Intelligence xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Intelligence

# The negative Flynn Effect: A systematic literature review

Edward Dutton [a,*], Dimitri van der Linden [b], Richard Lynn [c]

[a] Ulster Institute for Social Research, London, UK
[b] Erasmus University Rotterdam, Department of Psychology, Education, and Child Studies, The Netherlands
[c] University of Ulster, London, UK

ARTICLE INFO

*Article history:*
Received 18 January 2016
Received in revised form 9 September 2016
Accepted 13 October 2016
Available online xxxx

*Keywords:*
Negative Flynn Effect
IQ decline
Dysgenics
Immigration

ABSTRACT

The Flynn Effect (rising performance on intelligence tests in the general population over time) is now an established phenomenon in many developed and less developed countries. Recently, evidence has begun to amass that the Flynn Effect has gone into reverse; the so-called 'Negative Flynn Effect.' In this study, we present a systematic literature review, conducted in order to discover in precisely how many countries this reverse phenomenon has been uncovered. Using strict criteria regarding quality of the sample and the study, we found nine studies reporting negative Flynn Effects in seven countries. We also discuss several possible explanations for the negative Flynn Effect as an attempt to understand its most probable causes.

© 2016 Elsevier Inc. All rights reserved.

## 1. Introduction

Numerous studies have shown that scores on intelligence tests (i.e., IQ tests) have substantially increased during much of the twentieth century. This increase in IQ scores was first reported in the United States by Runquist (1936) and was later confirmed by Smith (1942) and Tuddenham (1948). Cattell (1951) reported a similar trend in England after it became apparent that the increase in IQ scores was a largely global phenomenon found in many different countries (summarized in Lynn, 2013).

This phenomenon has been designated *the Flynn Effect* after the review by Flynn (1984, 1987, 2012) who first documented it on the Stanford-Binet and Wechsler tests. A Flynn Effect is distinct from simple sample fluctuations in IQ scores in annually assessed cohorts. More specifically, the effect is understood to be established if there is an overall increase in IQ score, rather than merely an increase on one IQ domain (e.g., verbal or spatial ability), and if the increase is continuous over many years or over a relatively long period of time. In this regard, Flynn (1984) found there had been a 13.8 point increase in IQ scores between the years 1932 and 1978, amounting to a rise of 0.3 points per year or 3 points per decade. More recently, the Flynn Effect was further evidenced by calculations of IQ score gains between 1972 and 2006 on versions of the Stanford-Binet (SB), Wechsler Adult Intelligence Scale (WAIS), and Wechsler Intelligence Scale for Children (WISC) (Flynn, 2009). The Flynn Effect (Flynn, 2012) is not only occurring in developed countries, but it is increasingly being found in developing countries as well, such as Kenya, Turkey, Dominica, Saudi Arabia, China and Sudan (see Flynn, 2012). The effect has been found in Latin America, for example Brazil (Colom, Flores-Mendoza, & Abad, 2007) and Argentina (Flynn & Rossi-Case, 2012) and in the former Communist Bloc (e.g. Estonia; Must, Must, & Raudik, 2003). The average global increase of 3 IQ points per decade has also been confirmed in a recent meta-analysis by Pietschnig and Voracek (2015).

Although one criterion of the Flynn Effect is that it applies to the full IQ score, it should be noted, that it nevertheless is driven more strongly by rises on certain kinds of ability. Relative to scores produced by verbal tests, there have been greater gains in scores produced by nonverbal, performance-based measures like Raven's Progressive Matrices and Wechsler performance subtests (Flynn, 2012). These types of tests are associated with fluid intelligence, suggesting less of a rise in crystalized intelligence (which reflects the influence of education, such as vocabulary). A notable exception is the increasing scores produced by the Wechsler verbal subtest Similarities, although this subtest relates to elements of reasoning not required by the other subtests comprising the Wechsler Verbal IQ composite.

The Flynn Effect means that an individual in the most recent cohort will attain a higher IQ score when set against the norms of an earlier cohort than he will when set against his own. Accordingly, it means that the test will overestimate a subject's IQ by an increasing amount the longer the time-span is between the norming of the test and its administration. This has clear practical implications for any organization that seeks to use an IQ test for any purpose. Indeed, Flynn (2012) has highlighted the problem of US states having rules whereby those who are mentally retarded cannot be executed if found guilty of murder. If the test from which their IQ score was gleaned, such as a test at school, was not appropriately normed then it will not provide them with the correct IQ score and, accordingly, an intellectually impaired person

* Corresponding author.
*E-mail address:* ecdutton@hotmail.com (E. Dutton).

http://dx.doi.org/10.1016/j.intell.2016.10.002
0160-2896/© 2016 Elsevier Inc. All rights reserved.

Please cite this article as: Dutton, E., et al., The negative Flynn Effect: A systematic literature review, *Intelligence* (2016), http://dx.doi.org/10.1016/j.intell.2016.10.002

may be executed. This example, provided by Flynn, however, also implies that the increase in IQ scores may not be a genuine increase in intelligence. This notion is in line with more recent studies that have suggested that increases in IQ scores may not necessarily reflect an increase in the latent intelligence factor, *g* (te Nijenhuis, van Vianen, & van der Flier, 2007; te Nijenhuis & Van der Flier, 2013). All in all, the Flynn Effect is an important phenomenon in the study of intelligence and beyond.

Recently, however, a number of prominent studies have reported a reversal in the Flynn Effect, beginning in the latter decades of the twentieth century. A number of these studies (e.g. Sundet, Barlaug, & Torjussen, 2004) draw upon annually accrued conscript military data, which involves the vast majority of the nation's males in any given annual cohort. Accordingly, they allow us to conclude that there is a continuous decline in IQ scores over time and that this is a real phenomenon and not simply a blip.

