



A systematic review of the state of literature relating parental general cognitive ability and number of offspring



Charlie L. Reeve^{a,*}, Michael D. Heeney^a, Michael A. Woodley of Menie^{b,c}

^a Department of Psychological Science, The University of North Carolina at Charlotte, USA

^b Department of Psychology, Technische Universität Chemnitz, Germany

^c Center Leo Apostel at the Vrije Universiteit Brussel, Belgium

ARTICLE INFO

Keywords:

Intelligence
Cognitive ability
'g'
Reproductive success
Meta-analysis
Dysgenic trend

ABSTRACT

The purpose of this study is to conduct a systematic review of the literature on the relationship between general cognitive ability and fertility among modern humans. Our goals were to (a) evaluate the state of the extant literature, and (b) provide a quantitative summary of effect sizes to the extent possible (given the limitations of the literature). A thorough search identified 17 *unique datasets* that passed the inclusion criteria. Using a Random Effects Model to evaluate the data, the overall weighted effect was $r = -0.11$, although the data also indicated a sex effect (stronger correlations among females than males), and a race effect (stronger correlations among Black and Hispanic populations compared to Whites). Importantly, the data suggest the correlation has been increasing in strength throughout the 20th century (and early 21st). Finally, we discovered several notable limitations of the extant literature; limitations that currently prohibit a psychometric meta-analysis. We discuss these issues with emphasis on improving future primary studies to allow for more effective meta-analytic investigations.

1. Introduction

Several studies have documented an inverse relationship between cognitive ability and number of offspring in modern societies (sometimes termed “dysgenic fertility”, e.g. Lynn, 1996), which has, in turn, resulted in both predictions and observations of a declining ‘genotypic’ IQ (e.g., Kong et al., 2017; Lynn, 1996; Lynn & Harvey, 2008). More recently, research has linked this association to secular trends suggestive of long-term phenotypic declines in general cognitive ability (*g*) (Sarraf, 2017; Woodley of Menie et al., 2017). This trend, if present at a broad scale, would negatively affect occupational, educational, and social outcomes at the individual level, and our ability to solve social problems in an increasingly complex world (Neiss, Rowe, & Rodgers, 2002; Rindermann, 2012, 2018; Rindermann, Sailer, & Thompson, 2009). For example, Woodley of Menie and colleagues (e.g., Woodley, 2012; Woodley of Menie et al., 2017) reported evidence of a strong association between a decline in *g* (as measured using convergent phenotypic indicators) and a decline in per capita macro-innovation rates (i.e. the frequency of disruptive or ‘breakthrough’ innovations). Woodley of Menie (2015) also reported a small meta-analysis of studies reporting empirical estimates of IQ loss due to dysgenic reproduction in U.S. and U.K. populations. After correction for method artifacts, reliability and validity, the data revealed an aggregate *g* loss of -0.38 IQ

points per decade.

A decrease in our ability to innovate solutions to complex problems could have serious implications for the future of humanity. As such, the aim of this review is to enhance our understanding of the negative relationship between cognitive ability and number of offspring (NoO) by conducting a systematic review of, and a quantitative descriptive summary of the extant empirical literature. Our specific goals are to (a) provide a descriptive summary of the distribution of effect sizes found in the literature, and (b) highlight methodological concerns with the primary literature vis-a-vis potential meta-analytic studies. We begin by presenting multiple, complimentary theoretical perspectives leading to a hypothesis of a negative relationship between cognitive ability and number of offspring in modern societies. Following this, we report a quantitative summary of effect sizes found in the extant literature. Finally, we discuss the implications of our quantitative findings, and the state of the literature with respect to its ability to support potential formal meta-analyses.

1.1. General cognitive ability (*g*)

“Intelligence” may be best conceptualized as a collection of closely-related constructs, structured hierarchically in a tightly-knit nomological network, with *g* reflecting the critical core of intelligence, namely

* Corresponding author at: Department of Psychological Science, University of North Carolina at Charlotte, 9201 University City Boulevard, Charlotte, NC 28223-0001, USA.
E-mail address: clreeve@unc.edu (C.L. Reeve).

the ability to learn from and reason with novel information (Jensen, 1998; Reeve & Bonaccio, 2011). Empirical research on *g* has expanded across two primary dimensions to form what is commonly referred to as the *g*-nexus (Jensen, 1998). The vertical dimension assumes a focus on the biological and neurological bases for intelligence. Examples include the study of relationships between IQ and a range of heritable traits including reaction times, evoked potentials of the cerebral cortex, and brain pH (Herrnstein & Murray, 1994; Jensen, 1998). The horizontal dimension examines relationships between *g* and a range of personal, social, educational, occupational and health outcomes. Non-exhaustive examples of the horizontal line of inquiry include studies of the relationships between *g* and myriad indicators of psychological wellbeing (Lubinski & Benbow, 2000), physical health (Gordon, 1997; Gottfredson, 2004), religiosity (Razmyar & Reeve, 2013; Reeve, 2009), job performance (Gottfredson, 1997; Meisenberg, 2010), criminality and poverty (Gordon, 1997), and of most relevance to the current study, reproductive behavior (Reeve, Lyerly, & Peach, 2013; Lynn, 1999; Lynn & Harvey, 2008; Meisenberg, 2010; Peach, Lyerly, & Reeve, 2014; Retherford & Sewell, 1989).

1.2. Selection favoring lower *g*

According to evolutionary theory, *g* originally evolved as an adaptation to confer a survival advantage via the enhanced ability to adapt and survive in evolutionarily novel situations (Chiappe & MacDonald, 2005; Gordon, 1997; Jensen, 1998; Kanazawa, 2004). The advent and rise of civilization and technology lead to a modern environment which is radically different to our ancestral environment, particularly with respect to the degree of “evolutionary novelty” we encounter. As such, the behavioral advantage conferred to an individual by higher *g* is almost ubiquitous in modern society (where most of that environment is “evolutionarily novel”). However, there is one domain of modern life in which higher *g* may no longer confer the “Darwinian Advantage” it did in the ancestral environment; namely, human reproduction (i.e., increased probability of passing genes to the next generation).

