By their words ye shall know them: Evidence of genetic selection against general intelligence and concurrent environmental enrichment in vocabulary usage since the mid 19th century

Michael A. Woodley of Menie[1,2]*, Heitor B. F. Fernandes[3], Aurelio José Figueredo[4] and Gerhard Meisenberg[5]

[1] Department of Psychology, Technische Universität Chemnitz, Chemnitz, Germany, [2] Center Leo Apostel for Interdisciplinary Studies, Vrije Universiteit Brussel, Brussels, Belgium, [3] Departments of Psychology and Genetics, Federal University of Rio Grande do Sul, Porto Alegre, Brazil, [4] Department of Psychology, University of Arizona, Tucson, AZ, USA, [5] Department of Biochemistry, Ross University School of Medicine, Portsmouth, Dominica

It has been theorized that declines in general intelligence (g) due to genetic selection stemming from the inverse association between completed fertility and IQ and the Flynn effect co-occur, with the effects of the latter being concentrated on less heritable non-g sources of intelligence variance. Evidence for this comes from the observation that 19th century populations were more intellectually productive, and also exhibited faster simple reaction times than modern ones, suggesting greater information-processing ability and therefore higher g. This co-occurrence model is tested via examination of historical changes in the utilization frequencies of words from the highly g-loaded WORDSUM test across 5.9 million texts spanning the period 1850–2005. Consistent with predictions, words with higher difficulties (δ parameters from Item Response Theory) and stronger negative correlations between pass rates and completed fertility declined in use over time whereas less difficult and less strongly selected words, increased in use over time, consistent with a Flynn effect stemming in part from the vocabulary enriching effects of increases in population literacy. These findings persisted when explicitly controlled for word age, changing literacy rates and temporal autocorrelation. These trends constitute compelling evidence for the co-occurrence model.

Keywords: co-occurrence model, intelligence, Flynn effect, WORDSUM, vocabulary

Introduction

Ever since Galton (1869) forecast declining intelligence on the basis of the shifting demographics of the Victorian population, there has been controversy about the future of human intelligence.

Early use of IQ testing seemed to confirm Galton's (1869) predictions, as most studies found that IQ was inversely related to fertility, suggesting directional genetic selection for lower intelligence (Lynn, 2011) – a trend that persists into the present (Lynn and van Court, 2004; Meisenberg, 2010; Reeve et al., 2013; Kanazawa, 2014).

In the West, up until the early to mid 19th century, those with high levels of socioeconomic status, wealth, and education (all of which are proxies for intelligence; Herrnstein and Murray, 1994) had higher numbers of surviving offspring relative to those with comparatively lower levels (Clark, 2007; Skirbekk, 2008), suggesting that higher intelligence may have conferred fitness advantages on individuals having to cope with extremes of cold, disease outbreaks, and conflict (Woodley and Figueredo, 2013). Subsequent increases in global temperature, coinciding with the end of the Little Ice Age in the mid-19th century, reduced environmental harshness, boosting agricultural yields thus reducing ecological stress and conflict (see: Zhang et al., 2007, 2011 for a demonstration of the inverse historical relationship between temperature and conflict). This would have substantially relaxed selection against those with lower intelligence (Woodley and Figueredo, 2013). This was coupled with advances in medicine (which would have included better means of fertility control, hygiene, nutrition, and medication; Lynn, 2011), and also social innovations such as welfare, mass schooling, and universal healthcare. The combined effect of these was a demographic transition characterized by general reductions in fertility, which were most pronounced among those with higher intelligence (Lynn, 2011). This was mediated primarily by fertility control coupled with the increasing prevalence of opportunities to delay fertility (i.e., higher education, increasing status competition, etc., which disproportionately attenuated the fertility of high-IQ women relative to men; Low et al., 2002; Meisenberg, 2010).

Child mortality was historically concentrated among those with low socioeconomic status (Geary, 2000) and impacted 50% of all children born in some European regions during the Renaissance – dropping to around 1% in the modern era, starting in the 19th century (Volk and Atkinson, 2008). Reductions in child mortality therefore further boosted the reproductive success of those with low IQ relative to those with high IQ. Additionally, historically high levels of child mortality would also have functioned as a source of purifying selection against *de novo* Single Nucleotide Polymorphisms (SNPs) and other deleterious mutations, which based on present rates of accumulation (i.e., 70 *de novo* SNPs per diploid genome per generation; Kong et al., 2012) should be associated with reproductive failure rates of a similar predicted magnitude to those that were actually observed historically in the child mortality data (i.e., 88%; Keightley, 2012).

The presence of directional selection favoring lower IQ coupled with increasing levels of mutation accumulation stemming from the breakdown of purifying selection should have reduced population-level IQ in the West since the 19th century. Consistent with this expectation, early intelligence researchers predicted that time-series studies of populations would reveal substantial generational declines in intelligence (Lentz, 1927; Cattell, 1937). Despite these predictions, the first such studies revealed that IQ scores had in fact *risen* across time (i.e., Cattell, 1950). This apparent contradiction was even termed "Cattell's Paradox" (Higgins et al., 1962) after psychometrician Raymond B. Cattell, who in the 1930s was prominent in predicting declining intelligence due to the lower fertility of those with higher intelligence (i.e., Cattell, 1937).

Debate as to the reality of massive secular gains in IQ scores was put to rest in the 1980s by Flynn (1984, 1987), who documented a steady rise in IQ across countries and across IQ batteries averaging three points per decade. The preponderance of the data collected subsequently has reinforced Flynn's finding (Trahan et al., 2014; Pietschnig and Voracek, 2015). Few now dispute the reality of this 'Flynn effect' (Herrnstein and Murray, 1994), however, there remains much debate as to its causes (Williams, 2013; Pietschnig and Voracek, 2015).

In the 1990s Lynn (2011) proposed a solution to Cattell's Paradox based on the idea that while "genotypic intelligence" (i.e., the theoretical level of intelligence resulting from the action of genes alone) has been declining due to genetic selection, these declines have been massively offset by gains in "phenotypic intelligence" (i.e., the intelligence that results from the interaction between a population's genes and improved environments). Loehlin (1997) illustrated this with the analogy of rising tides (representing the phenotypic-IQ-boosting effects of improving environments) lifting leaky boats (representing the much smaller losses expected on the basis of selection for lower IQ).