More recently, a negative Flynn Effect has also been reported in a number of other countries, such as France (Dutton & Lynn, 2015). In addition, a series of studies have noted declines in proxies for intelligence, such as in the lengthening of reaction times between 1900 and 2000 (Woodley, te Nijenhuis, & Murphy, 2014), a decline in colour discrimination ability (Woodley, te Nijenhuis, & Murphy, 2015) and a rise in fluctuating facial asymmetry (Woodley & Fernandes, 2016), something which has been shown to be weakly negatively correlated with intelligence. These sets of findings by Woodley and colleagues seem to indicate that approximately a century ago, a decline in the latent factor of genuine intelligence began, which until now has largely been cloaked by an increase in scores on IQ tests. Yet, with the increase in IQ scores slowing down (see Pietschnig & Voracek, 2015) or even going into reverse, the decline in *g* may start to become apparent on IQ tests. Woodley et al. (2015) term this the 'co-occurrence model' which states that the Flynn Effect occurs on the more environmentally influenced and less g-loaded aspects of intelligence, while dysgenics concomitantly occurs on the more g-loaded and more heritable dimensions.

If there truly is a trend toward declining intelligence as indicated by a negative Flynn Effect then this would be a phenomenon with potentially serious implications. This is because there is evidence that national IQ is associated with economic prosperity (Lynn & Vanhanen, 2012), political stability (Lynn & Vanhanen, 2012) and scientific achievement (Rindermann & Thompson, 2011). Therefore in the present study, we wish to establish in precisely which countries there has been a reported negative Flynn Effect. We aim to do this by conducting an exhaustive and systematic literature search, as we will outline below. We then wish to establish its most likely causes and to address any possible anomalies. In order to establish this, we conducted a systematic literature review the details of which we will now present.

## 2. Method

### 2.1. Literature search

In conducting this review, we followed the PRISMA guidelines (http://www.prisma-statement.org/). We decided to use the following eligibility criteria for including studies in the review:

(1) The study related to a population sample, or a sample likely to be representative of the country's population.
(2) A negative Flynn Effect was noted over a considerable period of time, which was set at 5 years. It is appreciated that this border involves some degree of arbitrariness. This amount of time was selected, however, because, from the studies of which we were already aware showing a negative Flynn Effect, there had never been a rise in scores after 5 years of decline.
(3) The effect had to be on an IQ test, rather than on some proxy for this such as a scholastic test. This is very important, such that we were comparing like with like. The Flynn Effect was observed on IQ tests and this restriction had to be adhered to.
(4) It had to be an overall decline in IQ score on the test, rather than simply a decline restricted to one of its domains. Again, this is an essential dimension to the Flynn Effect itself and must therefore be employed in any analysis of the reversal of this effect.
(5) It had to be original research. Articles which cited other articles were ruled out.

Based on these criteria we searched for the literature using the following Prisma procedure. We first conducted the literature search using Google Scholar. Google Scholar was chosen because of its expansive nature. Unlike many academic search engines, it is not restricted to peer-reviewed journals, let alone high-impact journals. Thus, the use of this database would be more likely to allow us to uncover undergraduate and postgraduate theses, book content, and working papers. We searched for papers up until 2016.

The literature search strategy involved selecting the appropriate key words. We identified these from the titles and abstracts of already published literature on the Flynn Effect and through brainstorming synonymous or roughly synonymous terms.

The next step was the 'Data Management' of the Prisma procedure. In line with the Data management stage, all titles were copied into a Word Document, numbered, and marked with whether they were to be accepted or rejected and, if so, on what grounds.

Finally, we conducted the Prisma Selection Process: Based on the general criteria outlined above and on further analyses of the content of the articles (e.g. title, abstract, text), the three authors decided collectively, based on the above-described criteria, which papers were to be included and there had to be complete agreement for this to be the case.

In line with Prisma guidelines, we now outline our Literature Search Process: The term *'Anti-Flynn Effect'* yielded 139 articles or theses with this term in the text. Of these, seven articles were original articles that had found an overall negative Flynn Effect. However, this was reduced to 6 because Ronnlund, Carlstedt, Blomstedt, Nilsson, and Weinehall's (2013) negative Flynn Effect on Swedish conscript data between 1992 and 1993 was only over a year.

The term *'Negative Flynn Effect'* yielded 37 hits and 3 articles within our inclusion criteria, including one which the previous term had not uncovered. We rejected Pietschnig and Gittler's (2015) German data because the test it drew upon, namely the 3DC, is mainly a test of spatial intelligence.

*'Reverse Flynn Effect'* yielded 13 hits but none within our inclusion criteria.

The term *'Reversal of the Flynn Effect'* uncovered 129 hits of which 8 unique hits were within our inclusion criteria. Among those excluded was a study in Portugal (Schelini, Almeida, Duarte, & Primi, 2011) which found a negative Flynn Effect over five years among recruits to the Portuguese gendarmes. However, not only was this not a representative sample, but the two samples employed had significantly different sex compositions and sizes. This also uncovered Bandeira, Costa, and Arteche (2012) who found, overall, no generational gain in IQ in a study of children aged between 6 and 12 on both the Ravens and Draw-a-Man. Their data came from Porto Alegre in Brazil, which they argued was an elite area. But, again, this was not within our inclusion criteria.