Throughout human history, any form of industrialization initiates a shift away from a hunter-gather and agrarian lifestyles (common to our ancestors) where larger families were an asset for kin-group productivity and survival. As human culture and technology began to fundamentally change our daily environment into a largely “evolutionarily novel” one (Kanazawa, 2010), the traditional Darwinian advantage for larger numbers of offspring began to erode. For example, medicine began to reduce the infant mortality rate (thus, one did not need to have large numbers of offspring to ensure survival of at least a few), and large numbers of offspring were more likely to be an economic liability rather than a resource, for most people. Ergo, the ability to discern the personal advantage of limited reproduction (though clearly not a *genetic* reproductive advantage), and the ability to control one's reproduction independent of copulation, became a catalyst for reduced reproduction at the individual level. In other words, sexual reproduction became an “evolutionarily novel” proposition. Such theorems were coined the “Internal Relaxation/Reversal of Darwinian Selection” (IRDS) by Nyborg (2012), and are widely cited as the underlying drivers of the negative *g*-NoO relationships observed in a large and growing body of empirical literature (Lynn & Van Court, 2004; Herrnstein & Murray, 1994; Lynn & Harvey, 2008; Nyborg, 2012; Woodley & Figueredo, 2013).

Another theoretical perspective that makes similar predictions is *Population cycle theory* (Woodley of Menie et al., 2017). It proposes, that for most of human history, colder climates promote inter-group conflict because it makes vital resources scarcer (e.g., food, proper shelter, fertile ground). Such conditions should place a fitness premium on *g* because those who are better able to solve novel problems or learn critical skills faster will be better able to survive and prosper in difficult environments. Higher *g* also leads to more innovations that can give a population an edge in inter-group conflict. Conversely, warmer climates

are proposed to reduce this ecological stress on populations (relatively speaking), and thus the fitness value of *g* would be much lower (i.e., variance in *g* would have less impact on selection). Thusly, historically, we should see evidence of increasing *g* over time in populations living in colder climates, and a weaker increase in populations historically isolated in warmer climates (see also Lynn, 1991, and Rushton, 1995, for similar arguments). The prediction that evolutionary novelty increases with colder climates, and that both novelty and the increased selection pressures of cold climates influence reproductive trends has been corroborated (Kanazawa, 2008).

Similar to the previous theories, population cycle theory also suggests that greater ecological and social stability stemming from increased global temperature, coupled with advances in technology, over the last 200 years have significantly changed the pattern of selection on *g* by artificially raising the odds favoring reproduction of those with lower *g*, relative to those with higher *g*, who, as was mentioned previously, can use innovations such as contraception to attenuate their fertility. Under such conditions, one would expect to see those with lower *g* exhibiting higher fitness. To empirically test the population cycle theory, Woodley of Menie et al. (2017) tracked the utilization frequencies of the four high-difficulty words from WORDSUM across 400 years of Google Ngram viewer data. The utilization frequencies of these words served as a proxy vocabulary IQ test (which is highly *g*-loaded; Kan, Wicherts, Dolan, & van der Maas, 2013). First, their analyses showed that the utilization frequencies of the common factor among these words increases between 1500 and 1850, and declines precipitously thereafter. Second, and importantly, they were able to empirically predict this cycle using variations in (a) both the mean and variance in global temperature, and (b) strength of intergroup competition pressures.

Finally, it is critical to point out why our theories focus on *g* as the putative variable, rather than other non-*g* skills and abilities. First and foremost, the negative correlation between IQ and fertility has been shown to be a Jensen Effect. A Jensen Effect refers to the empirical finding that effect sizes (e.g., heritability coefficients, predictive validity coefficients, mean differences in average phenotypic intelligence, etc.) typically correlate significantly with the subtests' *g*-loadings – meaning that *g* can be said to moderate the association. Said differently, the Jensen effect refers to the finding that measures with higher *g*-saturation are better than lower *g*-saturated measures at differentiating between individuals and groups in outcomes that are influenced by IQ. A number of studies have found large and significant Jensen effects for the Black–White mean test score difference (e.g., te Nijenhuis & van den Hoek, 2016), inbreeding depression scores, evoked potentials, brain pH, reaction times, test heritabilities (Jensen, 1998; van Bloois, Geutjes, te Nijenhuis, & de Pater, 2009), and sex differences (Nyborg, 2005) among others. Jensen effects have also been found to explain variability on non-biological variables such as differences in retest effects (e.g., Reeve & Lam, 2007) and race differences in work criteria (Reeve & Bonaccio, 2009).

While the negative association between scores on cognitive tests and NoO has acquired increasing empirical support and scientific publicity, there remains some skepticism as to whether these effects are truly due to differences in *g*, partly because specific estimates of the fertility gradient appear to vary somewhat depending on the specific ability measure used. However, it has been shown that the magnitude of the fertility gradient is positively moderated by the *g*-loading of cognitive ability measures (making it a *Jensen Effect*). For example, Peach, Lyerly, and Reeve (2014) used the Project Talent database to conduct a correlated vectors analysis by correlating the vector of *g*-loadings of the ability subtests with the vector of computed fertility gradients (i.e., the correlations between each subtest and the number of biological children). The strength of the Jensen effect was $r = 0.89$ among the full sample. These findings indicate that the fertility gradient is strongly proportional to the *g*-loading of the test, thus confirming the hypothesis that *g* is the primary factor on which selection operates. More recently,

Woodley of Menie et al. (2017) confirmed this finding via a meta-analysis of studies reporting vector correlations involving the “IQ-fertility gradient,” revealing an aggregate vector correlation of 0.87 ($K = 6$, $N = 108,040$).

The natural synergy among these perspectives is clear. Both Population Cycle Theory and IRDS assert that improving living conditions set the stage for increasing the odds of survival and reproduction among those with fewer cognitive resources. How we fare in the world, educationally, occupationally, economically, and reproductively are all linked to the g -factor, to the degree that they reflect evolutionarily novelty. Finally, Population Cycle Theory predicts that selection should start to favor lower g around the mid-19th century for similar reasons (modern society coupled with a milder, less ecologically stressful climate gave low- g individuals high odds of survival and reproduction). Based on all these perspectives, we expect to find (a) the mean of the distribution of reported effect sizes (of the relation between measures of cognitive ability and NoO) will be negative, and (b) the magnitude of the inverse effect is increasing in strength over time. In addition, we report descriptive analyses for any sub-groups for which the data allowed.