Anomalies Lead to New Findings

Lynn's model fails to account for certain observations. For example, if intelligence has been rising overall, why were 19th century Western populations much more innovative on a per capita basis across a wide range of fields (science, technology, mathematics, literature, and philosophy) and also much more generative of geniuses than modern populations (Murray, 2003; Huebner, 2005; Simonton, 2013), despite there being more individuals alive today with greater access to education, hygiene, high-quality nutrition, and other proposed elicitors of the Flynn effect (Williams, 2013)? This may seem like a counter-intuitive proposition, as there are many examples of substantial scientific progress in the modern era (i.e., in fields such as computing, genetics, materials science etc). Nonetheless, while breakthroughs are still occurring, they are occurring at a lower rate than was the case in the past, as are the geniuses responsible for them.

These indices of innovation and genius demonstrating per capita declines are based on the historiometric method of collating notable developments and individuals across many different encyclopedic reference works, first proposed by Galton (1869). There is a striking degree of agreement among different reference works as to precisely what counts as a major innovation, and who was responsible for it, which suggests that this is a robust method for estimating secular trends (Murray, 2003).

Simulated historical trends in "genotypic intelligence" (the population level intelligence change that would be expected due to selection alone, absent the Flynn effect) predict changing rates of innovation and genius (Woodley, 2012; Woodley and Figueredo, 2013). Clearly, there are factors other than intelligence influencing the innovativeness and creativity of populations, such as the presence or absence of key cultural factors and 'low-hanging fruit,' that is, novel innovations and discoveries that are easy to make (e.g., Horgan, 1997; Cowen, 2011), however, the findings of Woodley (2012) and Woodley and Figueredo (2013) suggest that losses in intelligence occurring *despite* the Flynn

The Co-occurrence Model: Decreases in *g* and the Flynn Effect Occur Simultaneously

The general intelligence factor, or *g* factor (Spearman, 1904), represents the component of intelligence which all tests of mental ability collectively share – it measures the ability to cope with cognitive complexity and is thus extremely important, as it is better at predicting individual differences in various life outcomes than are the relatively narrower cognitive abilities (Jensen, 1998).

The method of correlated vectors (MCV) compares the *g* loading (i.e., the correlation between a particular ability measure and the *g* factor) of different subtests of an IQ battery with the magnitudes of a set of associated effect sizes by correlating the vector of one with the vector of the other. This indicates the extent to which a given correlate of IQ is associated with *g* or narrower cognitive abilities (Jensen, 1998). When using MCV to examine whether the magnitude of genetic selection against IQ is related to the *g* loadings of subtests, positive correlations have been found (Woodley and Meisenberg, 2013a; Peach et al., 2014). When the same analysis is attempted for the Flynn effect, the effect size is *negatively* related to subtest *g* loadings (te Nijenhuis and van der Flier, 2013). The vectors of genetic and biological factors such as subtest heritabilities, inbreeding depression, reaction times, and indicators of mutation load typically correlate positively with *g* loadings in MCV (Prokosch et al., 2005; Rushton and Jensen, 2010; te Nijenhuis et al., 2014b), whereas the vectors of environmental effects, such as intelligence gains among adopted children, gains via educational interventions and gains via retesting typically correlate negatively with *g* loadings (te Nijenhuis et al., 2007, 2014a, 2015).

This suggests that the *highly heritable g* factor has been declining historically due to genetic selection and accumulating mutations (thus accounting for the apparent high intellectual productivity of 19th century populations relative to modern ones) whereas more *trainable* and *less heritable* specialized abilities exhibiting lower *g* loadings have been increasing in populations over time in response to educational and environmental improvements. Thus, Flynn effects and declines in *g* due to genetic selection and mutation accumulation *co-occur*, albeit *hierarchically* in that selection and mutation affect the top of the latent cognitive ability hierarchy [i.e., Carroll's (1993) Stratum III] whereas the Flynn effect is restricted to the narrower abilities and test specificities at the bottom of the hierarchy [i.e., Carroll's (1993) Stratum I; Woodley and Figueredo, 2013].

Evidence for this model has recently been found in data indicating that performance on tests of simple reaction time has been declining since the late 19th century (Silverman, 2010; Woodley et al., 2014a). Simple reaction time, as a measure related to information processing speed, is a culturally neutral biological marker of *g* (Jensen, 2006, 2011). Changing population averages may thus reflect the effects of genetic selection and mutation accumulation on *g*. The change in mean reaction time when scaled in terms of *g*-equivalents, suggests a decline of −1.21 points per decade in the US and UK when the study-means are controlled for various sources of between-study methods variance (Woodley et al.,

2014a). This observed loss is similar to the predicted loss derived from combining the results of a meta-analysis of 10 estimates of *g* decline computed on the basis of negative IQ × fertility correlations (−0.39 points per decade) with the additional decline derived from a study examining the effects of paternal age on off-spring *g* as a proxy for the generational effects of *de novo* SNPs on *g* (−0.84 points per decade; combined loss = −1.23 points per decade; Woodley of Menie, 2015).

The co-occurrence model has also been tested using cohorts from the Netherlands born between 1950 and 1990. Using MCV on a sample of 63 subtests, it was found that the decline magnitudes among subtests showing *anti-* or "reverse" Flynn effects (i.e., IQ losses presumably due to selection and mutation) were borderline significantly and positively related to their *g* loadings, whereas those showing Flynn effects were not. The average *g* loading of the subtests exhibiting anti-Flynn effects was furthermore higher than that of the subtests exhibiting Flynn effects, consistent with predictions from the co-occurrence model (Woodley and Meisenberg, 2013b).

A New Test of the Co-occurrence Model

Here a novel test of the co-occurrence model is presented involving an examination of the temporal prevalence of vocabulary words across English-language texts over the past century and a half. Scores on tests of vocabulary, although themselves assessing a specific ability, are typically among the most *g*-loaded measures of IQ and are also among the most heritable (Kan et al., 2013). Nevertheless there is considerable heterogeneity in the difficulties of the vocabulary items that comprise these scales (Beaujean and Sheng, 2010), thus there is scope for testing predictions derived from the co-occurrence model using vocabulary measures at the item level.

