



## Were the Victorians cleverer than us? The decline in general intelligence estimated from a meta-analysis of the slowing of simple reaction time

Michael A. Woodley <sup>a,b,\*</sup>, Jan te Nijenhuis <sup>c,2,3</sup>, Raegan Murphy <sup>d,4</sup>

<sup>a</sup> Department of Psychology, Umeå University, Sweden

<sup>b</sup> Center Leo Apostel for Interdisciplinary Studies, Vrije Universiteit Brussel, Belgium

<sup>c</sup> Work and Organizational Psychology, University of Amsterdam, The Netherlands

<sup>d</sup> School of Applied Psychology, University College Cork, Ireland

### ARTICLE INFO

#### Article history:

Received 17 February 2013

Received in revised form 15 April 2013

Accepted 15 April 2013

Available online 8 May 2013

#### Keywords:

Dysgenic

Genetic *g*

IQ

Simple reaction time

Psychometric meta-analysis

### ABSTRACT

The Victorian era was marked by an explosion of innovation and genius, per capita rates of which appear to have declined subsequently. The presence of dysgenic fertility for IQ amongst Western nations, starting in the 19th century, suggests that these trends might be related to declining IQ. This is because high-IQ people are more productive and more creative. We tested the hypothesis that the Victorians were cleverer than modern populations, using high-quality instruments, namely measures of simple visual reaction time in a meta-analytic study. Simple reaction time measures correlate substantially with measures of general intelligence (*g*) and are considered elementary measures of cognition. In this study we used the data on the secular slowing of simple reaction time described in a meta-analysis of 14 age-matched studies from Western countries conducted between 1889 and 2004 to estimate the decline in *g* that may have resulted from the presence of dysgenic fertility. Using psychometric meta-analysis we computed the true correlation between simple reaction time and *g*, yielding a decline of  $-1.16$  IQ points per decade or  $-13.35$  IQ points since Victorian times. These findings strongly indicate that with respect to *g* the Victorians were substantially cleverer than modern Western populations.

© 2013 Elsevier Inc. All rights reserved.

## 1. Introduction

### 1.1. The Victorians

Queen Victoria of the United Kingdom reigned from 1837 to 1901. The Victorian era was a period of immense industrial, cultural, political, scientific, and military change in Western

Europe marked by an explosion of creative genius that strongly influenced all other countries in the world. In international relations there was a long period of peace, known as the *Pax Britannica*. Breakthroughs in science led to an escape from the Malthusian trap: increasing populations did not starve and longevity increased. The growth in economic efficiency before the Victorian era was a miniscule 1% per century (Clark, 2008), but started increasing spectacularly in the Victorian era. The height of the per capita numbers of significant innovations in science and technology and also the per capita numbers of scientific geniuses was clearly situated in the Victorian era; after which there was a decline (Huebner, 2005; Murray, 2003; Woodley, 2012; Woodley & Figueredo, 2013).

IQ scores are excellent predictors of job performance (Schmidt & Hunter, 1999) and high-IQ people are more productive and more creative (Jensen, 1998). A population with a higher intelligence will in general be more productive and

\* Corresponding author at: Department of Psychology, Umeå University, Sweden.

E-mail address: Michael.Woodley@psy.umu.se (M.A. Woodley).

<sup>1</sup> MAW conceived the analysis and drafted the manuscript.

<sup>2</sup> The two first authors contributed equally to this study; the order of the names is random.

<sup>3</sup> JTN conducted the analyses and contributed to subsequent drafts of the manuscript.

<sup>4</sup> RM collected and validated the data used in the analysis.

creative than a population with lower intelligence (Lynn & Vanhanen, 2012; Rindermann, Sailer, & Thompson, 2009). Were the Victorians therefore cleverer than us? Here we test this hypothesis using measures of reaction time (RT), which give a good indication of general intelligence (e.g. Johnson & Deary, 2011) in a meta-analytic study.

### 1.2. Measured IQ scores increase: The Flynn effect

At first sight, the case for a decrease in intelligence since Victorian times seems highly implausible. After all, there is now consensus that at least since World War II, IQ scores have been going up, the so-called Flynn effect. Flynn (1987, 2009) showed a worldwide increase in measured IQ scores of approximately 3 points a decade. Recent studies show similar gains in South Africa (te Nijenhuis, Murphy, & van Eeden, 2011) and much larger effects in South Korea (te Nijenhuis, Cho, Murphy, & Lee, 2012). These gains are thought to be due almost entirely to environmental improvements stemming from factors such as improved education, nutrition, hygiene, and exposure to cognitive complexity (Neisser, 1997). The Flynn effect has therefore been described as an increase in *phenotypic* intelligence, i.e. the intelligence that results from a combination of genes and environmental factors (Lynn, 2011).

### 1.3. The dysgenics paradox

Dysgenic trends result from socially valued and heritable traits, such as intelligence, declining within populations over time due to the effects of selection operating against those traits (Galton, 1869; Lynn, 2011). Before 1825 Western countries were in eugenic fertility, in that those with the highest levels of education and/or social status had the largest numbers of *surviving* offspring (Lynn, 2011; Skirbekk, 2008). The majority of these countries completed the transition into dysgenic fertility for these IQ proxies by around the middle of the 19th century (Lynn, 2011; Skirbekk, 2008).

The presence of a dysgenic effect on intelligence has proven difficult to detect via direct measurement, i.e. by comparing IQ scores of different age-matched generations on the same IQ battery. The earliest cross-sectional studies (1930s–1950s), attempting to quantify the decline actually found the opposite effect i.e. rising IQ scores (e.g. Cattell, 1950). This presented a paradox as studies from the same time period consistently found negative correlations between IQ and variables such as fertility and family size (Lynn, 2011; van Court & Bean, 1985). Given the observation that IQ is substantially heritable, this finding should have entailed declining rather than increasing IQ (Lynn, 2011; van Court & Bean, 1985). The failure to directly measure a dysgenic effect on IQ is now attributed to the Flynn effect: the strong secular rise in IQ simply masks the likely much weaker dysgenic decline in IQ (Lynn, 2011).