The term *'Decline in IQ'* yielded 1890 hits of which 2 were within our inclusion criteria. Using alternative words such as 'decline of IQ' or 'decline of the IQ' did not lead to unique and/or relevant hits above and beyond the first alternative.

The term *'Fall in intelligence'* unearthed 58 hits, of which none were relevant. Among those excluded were Cancian and Klein (2015) who found that the IQ score of US marine officer recruits had fallen by 2.6 points between 1980 and 2014. Clearly, this is not representative of the US population. 'Fall of intelligence,' (77 hits) 'fall of the intelligence' (26 hits) and 'fall in the intelligence' (11 hits) did not reveal additional

Please cite this article as: Dutton, E., et al., The negative Flynn Effect: A systematic literature review, *Intelligence* (2016), http://dx.doi.org/10.1016/j.intell.2016.10.002

ARTICLE IN PRESS



unique hits. The term *'End of the Flynn Effect'* unearthed 252 hits of which 7 were relevant and one (Sundet et al., 2004) a new relevant hit.

Reading the relevant articles revealed a further potential article in the form of Cotton et al. (2005). However, this found zero gains on the Coloured Progressive Matrices in Australia when comparing 1975 and 2003. Although this may imply that there has been a negative Flynn Effect, this does not constitute one. A small negative Flynn effect was found among the 8, 10 and 11 year olds but not among the 7, 9, and 10 year olds. We also found, using this method, Koivunen (2007), which reported a negative Flynn Effect in Finland up to 2001.

Parallel to the search in Google Scholar, we conducted the same search using Scopus. Due to the more limited nature of Scopus' sample, one highly relevant finding, in the form of a thesis, was absent. The only new potentially germane hit was Flynn (2012) *Are We Getting Smarter?* This revealed that on the CPM, among British 14.5–15.5 year olds between 1979 and 2008 IQ had decreased by 1.88 points. But when comparing 7.5–15.5 year olds over the same period it had increased by 4.26 points.

We also consulted the most recent meta-analysis of the Flynn Effect in general (Pietschnig & Voracek, 2015). This meta-analysis found most, but not all, of the relevant hits we had found. This meta-analysis, in itself, revealed a large number of Flynn Effect reversals or cessations, but they were either at the domain level only or on unrepresentative samples. For example, the Brazilian study was based on a very specific sample from a wealthy area of Brazil. The excluded studies were Brazil (2), Canada (1), Germany (2), Netherlands (2), South Africa (1), Spain (1), Sudan (1), UK (1), and USA (22). Although these studies did not reach the criteria for being included in the present review, in order to be complete, in the supplementary material, we have added information about those studies that showed declines in subgroups or decline in specific cognitive abilities (see Appendix 1). Having found these results, which we will report below, we also planned to obtain some preliminary information about the possible causes of the Negative Flynn Effect. The literature reveals a number of possible causes that have been proposed. These are:

(1) Immigration. It has been consistently found that non-Western ethnic minorities in Western countries, such as Sub-Saharan Africans and South Asians, on average, score lower on IQ tests than Europeans (Lynn, 2006). We tested the possible influence of immigration on negative Flynn effects by correlating the mean per capita immigration level between 1981 and 2015 with each country's magnitude of negative Flynn Effect. We took these data from Migration Policy (2016), which presents mean per capita immigration for each country at 5 year intervals from 1950 to 2015.
(2) Sex. Several studies indicated that there are sex differences in intelligence and that, in adult samples, males have a slight advantage of somewhere between 3 and 5 IQ-points ($\approx$0.3 *SD*) over females (Lynn & Irwing, 2004). In as far as this is indeed the case, a difference in sex ratio may be accompanied with a change in average IQ score.
(3) Parental age. Older age of fatherhood has been associated with reduced offspring intelligence, possibly caused by decreasing sperm quality (Saha et al., 2009). It has also been shown that increasing maternal age is negatively associated with IQ (e.g. Myrskylä, Silventoinen, Tynelius, & Rasmussen, 2013) and that maternal age is increasing at the same rate as paternal age in Western countries (e.g. Wehner, Kambskard, & Abrahamson, 2002). Data on age at first motherhood by country was available from the CIA World Factbook for 2013 (CIA World Factbook, 2016).
(4) Dysgenics. Lynn (2011) has argued that the negative association between intelligence and fertility in Western countries, especially among women, should eventually be reflected in declining IQ scores, for reasons we will explore more in our Discussion section. Lynn (2011) has calculated a dysgenic fertility ratio for most of the countries in which we have observed a negative Flynn Effect. This was based on the difference in fertility between more and less educated men and women. Unfortunately, these data are rather old. The Norwegian data was from the period 1925–1930 while the Dutch data was from 1920 to 1925. With all the other countries, it was from 1966 to 1967. Nevertheless, despite these limitations regarding the age of the data, they potentially imply that dysgenics may have begun earlier or proceeded more speedily in some countries than others and that this may have led to later differences in the extent and speed of the negative Flynn Effect. More specifically, the dysgenic fertility that may have caused a possible decline in genetic intelligence may not have been directly accompanied by a decline in IQ scores, as the latter would have also been prone to more environmental factors that have caused an increase in IQ. As such, it is useful to test what association they have with more recent negative Flynn Effects.