An excellent vocabulary measure is WORDSUM (developed by Thorndike, 1942) from the General Social Survey (GSS), which has been administered to 'household' samples of the American public on a regular basis since 1974. WORDSUM involves showing the respondent a card containing 10 target words. They must find the synonymous term or phrase among five alternatives.

Wolfle (1980) found that WORDUM performance correlated at 0.71 with full-scale IQ. When this correlation is corrected for the reliability of WORDSUM (0.73; Hout and Hastings, 2012), and psychometric validity (0.90; Jensen, 1998), it rises to 0.93, indicating a very high *g* loading, as is typical of vocabulary measures (Kan et al., 2013).

Attempts at determining whether there is any kind of a secular trend toward changing overall performance on this measure in the GSS have yielded inconsistent results (Haung and Hauser, 1998; Beaujean and Sheng, 2010; Flynn, 2012). Studies by Bowles et al. (2005) and Cor et al. (2012) found that WORDSUM words could be grouped into two classes based on difficulty. Both groups of researchers also found that earlier-born cohorts exhibited higher difficult-vocabulary knowledge relative to more recently-born ones, suggesting declining performance with respect to difficult words. This is consistent with the co-occurrence model, as it is performance on the most difficult (and therefore most *g*-loaded) words that is declining. These trends would furthermore be congruent with the presence of persistent negative associations

between IQ and fertility on WORDSUM vocabulary knowledge, which have been found in GSS birth cohorts dating back to 1880–1899 (van Court and Bean, 1985; Lynn and van Court, 2004).

Studies involving large lexical databases, such as Haung and Hauser (1998), who attempted to determine changes in the frequencies of WORDSUM target words across a 1 million word database, and Roivainen (2014), who used Google Ngram Viewer (Michel et al., 2011) to track the frequencies of WORDSUM, WAIS, WAIS-R, WISC, and WISC-R vocabulary words, have also found evidence for declining usage frequencies, especially among more difficult words. Here, the degree to which WORDSUM item-level difficulties, the negative correlation between item pass rates and fertility (a measure of the strength of genetic selection associated with each item), and changing levels of population literacy predict changes in word prevalence across texts from the mid 19th century to the present, is investigated, while controlling for various factors. Based on the co-occurence model, it is predicted that more difficult words should be declining in usage over time and that this decline should in part be predicted by genetic selection. Conversely, there should be a Flynn effect on easier words owing in part to the effects of increasing population level literacy enriching people's vocabularies coupled with increasing demand for literature containing less cognitively demanding words.

In the present study Google Ngram Viewer is employed in tracking WORDSUM word frequencies. The Ngram viewer provides a database of more than 5 million texts (newspapers, works of fiction, non-fiction, technical works, etc.), comprising more than 500 billion words that can be searched using the target WORDSUM words, thus revealing their year-on-year frequencies. The database has considerable reach in time also – spanning from 1500 to nearly the present.

One major advantage of examining the prevalence of WORDSUM words across texts is that it can be reasonably assumed that the authors of the texts were using these words correctly – hence appearance in print is tantamount to the authors effectively 'passing' that item in WORDSUM. This is potentially important as scores on psychometric tests with multiple-choice-type answer formats are known to be inflated by factors relating to test wiseness such as guessing (Brand, 1987; Must and Must, 2013). It has been found that people are more likely to utilize guessing on more *g*-loaded measures of ability (Woodley et al., 2014b). Secular gains due purely to increased guessing therefore potentially weaken the capacity for psychometric tests to directly detect declines on *g* due to genetic selection and mutation accumulation as the effects are occurring on the same variance component (Woodley et al., 2014b). Tracking word usage trends across a representative corpus of written texts therefore yields potentially more *ecologically-valid* data on secular trends, as guessing and other factors associated with the 'artificiality' of the testing environment cannot be influencing these trends.

Materials and Methods

Google Ngram Viewer (Michel et al., 2011) generates estimates of the year-on-year prevalence of the 10 WORDSUM target words across a very large sample of English-language texts scanned by Google. All years between 1850 and 2005 were considered. 1850 was used as the lower cutoff year because in the majority of Western countries fertility began to negatively relate to IQ proxies such as socioeconomic status and education by the middle of the 19th century (Skirbekk, 2008). Ngram Viewer's algorithm normalizes the frequency of words by the number of books published in each year; therefore the yearly word frequencies are not biased due to increasing amounts of text. All searches were case insensitive (permitting both upper and lower case variants of the same word to be included in the search), and were restricted to the lexical category, grammatical conjugation, and grammatical number specified in WORDSUM, thus Ngram was only searched for the specific WORDSUM target word, rather than variants.

Item difficulties (δ) for the 10 target words were obtained from Beaujean and Sheng (2010), calculated using item response theory (IRT) on individual-level GSS-respondent data totaling 25,555 participants. This is a measure of how difficult each word was to "pass" on the test, and roughly reflects the relative proportion of the calibration sample that got each word correct.

The correlations between fertility and item-level pass rates (henceforth *r*FPR) were also computed for each word among the white subset of the GSS that had completed fertility for males and females separately and then averaged (i.e., the subset aged 41 and older; $N = 4,252$ for females, $N = 3,542$ for males, total $N = 7,794$). Using the white subsample ensured that the fertility patterns among the demographic that historically contributed disproportionately to literature were captured (Murray, 2003). Each *r*FPR value was disattenuated for unreliability by dividing it by the square-root of the item-level reliabilities reported in Hout and Hastings (2012, Figure 2C, p. 12; these were obtained visually from the figure). All of these coefficients are *negative* in direction indicating that *r*FPR was inversely related to number of children. Furthermore, MCV demonstrates that *r*FPR is *inversely proportional* to the difficulty of each word ($r = -0.71$, $p < 0.05$, $N = 10$), indicating that the *magnitude* of item-level *r*FPR is *positively* related to item difficulty and that genetic selection is strongest on more *g*-loaded measures, replicating previous findings (Woodley and Meisenberg, 2013a; Peach et al., 2014).