Nonetheless attempts have been made to estimate the theoretical rate of dysgenic change in IQ based on the magnitude of the negative correlation between fertility and IQ (see: Lynn, 2011 for an overview of these studies). These estimates, which range from a low of  $-.12$  (Retherford & Sewell, 1988) to a high of approximately  $-1.6$  points per decade (Lentz, 1927), are however *inferred* rather than *observed* declines. So, dysgenic

effects appear to be unmeasurable directly using standard IQ tests.

### 1.4. Genotypic IQ decreases

Other research has examined whether dysgenic effects have a genetic component by testing for so-called Jensen effects (Rushton, 1998). When looking at the subtests of an IQ battery these subtests range from high complexity (high loadings on the *g* factor of intelligence) to low complexity (low loadings on the *g* factor). Jensen effects refer to the tendency for the test's *g* loadings to positively correlate with the size of the effect of other variables on the same subtests. So, subtests with high *g* loadings go with strong effects and subtests with low *g* loadings go with weak effects. Jensen effects exist on genetic variables, such as heritability, inbreeding depression, and it's opposite, hybrid vigour (Jensen, 1998; Rushton & Jensen, 2010). Clear Jensen effects have also been found for dysgenic fertility (Woodley & Meisenberg, 2013). This indicates that dysgenic fertility is predominantly a genetic effect: i.e. *genotypic* IQ or more accurately 'genetic *g*' (Rushton & Jensen, 2010) decreases. However, the Flynn effect is clearly not a Jensen effect, as it exhibits a modest, negative correlation with subtest *g* loadings (te Nijenhuis & van der Flier, 2013–this issue). In summary therefore the pattern of genetic effects such as heritabilities on the subtests of an IQ battery are highly similar to the pattern in dysgenic effects, however both show no resemblance to the pattern in the Flynn effect.

### 1.5. Reaction time as a high-quality measure of general intelligence

Galton (1883) was the first to suggest that RT might be an elementary cognitive measure as it appeared to be an indicator of speed of mental processing. Subsequent research has confirmed many key predictions of the speed-of-processing theory of intelligence via the demonstration of robust correlations between measures of RT and IQ (see: Jensen, 2006 for an overview). Moreover, there is a Jensen effect on RT, as more *g*-loaded subtests of an IQ battery correlate more strongly with RT measures than do less *g*-loaded ones (Jensen, 1998, pp. 234–238). This has led Jensen (1998, 2006, 2011) to suggest that RT is in fact a biological marker of mechanisms fundamental to the operation of general intelligence, such as neurophysiological efficiency. Furthermore, RT is a 'ratio-scale' measure of intelligence meaning that it has a true zero (analogously to the Kelvin scale in temperature measurement). This means that RT can be used to meaningfully compare historical and contemporary populations in terms of levels of general intelligence (Jensen, 2011).

Even the most simple measure of RT (i.e. the time that it takes for an individual to respond to a sensory stimulus) appears to be robustly associated with IQ. Rijdsdijk, Vernon, and Boomsma (1998) for example investigated the relationship between simple RT and IQ in a genetic analysis using twins. Simple RT and IQ as measured using the Raven's Advanced Progressive Matrices were found to exhibit identical levels of heritability (.58 and .58, respectively) and furthermore the phenotypic correlation between the two of  $-.21$  (increasing IQ goes with decreasing RT speed, hence the correlation is negative) was completely mediated by *common* genetic factors. Another relevant study is that of Deary, Der, and Ford (2001) who set out to generate benchmark estimates for the correlation

between IQ and various RT measures (including simple) in a population-representative sample yielding a correlation between the two of  $-.31$ , indicating a substantive relationship.

### 1.6. A secular slowing of reaction time

Silverman (2010) reviews simple RT studies conducted between the 1880s and the present day. In Silverman's (2010) study, Galton's estimates collected between 1884 and 1893 (as reported in Johnson et al., 1985) were compared with twelve studies from the modern era (post 1941). Galton's measures indicated a simple visual RT mean of 183 milliseconds (ms) for a large sample of 2522 young adult males (aged between 18 and 30), along with a mean of 187 ms for a sample of 888 equivalently aged females. These means seem to be representative of the period as a 1911 review of various studies conducted in the 19th century (Ladd & Woodworth, 1911), which did not include Galton's measures, found an RT range of 151–225 ms (mean 190 ms), using different instrumentation to that employed by Galton (1889). Moreover, Silverman was also able to comprehensively rule out lack of socioeconomic diversity, as Galton's samples were diverse enough to be stratified into seven male and six female occupational groups (Johnson et al., 1985).

Twelve modern (post 1941) simple RT studies by contrast revealed considerably slower RTs for both males (mean 250.43 ms) and females (mean 277.71 ms) in a combined sample of 6929. In comparing the 19th-century measures with the modern ones, Silverman found that in all but one comparison, the differences were statistically significant. Furthermore age was not a confounding factor as Silverman matched studies across time based on age range.

### 1.7. Estimating the dysgenic effect for *g*

The studies of Deary et al. (2001) and Rijdsdijk et al. (1998) combine to indicate that the simple RT/IQ correlation is substantial at the population level, and that furthermore the association between the two is completely mediated by common genetic factors. Hence, given the strong Jensen effects on both dysgenic effects (Woodley & Meisenberg, 2013) and simple RT (Jensen, 1998) a secular increase in simple RT latency is in fact an expected outcome of a dysgenic decline in 'genetic *g*'. Based on this it should be possible to estimate the degree to which 'genetic *g*' has declined in Western populations due to dysgenic pressures, since the 1880 s using Silverman's (2010) data.

