kan, K.-J., Wicherts, J. M., Dolan, C. V., & van der Maas, H. L. J. (2013). On the nature and nurture of intelligence and specific cognitive abilities: The more heritable, the more culture dependent. *Psychological Science, 24*, 2420–2428.

Kell, H. J., Lubinski, D., Benbow, C. P., & Steiger, J. H. (2013). Creativity and technical innovation: Spatial ability's unique role. *Psychological Science, 24*, 1831–1836.

Ketelaar, T., & Ellis, B. J. (2000). Are evolutionary explainations unfalsifiable? Evolutionary psychology and the Lakatosian philosophy of science. *Psychological Inquiry, 11*, 1–21.

Kondrashov, A. S. (2017). *Crumbling genome: The impact of deleterious mutations on humans*. Wiley Blackwell.

Kong, A., Frigge, M. L., Thorleifsson, G., Stefansson, H., Young, A. I., Zink, F., … Stefansson, K. (2017). Selection against variants in the genome associated with

educational attainment. *Proceedings of the National Academy of Sciences, USA, 114*, E727–E732.
Kuntsi, J., Eley, T. C., Taylor, A., Hughes, C., Asherson, P., Caspi, A., & Moffitt, T. E. (2004). Co-occurrence of ADHD and low IQ has genetic origins. *American Journal of Medical Genetics, Part B: Neuropsychiatric Genetics, 124b*, 41–47.
Lakatos, I. (1978a). Falsification and the methodology of scientific research programmes. In J. W. G. Currie (Ed.), *Vol. 1. The methodology of scientific research Programmes. Philosophical papers* (pp. 8–101). Cambridge University Press.
Lakatos, I. (1978b). Science and pseudoscience. In G. C. J. Worrall (Ed.), *Vol. 1. The methodology of scientific research Programmes. Philosophical papers* (pp. 1–7). Cambridge University Press.
Larson, G. E., Haier, R. J., LaCasse, L., & Hazen, K. (1995). Evaluation of a "mental effort" hypothesis for correlations between cortical metabolism and intelligence. *Intelligence, 21*, 267–278.
Lentz, T., Jr. (1927). Relation of I.Q. to size of family. *Journal of Educational Psychology, 18*, 486–496.
Loehlin, J. C. (1997). Dysgenesis and IQ: What evidence is relevant? *American Psychologist, 39*, 1236–1239.
Lynn, R. (1982). IQ in Japan and the United States shows a growing disparity. *Nature, 297*, 222–223.
Lynn, R. (1990). The role of nutrition in secular increases in intelligence. *Personality & Individual Differences, 11*, 273–285.
Lynn, R. (1996). Dysgenics: Genetic deterioration in modern populations. *Praeger*.
Lynn, R. (2009). What has caused the Flynn effect? Secular increases in the Development Quotients of infants. *Intelligence, 37*, 16–24.
Lynn, R. (2013). Who discovered the Flynn effect? A review of early studies of the secular increase of intelligence. *Intelligence, 41*, 765–769.
Madison, G., Woodley of Menie, M. A., & Sänger, J. (2016). Secular slowing of auditory simple reaction time in Sweden (1959–1985). *Frontiers in Human Neuroscience, 10*, 407.
Meisenberg, G. (2010). The reproduction of intelligence. *Intelligence, 38*, 220–230.
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.
Mingroni, M. A. (2014). Future efforts in Flynn effect research: Balancing reductionism with holism. *Journal of Intelligence, 2*, 122–155.
Murray, C. (2003). *Human accomplishment: The pursuit of excellence in the arts and sciences, 800 BC to 1950*. Harper Collins.
Must, O., & Must, A. (2018). Speed and the Flynn effect. *Intelligence, 68*, 37–47.
Must, O., Must, A., & Raudik, V. (2003). The Flynn Effect for gains in literacy found in Estonia is not a Jensen Effect. *Personality & 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 & Individual Differences, 7*, 201–207.
Nalls, M. A., Sanchez-Simon, J., Gibbs, J. R., Paisan-Ruiz, C., Bras, J. T., Tanaka, T., … Singleton, A. B. (2009). Measures of autozygosity in decline: Globalization, urbanization, and its implications for medical genetics. *PLoS Genetics, 5*, Article e1000415.
Nave, G., Jung, W. H., Karlsson Linnér, R., Kable, J. W., & Koellinger, P. D. (2019). Are bigger brains smarter? Evidence from a large-scale preregistered study. *Psychological Science, 30*, 43–54.