A potentially confounding factor influencing word frequencies is word age. Curzan (2009) has argued that older words are likely to be better known to people who read as changes in written language lag behind changes in spoken language. An implication of this is that older words will therefore have higher usage frequencies than younger words simply because they are more established in text, which could influence year-on-year usage frequencies independently of cognitive or other changes. Thus word age is a potentially important factor to control. This was operationalized using data from Merriam-Webster (2010), the *Random House Unabridged Dictionary* (Flexner, 2003), and the *Online Etymology Dictionary* (Harper, 2010) on the year in which the word first appeared in its modern-usage context in the English language (see **Table 1**). Words for which only the decade of first use was available instead of the precise year were assigned the median year of the respective decade.

Another important contributing factor is literacy. As more people become literate, literature moves out of the domain of

TABLE 1 | WORDSUM item data used in the analyses.

WORDSUM (Word code)	Item difficulty(δ)	Item fertility pass rate correlation (rFPR)*	Item-level reliability	Word debut (Year of first use in the Anglo phone literature)
A	−1.17	−0.039	0.60	1275
B	−1.27	−0.038	1.00[†]	1726
C	1.26	−0.062	0.90	1742
D	−1.13	−0.022	0.75	1602
E	−0.67	−0.082	0.97	1492
F	−0.83	−0.052	0.57	1387
G	0.74	−0.058	0.63	1287
H	0.82	−0.079	0.88	1666
I	−0.92	−0.031	0.74	1437
J	1.04	−0.100	0.76	1546

*All rFPR correlations are statistically significant at the $p \leq 0.05$ level for an N of 7,794. [†]Hout and Hastings (2012) assign an item-level reliability of >1 to word B. This may have resulted from model overfitting therefore the word is assigned a perfect reliability (necessitating no correction) instead. In order to safeguard future General Social Survey (GSS) waves utilizing WORDSUM, the target words are kept confidential. They are therefore represented here with letter codes corresponding to the order in which the items are administered.

a literate elite, into a mass-market dominated by newly literate people of more modest intelligence. Therefore increasing literacy might drive down the demand for cognitively demanding words in addition to enriching people's vocabularies with respect to less difficult words (consistent with this expectation Must et al. (2003) found that secular literacy gains are more pronounced on less cognitively demanding measures of literacy). Changing literacy was operationalized using a year-on-year average measure of both male and female written literacy rates sourced from the British Library (2001). These data cover the period from 1850 to 1900, when literacy rates reach approximately 99% for both sexes, and apply to the UK, however, historical literacy trends are broadly paralleled across Western countries (Vanhanen, 2009). Written literacy likely also overestimates the proportion of illiterates, as more people historically attained reading literacy than written literacy (British Library, 2001), therefore it constitutes a conservative measure. Given that near maximal literacy (99%) was attained by 1900, this value was assigned to all subsequent years.

As serial measurements with the same experimental subject or unit of analysis (such as the time series of word frequencies) are likely to be auto-correlated, analyses with repeated measurements must account for this. The method of Multi-Level Modeling (MLM) was employed.

The Restricted Maximum Likelihood (REML) procedure was used to estimate the variance and covariance parameters, as this is recommended for fitting mixed models due to nuisance parameters (such as negative estimates of variances that are very close to zero) having no effect upon the estimation (Corbeil and Searle, 1976). The significance tests were conducted with Hierarchical Type I Sum of Squares so that the effect of each predictor was residualized against the effects of prior predictors, according to the theoretically specified order (see **Table 2**). This permitted the effects of time with word difficulty and with rFPR upon word utilization frequency to be tested after accounting for their component main effects, as well as after statistically controlling for word age, population literacy and their interactions with other predictors, in addition to within-subject temporal autocorrelations.

TABLE 2 | Hierarchical Type I Sum of Squares tests of fixed effects[a] upon word usage frequencies in the literature.

Predictor	Semipartial r
Word debut year	−0.481*
Literacy	0.020*
Ln(Time)	0.049*
Item difficulty(δ)	−0.186*
rFPR	0.006*
Literacy × Item difficulty(δ)	−0.062*
rFPR × Item difficulty(δ)	−0.105*
Ln(Time) × Item difficulty(δ)	−0.066*
Ln(Time) × rFPR	0.010*
Ln(Time) × rFPR × Item difficulty(δ)	0.032*

δ, IRT-based word difficulty; *$p < 0.01$. [a]Intercepts of the words were allowed random effects. Effects were estimated with restricted maximum likelihood algorithms, under the assumptions of compound symmetry, first-order autoregressive, and unstructured covariance structures.

To correct estimates for within-subject temporal autocorrelation several covariance structures have been developed using *random-effects parameters* – additional unknown random variables assumed to affect the variability of the data. Jennrich and Schluchter (1986), proposed a general class of mixed linear models with structured and unstructured (*UN*) within-subject covariances, which can control for the possible effects of temporal autocorrelation. As there was no *a priori* hypothesis regarding the expected covariance structures among these temporal autocorrelations, the model was estimated under the assumption of an *UN* covariance matrix among the residuals, which constitutes the least restrictive set of assumptions that may be applied. Using *UN*, different parameters are estimated for the variance of each variable as well as different covariance parameters for each repeated measurement sequence (Jennrich and Schluchter, 1986). The nil χ^2 and associated degrees of freedom were due to the extremely small magnitudes of these residual covariances (Residual$_{UN}$ = 8.117×10^{-8}) and the "saturated" *UN* model of

the residual covariance matrix, respectively. Thus, as compared with an equivalent General Linear Model (*GLM*) using Ordinary Least Squares (*OLS*), the parameter estimates for the effect sizes were within rounding error of each other.

In this MLM, the criterion was the absolute yearly frequency of word usage from 1850 to 2005 (operationalized as the percentage represented by each word of the total number of words then employed), and predictors were first year of word usage (*Word Debut*), the annual literacy rates of the UK population from 1850 to 2005, the natural logarithm (Ln) of time (to model the expected curvilinearity using the minimal model degrees of freedom), the item difficulties (δ) for each word, the rFPR for each word, the two-way interaction between population literacy and δ, the two-way interaction between rFPR and δ, the two-way interaction between δ and LN(time), the two-way interaction between rFPR and LN(time) and the three-way interaction between δ, rFPR and LN(time). Also estimated were the possible effects of any residual differences between the usages of words that were *not* attributable to δ by testing their main effects (operationalized as CLASS variables) and their temporal interactions at the end of each of the hierarchical MLMs constructed.