## 3. Results

Based on the literature search detailed in the Method section, and within our inclusion criteria, we found reports of a negative Flynn Effect in seven countries. The results can be seen in Table 1. The nine articles included, draw upon the following tests:

(1) Sundet et al. (2004) used the General Ability Test, an IQ test developed by the Norwegian army in 1954. It is composed of Words, Numbers and Shapes and conscripts are given a GA (General Ability) score, which corresponds to an IQ score.
(2) Woodley and Meisenberg's (2013) meta-analysis of tests of Dutch adults used the GATB = General Aptitude Test Battery. This measures 9 different 'aptitudes' among which are verbal aptitude, numerical aptitude and spatial aptitude.
(3) Teasdale and Owen's (2008) study drew upon the Borge Prien's Prove, which is an IQ test used by the Danish army on recruits since 1961. It is comprised of logical, verbal, numerical and spatial reasoning tests.
(4) Shayer and Ginsburg's (2007, 2009) studies drew upon the Piagetian test: An IQ test developed for children. Piaget's theory focuses on interviewing the subjects to discover why they answered in a particular way.
(5) Dutton and Lynn's (2013) study drew upon annual average results of the Finnish Peruskoe, which literally translates as 'Basic test.' This is an IQ test developed by the Finnish army composed of Numbers, Words and Shapes tests. These results were reported in Koivunen (2007) up to 2001, a thesis which was sent to them by a Finnish army researcher, as well as in correspondence with the same Finnish army researcher for 2008–9.
(6) Korgesaar's (2013) Estonian study drew upon the Raven Standard Progressive Matrices (SPM) test, which is a widely-accepted test of general intelligence and, as such, the study is within our inclusion criteria.
(7) (7) Dutton and Lynn (2015) drew upon the French WAIS (Wechsler Adult Intelligence Test) IV manual.

From Table 1 it can be seen that in the majority of the studies the decline ranges between 0.38 and 4.30 IQ points per decade. The Estonian study seems to be somewhat of an outlier with a decline of 8.4 IQ points per decade. Taking the un-weighted average of all the studies, the mean decline per decade in the studies would be 3.18 points. When excluding the rather high value of Estonia, the average decline in the remaining seven studies becomes 2.44 IQ points per decade.

Please cite this article as: Dutton, E., et al., The negative Flynn Effect: A systematic literature review, *Intelligence* (2016), http://dx.doi.org/10.1016/j.intell.2016.10.002

**Table 1**
Negative Flynn Effect per country.

| Country | Age | Test | Years | Type | IQ (decline per decade) | Reference |
|---|---|---|---|---|---|---|
| Norway | 18–19 | General ability | 1996–2002 | All conscripts in every year | 0.38 | Sundet al., 2004 |
| Denmark | 18–19 | Borge Priene's Prove | 1998–2003/4 | All conscripts in every year | 2.70 | Teasdale & Owen, 2008 |
| Britain | 11–12 | Piagetian | 1975/2003 | 10,023 over 5 cohorts: 1975, 2000, 2001, 2002, 2003 (each cohort roughly equal in size) | 4.30 | Shayer & Ginsburg, 2007 |
| Britain | 13–14 | Piagetian | 1976/2006 | 2006: N 446, 2007: N 357 (total: 793) | 2.50 | Shayer & Ginsburg, 2009 |
| Netherlands | Adults | GATB | 1975/2005 | Meta-analysis | 1.35 | Woodley & Meisenberg, 2013 |
| Finland | 18–19 | Peruskoe | 1998–2009 | All conscripts 1998–2001 and 2008/9 | 2.0 | Dutton & Lynn, 2013; Koivunen, 2007 |
| France | Adults | WAIS III & IV | 1999/2008–9 | Two representative groups of 79 | 3.8 | Dutton & Lynn, 2015 |
| Estonia | 18–19 | Raven SPM | 2001/2005/2012 | Representative student sample: 2001: 573, 2005: 417, 2012: 338 | 8.4 | Korgesaar, 2013 |

### 3.1. Possible explanations

As part of our analysis, we decided to discuss and, where possible, to test a number of possible causes of the negative Flynn Effect. These were as follows:

#### 3.1.1. Sex ratio

One of the possible explanations for a change in IQ scores is the sample sex ratio. This notion is based on the literature suggesting that there may be sex differences in average IQ. For example, Lynn and Irwing (2004) have suggested that in adults, women score, on average 3 to 5 IQ points lower than men. If this is indeed the case then a change in the sex ratio of samples may accompany a change in IQ. However, in the studies in the present review, sex ratio is unlikely to have caused the effects found. First, the Danish and Finnish samples consist of males only. In addition, the samples that do have mixed sex are very similar over time regarding sex ratios. For example, the Estonian 2001 sample is 43% male while the 2012 sample is 42% male. For the following samples: - Shayer and Ginsburg (2007, 2009), Woodley and Meisenberg (2013), and Dutton and Lynn (2015) - the sex ratio could not be reliably retrieved. However, given the information described above, it can be concluded that sex ratio is unlikely to be a determining factor in the change in IQ scores that we report in this review.

#### 3.1.2. Immigration

A number of studies have found that immigrants from developing countries have average IQs that are significantly lower than the European average (e.g. Lynn, 2006). Therefore, ideally, we would be able to correlate the immigration figures per capita and per year for each country where the negative Flynn Effect has been found with the extent of IQ decline. However, this would be very complicated because immigrants from different countries will have different IQs, will go on to attain differential socioeconomic statuses, and potentially be reacted to differently by hosts due to a variety of factors, including religion. Focussing on the Nordic data, Sundet (correspondence quoted in Dutton, 2014, p.243) has noted that: "Men from (South) Asian and African countries have around 5–6 IQ points lower than non-immigrants. Yet, they seem to comprise no more than around 2–3% of the conscripts in this period. If there would be any effect then this would deflate the total mean IQ by around a maximum of 0.1–0.2 IQ points." As such, it simply cannot fully explain the decline. Also, conscript data from Finland is particularly important in assessing the immigration hypothesis. Finland did not experience any significant third world immigration until around 1992 (see Dutton & Lynn, 2013). However, the conscripts in 1997 would have been mainly born in 1978 when the non-white population of Finland was vanishingly small.