### 1.8. Research questions

This leads to the following two research questions. 1) How strong is the secular slowing of simple RT? 2) How strong is the decadal *g* decline based on simple RT measures?

## 2. Methods

The data on simple RT used here, with the exception of one study (Thompson, 1903), comes from Silverman (2010) and sources contained therein. Silverman carried out various analyses on simple visual reaction time measures and is an excellent source. He describes a thorough meta-analytical search yielding the means for 13 different studies, involving samples

of equivalent age ranging in time from between 1889 and 2004. He also mentions the review by Ladd and Woodworth (1911) of eight early studies of reaction time from the 19th century, which indicate the representativeness of Galton's simple RT estimates. Like Silverman we do not include the results of these studies in our final analysis as there are too few details provided that would permit their suitability to be determined, based on the inclusion rules. This leaves all 13 age-matched studies used in Silverman's (2010) analysis, along with one other 19th century simple RT study from the US (Thompson, 1903). We were directed to the Thompson study by Silverman (pers. com), on the basis that even though he missed it in his 2010 study, it nonetheless satisfies his inclusion rules and should be included on that basis.

### 2.1. General inclusion rules

We take our general inclusion rules from the meta-analysis by Silverman (2010). First, the samples consisted of people recruited from the general population and whose ages ranged from about 18 to 30 years. Second, the study sample had to be free of neurotoxins as these are a known inhibitor of RT performance. Third, given that Galton's sample was British the studies had to have been conducted in a Western country. Fourth, the study samples had to be 20 or larger in size for each sex. Fifth, the delivery of the stimulus was not predictable, which ruled out studies in which the interval between stimuli was fixed or increased or decreased according to a regular pattern. Sixth, the response to the stimulus had to be manual in nature, such as pressing or releasing a button or key. Seventh, to generate the response, the arm did not have to be moved (this restriction was based on the consideration that if the arm must be moved, RT is necessarily lengthened, and the *g*-loadedness of the estimate potentially reduced due to the addition of a non-cognitive 'movement time' component to the measures, Jensen, 2006). Eighth, the RT measure had to be representative of the total set of RTs. This restriction eliminated studies in which RT was measured in terms of the best RTs or the longest or shortest RT.

As sex-differences data were not available for each study, we generate weighted averages for studies reporting sex differences, thus we produce a single RT mean for each study. Finally it must be noted that reaction time measures tend to show strongly skewed distributions (see: Jensen, 2006). For skewed distributions the median would be the better measure, but because not all reaction time studies reported the median, we choose the mean instead. Table 1 reports all data used in this study.

### 2.2. Psychometric meta-analysis

Regression with year is used to generate trend-weighted estimates of 19th-century (1889 – mean year of Galton's study) and modern (2004 – the year of the most recent study in the collection) RT means. The population-representative study of Deary et al. (2001) is used for obtaining benchmark estimates of the simple RT/IQ correlation, along with estimates of standard deviations. Psychometric meta-analysis (Hunter & Schmidt, 1990, 2004) can be used to correct for statistical artefacts that typically alter the value of outcome measures. There are five such artefacts that need controlling. These include sampling error, reliability of the first variable, reliability of the second variable, restriction of range, and deviation

**Table 1**

13 simple RT studies used in Silverman (2010) and Thompson (1903) along with 16 simple RT means, sample sizes, collection/publication year and references.

| Testing year and country                    | Males (N)                | Females (N)              | Sample size weighted mean (total N) | Reference                                       |
|---|--------------------------|--------------------------|-------------------------------------|---|
| 1888.5 <sup>a</sup> (1884–1893) (UK)        | 183 (2522)               | 187 (888)                | 184 (3410)                          | Galton's data in Johnson et al. (1985)          |
| 1894.5 (1889–1900) (USA)                    | 199 (24)                 | 217 (25)                 | 208.2 (49)                          | Thompson (1903)                                 |
| 1941 (USA)                                  | 197 (47)                 | n.a                      | 197 (47)                            | Seashore, Starmann, Kendall, and Helmick (1941) |
| 1941 (USA)                                  | 203 (47)                 | n.a                      | 203 (47)                            | Seashore et al. (1941)                          |
| 1945 (UK)                                   | 286 (76)                 | n.a                      | 286 (76)                            | Forbes (1945)                                   |
| 1970 (Canada)                               | 236 (40)                 | 263 (40)                 | 249.5 (80)                          | Lefcourt and Siegel (1970)                      |
| 1990 (Finland)                              | 199 (893)                | n.a                      | 199 (893)                           | Taimela (1991)                                  |
| 1987 (Finland)                              | 183 (123)                | n.a                      | 183 (123)                           | Taimela, Kujala, and Osterman (1991)            |
| 1993 (USA)                                  | 260 (80)                 | 285 (140)                | 275.9 (220)                         | Anger et al. (1993)                             |
| 1993 (USA)                                  | 250 (73)                 | 280 (163)                | 270.7 (236)                         | Anger et al. (1993)                             |
| 1999 (UK)                                   | 306 (64)                 | n.a                      | 306 (64)                            | Smith et al. (1999)                             |
| 2002 (UK)                                   | 324 (24)                 | n.a                      | 324 (24)                            | Brice and Smith (2002)                          |
| 1999.5 <sup>a</sup> (1999–2000) (Australia) | 214 (1163)               | 224 (1241)               | 219.2 (2404)                        | Jorm, Anstey, Christensen, and Rodgers (2004)   |
| 2004 (Canada)                               | 253 (171)                | 268 (198)                | 261.1 (369)                         | Reed, Vernon, and Johnson (2004)                |
| 1987.5 (1987–1988) (UK)                     | 295 (254.5) <sup>b</sup> | 306 (288.5) <sup>b</sup> | 300.8 (543)                         | Deary and Der (2005a)                           |
| 1984.5 (1984–1985) (UK)                     | 300 (834)                | 318 (1023)               | 309.9 (1857)                        | Der and Deary (2006)                            |

*Additional.* We went back to Johnson et al. (1985) and cross-referenced it with Silverman (2010). The total N for females should be 888 rather than 302. We changed the above N to reflect the correct females sample size.