Nettelbeck, T. (2014). Smarter but slower? A comment on Woodley, te Nijenhuis & Murphy (2013). *Intelligence, 42*, 1–4.
te Nijenhuis, J., & van der Flier, H. (2013). Is the Flynn effect on g?: A meta-analysis. *Intelligence, 41*, 802–807.
Panizzon, M. S., Vuoksimaa, E., Spoon, K. M., Jacobson, K. C., Lyons, M. J., Franz, C. E., … Kremen, W. S. (2014). Genetic and environmental influences on general cognitive ability: Is g a valid latent construct? *Intelligence, 43*, 65–76.
Parker, S. (2014). Were the Victorians cleverer than us? Maybe, maybe not. *Intelligence, 47*, 1–2.
Parks, R. W., Crockett, D. J., Tuokko, H., Beattie, B. L., Ashford, J. W., Coburn, K. L., … McGeer, E. G. (1989). Neuropsychological "systems efficiency" and positron emission tomography. *The Journal of Neuropsychiatry & Clinical Neurosciences, 1*, 269–282.
Peach, H., Lyerly, J. E., & Reeve, C. L. (2014). Replication of the Jensen effect on dysgenic fertility: An analysis using a large sample of American youth. *Personality and Individual Differences, 71*, 56–59.
Pietschnig, J., Deimann, P., Hirschmann, N., & Kastner-Koller, U. (2021). The Flynn effect in Germanophone preschoolers (1996–2018): Small effects, erratic directions, and questionable interpretations. *Intelligence, 86*, Article 101544.
Pietschnig, J., & Gittler, G. (2015). A reversal of the Flynn effect for spatial perception in German-speaking countries: Evidence from a cross-temporal IRT-based meta-analysis (1977–2014). *Intelligence, 53*, 145–153.
Pietschnig, J., & Gittler, G. (2017). Is ability-based emotional intelligence impervious to the Flynn effect? A cross-temporal meta-analysis (2001-2015). *Intelligence, 61*, 37–45.
Pietschnig, J., Penke, L., Wicherts, J. M., Zeiler, M., & Voracek, M. (2015). Meta-analysis of associations between human brain volume and intelligence differences: How strong are they and what do they mean? *Neuroscience & Biobehavioral Reviews, 57*, 411–432.
Pietschnig, J., & Voracek, M. (2015). One century of global IQ gains: A formal meta-analysis of the Flynn effect (1909-2013). *Perspectives on Psychological Science, 10*, 282–306.
Pietschnig, J., Voracek, M., & Gittler, G. (2019). Is the Flynn effect related to migration? Meta-analytic evidence for correlates of stagnation and reversal of generational IQ test score changes. *Politische Psychologie, 2*, S267–S283.
Protzko, J. (2020). Kids These Days! Increasing delay of gratification ability over the past 50 years in children. *Intelligence, 80*, Article 101451.
Ree, J. M., & Carretta, T. R. (2022). Thirty years of research on general and specific abilities: Still not much more than *g*. *Intelligence, 91*, Article 101617.
Reeve, C. L., Heeney, M. D., & Woodley of Menie, M. A. (2018). A systematic review of the state of literature relating parental general cognitive ability and number of offspring. *Personality & Individual Differences, 134*, 107–118.
Rindermann, H., & Becker, D. (2018). FLynn-effect and economic growth: Do national increases in intelligence lead to increases in GDP? *Intelligence, 69*, 87–93.
Rindermann, H., Sailer, M., & Thompson, J. (2009). The impact of smart fractions, cognitive ability of politicians and average competence of peoples on social development. *Talent Development and Excellence, 1*, 3–25.
Roesell, F. P. (1937). Comparative mental ability of high school pupils in three Minnesota towns in 1920 and 1934. *Minnesota Studies in Articulation*, 122–128.
Roivainen, E. (2014). Changes in word usage frequency may hamper intergenerational comparisons of vocabulary skills: An Ngram analysis of Wordsum, WAIS, and WISC test items. *Journal of Psychoeducational Assessment, 32*, 83–87.
Runquist, E. A. (1936). Intelligence test scores and school marks in 1928 and 1933. *School and Society, 43*, 301–304.
Rushton, J. P. (1999). Secular gains in IQ not related to the *g* factor and inbreeding depression – unlike Black-White differences: A reply to Flynn. *Personality & Individual Differences, 26*, 381–389.