Despite the limitations outlined above and the small $N$ (=7) for country level, we decided to calculate the correlation between the IQ decline per decade and average immigration between 1950 and 2015 (Migration Policy, 2016). When including all countries in the review, the correlation was virtually zero ($r(7) = 0.033$, $p = 0.94$). However, it has to be noted that Estonia did not only appear to be an outlier regarding the IQ decline, but also regarding immigration, because it was the only country that showed negative immigration numbers. When conducting the calculations again, but this time excluding Estonia, the correlation became $r(6) = 0.802$ and reached marginal significance ($p = 0.055$). Nevertheless, we have already noted that, based on the percentage of immigrants in the most reliable samples, immigration is unlikely to have a large influence. Accordingly, this association may be underpinned by a factor, which underlies both dysgenics and high immigration levels, such as degree of societal development or putative amount of time since industrialization. That is to say, highly developed countries may, because they are highly developed, have high levels of immigration.

#### 3.1.3. Dysgenics

A third possibility that has been suggested is a so-called process of dysgenics on intelligence (e.g. Dutton & Lynn, 2015) which has been argued to be occurring due to the negative association between intelligence and fertility in modern populations (Lynn, 2011). Higher fertility of less intelligent parents had already been noted anecdotally in Galton (1869). Based on this, one would expect that IQ would already have shown a relatively large decline since the beginning of the 20th century, especially at its heritability has been estimated at around 0.8 (Lynn, 2011). This, however, clearly did not happen as the Flynn Effect showed that IQ in fact went up for almost a century.

One possible explanation for this is that a possible dysgenic effect may be obscured by a strong increase in environmental quality that has boosted phenotypic IQ. Specifically, there is some agreement that the Flynn Effect may not fully reflect a genuine rise in intelligence, because it occurs on the less *g*-loaded parts of the tests (Lynn, 2011; te Nijenhuis et al., 2007; te Nijenhuis & Van der Flier, 2013). Rather, it may reflect a rise in pattern spotting ability (which is correlated with intelligence), something precipitated, possibly, by living in more educated society, as argued by Flynn (2012). He argues that modern society makes us increasingly look at the world through 'scientific spectacles' and, accordingly, examine it analytically, boosting performance IQ, especially on similarities. As this is ultimately underpinned by intelligence, it would have a genotypic limit and if genotypic intelligence were declining then the imperfect nature of the IQ test as a measure of intelligence would mask this, but only up to the genotypic limit. Once this limit was reached, any genotypic decline in IQ would become visible on the IQ tests. As mentioned above, this is known as the Co-occurrence Model. It is possible that this is what has happened because dysgenic fertility - a negative association between intelligence and numbers of children - has been observed in Denmark, Sweden, Finland, and a number of other countries reviewed in Lynn (2011).

Indeed, if Flynn's 'scientific spectacles' explanation is accurate then we would expect to see, prior to an overall negative Flynn Effect, a

Please cite this article as: Dutton, E., et al., The negative Flynn Effect: A systematic literature review, *Intelligence* (2016), http://dx.doi.org/10.1016/j.intell.2016.10.002

negative effect on verbal and mathematical IQ concomitant with a positive effect on other parts of the test. This is, indeed, what we see in the studies we excluded. Khaleefa, Sulman, and Lynn (2009) found that Sudanese Full-scale IQ increased 2.05 points per decade between 1987 and 2007, but Verbal IQ decreased by 1.65 points over the period. Colom, Andres-Pueyo, and Juan-Espinosa (1998) reported a decline in Spanish verbal reasoning (male and female −0.3) and mathematical reasoning (male −2.4; female −2.1) between 1979 and 1995 but a rise on abstract reasoning (and also Ravens) sufficient to create an overall Flynn Effect.

Besides such differential effects on subtests, we would also expect to see a slowing down of the Flynn Effect before it ultimately ceased, because the Flynn Effect itself would be partly g-loaded (with g in decline) and there would be a limit to the extent to which the environment can raise IQ scores. The meta-analysis of the Flynn Effect by Pietschnig and Voracek (2015) does indeed show that IQ gains since the 1980s had considerably slowed down. The gains were also increasingly non-linear in this period.

*3.1.4. Maternal age*

Finally, we tested maternal age, which could also be regarded as a proxy for age at which people become parents more generally. Unfortunately, we could not find data, by country, on average age of first fatherhood by year, only averages for different years, rendering the datasets incomparable. We did, however, find the average maternal age per country ($N = 7$) to negatively correlate to the level of IQ decline ($r = -0.88$). Despite the small sample size, the correlation reached significance ($p < 0.01$). Nevertheless, as the correlation was negative this seems to indicate that among the countries showing a negative Flynn Effect, the ones with the highest average maternal age show smaller declines in IQ. As such, it seems unlikely that maternal age is a factor in IQ decline.

## 4. Discussion

In the present study we conducted an exhaustive literature search in order to find studies reporting a negative Flynn Effect. Based on a set of strict criteria which we detailed in the Method section, we found a series of articles reporting a decline in general IQ scores over a period of at least five years. Notably, the most useful studies were those from Norway, Denmark and Finland because they involve the vast majority of males in each annual cohort and provided clear evidence of a year-on-year negative Flynn Effect.