Due to small time intervals between measurements (1 year), parameter estimates for both time predictors and for the predictors involving the interaction of time with difficulty were very small. To compensate for this the estimates were recalculated by dividing all the years by 1000, changing the metric to occurrences per millennia. The advantage of this simple linear transformation (by which semipartial correlations and associated significance tests are not affected) is that the numerical magnitudes of the parameter estimates are visually increased, making the results more legible. All MLMs were conducted using PROC MIXED in SAS 9.3, and semipartial correlation coefficients were estimated using a beta version of UniMult 2 (for documentation on UniMult 1, see Gorsuch, 1991).

Results

Table 2 displays the semipartial correlations between word usage frequencies over time accounted for with each predictor, as assigned by hierarchical partitioning over variance (Type I Sums of Squares); all of these effect sizes were statistically significant at $p < 0.001$. As seen in **Figure 1**, words that are more difficult and for which the negative association between fertility and pass-rate is stronger present a more persistently negative trend over time. Easier words by contrast present a more persistently positive trend over time.

Discussion

More difficult words presented *sharper* historical declines in usage over the 1850–2005 period, as indicated by the *negative* effect of the interaction between δ and LN(time). In line with this, the *positive* effect of the interaction between rFPR for each word and LN(time) indicates that words for which pass rates are *more negatively* associated with fertility are decreasing in usage over time, which is consistent with those being the *more difficult* words, as indicated by the MCV results presented above. Also consistent with predictions was the *positive* three-way interaction among each word's δ, rFPR, and LN(time), which indicates that words that are both *more difficult* and for which pass rates are *more negatively* associated with fertility are *decreasing* over time. This suggests that δ had a *more negative* impact on usage of more difficult words over this historical period in cases where pass rates

FIGURE 1 | Temporal trends of the 10 WORDSUM words based on the fitted mixed models. *Solid lines* represent more difficult and *dashed lines* less difficult words. *Black lines* represent words that are more associated and *gray lines* less associated with rFPR (strength of association was based on a median split). WORDSUM codes are displayed to the right.
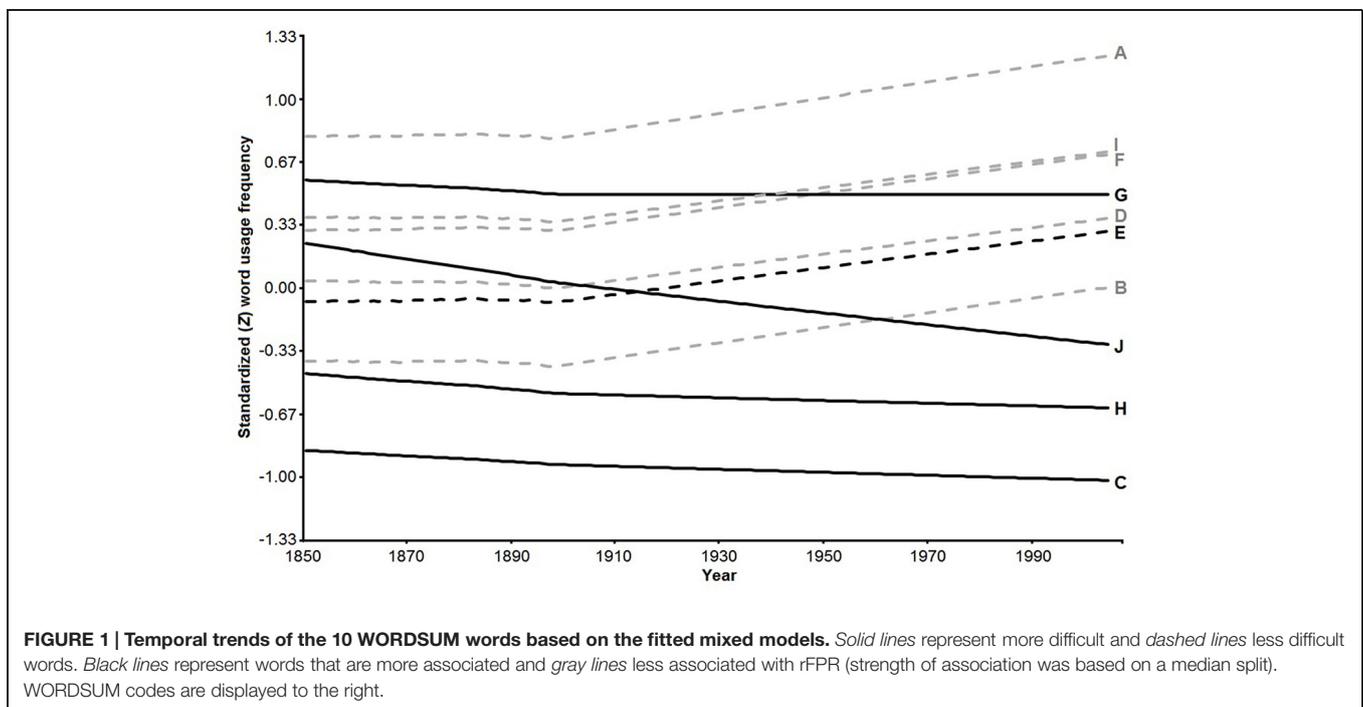
are *more negatively* associated with fertility. Therefore the *positive* two-way interaction between *r*FPR for each word and LN(time) applies primarily to the *more difficult* words. Generally speaking, words that are *more difficult* and for which pass rates are *more negatively* associated with fertility present a *sharper* historical decline in usage over the period under scrutiny; words that are *less difficult* and that exhibit a weaker association between fertility and pass rates present a *shallower* historical decline in usage over the same period. Also importantly, the *positive* main effect of time indicates a secular trend toward rising usage of all words over time, irrespective of population literacy. Additionally, the interactions discussed above clearly indicate that the less difficult words are increasing in usage over time, consistent with a Flynn effect, as can be seen in **Figure 1**.

The negative main effect of the δ parameter itself on word usage frequency simply indicates that more difficult words are generally used less frequently. Accordingly, the *small but positive* effect of *r*FPR on each word upon word usage indicates that words for which pass rates are *more negatively* associated with fertility are being used *less frequently* at the historical baseline than others. The *negative* effect of the residual interaction between δ and *r*FPR for each word upon word usage indicates counterintuitively that item difficulty had a *more negative* impact on overall word usage (i.e., independent of passing time) in words for which *r*FPR is weaker, unlike the time trends and main effects identified.