<sup>a</sup> When a range of years is given the average is taken.

<sup>b</sup> In these studies between 254–255 males and 288–289 females were used – hence the Ns are averaged.

from perfect construct validity. These corrections are used to determine the true correlation between *g* and simple RT, and hence the rate of *g* decline between 1889 and 2004.

### 2.3. Meta-regression

Meta-regression is a method for examining the influence of one or more covariates on the outcome effects. We carried out meta-regression in which we regressed the effect size – the mean RT of a study – on the covariate – the year of the study, using the software available on [www.stattools.net](http://www.stattools.net). We carried out a random-effects meta-regression, because it is generally considered to be the more appropriate technique in most studies (Borenstein, Hedges, Higgins, & Rothstein, 2009), and computed Tau<sup>2</sup> using the Empirical Bayes Estimate (see: Thompson & Sharp, 1999).

Each study needs to be given a specific weight, and in this case we took the Standard Error of the Mean (SEM) score of the study. SEM is the standard deviation of the sample-mean's estimate of a population mean. SEM is usually estimated by the sample estimate of the population standard deviation (sample standard deviation) divided by the square root of the sample size:  $SEM = s/\sqrt{N}$ .

Where:

*s* is the sample standard deviation, and  
*N* is the size of the sample.

The study of Deary et al. (2001) reports a good approximation of the population SD of a simple RT measure. Adding to the study of Deary et al. (see above) we estimate that the population SD = 160.4. The population SD is of course much better than the sample SDs, which are merely estimates of the population SD, and which vary substantially among themselves introducing additional error. So, we decided to use the value of the population SD in the computation of the standard errors of all the individual studies in our meta-analysis, using the formula  $SEM = 160.4/\sqrt{n}$ .

## 3. Results

### 3.1. Estimation of decline in average reaction times

In Fig. 1 the simple RT means for all 16 effects were regressed against year so as to determine the overall temporal trend. The trend beta coefficient, computed using a random-effects meta-regression, equals .670 and is significant at  $p = .02$ .

The difference between the meta-regression trend-weighted present (2004) simple RT mean (270.18 ms) and the trend-weighted 1889 mean (193.17 ms) is 77 ms.

### 3.2. Estimation of the population SD of reaction time and IQ

The study by Deary et al. (2001) is an attempt to generate a benchmark estimate of various parameters relating to a variety of RT measures and IQ based on a sample broadly representative of, in this case, the Scottish population ( $N = 900$ ). The sample was drawn from the West of Scotland Twenty-07 Study, which is a population-based cohort study and was obtained using a two-stage random sampling strategy. The mean age of the participants was 56 so the sample is representative of cohorts that are older than those used by Silverman. RT (both simple and choice) was measured using a 'Hick'-style device and IQ was measured using the 65 items of deductive reasoning constituting the numeric and verbal sections of the Alice Heim Group Ability Test Part I (AH4 Part I).

The study reports a simple RT SD of 119.7 ms. This value is much higher than the SDs in the individual samples (which range from 15 [Lefcourt & Siegel, 1970] to 90.6 [Deary & Der, 2005a]) indicating a strong restriction of range in virtually all samples in Silverman's study. When comparing the SD values of this Scottish sample on the AH4 total score (11.3) with the SD values of the samples in the AH4 manual it is clear that the value from the Scottish sample is much lower, indicating that it underestimates the true population value. The samples in the AH4 manual are not nationally representative (Alexopoulos,

1998, p. 645). However, quite a few samples of children and young adults were tested and some of the samples are quite large.

We compare the *SD* values for simple RT with the *SD* values of the AH4 from the manual, so as to correct the former for range restriction. This is achieved by comparing the *SD* of the Deary et al. (2001) sample to *SD*s of young adults (seventeen-plus-years-old) and of young children from the manual, it is apparent that all samples of seventeen-plus-years-old have *SD*s that are substantially larger. The sample size-weighted *SD* is 14.3, indicating that the range in the Scottish sample is at least 21% too small. However, these samples are still not truly nationally representative, hence the population *SD*s will most likely be even larger. The best approximations of nationally representative samples in the manual are the two large samples of, respectively, eleven-year-olds and twelve-year-olds from comprehensive schools. This is because after primary school, children are allocated secondary education which is most optimal for their IQ level, leading to IQs that are more homogeneous in secondary than in primary education. We therefore computed the sample-size weighted mean *SD* of the two large samples of eleven-year-olds and twelve-year-olds from comprehensive schools, which yielded an *SD* with a value of 17.2. This suggests that the simple RT *SD* in the Scottish sample is similarly underestimated by no less than 34%, meaning that the *SD* is not 119.7, but 160.4. We take this value of 160.4 as the best estimate of the population *SD* of simple RT.

### 3.3. Estimate of the true correlation between reaction time and IQ/g

Deary et al. (2001) estimate the correlation between simple RT and IQ in their population-representative sample at  $-.31$ . However, in contrast with the meta-analysis by Jensen (1987) Deary et al. do not correct for measurement artefacts. On this basis there is likely to be measurement error in the correlation, and the value of  $-.31$  is an underestimate of the true correlation. We therefore correct for unreliability in the simple RT and IQ measures, restriction of range in the IQ

measure and imperfect measurement of the construct of *g* (Hunter & Schmidt, 2004; Jensen, 1998, pp. 380–383).