2. The quantitative summary

To recap, our goals were to provide a quantitative summary of effect sizes to the extent possible (given the limitations of the literature), and second, to use the results of the literature search itself to identify limitations and recommend improvements to the primary literature. We discuss some of the limitations here to clarify our decisions in reporting; but reserve most of the discussion of limitations of the literature for the main discussion.

3. Methods

3.1. Search strategy

To find existing research we utilized PsycINFO, PsycArticles, MEDLINE, Health Source Nursing Edition, Health Source Consumer Edition, Academic Search Complete and ProQuest Dissertation and Theses. These databases were searched using variants and combinations of keywords identified in a subset of articles including: “intelligence”, “IQ”, “cognitive ability”, “GMA”, “reproduc*”, “dysgenic”, “fertility”, “number of children, and “birth rate”. Following this, forward and backward searches using the Social Science Citation Index were performed, followed by manual reviews of the reference sections from all articles obtained. Finally, prominent experts in the field were contacted via email to identify any unpublished data or any studies currently in press. No additional suggestions were received, no respondent reported any unpublished data, and no additional unpublished reports were discovered. We do not claim a truly exhaustive list of effect sizes; it is possible we missed some effect sizes to be found in unpublished (or obscure) papers.

3.2. Inclusion/exclusion criteria

We included studies published in English language journals that were available as of November 5th, 2016. Studies needed to have related quantitative measures of cognitive abilities or intellectual achievement tests (e.g., Program for International Student Assessment) with the total number of biological children produced by that person. Upon collecting the initial pool of articles, studies were screened for relevance and reported effect sizes in the form of bivariate correlations, or other effect-size statistics that could be converted to bivariate correlations keeping with guidelines outlined in Peterson and Brown (2005).

Fig. 1 depicts the literature search and screening process. We began with 737 articles from the initial electronic search. After reviewing the

abstracts, 580 articles that clearly did not address our focal question. Examples of these articles include studies of fertility treatment efficacy, impacts on offspring born to mothers with various health conditions, studies of birth order and intelligence, and the relations between traits and behaviors of teenage parents with offspring development. If the abstract provided any suggestion that the relation between cognitive ability and number of offspring was discussed, the article was retained for further review. While it may seem like we excluded a large number of studies at this stage, it is worth noting that this is due to the fact that we purposely cast a very large net in our initial electronic search in order to maximize the likelihood of finding all relevant data; in essence we collected the haystack so as to ensure we found all the needles.

The remaining 157 articles were reviewed for comparison to the remaining inclusion criteria. At this stage, 41 articles were excluded because they were reviews, or theoretical papers, that provided no novel empirical results (references in these review articles were cross-checked). Next, another 40 articles were excluded because they reported relations that were distinctly different from the relationship of interest (i.e., did not provide an effect size estimate of the relation between cognitive ability and any estimate of NoO). Finally, another 20 studies were excluded because they used a different unit of analysis (e.g., birth rates for geo-political units or geographic regions).

Of the remaining 41 articles, fifteen articles (all primarily in the early 20th century) were excluded because they did not measure the intelligence of the potential parent, but rather used the cognitive-ability scores of the offspring themselves as a proxy for parental intelligence. Four more articles based on cognitively impaired groups were excluded. Finally, ten of the remaining articles were removed because they reported analyses from the same dataset as other articles in our pool. In each case, we excluded the article that reported fewer effect sizes (i.e., fewer cuts of the data). Characteristics of the final pool of studies are presented in Table 1. Our data is available on-line (Heeney & Reeve, 2017).

3.3. Coding procedures

Various characteristics of the studies and effect sizes were coded for potential sub-group analyses. Coded data are maintained in an Excel table that includes the following variables: bivariate correlation or (or convertible effect-size statistics) relating g /IQ and measures of NoO, and other numeric indicators such as mean sample age, year in which the criterion data were collected, and sample size. Additionally, this table includes potential moderator variables such as sex, race, geographic region and indicators of study characteristics.

Effects were recorded at the most granular level available (for example, Black Female/Black Male/White Female/White Male) to facilitate the most fine-grained presentation of the data. Effect-size statistics that are convertible to bivariate correlations were converted using the Practical Meta-Analysis Effect Size Calculator that accompanies Lipsey and Wilson (2001). Sex of the sample was coded dichotomously. In cases where sex or race were not indicated ($k = 2$ and $k = 5$ respectively), the categorical value “All” was coded into the database. Reported effects were also coded for a geographic region. Coding of the articles was first accomplished by the second author (expertise in meta-analyses). The lead author (expertise with the g -nexus) then independently coded 20% of articles. These were crossed checked for discrepancies. As only a single discrepancy was found (which the authors quickly resolved by discussion), the coding was deemed reliable. The database is publicly available on Mendeley Data (< < link omitted for blind review process > >)

3.4. Quantitative analyses

SAS Version 9.4 (SAS Institute, 2017) was used to import the coded Excel database and stage the data for calculation of the descriptive statistics of the effect size distribution. Though not a meta-analysis, we

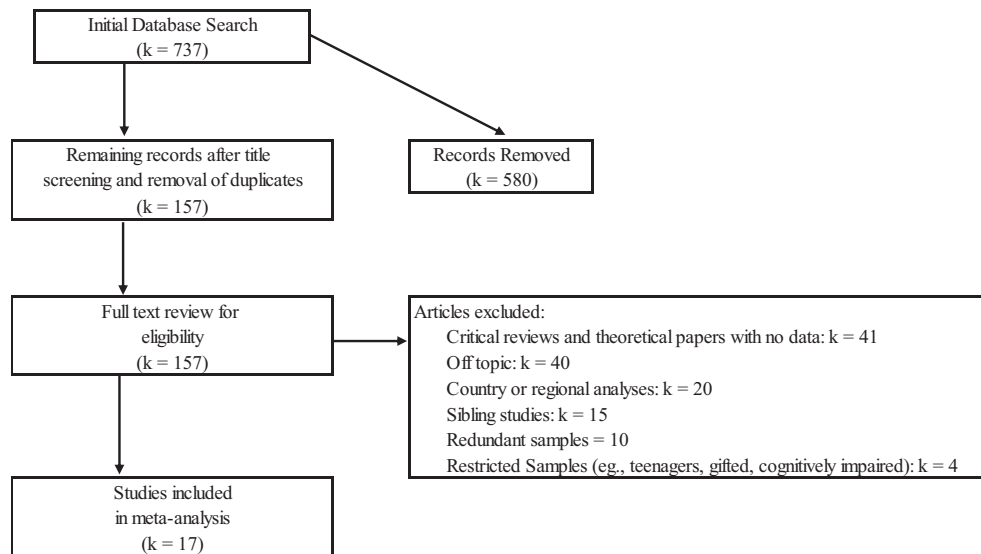


Fig. 1. Flowchart for literature selection procedure.