Sarraf, M. A. (2017). Review of historical variability in heritable general intelligence: Its evolutionary origins and socio-cultural consequences. *Personality & Individual Differences, 109*, 238–241.
Sauce, B., & Matzel, L. D. (2018). The paradox of intelligence: Heritability and malleability coexist in hidden gene-environment interplay. *Psychological Bulletin, 144*, 26–47.
Schaie, K. W., & Strother, C. R. (1968). A cross-sequential study of age changes in cognitive behavior. *Psychological Bulletin, 70*, 671–680.
Schmidt, F. L., & Hunter, J. E. (1999). Theory testing and measurement error. *Intelligence, 27*, 183–198.
Sesardic, N. (2005). *Making sense of heritability*. Cambridge University Press.
Silverman, I. W. (2010). Simple reaction time: It is not what it used to be. *The American Journal of Psychology, 123*, 39–50.
Simonton, D. K. (2013). Scientific genius is extinct. *Nature, 493*, 602.
Skirbekk, V. (2008). Fertility trends by social status. *Demographic Research, 18*, 145–180.
Smith, S. (1942). Language and non-verbal test performance of racial groups in Honolulu before and after a fourteen year interval. *The Journal of General Psychology, 26*, 51–93.
Spearman, C. (1904). "General intelligence": objectively determined and measured. *The American Journal of Psychology, 15*, 201–293.
Srinivasan, S., Bettella, F., Frei, O., Hill, W. D., Wang, Y., & Andreassen, O. A. (2018). Enrichment of genetic markers of recent human evolution in educational and cognitive traits. *Scientific Reports, 8*, 12585.
Sundet, J. M. (2014). The Flynn Effect in families: Studies of register data on Norwegian military conscripts and their families. *Journal of Intelligence, 2*, 106–118.
Sundet, J. M., Borren, I., & Tambs, K. (2008). The Flynn effect is partly caused by changing fertility patterns. *Intelligence, 36*, 183–191.
Trahan, L. H., Stuebing, K. K., Fletcher, J. M., & Hiscock, M. (2014). The Flynn effect: A meta-analysis. *Psychological Bulletin, 140*, 1332–1360.
Tuddenham, R. D. (1948). Soldier intelligence in World Wars I and II. *American Psychologist, 3*, 54–56.
Uchiyama, R., Spicer, R., & Muthukrishna, M. (2021). Cultural evolution of genetic heritability. *The Behavioral and Brain Sciences*, 1–147.
Urbach, P. F. (1974). Progress and degeneration in the "IQ Debate" (I)*. *The British Journal for the Philosophy of Science, 25*, 99–135.
Verhaeghen, P. (2014). *The elements of cognitive aging: Meta-analyses of age-related differences in processing speed and their consequences*. Oxford University Press.
Voracek, M. (2006). Phlogiston, fluid intelligence, and the Lynn–Flynn effect. *The Behavioral and Brain Sciences, 29*, 142–143.
Wheeler, L. R. (1942). A comprehensive study of the intelligence of East Tennessee mountain children. *Journal of Educational Psychology, 33*, 321–334.
Wicherts, J. M. (2017). Psychometric problems with the method of correlated vectors applied to item scores (including some nonsensical results). *Intelligence, 60*, 26–38.
Wicherts, J. M., Dolan, C. V., Hessen, D. J., Oosterveld, P., van Baal, G. C. M., Boomsma, D. I., & Span, M. M. (2004). Are intelligence tests measurement invariant over time? Investigating the nature of the Flynn effect. *Intelligence, 32*, 509–537.
Wongupparaj, P., Wongupparaj, R., Kumari, V., & Morris, R. G. (2017). The Flynn effect for verbal and visuospatial short-term and working memory: A cross-temporal meta-analysis. *Intelligence, 64*, 71–80.
Woodley, M. A. (2011). Heterosis doesn't cause the Flynn effect: A critical examination of Mingroni (2007). *Psychological Review, 118*, 689–693.
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–Fynn effect. *Personality & 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*. The University of Buckingham Press.
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 time. *Intelligence, 41*, 843–850.