The *small but positive* main effect of literacy is the general baseline for the usage of all words in the sample, which *increased* as literacy increased, consistent with a Flynn effect. This effect is likely due to the general trend toward universal literacy (reaching an asymptote at 99% around 1900) amongst the reading public and concomitant vocabulary enrichment, especially with respect to less cognitively demanding words. The *negative* effect of the literacy $\times$ δ interaction indicates that more difficult words tended to *decrease* in usage with increasing literacy. This is likely also a consequence of the trend toward universal literacy, which presumably involved a higher proportion of persons of lower ability gaining access to literacy over time and demanding less difficult words in literature.

A large proportion of variance was attributable to the recency of each word's introduction into the language, which was simply a control variable to statistically adjust for the effect of word age, but was not theoretically relevant to the hypotheses. The *negative* main effect of word debut year indicates that the *more recently* introduced the word, the *lower* the word usage, which is consistent with Curzan's (2009) observations.

Conclusion

These findings provide compelling evidence for the co-occurrence model, adding to the nomological net from which these and other predictions have been derived (Woodley et al., 2014a). They furthermore add weight to the argument that variance components should be taken seriously when considering the pattern of secular trends among indicators of IQ. The presence of oppositely directed trends on different WORDSUM items might even account for the results of studies finding no consistency in secular trends at the level of full-scale WORDSUM scores in GSS waves dating back to the 1970s (Beaujean and Sheng, 2010). These results replicate the findings of other studies on WORDSUM indicating performance declines on harder words, coupled with null-trends or improvements on less difficult ones, both at the level of the GSS survey waves (Bowles et al., 2005; Cor et al., 2012) and also among lexical databases (Haung and Hauser, 1998; Roivainen, 2014). The findings also build on this previous research by showing direct contributions to the decline in frequency among difficult words stemming from genetic selection.

It is predicted by the co-occurrence model that specific cognitive abilities should be increasing over the same period of time as *g* is declining. There are good theoretical reasons to expect that human social evolution should favor the role of cognitive specialization among individuals into different socioecological micro-niches, thus reducing intraspecific competition and enhancing the productivity of the social group by the economic mechanisms of Ricardo's Law of Comparative Advantage (Woodley et al., 2013; Cabeza de Baca and Figueredo, 2014). Consistent with this is the finding of a direct contribution to the increase in frequency among less difficult words stemming from literacy, a theorized elicitor of the Flynn effect (i.e., Must et al., 2003; Marks, 2010). This is suggestive of increasing specialization with respect to the non-*g* ability variance associated with vocabulary.

It must finally be noted, however, that only the effect of genetic selection on word usage frequency over time was modeled here. Generational mutation accumulation is another factor likely influencing population *g* and potentially therefore word-usage frequency (Woodley of Menie, 2015). As mutation accumulation was not directly modeled, the effects of selection on word usage frequency likely underestimate the full effect of genetic changes taking place within Western populations over the time period covered in the present analysis.

References

Beaujean, A. A., and Sheng, Y. (2010). Examining the Flynn effect in the general social survey vocabulary test using item response theory. *Pers. Individ. Dif.* 48, 294–298. doi: 10.1016/j.paid.2009.10.019

Bowles, R. P., Grimm, K. J., and McArdle, J. J. (2005). A structural factor analysis of vocabulary knowledge and relations to age. *J. Gerontol. B Psychol. Sci. Soc. Sci.* 60, 234–241. doi: 10.1093/geronb/60.5.P234

Brand, C. R. (1987). Intelligence testing: bryter still and bryter? *Nature* 328, 110. doi: 10.1038/328110a0

British Library (2001). *Aspects of the Victorian Book: Introduction*. Available at: http://www.bl.uk/collections/early/victorian/pr_intro.html, retrieved on 10/01/2015.

Cabeza de Baca, T., and Figueredo, A. J. (2014). The cognitive ecology of Mexico: climatic and socio-cultural effects on life history strategies and cognitive abilities. *Intelligence* 47, 63–71. doi: 10.1016/j.intell.2014.08.007

Carroll, J. B. (1993). *Human Cognitive Abilities: A Survey of Factor-Analytic Studies.* Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511571312

Cattell, R. B. (1937). *The Fight for Our National Intelligence.* London: P.S. King & Son.

Cattell, R. B. (1950). The fate of national intelligence: test of a thirteen-year prediction. *Eugen. Rev.* 42, 136–148. doi: 10.1016/j.intell.2014.03.003

Clark, G. (2007). *A Farewell to Alms: A Brief Economic History of the World.* Princeton, NJ: Princeton University Press.

Cor, M. K., Haertel, E., Krosnick, J. A., and Malhorta, N. (2012). Improving ability measurement in surveys by following the principles of irt: the wordsum vocabulary test in the general social survey. *Soc. Sci. Res.* 41, 1003–1016. doi: 10.1016/j.ssresearch.2012.05.007

Corbeil, R. R., and Searle, S. R. (1976). A comparison of variance component estimators. *Biometrics* 71, 779–791. doi: 10.2307/2529264

Cowen, T. (2011). *The Great Stagnation: How America Ate All the Low-Hanging Fruit of Modern History, Got Sick, and Will (Eventually) Feel Better.* New York: A Penguin eSpecial from Dutton.

Curzan, A. (2009). "Historical corpus linguistics and evidence of language change," in *Corpus Linguistics,* eds A. Luedeling and M. Kytö (Berlin: Gruyter), 1091–1108.

Flexner, S. B. (Ed.). (2003). *Random House Webster's Unabridged Dictionary,* 2nd Edn. New York, NY: Random House.

Flynn, J. R. (1984). The mean IQ of Americans: massive gains 1932 to 1978. *Psychol. Bull.* 95, 29–51. doi: 10.1037/0033-2909.95.1.29

Flynn, J. R. (1987). Massive IQ gains in 14 nations: what IQ tests really measure. *Psychol. Bull.* 101, 171–191. doi: 10.1037/0033-2909.101.2.171

Flynn, J. R. (2012). *Are we Getting Smarter? Rising IQ in the Twenty-First Century.* Cambridge: Cambridge University Press.