did employ some basic concepts for our analyses. In particular, a shifting unit of analysis approach was applied as prescribed by Cooper (2010), to aggregate effects appropriately before conducting each separate analysis. For example, in staging the input for calculating the overall effect, the process generates seventeen study-level effects by calculating a sample-weighted average of all inputs by study and aggregating a total corresponding sample size for each effect. Then for the “moderator analysis” (e.g., examining by sex or race) the process begins again with the raw data and calculates weighted effects (with corresponding total N) for each sex. For each moderator analysis, the raw data were re-aggregated according to this methodology for each category. Thus, the number of relevant effect sizes varies depending on the specific sub-categories reported.

Once the data were staged, descriptive analyses were conducted using Comprehensive Meta-analysis (CMA) Version 3 (Borenstein, Hedges, Higgins, & Rothstein, 2015). Due to limitations in the primary data (discussed further below), we were unable to apply normal psychometric corrections for statistical artifacts as prescribed by Schmidt and Hunter (2015), which would be the ideal. The initial analysis estimated the weighted mean effect size and its distribution across studies. We utilized a Random Effects Model because effects are expected to be heterogeneous across studies due to potentially diverse measures of g (i.e., differential g -loadings) and sample characteristics discussed above. Within-group comparisons at the study level are reflected in a statistic of within-group variation, Q_w . Between-group variation is indexed with Q_b . Both Q -statistics follow a chi-square distribution similar to those applied in the analysis of individual samples. Consistent with recommendations of the Cochrane collaborative (Higgins & Green, 2008) the present study adopted an alpha level of 0.10.

3.5. Analysis for publication bias

Publication bias can result when studies yielding null findings fail to be reported. To the extent that this occurs, conventional literature searches may overlook these studies causing a potential bias in the distribution of effect sizes examined (Borenstein, Hedges, Higgins, & Rothstein, 2009). Although there is no way to account for publication bias directly, there are techniques to estimate the potential for this to occur based on the studies that were identified through the literature search. For example, Rosenthal's Fail-Safe N calculates the number of studies required with null results to render the p -value of the overall effect to be > 0.05 . Orwin's Fail-Safe N is considered a more conservative test of publication bias because it does not assume null results;

rather, it allows for selection of a trivial effect (Lipsey & Wilson, 2001). We chose an effect size equal to half of the overall point-estimate as the “trivial” criterion. A third method of detecting publication bias is through the use of a funnel plot (see Fig. 4). Publication bias is evident when the plot depicts an asymmetrical distribution of effects about the overall point-estimate (represented by a vertical line), suggesting that studies of smaller sample sizes (and therefore greater standard error) and with large effects are favored by publishers due to favorable outcomes (Borenstein, Hedges, Higgins, & Rothstein, 2009). However, in the current data, most effect sizes were based on very large samples (average $N = 21,666$). Thus, the use of the funnel plot in this case is quite limited as almost all of the observations lie at the tip of the funnel (Fig. 3).

4. Results

A forest plot of the seventeen study-level effects and their 95% confidence intervals is presented in Fig. 2. All studies but one (Bajema, 1968) reflect a negative effect size, which is also among three studies that fail to reflect a significant difference from zero at $p < .05$. As we are examining distributions of correlations as our effect size indicators, we refer to Cohen's (1969) traditional effect size guidelines for interpreting the average effect sizes seen here. Namely, the values for Pearson product moment correlations that correspond to the traditional D values of 0.2, 0.5, and 0.8. for so-called “small,” “medium,” and “large” effects, respectively, are $r \geq 0.10$, $r \geq 0.24$, and $r \geq 0.37$ (Rosenthal & Rosnow, 1991, p. 446). However, we also caution against over-interpretation of Cohen's guidelines. Whether a given effect (of any size) portrays an important impact on human affairs is determined by many other factors including the degree of variance in the outcome, severity of outcome, potential for accumulation of effects over time, etc. (Abelson, 1985). While Cohen's suggestions provide a standard convention for translating statistical coefficients into verbal descriptions, it must be noted that these general, context-free standards do not take into consideration substantive considerations. “Small” effects can have very large consequences, particularly when applied to large populations and can cumulate over time.

The overall correlation between cognitive ability scores and the number of children born is $r = -0.11$, with a 95% confidence interval of $r = -0.08$ to $r = -0.13$. This indicates there is a significant, negative relationship between parental cognitive ability and the number of children born to that parent. The top of Table 2 indicates that, consistent with expectations, study level effect sizes are heterogeneous, χ^2