### 3.4. Reliability in the simple RT and IQ measures

Deary et al. (2001, p. 397) suggest a test–retest reliability of both the simple RT and IQ measures of .85. We use this value for our corrections. Correcting for unreliability means dividing the observed value by the square root of the reliability, which yields a correction factor in both cases of 1.09.

### 3.5. Restriction of range

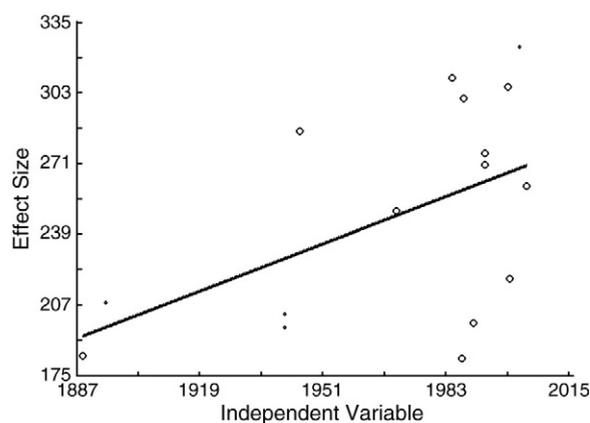
The value of the correlation between the IQ measure and the simple RT measure is attenuated by range restriction in the sample. The solution to variation in range is to define a reference population and express all correlations in terms of it (Hunter & Schmidt, 1990, pp. 47–49). The next step is to compute what the correlation in a given population would be if the *SD* were the same as in the reference population. The *SD*s can be compared by dividing the *SD* of the study population by the *SD* of the reference group, that is  $u = SD_{\text{study}} / SD_{\text{ref}}$ . Previously we showed that the Scottish sample is likely strongly affected by range restriction, yielding a value of  $u = 119.7/160.4 = .75$ . Correcting for restriction of range means dividing the observed correlation by the value of *u*, yielding a correction factor of 1.33.

### 3.6. Imperfectly measuring the construct of *g*

The deviation from perfect construct validity in *g* attenuates the values of the correlation between the IQ test and the simple RT measure. In making up any collection of cognitive tests, we do not have a perfectly representative sample of all possible cognitive tests. Therefore any one limited sample of tests will not yield exactly the same *g* as another such sample. The sample values of *g* and therefore also correlations involving measures of *g* are attenuated by psychometric sampling error, but the fact that *g* is very substantially correlated across different test batteries implies that the differing obtained values of *g* can all be interpreted as estimates of a “true” *g* (e.g. Johnson, Bouchard, Krueger, McGue, & Gottesman, 2004).

The more tests and the higher their *g* loadings, the higher the *g* saturation of the composite score. The Wechsler tests have a large number of subtests with quite high *g* loadings, yielding a highly *g*-saturated composite score. The *g* score of the Wechsler tests correlates more than .95 with the tests’ IQ score (Jensen, 1998, pp. 90–91). However, shorter batteries with a substantial number of tests with lower *g* loadings will lead to a composite with somewhat lower *g* saturation. The average *g* loading of an IQ score as measured by various standard IQ tests lies in the  $+.80$ s (Jensen, 1998, ch. 10). When this value is taken as an indication of the degree to which an IQ score is a reflection of “true” *g*, it can be estimated that a tests’ *g* score correlates about .85 with “true” *g*. As *g* loadings represent the correlations of tests with the *g* score, it is likely that most empirical *g* loadings will underestimate “true” *g* loadings. To limit the risk of overcorrection a conservative value of .90 can be used as a basis for a classical test battery like the Wechsler.

As the AH4 Part 1 consists solely of items that measure fluid intelligence it is expected that the *g* loadedness of the sum



**Fig. 1.** Simple RT mean vs. year for 16 effects. The size of the bubbles is categorically determined by sample size with small bubbles representing studies with *N* values  $<50$  and large bubbles representing *N* values  $\geq 50$ . The scatter is fitted to a linear function, so as to illustrate the secular trend, and is weighted based on a random-effects meta-regression model.

score is high. The manual of the AH4 (p. 10) shows correlations ranging from .60 to .76 between the total score on the AH4 and the total score on other IQ tests, including the Raven's Progressive Matrices, with higher values for the larger samples. This is actually higher than the mean correlation of .67 between the total scores of various standard intelligence tests reported elsewhere in the literature (Jensen, 1980, pp. 314–315). The manual reports a factor analysis of the intercorrelations between the various sum scores of similar items showing a strong general factor running through the whole test, each sum score correlating highly with the general factor – values of  $r$  lying between .80 and .86 (Heim, 1970, p.9). So, it appears the  $g$  loadedness of the AH4 Part 1 is similar to the  $g$  loadedness of a classical battery such as the Wechsler. Therefore, the correction for imperfectly measuring the construct  $g$  should be modest: 10%, hence a correction factor of 1.10.

In sum, the observed correlation in the Scottish data between the IQ test and the simple RT test is  $-.31$ . The correction factor for unreliability in both the simple RT and IQ measures is 1.09, the correction factor for restriction of range is 1.33, and the correction factor for imperfectly measuring the construct of  $g$  is 1.10. Applying the four corrections to the value of the correlation yields a true correlation  $\rho = -.54$ . This demonstrates that the  $g$  loadedness of simple RT is quite a bit larger than the observed correlations suggest.