Building on the results of this systematic literature search, we discussed four possible causes of the negative Flynn Effect. First, we took into account sex ratio. We concluded that this sampling issue is very unlikely to play a role because the negative Flynn Effect can be found in samples of almost the entire male cohorts in certain countries and where we have a mixed-sex sample the sex-balance of the cohorts is roughly the same.

The second possibility we considered was immigration. We argued that in some samples (e.g., Finland, Dutton & Lynn, 2013; Koivunen, 2007) immigration could not entirely have caused the decline in IQ scores because the level of immigrants in those cases was too small to have any significant effect on average IQ.

The third option we considered was dysgenics. Unfortunately, we could not extract dysgenic data for all the countries in which we found a negative Flynn Effect. Therefore, we could not reliably analyse the relationship with IQ decline. Nevertheless, the notion that dysgenics may play a role has been suggested in several other studies. For example it would be in line with Woodley et al. (2014) who have presented evidence that reaction times have declined across the twentieth century in a number of Western countries and they show that reaction times robustly correlate with intelligence. Woodley et al.'s (2014) studies on reaction times have been heavily criticised (e.g. Woods, Wyma, Yund, et al., 2015). However, it can be countered that Woodley and others have found corroborative evidence in other proxies for IQ, such as colour discrimination (Woodley & Fernandes, 2016), meaning that the simplest explanation is that these studies are tapping into the same phenomenon. Moreover, Woodley et al. (2015) have responded to the most recent criticisms with further evidence for their case. It might also be countered that their evidence would predict IQ losses throughout the twentieth century, rather than beginning at the end of it. But, as already discussed, the argument is that these would be being covered-up by massive rises on a particular intelligence sub-ability that was partly related to *g*. Eventually, the ceiling of this ability would be reached and the losses would start to reveal themselves on IQ tests. In addition, Woodley and Fernandes (2016) have shown that the Flynn Effect is not primarily occurring on general and heritable intelligence factor *g*, whereas the negative Flynn Effect does seem to occur on *g*. The negative Flynn Effect displays a Jensen Effect, and is mainly occurring on the more heritable abilities (Woodley of Menie & Dunkel, 2015). We would suggest that these different lines of evidence favour Flynn's argument that the Flynn Effect is an environmental phenomenon. The more educated and science-focused modern world forces us to adopt scientific spectacles causing us to reach our phenotypic limit on certain intelligence abilities, even though *g* is declining due to dysgenic fertility. This shows up as a secular rise in IQ scores, cloaking the decline in g, but this has reached its limit, so the decline in *g* is now showing up on the IQ tests as a negative Flynn Effect.

Fourth, and finally, we discussed maternal age. We tested the possibility that this may play a role in intelligence decline, but the direct correlation between IQ decline and the country's average maternal age was negative (−0.88). Thus, at this stage, maternal age is not persuasive as an argument for the IQ decline.

In the present study, we must emphasize that even though we identified several studies showing a decline in IQ, there is currently still a much larger pool of studies showing an increase in IQ. For example, Flynn (2012) has reported a positive Flynn Effect in the USA among adults between 1995 and 2001 of 3.06 points per decade, when comparing the WISC III and the WISC IV. Among US children, between 1989 and 2001, he reports a rise of 3.36 points per decade based on the same tests. Moreover, in their recent meta-analysis, Pietschnig and Voracek (2015) confirmed the 3 points per decade average increase in IQ. Therefore, it is imperative to discuss our negative Flynn Effects in the light of the positive Flynn Effects findings in the literature.

A first possible explanation is that the studies we report are simply outliers in a much larger set of longitudinal studies on IQ. However, given that some of the studies we reported drew upon very large and population-representative samples, this explanation seems rather unlikely.

A second possibility is that a phenotypic increase in IQ may have largely overshadowed a possible genotypic decline in IQ and that the latter is only more recently starting to show up in several representative datasets. If we are correct that the positive Flynn Effect is primarily an environmental effect then we should not be especially surprised that it operates differently in different parts of the world. For example, if the positive Flynn Effect would be partly the result of an increase in the 'scientific worldview' of the population, then the Flynn Effect may vary in line with the level that a country has reached in adopting such a view. Specifically, factors that may influence the extent to which the population adopts this 'scientific-spectacles' view, such as education, wealth, and, possibly, religion, differ between countries.

Regarding the trade-offs between forces that drive Flynn Effects and forces that drive negative Flynn Effects, it may not be a coincidence that the countries in which we found negative Flynn Effects are some of the wealthiest countries with the best distribution of wealth indices. It is possible that because of these optimizing factors, those countries are among the first to reach the boundaries of the environmental effects on IQ. Consequently, they may also be among the first to indicate a decline in IQ scores. Future research may want to elaborate on these possibilities and focus more strongly on disentangling the various explanations for Flynn versus negative Flynn Effects.

Please cite this article as: Dutton, E., et al., The negative Flynn Effect: A systematic literature review, *Intelligence* (2016), http://dx.doi.org/10.1016/j.intell.2016.10.002

ARTICLE IN PRESS



## 5. Limitations

There are a number of limitations to this analysis. Ideally, in order to establish and understand the causes of the negative Flynn Effect, we need large samples, annual cohorts over a longer period of time, and, where possible, the ability to rule out potential confounding factors such as sex and immigration. We only have this for Finland and, to a lesser extent, Denmark and Norway. Here, we have, from the military conscript data, samples which are almost the entire male population of a certain age, year-on-year. The impact of immigration is likely to be very small, as we have seen, and this is especially so in the case of Finland. Ideally, we would need more samples of this quality and also female samples of this quality as it is possible that the negative Flynn Effect is working at a different rate among females.