Table 1
Study characteristics.^{a,b}

Author(s)	N	Region	Study name or description	Test	g-score	Parents only?		Fertility completed	
						A	B	A	B
Bajema (1968)	437	US	Kalamazoo Fertility Study	Termin Group Intelligence Test	No	No	Yes	n/a	
Chen, Chen, Liao, and Chen (2013)	73	Asia	Sample of Taiwan adults	WAIS-III (Taiwan Version)	No	No	Yes	Yes	
Conrad and Jones (1932)	230	US	Rural New England sample; two cohorts.	Army Alpha	No	Yes	Yes	No	
Higgins, Reed, and Reed (1962)	2032	US	Minnesota State School and Hospital Study	Variety of tests based on availability	No	Yes	n/a	n/a	
Kanazawa (2014)	4973	Europe	National Child Development Study	Variety of tests	Yes	No	Yes	n/a	
Lynn and Van Court (2004)	6244	US	General Social Survey (GSS). Study consists of multiple cohorts.	General Social Survey (GSS) Vocabulary Scores	No	No	Yes	Yes	
Meisenberg, Lawless, Lambert, and Newton (2006)	570	Dominica	Dominica samples; Study of two cohorts.	Raven's and Vocabulary Test	No	No	Yes	Yes	
Meisenberg and Kaul (2010)	6370	US	National Longitudinal Study of Youth 1979 (NLSY79)	Armed Services Vocational Aptitude Battery (ASVAB).	Yes	No	No	n/a	
Reeve, Lyerly, and Peach (2013)	317,755	US	Project Talent Database	Project Talent Ability Battery	Yes	No	No	No	
Vining, (1982, 1995)	4150	US	National Longitudinal Study of Labor Market Experience.	Variety of aptitude tests based on availability	No	No	No	No	
von Stumm, Batty, and Deary (2011)	5647	Europe	The Aberdeen children's Study of 1950s.	Moray House Verbal Reasoning Tests I and II; and Math and English tests.	No	No	Yes	n/a	
Wang, Fuerst, and Ren (2016)	4770	Asia	China Family Panel Studies (CFPS) Dataset	Immediate Word Recall, Delayed Word Recall and Number Series Tests	No	No	Yes	Yes	
Willoughby (1928)	195	US	Sample of California families	A series of cognitive ability tests	No	Yes	No	n/a	
Willoughby and Coogan (1940)	373	US	Providence RI high school records combined with records from the State Bureau of Vital Statistics	Unknown	No	No	No	No	
Woodley of Menie et al. (2015)	3520	US	Midlife in the United States: A National Longitudinal Study of Health and Well-Being	Brief Test of Adult Cognition (BTACT)	Yes	No	Yes	Yes	
Woodley of Menie, Schwartz, and Beaver (2016)	1886	US	Add Health	Picture Vocabulary Test; Peabody Picture Test - Revised	No	No	No	No	

^a Defines completed fertility as a sample mean age of at least 45 years.

^b Defines completed fertility as a sample mean age of at least 50. "n/a" indicates insufficient information available to make determination.

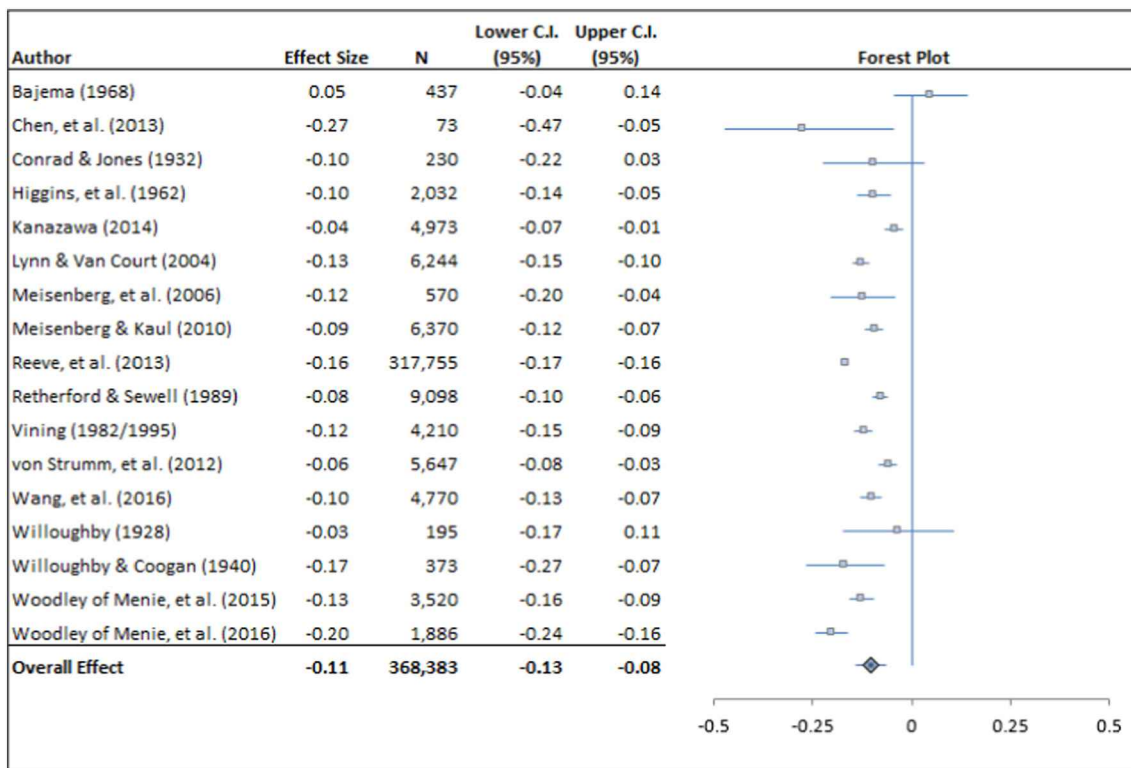


Fig. 2. Study-level effects, N, confidence intervals and forest plot. Note. Template provided by [Neyeloff, Fuchs, & Moreira, 2012](#).

(16) = 292.17, $p < .001$, confirming that there is significant variability in the study-level effects. Importantly, the I^2 index of 94.52%, which can be loosely conceptualized as a ratio of “signal-to-noise” ([Borenstein, Hedges, Higgins, & Rothstein, 2009](#)), supports the conclusion that the observed heterogeneity in study-level effects reflects primarily “real” variation rather than merely measurement error variance. Hence, we continued on to explore sub-group analyses.