Galton, F. (1869). *Hereditary Genius.* London: MacMillan. doi: 10.1037/13474-000

Geary, D. C. (2000). Evolution and proximate expression of human paternal investment. *Psychol. Bull.* 126, 55–77. doi: 10.1037/0033-2909.126.1.55

Gorsuch, R. L. (1991). *Unimult: for Univariate and Multivariate Data Analysis.* Altadena, CA: UniMult.

Harper, D. (2010). *Online Etymology Dictionary.* Available at: http://www.etymonline.com, retrieved on 10/01/2015.

Haung, M.-H., and Hauser, R. M. (1998). "Trends in black-white test-score differentials: II. The WORDSUM vocabulary test," in *The Rising Curve: Long-Term Gains in IQ and Related Measures,* ed. U. Neisser (New York, NY: The APA Press), 303–334.

Herrnstein, R. J., and Murray, C. (1994). *The Bell Curve: Intelligence and Class Structure in American Life.* New York, NY: Free Press Paperbacks.

Higgins, J. V., Reed, E. W., and Reed, S. C. (1962). Intelligence and family size: a paradox resolved. *Eugen. Q.* 9, 84–90. doi: 10.1080/19485565.1962.9987508

Horgan, J. (1997). *The End of Science: Facing the Limits Of Knowledge in the Twilight of the Scientific Age.* New York, NY: Broadway Book.

Hout, M., and Hastings, O. P. (2012). Reliability and Stability Estimates for the Gss Core Items From the Three-Wave Panels, 2006–2010. GSS Methodological Report 119, Chicago: NORC.

Huebner, J. (2005). A possible declining trend for worldwide innovation. *Technol. Forecast. Soc. Change* 72, 980–986. doi: 10.1016/j.techfore.2005.01.003

Jennrich, R. I., and Schluchter, M. D. (1986). Unbalanced repeated-measures models with structured covariance matrices. *Biometrics* 42, 805–820. doi: 10.2307/2530695

Jensen, A. R. (1998). *The G Factor: The Science of Mental Ability.* Westport, CT: Praeger Publishers.

Jensen, A. R. (2006). *Clocking the Mind: Mental Chronometry and Individual Differences.* Oxford: Elsevier.

Jensen, A. R. (2011). The theory of intelligence and its measurement. *Intelligence* 39, 171–177. doi: 10.1016/j.intell.2011.03.004

Kan, K. J., Wicherts, J. M., Dolan, C. V., and van der Maas, H. L. J. (2013). On the nature and nurture of intelligence and specific cognitive abilities: the more heritable, the more culture dependent. *Psychol. Sci.* 24, 2420–2428. doi: 10.1177/0956797613493292

Kanazawa, S. (2014). Intelligence and childlessness. *Soc. Sci. Res.* 48, 157–170. doi: 10.1016/j.ssresearch.2014.06.003

Keightley, P. D. (2012). Rates and fitness consequences of new mutations in humans. *Genetics* 190, 295–304. doi: 10.1534/genetics.111.134668

Kong, A., Frigge, M. L., Masson, G., Besenbacher, S., Sulem, P., Magnusson, G., et al. (2012). Rate of de novo mutations and the importance of father's age to disease risk. *Nature* 488, 471–475. doi: 10.1038/nature11396

Lentz, T. (1927). Relation of I.Q to size of family. *J. Educ. Psychol.* 18, 486–496. doi: 10.1037/h0072940

Loehlin, J. C. (1997). Dysgenesis and IQ: what evidence is relevant? *Am. Psychol.* 52, 1236–1239. doi: 10.1037/0003-066X.52.11.1236

Low, B., Simon, C. P., and Anderson, K. G. (2002). An evolutionary ecological perspective on demographic transitions: modeling multiple currencies. *Am. J. Hum. Biol.* 14, 149–167. doi: 10.1002/ajhb.10043

Lynn, R. (2011). *Dysgenics: Genetic Deterioration in Modern Populations,* Revised Edn. London: Ulster Institute for Social Research.

Lynn, R., and van Court, M. (2004). New evidence of dysgenic fertility for intelligence in the United States. *Intelligence* 32, 193–201. doi: 10.1016/j.intell.2003.09.002

Marks, D. F. (2010). IQ variations across time, race, and nationality: an artifact of differences in literacy skills. *Psychol. Rep.* 106, 643–664. doi: 10.2466/pr0.106.3.643-664

Meisenberg, G. (2010). The reproduction of intelligence. *Intelligence* 38, 220–230. doi: 10.1016/j.intell.2010.01.003

Merriam-Webster. (2010). *The Merriam Webster Dictionary.* New Delhi: Goyal.

Michel, J.-B., Shen, K. Y., Aiden, A. P., Veres, A., Gray, M. K., The Google BooksTeam., et al. (2011). Quantitative analysis of culture using millions of digitized books. *Science* 331, 176–182. doi: 10.1126/science.1199644

Murray, C. (2003). *Human Accomplishment: The Pursuit of Excellence in the Arts and Sciences, 800 BC to 1950.* New York, NY: Harper Collins.

Must, O., and Must, A. (2013). Changes in test-taking patterns over time. *Intelligence* 41, 791–801. doi: 10.1016/j.intell.2013.04.005

Must, O., Must, A., and Raudik, V. (2003). The Flynn effect for gains in literacy found in Estonia is not a Jensen effect. *Pers. Individ. Dif.* 34, 1287–1292. doi: 10.1016/S0191-8869(02)00115-0

Peach, H., Lyerly, J. E., and Reeve, C. L. (2014). Replication of the Jensen effect on dysgenic fertility: an analysis using a large sample of American youth. *Pers. Individ. Dif.* 71, 56–59. doi: 10.1016/j.paid.2014.07.017

Pietschnig, J., and Voracek, M. (2015). One century of global iq gains: a formalmeta-analysis of the flynn effect (1909–2013). *Perspect. Psychol. Sci.* (in press).