Woodley, M. A., te Nijenhuis, J., & Murphy, R. (2014). Is there a dysgenic secular trend towards slowing simple reaction time? Responding to a quartet of critical commentaries. *Intelligence, 46*, 131–147.

Woodley, M. A., Younuskunju, S., Balan, B., & Piffer, D. (2017). Holocene selection for variants associated with general cognitive ability: Comparing ancient and odern genomes. *Twin Research and Human Genetics, 20*, 271–280.

Woodley of Menie, M., & Dunkel, C. S. (2015). In France, are secular IQ losses biologically caused? A comment on Dutton and Lynn (2015). *Intelligence, 53*, 81–85.

Woodley of Menie, M., Fernandes, H., Te Nijenhuis, J., Peñaherrera-Aguirre, M., & Figueredo, A. (2017). General intelligence is a source of individual differences between species: Solving an anomaly. *The Behavioral and Brain Sciences, 40*, E223.

Woodley of Menie, M., Fernandes, H. B. F., Figueredo, A. J., & Meisenberg, G. (2015). By their words ye shall know them: Evidence of genetic selection against general intelligence and concurrent environmental enrichment in vocabulary usage since the mid 19th century. *Frontiers in Psychology, 6*.

Woodley of Menie, M., te Nijenhuis, J., & Murphy, R. (2015a). Are signal luminances and confounded stimuli a source of slowing simple reaction time and cross-study heterogeneity? A response to Parker (2014). *Intelligence, 49*, 23–24.

Woodley of Menie, M., te Nijenhuis, J., & Murphy, R. (2015b). The Victorians were still faster than us. Commentary: Factors influencing the latency of simple reaction time [general commentary]. *Frontiers in Human Neuroscience, 9*.

Woodley of Menie, M. A., & Fernandes, H. B. F. (2015). Do opposing secular trends on backwards and forwards digit span evidence the co-occurrence model? A comment on Gignac (2015). *Intelligence, 50*, 125–130.

Woodley of Menie, M. A., & Fernandes, H. B. F. (2016a). The secular decline in general intelligence from decreasing developmental stability: Theoretical and empirical considerations. *Personality and Individual Differences, 92*, 194–199.

Woodley of Menie, M. A., & Fernandes, H. B. F. (2016b). Showing their true colours: Possible secular declines and a Jensen effect on colour acuity—More evidence for the weaker variant of Spearman's other hypothesis. *Personality & Individual Differences, 88*, 280–284.

Woodley of Menie, M. A., Fernandes, H. B. F., Kanazawa, S., & Dutton, E. (2018). Sinistrality is associated with (slightly) lower general intelligence: A data synthesis and consideration of the secular trend in handedness. *Homo: Journal of comparative. Human Biology, 69*, 118–126.

Woodley of Menie, M. A., Figueredo, A. J., Fernandes, H. B. F., Peñaherrera-Aguirre, M., Sarraf, M. A., & Hertler, S. (2017). The rhythm of the west: A biohistory of the modern era AD 1600 to the present. *Journal of Social, Political and Economic Studies Monograph No., 37*.

Woodley of Menie, M. A., & Peñaherrera-Aguirre, M. (2021). General intelligence as a major source of cognitive variation among individuals of three species of lemur, uniting *g* with *G*. *Evolutionary Psychological Science*. https://doi.org/10.1007/s40806-021-00304-x

Woodley of Menie, M. A., Peñaherrera-Aguirre, M., Fernandes, H. B. F., Becker, D., & Flynn, J. R. (2016). It's getting bigger all the time: Estimating the Flynn effect from secular brain mass increases in Britain and Germany. *Learning and Individual Differences, 45*, 95–100.

Woodley of Menie, M. A., Peñaherrera-Aguirre, M., Fernandes, H. B. F., & Figueredo, A. J. (2018). What causes the anti-Flynn effect? A data synthesis and analysis of predictors. *Evolutionary Behavioral Sciences, 12*, 276–295.

Woodley of Menie, M. A., Rindermann, H., Pallesen, J., & Sarraf, M. A. (2019). How intelligence affects fertility 30 years on: Retherford and Sewell revisited - with polygenic scores and numbers of grandchildren. *Twin Research and Human Genetics, 22*, 147–153.

Woodley of Menie, M. A., Sarraf, M. A., Peñaherrera-Aguirre, M., Fernandes, H. B. F., & Becker, D. (2018). What caused over a century of decline in general intelligence? Testing predictions from the genetic selection and neurotoxin hypotheses. *Evolutionary Psychological Science, 4*, 272–284.

Woodley of Menie, M. A., Schwartz, J. A., & Beaver, K. M. (2016). How cognitive genetic factors influence fertility outcomes: A mediational SEM analysis. *Twin Research and Human Genetics, 19*, 628–637.

Woodley of Menie, M. A., te Nijenhuis, J., Fernandes, H. B. F., & Metzen, D. (2016). Small to medium magnitude Jensen effects on brain volume: A meta-analytic test of the processing volume theory of general intelligence. *Learning and Individual Differences, 51*, 215–219.

Woods, D. L., Wyma, J. M., Yund, E. W., Herron, T. J., & Reed, B. (2015). Factors influencing the latency of simple reaction time. *Frontiers in Human Neuroscience, 9*, 131.

Xu, G., Strathearn, L., Liu, B., Yang, B., & Bao, W. (2018). Twenty-year trends in diagnosed attention-deficit/hyperactivity disorder among US children and adolescents, 1997-2016. *JAMA Network Open, 1*, Article e181471.

Zaboski, B. A., Kranzler, J. H., & Gage, N. A. (2018). Meta-analysis of the relationship between academic achievement and broad abilities of the Cattell-Horn-Carroll theory. *Journal of School Psychology, 71*, 42–56.