A second limitation can be found in our ability to test hypotheses as to the possible causes of the negative Flynn Effect. As discussed, some of these data have had to be drawn upon because they are the best that we have. Also, because a negative Flynn Effect has only been found in a small number of countries, we are left with a limitation of power, leaving it difficult to have confidence in any correlational finding in this study. In some cases, such as the immigration correlation, it raises the question of whether the correlation reflects immigration causing a country to have a negative Flynn Effect or whether countries that are more developed, and so have a negative Flynn Effect, are more likely to have high levels of immigration.

## 6. Concluding statements

Although, the Flynn Effect is a well-known phenomenon that has been replicated in many samples and in various countries, several theories predict a negative Flynn Effect (e.g., Lynn, 2011; Woodley, te Nijenhuis, & Murphy, 2013). In the current study, we identified the high quality samples reported in the literature that have reported such a negative Flynn Effect. Even though identifying these studies may be a crucial step in this area, future research should further scrutinize the evidence in light of other studies that continue to report an increase in IQ scores. For example, it is unclear why some Western countries, such as the USA, continue to display a positive Flynn Effect.

With regard to some of the theories we discussed, it seems to us that the simplest explanation for the negative Flynn Effect has already been presented by scholars such as Woodley et al. (2013). The Industrial Revolution had two key and related effects on IQ. Firstly, it set off massive gains on IQ scores by establishing an environment which compelled us to think in more scientific way, compelled us to become more educated, and saturated us with knowledge, information, and novel problems. This process would have pushed certain latent intelligence abilities to their limit and, if this occurred at a sufficient pace, then the imperfect nature of the IQ test means that it would overwhelm whatever else was happening and show up as an IQ rise. However, the Industrial Revolution concomitantly set off a quite different process. It led to the development of reliable contraception and low child mortality, meaning people only needed to have very few children and the more intelligent would be more able to achieve this. Accordingly, the previous positive correlation between intelligence and fertility became negative (Lynn, 2011). Moreover, this positive correlation was a so-called Jensen Effect: it was occurring on general intelligence and on the more heritable dimensions of intelligence. The Flynn Effect would disguise this on IQ tests but once the ceiling of the Flynn Effect was reached, we would see it slowing down and then going into reverse as a negative Flynn Effect. There would, perhaps, be two dimensions to this. One would be purely genetic and the other would be environmental: as the society becomes less intelligent the degree to which it innovates an increasingly cognitively stimulating environment would, set off the by the Industrial Revolution, slow down, meaning it would outpace the decline g to a lesser and lesser extent until it stopped outpacing it altogether, manifesting itself as a negative Flynn Effect.

Supplementary data to this article can be found online at doi:10.1016/j.intell.2016.10.002.

## Acknowledgements

We would like to thank Jakob Pietschnig, Jim Flynn, and an anonymous reviewer for their useful comments and suggestions on previous versions of this article.

## References

Bandeira, D. R., Costa, A., & Arteche, A. (2012). The Flynn Effect in Brazil: Examining generational changes in the draw-a-person and in the Raven's colored progressive matrices. *Revista Latinoamericana de Psicologia*, *44*, 9–18.
Cancian, M., & Klein, M. (2015). *Military officer quality in the all-volunteer force.* National Bureau of Economic Research (Working Paper 21372).
Cattell, R. B. (1951). The fate of national intelligence: Test of a thirteen year prediction. *Eugenics Review*, *17*, 136–148.
CIA World Factbook (2016). https://www.cia.gov/library/publications/download/download-2013/index.html
Colom, R., Andres-Pueyo, A., & Juan-Espinosa, M. (1998). Generational IQ gains: Spanish data. *Personality and Individual Differences*, *25*, 927–935.
Colom, R., Flores-Mendoza, C., & Abad, F. (2007). Generational changes on the draw-a-man test: A comparison of Brazilian urban and rural children tested in 1930, 2002 and 2004. *Journal of Biosocial Science*, *39*, 79–89.
Cotton, S., Kiely, P., Crewther, D., Thompson, B., Laycock, R., & Crewther, S. (2005). A normative and reliability study of the Raven's coloured matrices for primary school aged children from Victoria, Australia. *Personality and Individual Differences*, *39*, 647–659.
Dutton, E. (2014). *Religion and intelligence: An evolutionary analysis.* London: Ulster Institute for Social Research.
Dutton, E., & Lynn, R. (2013). A negative Flynn Effect in Finland, 1997–2009. *Intelligence*, *41*, 817–820.
Dutton, E., & Lynn, R. (2015). A negative Flynn Effect in France, 1999–2008/9. *Intelligence*, *51*, 67–70.
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. (2009). *What is intelligence? Beyond the Flynn Effect.* New York: Cambridge University Press.
Flynn, J. R. (2012). *Are we getting smarter? Rising IQ in the twenty first century.* Cambridge: Cambridge University Press.
Flynn, J. R., & Rossi-Case, L. (2012). IQ gains in Argentina between 1964 and 1998. *Intelligence*, *40*, 145–150.
Galton, F. (1869). *Hereditary genius.* London: MacMillan.
Khaleefa, O., Sulman, A., & Lynn, R. (2009). An increase of intelligence in Sudan, 1987–2007. *Journal of Biosocial Science*, *41*, 279–283.
Koivunen, S. (2007). *Suomalaismiesten kognitiivisen kykyprofiilin muutokset 1988–2001: Flynnin efektiä suomalaisessa aineistossa?* (Pro Gradu) Jyväskylä University.
Korgesaar, M. (2013). *Flynni Efekti Esinemine Eesti Abiturientide Seas Raveni Testi Pohjal.* (Seminaritöö) Tartu Ulikool.
Lynn, R. (2006). *Race differences in intelligence: An evolutionary analysis.* Augusta, GA: Washington Summit Publishers.
Lynn, R. (2011). *Dysgenics: Genetic deterioration in modern populations.* London: Ulster Institute for Social Research.
Lynn, R. (2013). Who discovered the Flynn Effect? A review of early studies of the secular increase of intelligence. *Intelligence*, *41*, 765–769.
Lynn, R., & Irwing, P. (2004). Sex differences on the progressive matrices: A meta-analysis. *Intelligence*, *32*, 481–498.
Lynn, R., & Vanhanen, T. (2012). *Intelligence: A unifying construct for the social sciences.* London: Ulster Institute for Social Research.
Migration Policy (2016). http://www.migrationpolicy.org/programs/data-hub/international-migration-statistics
Must, O., Must, A., & Raudik, V. (2003). The secular rise in IQs: In Estonia the Flynn Effect is not a Jensen Effect. *Intelligence*, *31*, 461–471.
Myrskylä, M., Silventoinen, K., Tynelius, P., & Rasmussen, F. (2013). Is later better or worse? Association of advanced parental age with offspring cognitive ability among half a million young Swedish men. *American Journal of Epidemiology*, *177*, 649–655.
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., & Voracek, M. (2015). One century of global IQ gains: A formal meta-analysis of the Flynn Effect (1909–2013). *Perspective on Psychological Science*, *10*, 282–306.
Rindermann, H., & Thompson, J. (2011). Cognitive capitalism : The effect of cognitive ability on wealth, as mediated through scientific achievement and economic freedom. *Psychological Science*, *22*, 754.
Ronnlund, M., Carlstedt, B., Blomstedt, Y., Nilsson, L., & Weinehall, L. (2013). Secular trends in cognitive test performance: Swedish conscript data 1970–1993. *Intelligence*, *41*, 19–24.
Runquist, E. A. (1936). Intelligence test scores and school marks in 1928 and 1933. *School & Society*, *43*, 301–304.