### 3.7. Using effect sizes for reaction time to compute effect sizes for IQ/ $g$

Our measure is an imperfect reflection of  $g$ : its true, absolute correlation with  $g$  is .54 hence our measure's true  $g$ -loadedness is .54, similar to that of certain subtests of an IQ battery, whereas the true  $g$ -loadedness of  $g$  is by definition 1.00. In other words, simple RT measure 54% of the  $g$  factor. As our interest is in the decline in  $g$  we need to extrapolate our findings from a measure with a  $g$  loading of .54 to a measure with a  $g$  loading of 1.00. We therefore divide the effect size ( $d$ ) for simple RT by the value of .54, as the  $d$  for the simple RT measure between 1889 and 2004 is .48 (77/160.4). With a  $g$  loading of .54 this yields an equivalent  $d = .48/.54$ , which results in a correction factor of .89 for the total score on a broad IQ battery. This means that 'genetic  $g$ ' (recall that the most heritable IQ measures are also the most  $g$  loaded; Rushton & Jensen, 2010) has decreased by  $-13.35$  IQ points since Galton carried out his studies; this is a decline of  $-1.16$  IQ points per decade between 1889 and 2004.

### 3.8. Regression line

We carried out a meta-regression, where we test the hypothesis that the year of the study (year) predicts the mean simple RT of a study, so the regression formula is simple RT =  $a + b$  (year). This resulted in a regression line according to the formula:

$$\text{Simple RT} = -1071.7 + .6696(\text{year})$$

With the standard error (SE) of the regression coefficient  $b = .307$  and the 95% confidence interval of  $b$  from .0676 to 1.2716, it is clear that the regression coefficient does not traverse 0 implying that the value is significant. A precise estimate of the significance is found by calculating the absolute value of the

ratio of the  $b$  coefficient to the SE of  $b$ , which results in a  $z$  value of 2.18 with an associated probability value of .02. The mean simple RT becomes *significantly* slower as time goes by.

Fig. 1 shows the meta-regression weighted scatter of the means of the individual studies in the meta-analysis over time. It shows that there are a couple of data points that are at quite a distance from the regression line, but they frequently concern samples with smaller  $N$  values. Most of the larger studies are close or relatively close to the regression line. Our analyses confirm that the regression formula explains the variance between the data points quite well. The residual error of the Sum of Squares ( $Q_e$ ) is 13.3748 ( $df = 14$ ;  $p = .50$ ), which is non-significant. This means that after taking the year of study into account not a lot of heterogeneity is left over, so, there is little room left for additional moderators. We conclude that the year of the study has a clear influence on the mean simple RT of a study, in that the mean simple RT becomes slower over time.

## 4. Discussion

The Victorian era was characterized by great accomplishments. As great accomplishment is generally a product of high intelligence, we tested the hypothesis that the Victorians were actually cleverer than modern populations. We used a robust elementary cognitive indicator of general intelligence, namely measures of simple RT.

In the present study we used the data on the secular increase in simple RT described in a meta-analysis of 16 age-matched studies from Western countries conducted between 1889 and 2004 to generate estimates of the rate of IQ decline. The decline estimate of  $-1.16$  IQ points per decade from the present study falls within the range of those produced in previous studies employing the magnitude of the dysgenic effect on IQ as the basis for estimating declines (i.e.  $-.12$  to approximately  $-1.6$  points per decade). Our estimate is the first to be based on the use of real data rather than inference, however.

Whilst the dysgenic model is a plausible cause of the decline in RT performance Silverman (2010) does not address this potential cause, and instead offers other suggestions.

Silverman's first suggestion is that an ambient, population-wide increase in neurotoxic load stemming from persistent exposure to substances such as lead, may be responsible for the slowdown in simple RTs. Studies indicate however that the depressant effect of neurotoxins on IQ are typically *least* pronounced on the strongest measures of  $g$  (Lezak, 1983). As we are *only* considering the decline in simple RT that is due to the decline in 'genetic  $g$ ', this is grounds for ruling out contributions from neurotoxins, as whatever is diminishing 'genetic  $g$ ' should be a Jensen effect. This makes dysgenic fertility the prime candidate (Woodley & Meisenberg, 2013).

Silverman's other suggested cause of the decline is that the trend has resulted from those with poorer health and slower simple RTs surviving into adulthood more so in the modern era than in the past, and that it is the increasing numbers of such individuals that has diminished simple RT performance over time. We argue that this observation is fully compatible with the dysgenic model. One of the papers that Silverman cites in support of the association between health and RT is Deary and Der (2005b), who found that  $g$  mediates this

relationship. One source of correlations amongst these diverse traits is pleiotropic mutation-load (Miller, 2000). Pleiotropy describes the tendency for mutations to have general rather than isolated effects on different traits. For example a mutation which reduces myelination of neurons might simultaneously diminish both IQ and RT performance, as less myelinated neurons are less able to carry signals efficiently, therefore such neurons will be less efficient at processing information in the brain (Holm, Ullén, & Madison, 2011; Miller, 1994). Owing to the fact that they have general physiological effects, such mutations can diminish health also (Arden, Gottfredson, & Miller, 2009). As a consequence of this, if dysgenic fertility is favouring the carriers of mutant alleles that reduce 'genetic *g*' and RT performance, the frequency of certain diseases and disorders should increase also. Indeed there is evidence that this may well be occurring (Lynn, 2011).

There are some limitations to this study. Although Silverman used stringent selection criteria the trend may nonetheless be influenced by methodological artefacts and sample peculiarities. This is a potentially important issue as there appears to be a substantial discrepancy between the test-retest coefficients in Galton's data reported by Johnson et al. (1985), i.e. .21 for people tested within a year ( $N = 421$ ) and .17 for people retested over any time interval ( $N = 1069$ ), and the equivalent suggested coefficient of the 'Hick'-style device employed in our reference study (.85; Deary et al., 2001). Given the large  $N$  used by Silverman (2010) in establishing the Galton simple RT means, it is unlikely that even relatively low reliability at the individual level would seriously compromise the accuracy of the group mean of Galton's data. This is especially likely to be the case given the apparent representativeness of Galton's mean relative to other contemporaneous studies of simple RT, some of which employed likely much better quality instrumentation than that used by Galton (1889), such as the electro-mechanical Hipp chronoscope (Ladd & Woodworth, 1911; Thompson, 1903). It should also be emphasized that whilst our value of a  $-13.35$  IQ point decline is an estimate based on the best meta-analytical data available, a simple inspection of our figure shows there is a non-negligible amount of scatter around the regression line. The real magnitude of the effect might therefore be several IQ points lower or even higher.