Results are presented in [Table 2](#). First, because this relation is thought to be a Jensen Effect (i.e., due to g), we examined whether there are meaningful differences in effects between studies that use raw test scores vs. factor-analytically derived measures of g . The concern is that observed scores are not as g -saturated as g -scores, and thus could

under-estimate the relation. However, the results do not show a meaningful difference, suggesting that the use of raw test scores may not be a significant methodological concern generally speaking (obviously, the validity of any specific test used should always be considered). However, the limited number of studies that used factor-analytically derived g -scores ($K = 3$) may be a factor behind the non-significant result. A more important and critical discovery, in our opinion, was the inability to examine differences in g -saturation with more fidelity. Unfortunately, we quickly discovered during our review of articles that most studies failed to provide clear reports of the exact measures used, exactly how scores were computed, or their g -saturation. This information is necessary to allow one to systematically

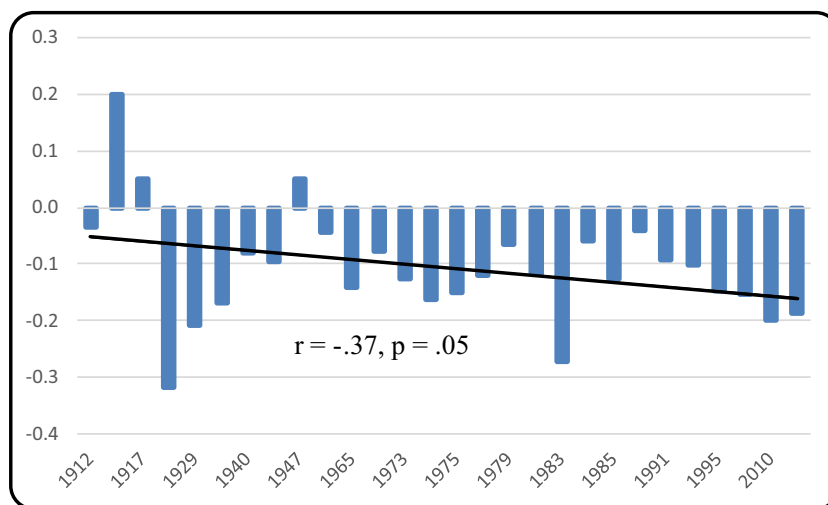


Fig. 3. Distribution of effects scaled to year when parent is age 30. Note. Each bar reflects the effect size of the g -NoO gradient reported by a study. The dates along the axis reflect the year in which the sample of that study was aged 30 (on average).

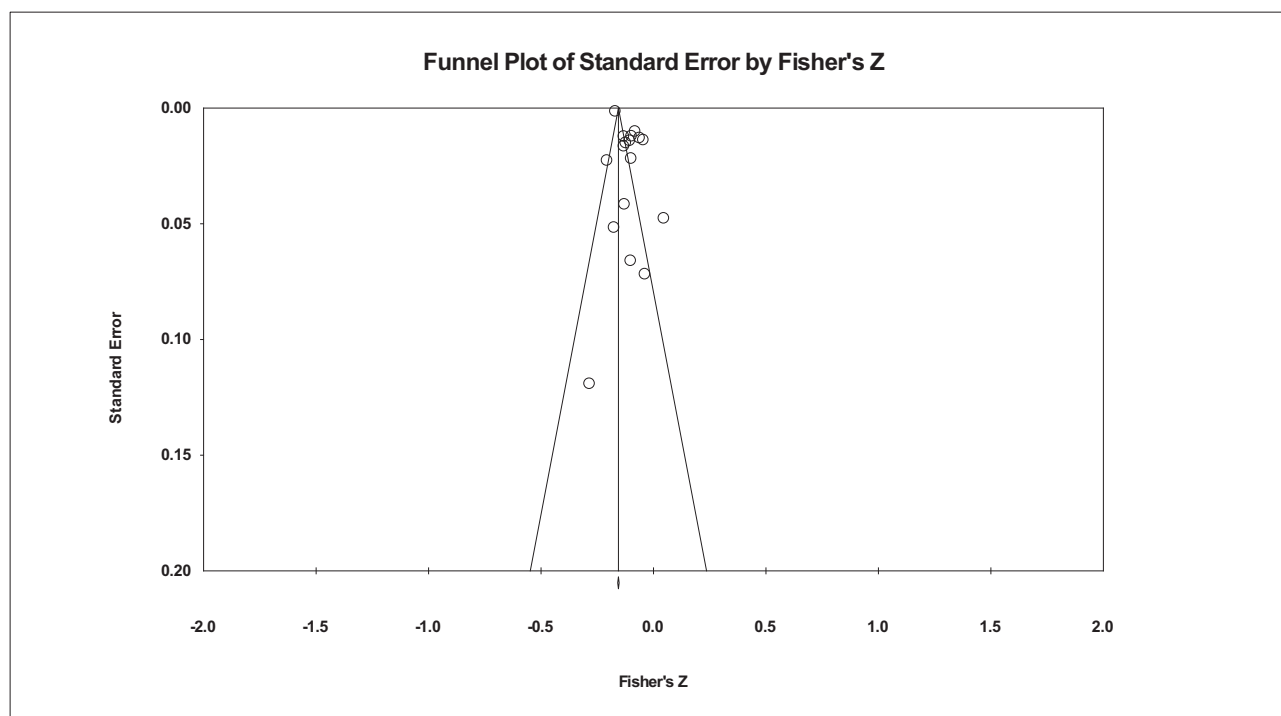


Fig. 4. Funnel plot of standard error by Fisher's Z for study level effects between cognitive ability and fertility.

correct for this artifact.

Second, we examined whether the nature of the sample moderated the effects. Three studies (Conrad & Jones, 1932; Higgins, Reed, & Reed, 1962; Willoughby, 1928) limit the samples to families with at least one child. By excluding participants with no children, this

sampling methodology has the potential to attenuate the observed effects. As such, we created a dichotomous variable to distinguish between these studies and studies that allow participants with no children. The results show the weighted effect for the parent-only group ($r = -0.08, p < .10$) is observed to be smaller than the group that

Table 2
Overall point estimate and moderator analyses for methodological artifacts.