Please cite this article as: Dutton, E., et al., The negative Flynn Effect: A systematic literature review, *Intelligence* (2016), http://dx.doi.org/10.1016/j.intell.2016.10.002

ARTICLE IN PRESS



Saha, S., Barnett, A. G., Foldi, C., Burne, T. H., Eyles, D. W., et al. (2009). Advanced paternal age is associated with impaired neurocognitive outcomes during infancy and childhood. *PLoS Medicine*, *6*, e1000040. http://dx.doi.org/10.1371/journal.pmed.1000040.

Schelini, P., Almeida, L., Duarte, C., & Primi, R. (2011). Efeito Flynn: Estudo com adultos Portuegesos. *Psicologia Educação e Cultura*, *15*, 407–415.

Shayer, M., & Ginsburg, D. (2007). Thirty years on - a large anti-Flynn Effect? The Piagetian test volume & heaviness norms 1975–2003. *Journal of Educational Psychology*, *77*, 25–41.

Shayer, M., & Ginsburg, D. (2009). Thirty years on—A large anti-Flynn Effect? 11: 13-year olds. Piagetian tests of formal operations norms 1976–2006/7. *British Journal of Educational Psychology*, *79*, 409–418.

Smith, S. (1942). Language and non-verbal test performance of racial groups in Honolulu before and after a fourteen year interval. *Journal of General Psychology*, *26*, 51–93.

Sundet, J. M., Barlaug, D. G., & Torjussen, T. M. (2004). The end of the Flynn effect? A study of secular trends in mean intelligence test scores of Norwegian conscripts during half a century. *Intelligence*, *32*, 349–362.

te Nijenhuis, J., & Van der Flier, H. (2013). Is the Flynn Effect on *g*? A meta-analysis. *Intelligence*, *41*, 802–807.

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

Teasdale, T. W., & Owen, D. R. (2008). Secular declines in cognitive test scores: A reversal of the Flynn effect. *Intelligence*, *36*, 121–126.

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

Wehner, C., Kambskard, M., & Abrahamson, P. (2002). *Demography of the family: The case of Denmark*. UK: University of York.

Woodley, M., & Fernandes, H. (2016). Showing their true colors: Secular declines and a Jensen Effect on color acuity – More evidence for the weaker variant of Spearman's other hypothesis. *Personality and Individual Differences*, *88*, 280–284.

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

Woodley of Menie, M. A., & Dunkel, C. (2015). In France, are secular IQ losses biologically caused? A comment on Dutton and Lynn. *Intelligence*, *53*, 81–85. http://dx.doi.org/10.1016/j.intell.2015.08.009.

Woodley, M., te Nijenhuis, J., & Murphy, R. (2013). Were the Victorians cleverer than us? The decline in general intelligence from a meta-analysis of the slowing of simple reaction time. *Intelligence*, *41*, 843–850.

Woodley, M., 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., te Nijenhuis, J., & Murphy, R. (2015). The Victorians were still faster than us. Commentary: Factors influencing the latency of simple reaction time. *Frontiers in Human Neuroscience*, *9*, 452.

Woods, D., Wyma, J., Yund, E., et al. (2015). Factors influencing the latency of simple reaction time. *Frontiers in Human Neuroscience*, *9*, 31. http://dx.doi.org/10.3389/fnhum.2015.00131.

Please cite this article as: Dutton, E., et al., The negative Flynn Effect: A systematic literature review, *Intelligence* (2016), http://dx.doi.org/10.1016/j.intell.2016.10.002