Prokosch, M. D., Yeo, R. A., and Miller, G. F. (2005). Intelligence tests with higher g loadings show higher correlations with body symmetry: evidence for a general fitness factor mediated by developmental stability. *Intelligence* 33, 203–213. doi: 10.1016/j.intell.2004.07.007

Reeve, C. L., Lyerly, J. E., and Peach, H. (2013). Adolescent intelligence and socioeconomic wealth independently predict adult marital and reproductive behavior. *Intelligence* 41, 358–365. doi: 10.1016/j.intell.2013.05.010

Roivainen, E. (2014). Changes in word usage frequency may hamper comparisons of vocabulary skills: an Ngram analysis of Wordsum, WAIS and WISC test items. *J. Psychoeduc. Assess.* 32, 83–87. doi: 10.1177/0734282913485542

Rushton, J. P., and Jensen, A. R. (2010). The rise and fall of the Flynn effect as a reason to expect a narrowing of the Black–White IQ gap. *Intelligence* 38, 213–219. doi: 10.1016/j.intell.2009.12.002

Silverman, I. W. (2010). Simple reaction time: it is not what it used to be. *Am. J. Psychol.* 123, 39–50. doi: 10.5406/amerjpsyc.123.1.0039

Simonton, D. K. (2013). After Einstein: scientific genius is extinct. *Nature* 493, 602. doi: 10.1038/493602a

Skirbekk, V. (2008). Fertility trends by social status. *Demogr. Res.* 18, 145–180. doi: 10.4054/DemRes.2008.18.5

Spearman, C. (1904). "General intelligence," objectively determined and measured. *Am. J. Psychol.* 26, 51–93. doi: 10.2307/1412107

te Nijenhuis, J., Jongeneel-Grimen, B., and Armstrong, E. L. (2015). Are adoption gains on the g factor? A meta-analysis. *Pers. Individ. Dif.* 73, 56–60. doi: 10.1016/j.paid.2014.09.022

te Nijenhuis, J., Jongeneel-Grimen, B., and Kirkegaard, E. O. W. (2014a). Are Headstart gains on the g factor? A meta-analysis. *Intelligence* 46, 209–215. doi: 10.1016/j.intell.2014.07.001

te Nijenhuis, J., Kura, K., and Hur, Y. M. (2014b). The correlation between g loadings and heritability in Japan: a meta-analysis. *Intelligence* 46, 275–282. doi: 10.1016/j.intell.2014.07.008

te Nijenhuis, J., and van der Flier, H. (2013). Is the Flynn effect on g? A meta-analysis. *Intelligence* 41, 802–807. doi: 10.1016/j.intell.2013.03.001

te Nijenhuis, J., van Vianen, A. E. M., and van der Flier, H. (2007). Score gains on g loaded tests: no g. *Intelligence* 35, 283–300. doi: 10.1016/j.intell.2006.07.006

Thorndike, R. L. (1942). Two screening tests of verbal intelligence. *J. Appl. Psychol.* 26, 128–135. doi: 10.1037/h0060053

Trahan, L. H., Stuebing, K. K., Fletcher, J. M., and Hiscock, M. (2014). The Flynn effect: a meta-analysis. *Psychol. Bull.* 140, 1332–1360. doi: 10.1037/a0037173

van Court, M., and Bean, F. D. (1985). Intelligence and fertility in the United States: 1912–1982. *Intelligence* 9, 23–32. doi: 10.1016/0160-2896(85)90004-2

Vanhanen, T. (2009). *Index of Power Resources (IPR) 2007 [Computer File]. FSD2420, Version 1.0 (2009-05-14)*. Tampere: Finnish Social Science Data Archive.

Volk, A. A., and Atkinson, J. (2008). Is child death the crucible of evolution? *J. Soc. Evol. Cult. Psychol.* 2, 247–260. doi: 10.1037/h0099341

Williams, R. L. (2013). Overview of the Flynn effect. *Intelligence* 41, 753–764. doi: 10.1016/j.intell.2013.04.010

Wolfle, L. M. (1980). The enduring effects of education on verbal skills. *Sociol. Educ.* 53, 104–114. doi: 10.2307/2112492

Woodley, M. A. (2012). The social and scientific temporal correlates of genotypic intelligence and the Flynn effect. *Intelligence* 40, 189–204. doi: 10.1016/j.intell.2011.12.002

Woodley, M. A., and Figueredo, A. J. (2013). *Historical Variability in Heritable General Intelligence: It's Evolutionary Origins and Socio-Cultural Consequences*. Buckingham: The University of Buckingham Press.

Woodley, M. A., Figueredo, A. J., Ross, K. C., and Brown, S. D. (2013). Four successful tests of the cognitive differentiation-integration effort hypothesis. *Intelligence* 41, 832–842. doi: 10.1016/j.intell.2013.02.002

Woodley, M. A., and Meisenberg, G. (2013a). A Jensen effect on dysgenic fertility: an analysis involving the National Longitudinal Survey of Youth. *Pers. Individ. Dif.* 55, 279–282. doi: 10.1016/j.paid.2012.05.024

Woodley, M. A., and Meisenberg, G. (2013b). In the Netherlands the anti-Flynn effect is a Jensen effect. *Pers. Individ. Dif.* 54, 871–876. doi: 10.1016/j.paid.2012.12.022

Woodley, M. A., te Nijenhuis, J., and Murphy, R. (2014a). Is there a dysgenic secular trend towards slowing simple reaction time? Responding to a quartet of critical commentaries. *Intelligence* 46, 131–147. doi: 10.1016/j.intell.2014.05.012

Woodley, M. A., te Nijenhuis, J., Must, O., and Must, A. (2014b). Controlling for increased guessing enhances the independence of the Flynn effect from g: the return of the Brand effect. *Intelligence* 43, 27–34. doi: 10.1016/j.intell.2013.12.004

Woodley of Menie, M. A. (2015). How fragile is our intellect? Estimating losses in general intelligence due to both selection and mutation accumulation. *Pers. Individ. Dif.* 75, 80–84. doi: 10.1016/j.paid.2014.10.047

Zhang, D. D., Brecke, P., Lee, H. F., He, Y.-Q., and Zhang, J. (2007). Global climate change, war, and population decline in recent human history. *Proc. Natl. Acad. Sci. U.S.A.* 104, 19214–19219. doi: 10.1073/pnas.0703073104

Zhang, D. D., Lee, H. F., Wang, C., Li, B., Pei, Q., Zhang, J., et al. (2011). The causality analysis of climate change and large-scale human crisis. *Proc. Natl. Acad. Sci. U.S.A.* 108, 17296–17301. doi: 10.1073/pnas.1104268108