Analysis	<i>k</i>	Weighted <i>r</i>	95% CI for <i>r</i>	<i>Q_b</i>	<i>Q_w</i>	<i>I</i> ²
Overall point-estimate	17	-0.11***	-0.13	-0.08		94.52
Type of score	17			0.01	292.17***	
Raw score	13	-0.10***	-0.14	-0.07		78.44
g-Score	4	-0.11***	-0.16	-0.06		97.29
Type of sample	17			0.34	110.64***	
Parents only	3	-0.08†	-0.16	0.00		0.00
Incl no children	14	-0.11***	-0.14	-0.08		95.37
Fertility^a	19			6.25*	281.03***	
Incomplete	10	-0.14***	-0.17	-0.10		92.21
Complete	9	-0.07***	-0.11	-0.04		78.78
Female only	16			1.21	115.52***	
Incomplete	8	-0.14***	-0.17	-0.11		72.92
Complete	8	-0.12***	-0.15	-0.09		62.18
Male only	17			2.33	37.70***	
Incomplete	8	-0.10**	-0.16	-0.03		95.61
Complete	9	-0.03	-0.09	0.03		70.57
Fertility^b	12			26.49***	27.18***	
Incomplete	7	-0.16***	-0.17	-0.16		0.00
Complete	5	-0.11***	-0.13	-0.09		15.35
Female only	9			0.90	5.63	
Incomplete	4	-0.17***	-0.20	-0.13		42.66
Complete	5	-0.14***	-0.18	-0.11		13.36
Male only	9			6.78**	4.62	
Incomplete	4	-0.15***	-0.18	-0.11		73.98
Complete	5	-0.07**	-0.11	-0.03		0.00

*** $p < .001$.

** $p < .01$.

* $p < .05$.

† $p < .10$.

^a Indicates the sample's mean age is at least 45 years.

^b Indicates the sample's mean age is at least 50 years and those aged 35–49 removed.

includes adults with no children in the sample ($r = -0.11$, $p < .001$), but due to the limited number of studies that include only parents ($K = 3$), the difference is not statistically significant.

In reviewing the studies, it becomes clear that a key difference across studies was the timing of the measurement of the dependent variable. Some studies report effects based on samples of individuals who are still in their child-bearing years whereas others wait to assess until most or all of the sample is likely past the age of bearing children. The expectation is that the effect sizes from samples measured too early would likely be somewhat larger than effects seen in samples who have “completed” fertility. The key issue is that delays in childbearing, perhaps in favor of school or early career opportunities, may make the effect seem larger to the degree that *delays*, but not actually decreases, in having children is associated with ‘g’. It is also possible that by measuring the outcome too early, researchers have truncated the upper tail of the distribution causing range restriction (e.g., some people may continue to have offspring late into life).

As such, two variables were created to code for this issue. Both approaches define the age of the sample using the best measure of central tendency available (the mean by default, and if not reported, the midpoint of the reported range). The first approach simply defines “completed fertility” as samples with an average age of 45 years old or more. This value corresponds to the results from [Fieder and Huber \(2007\)](#) who showed that 99.7% of women and 96.5% of men in a representative Swedish sample have completed their lifetime reproduction by 45. The second approach uses something akin to an extreme groups design by removing studies with an “indeterminate” mean age (range 35 to 49 years old). This leaves one group of effects based on “completed fertility” defined as a sample with a mean age of 50 years or older, and the “incomplete fertility” group defined as a mean age < 35 years old. Analysis of the data using both versions of our coding system yielded significant differences in mean effect sizes. When “incomplete vs. completed fertility” groups are classified as less (or more) than 45 years old, there was a significant difference in the mean effect sizes. When we excluded the data from samples with indeterminate completion status, the differences were more pronounced. These differences also appeared when data were analyzed for men and women separately (though the difference was not significant among men when the “indeterminate” studies were included).

Analyses of sex and race subgroups are shown in [Table 3](#). Overall, we obtained results that are generally consistent with expectations. The weighted effect for women ($r = -0.13$, $p < .001$) is over double that for men ($r = -0.06$, $p < .01$) and the difference is statistically significant ($\chi^2(28) = 5.96$, $p < .01$), although there is still significant variation in effect sizes within the two groups (women, $\chi^2(13) = 52.70$, $p < .001$; men, $\chi^2(14) = 362.50$, $p < .001$).

We also found significant differences in effect sizes between each of the four identified racial groups ($\chi^2(15) = 21.87$, $p < .001$). Effect sizes appear to be larger in samples of Blacks ($r = -0.17$, $p < .001$) and Hispanics ($r = -0.23$, $p < .001$) compared to samples of Whites ($r = -0.07$, $p < .001$) and Asians ($r = -0.11$, $p < .001$). Further analysis suggests that these differences are concentrated primarily among women ($\chi^2(14) = 56.87$, $p < .001$), with no statistically significant differences across racial groups for men ($\chi^2(14) = 4.42$, $p > .10$).

Since the timing of the shift towards the *g*-fertility gradient becoming negative has been largely attributed to nineteenth-century industrialization in the Western world ([Lynn, 1996](#); [Nyborg, 2012](#); [Woodley of Menie et al., 2017](#)), this may suggest that differing levels of present-day industrialization would exert an influence on the relationship between intelligence and reproductive behavior. From a theoretical perspective ([Kanazawa, 2010](#)), populations living in less industrialized nations would experience relatively less evolutionary novelty in the reproductive process than those living in fully industrialized societies. Thus, we would expect the IQ-fertility relationship to be stronger in industrialized societies than non-industrialized

Table 3
Results for substantive moderators.

Analysis	k	Weighted r	95% CI for r	Qb	Qw	I ²
Sex	29			5.96*		
Female	14	-0.13***	-0.17 -0.09		52.70***	75.33
Male	15	-0.06**	-0.10 -0.02		362.50***	96.14
Race	16			21.87***		
Asian	2	-0.11***	-0.18 -0.05		2.21	54.65
Black	3	-0.17***	-0.22 -0.12		4.84†	58.67
Hispanic	1	-0.23***	-0.31 -0.14		0.00	0.00
White	10	-0.07***	-0.09 -0.04		35.63***	74.74
Race (female only)	15			56.87***		
Asian	2	-0.14***	-0.19 -0.10		1.83	45.45
Black	3	-0.25***	-0.29 -0.21		0.53	0.00
Hispanic	1	-0.27***	-0.35 -0.19		0.00	0.00
White	9	-0.09***	-0.11 -0.08		10.03	20.22
Race (male only)	15			4.42		
Asian	2	-0.08	-0.17 0.02		0.37	0.00
Black	2	-0.01	-0.10 0.08		0.09	0.00
Hispanic	1	-0.17†	-0.30 -0.04		0.00	0.00
White	10	-0.05†	-0.08 -0.01		44.10***	79.59
Geographic region	17			3.43		
Asia	2	-0.12**	-0.21 -0.04		2.21	54.65
Dominica	1	-0.12†	-0.24 0.00		0.76	0.00
Europe	2	-0.05	-0.11 0.02		0.00	0.00
U.S.	12	-0.11***	-0.14 -0.08		151.66***	92.75
Date of study	18			0.04		
Post WWII	13	-0.11***	-0.14 -0.07		279.23***	95.70
Pre WWII	5	-0.10**	-0.16 -0.03		2.96	0.00
Date of study	13			6.90**		
Post Roe	6	-0.14***	-0.17 -0.11		50.23***	90.05
Pre Roe	7	-0.08***	-0.11 -0.04		11.64†	48.44
Females	10			19.52***		
Post Roe	5	-0.15***	-0.15 -0.15		7.44	46.23
Pre Roe	5	-0.10***	-0.12 -0.07		0.26	0.00
Males	11			1.51		
Post Roe	5	-0.10***	-0.17 -0.04		108.77***	96.32
Pre Roe	6	-0.04	-0.11 0.03		7.12	0.21