In conclusion however these findings do indicate that with respect to 'genetic *g*' the Victorians were indeed substantially cleverer than modern populations.

## Acknowledgements

We would like first and foremost to thank Bruce Charlton for inspiring this study and for constructively critical comments on earlier drafts of this manuscript. He was not only the first person to propose a relationship between declining reaction times and dysgenic fertility, but he was the first to attempt an estimation of the decline using Silverman's data on his blog *Charlton's Miscellany* (see: Charlton, 2012 in the references for a URL to the original). We would also like to thank Irwin Silverman for sharing his expertise on reaction time testing with us and supplying an additional data point for our meta-analysis. Finally, we would like to thank Richard Lynn, Gerhard Meisenberg, and Guy Madison for comments, which in all cases enhanced this manuscript.

## References

- Alexopoulos, D. S. (1998). Factor structure of Heim's AH4. *Perceptual and Motor Skills*, 86, 643–646.
- Anger, W. K., Cassitto, M. G., Liang, Y. -X., Amador, R., Hooisma, J., Chrislip, D. W., et al. (1993). Comparison of performance from three continents on the WHO-recommended Neurobehavioral Core Test Battery (NCTB). *Environmental Research*, 62, 125–147.
- Arden, R., Gottfredson, L. S., & Miller, G. (2009). Does a fitness factor contribute to the association between intelligence and health outcomes? Evidence from medical abnormality counts among 3654 US Veterans. *Intelligence*, 37, 581–591.
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Chichester, UK: Wiley.
- Brice, C. F., & Smith, A. P. (2002). Effects of caffeine on mood and performance: A study of realistic consumption. *Psychopharmacology*, 164, 188–192.
- Cattell, R. B. (1950). The fate of national intelligence: Test of a thirteen-year prediction. *The Eugenics Review*, 42, 136–148.
- Charlton, B. G. (2012). *Taking on board that the Victorians were more intelligent than us*. Bruce Charlton's miscellany (<http://charltonteaching.blogspot.co.uk/2012/06/taking-on-board-that-victorians-were.html>)
- Clark, G. (2008). *A farewell to alms: A brief economic history of the world*. Princeton, NJ: Princeton University Press.
- Deary, I. J., & Der, G. (2005a). Reaction time, age, and cognitive ability: Longitudinal findings from age 16 to 63 years in representative population samples. *Aging, Neuropsychology and Cognition*, 12, 187–213.
- Deary, I. J., & Der, G. (2005b). Reaction time explains IQ's association with death. *Psychological Science*, 16, 64–69.
- Deary, I. J., Der, G., & Ford, G. (2001). Reaction times and intelligence differences: A population-based cohort study. *Intelligence*, 29, 389–399.
- Der, G., & Deary, I. J. (2006). Age and sex differences in reaction time in adulthood: Results from the United Kingdom Health Lifestyle Survey. *Psychology and Aging*, 21, 62–73.
- Flynn, J. R. (1987). Massive IQ gains in 14 nations: what IQ tests really measure. *Psychological Bulletin*, 101, 171–191.
- Flynn, J. R. (2009). *What is intelligence? Beyond the Flynn effect* (expanded ed.). Cambridge, UK: Cambridge University Press.
- Forbes, G. (1945). The effect of certain variables on visual and auditory reaction times. *Journal of Experimental Psychology*, 35, 153–162.
- Galton, F. (1869). *Hereditary genius*. London, UK: Macmillan Everyman's Library.
- Galton, F. (1883). *Inquiries into human faculty and its development*. London, UK: Macmillan Everyman's Library.
- Galton, F. (1889). An instrument for measuring reaction time. *Report of the British Association for the Advancement of Science*, 59, 784–785.
- Heim, A. W. (1970). *AH4 group test of general intelligence manual*. Windsor: NFER.
- Holm, L., Ullén, F., & Madison, G. (2011). Intelligence and temporal accuracy of behaviour: Unique and shared associations with reaction time and motor timing. *Experimental Brain Research*, 214, 175–183.
- Huebner, J. (2005). A possible declining trend for worldwide innovation. *Technological Forecasting and Social Change*, 72, 980–986.
- Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis: Correcting error and bias in research findings*. Newbury Park, CA: Sage.
- Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis (2nd Ed.): Correcting error and bias in research findings*. Thousand Oaks, CA: Sage.
- Jensen, A. R. (1980). *Bias in mental testing*. New York, NY: The Free Press.
- Jensen, A. R. (1987). Process differences and individual differences in some cognitive tasks. *Intelligence*, 11, 107–136.
- Jensen, A. R. (1998). *The g factor: The science of mental ability*. Westport, CT: Praeger.
- Jensen, A. R. (2006). *Clocking the mind: Mental chronometry and individual differences*. Oxford, UK: Elsevier.
- Jensen, A. R. (2011). The theory of intelligence and its measurement. *Intelligence*, 39, 171–177.
- Johnson, W., Bouchard, T. J., Jr., Krueger, R. F., McGue, M., & Gottesman, I. I. (2004). Just one g: Consistent results from three test batteries. *Intelligence*, 32, 95–107.
- Johnson, W., & Deary, I. (2011). Placing inspection time, reaction time, and perceptual speed in the broader context of cognitive ability: The VPR model in the Lothian Birth Cohort 1936. *Intelligence*, 39, 405–417.
- Johnson, R. C., McClearn, G., Yuen, S., Nagosha, C. T., Abern, F. M., & Cole, R. E. (1985). Galton's data a century later. *American Psychologist*, 40, 875–892.
- Jorm, A. F., Anstey, K. J., Christensen, H., & Rodgers, B. (2004). Gender differences in cognitive abilities: The mediating role of health state and health habits. *Intelligence*, 32, 7–23.
- Ladd, G. T., & Woodworth, R. S. (1911). *Physiological psychology*. New York, NY: Scribner.

- Lefcourt, H. M., & Siegel, J. M. (1970). Reaction time behaviour as a function of internal-external control of reinforcement and control of test administration. *Canadian Journal of Behavioural Sciences*, 2, 253–266.
- Lentz, T. (1927). Relation of IQ to size of family. *Journal of Educational Psychology*, 18, 486–496.
- Lezak, M. D. (1983). *Neuropsychological assesment*. New York, NY: Oxford University Press.
- Lynn, R. (2011). *Dysgenics: Genetic deterioration in modern populations* (revised ed.). London, UK: Ulster Institute for Social Research.
- Lynn, R., & Vanhanen, T. (2012). *Intelligence: A unifying construct for the social sciences*. London, UK: Ulster Institute for Social Research.
- Miller, E. M. (1994). Intelligence and brain myelination: A hypothesis. *Personality and Individual Differences*, 17, 803–832.
- Miller, G. F. (2000). Mental traits as fitness indicators: Expanding evolutionary psychology's adaptationism. *Annals of the New York Academy of Sciences*, 907, 62–74.
- Murray, C. (2003). *Human accomplishment: The pursuit of excellence in the arts and sciences, 800 BC to 1950*. New York, NY: Harper Collins.
- Neisser, U. (Ed.). (1997). *The rising curve. Long-term gains in IQ and related measures*. Washington DC: American Psychological Association.
- Reed, T. E., Vernon, P. A., & Johnson, A. M. (2004). Sex difference in brain nerve conduction velocity in normal humans. *Neuropsychologica*, 42, 1709–1714.
- Retherford, R. D., & Sewell, W. H. (1988). Intelligence and family size reconsidered. *Social Biology*, 35, 1–40.
- Rijsdijk, F. V., Vernon, P. A., & Boomsma, D. I. (1998). The genetic basis of the relation between speed-of-information-processing and IQ. *Behavioral Brain Research*, 95, 77–84.
- 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 & Excellence*, 1, 3–25.
- Rushton, J. P. (1998). The "Jensen effect" and the "Spearman-Jensen hypothesis" of Black-White IQ differences. *Intelligence*, 26, 217–225.
- Rushton, J. P., & Jensen, A. R. (2010). The rise and fall of the Flynn effect as a reason to expect the narrowing of the Black-White gap. *Intelligence*, 38, 213–219.
- Schmidt, F. L., & Hunter, J. E. (1999). Theory testing and measurement error. *Intelligence*, 27, 183–198.
- Seashore, R. H., Starmann, R., Kendall, W. E., & Helmick, J. S. (1941). Group factors in simple and discrimination reaction times. *Journal of Experimental Psychology*, 29, 346–394.
- Silverman, I. W. (2010). Simple reaction time: It is not what it used to be. *The American Journal of Psychology*, 123, 39–50.
- Skirbekk, V. (2008). Fertility trends by social status. *Demographic Research*, 18, 145–180.
- Smith, A., Sturgess, W., Rich, N., Brice, C., Collison, C., Bailey, J., et al. (1999). The effects of idazoxan on reaction times, eye movements and the mood of healthy volunteers and patients with upper respiratory tract illness. *Journal of Psychopharmacology*, 13, 148–151.
- Taimela, S. (1991). Factors affecting reaction-time testing and the interpretation of results. *Perceptual and Motor Skills*, 73, 1195–1202.
- Taimela, S., Kujala, U. M., & Osterman, K. (1991). The relation of low grade mental ability to fractures in young men. *International Orthopedics*, 15, 75–77.
- te Nijenhuis, J., Cho, S. H., Murphy, R., & Lee, K. H. (2012). The Flynn effect in Korea: Large gains. *Personality and Individual Differences*, 53, 147–151.
- te Nijenhuis, J., Murphy, R., & van Eeden, R. (2011). The Flynn effect in South Africa. *Intelligence*, 39, 465–467.
- te Nijenhuis, J., & van der Flier, H. (2013). Is the Flynn effect on g? A meta-analysis. *Intelligence*, 41, 802–807 (this issue).
- Thompson, H. B. (1903). *The mental traits of sex. An experimental investigation of the normal mind in men and women*. Chicago, IL: The University of Chicago Press.
- Thompson, S. G., & Sharp, S. J. (1999). Explaining heterogeneity in meta-analysis: A comparison of methods. *Statistics in Medicine*, 18, 2693–2708.
- van Court, M., & Bean, F. D. (1985). Intelligence and fertility in the United States: 1912–1982. *Intelligence*, 9, 23–32.
- Woodley, M. A. (2012). The social and scientific temporal correlates of genotypic intelligence and the Flynn effect. *Intelligence*, 40, 189–204.
- Woodley, M. A., & Figueredo, A. J. (2013). *Historical variability in heritable general intelligence: Its evolutionary origins and socio-cultural consequences*. Buckingham, UK: The University of Buckingham Press.
- Woodley, M. A., & Meisenberg, G. (2013). A Jensen effect on dysgenic fertility: An analysis involving the National Longitudinal Survey of Youth. *Personality and Individual Differences*, 55, 279–282.