*** $p < .001$.

** $p < .01$.

* $p < .05$.

† $p < .10$.

ones, to the extent culture is independent of *g*. Such cross-country comparisons could also highlight the influence of culture on fertility independent of *g*. For example, previous research has also shown that national religiosity has a large effect on fertility rates, and less industrialized nations tend to have high religiosity rates ([Reeve, 2009](#)). Thus, this cultural factor could potentially moderate the *g*-fertility gradient.

Differences due to location are reported in [Table 3](#). The weighted effect size for the two European studies was less than half of those for the other geographic regions. However, this analysis is merely cursory as the extremely limited number of studies from regions outside of the United States was highly limited (Asia = 2, Dominica = 1, Europe = 2). We report the specific results for the sake of completeness, but we caution against strong inferences given the limited data. Rather, the most significant finding here is the need for more data from outside the U.S.

The studies included in this analysis range from the early 20th century ([Willoughby, 1928](#)) to current day ([Wang, Fuerst, & Ren, 2016](#); [Woodley of Menie, Schwartz, & Beaver, 2016](#)). This provides us with two alternative approaches to exploring the *g*-fertility gradient over time. First, two dichotomous variables were derived to reflect what we thought would be the most relevant historical events for the time frame under study: 1) pre vs. post World War II, and 2) pre vs. post Roe vs. Wade (for U.S. studies only). Second, we conducted a correlational analysis between the date of the data collection and the size of the *g*-

fertility gradient.

Of the two variables that classified historical events, only the pre vs. post Roe vs. Wade moderator demonstrated a statistically significant difference ($\chi^2(13) = 6.90, p < .01$), with a much stronger effect observed post Roe vs. Wade ($r = -0.14, p < .001$). While the effect sizes were higher for both male and female samples after that landmark decision, further analysis shows the difference is significant only among women ($\chi^2(10) = 19.52, p < .001$).

Next, we computed a bivariate correlation between the date of study (scaled according to the age of each sample) and effect size produced by the study. This analysis yields a strong inverse association ($r = -0.37$), indicating the magnitude of the effect is increasing over time with a large magnitude effect size (see Fig. 2). However, one early study by Conrad and Jones (1932) reported a large, positive association for a group of males only. Empirically, this estimate is an outlier; hence, to be conservative we explored the effect of removing this datum. With that study removed, the correlation between date of study and effect size diminishes noticeably, but still yields a moderate inverse effect ($r = -0.22$). Accordingly, it appears the gradient is growing stronger over time at a moderate pace. While this finding is contradictory to Lynn (2011) who examined the correlation between IQ and number of siblings in the UK, finding a weakening over 60 years, and also to Wang, Fuerst, and Ren (2016), who found that in China the effect was weakening over time when both completed fertility and sibling numbers were used, it is consistent with the more recent finding from Sweden showing a possible slight strengthening of selection against g over several decades among mixed sex cohorts (Madison, Woodley of Menie, & Sanger, 2016). To better explore whether this effect is influenced by “early” studies that might reflect significantly different social conditions, we recomputed the correlation, restricting the sample to just those occurring after WW2. This arguably provides an estimate of the trend under recent social conditions. That analysis yields $r = -0.56$. Thus, the secular trend is not likely an artifact of a few early studies.

Finally, we computed two fail-safe checks to evaluate the possibility of a “file drawer bias”. Rosenthal’s Fail-Safe N indicates that 7552 studies with null findings would be required to reduce the overall effect to a level that is non-significant at $p > .05$. Orwin’s Fail-Safe N indicates that it would require 36 null studies (more than double the number available) to reduce the overall effect to less than $r = -0.05$ (which is about half of the observed weighted effect).

5. Discussion & critique of the literature

The overall goal of the current study was to systematically analyze the empirical literature on the relation between general cognitive ability and reproduction. To that end, a thorough review of the literature revealed 17 independent studies that provided sufficient information for inclusion in at least a basic level meta-analysis. Our results indicate, that on average, there is a small but significant and meaningful negative g -fertility gradient present among modern human populations ($r = -0.11$). Although most of the data analyzed to date stems from North America and Europe, the small number of studies outside this region suggest the trend is likely operating in other regions also. These findings are consistent with multiple theoretical perspectives as described in the introduction.

As expected, we also discovered significant variation in the magnitude of the effect. We sought to test several factors as potential sources of this variation, both substantive and methodological. One key result is that the selection trend appears to be stronger among women than men. This study cannot empirically determine why the difference exists, but we can posit potential explanations. One explanation may be more intelligent women are more deliberate in their decisions regarding childbearing than equally intelligent men. This proposition is supported by Kanazawa (2014), who found that more intelligent young men and women at 23 are more likely to desire to remain childless for life than less intelligent young men and women do, but, by age 47, intelligent

men were no more likely to have remained childless than less intelligent men, whereas more intelligent women did in fact remain childless than less intelligent women. More intelligent women, it seems, were able to implement their desire to remain childless for life, whereas more intelligent men did not do so. Additionally, Kanazawa (personal communication) analyzed the GSS data for the years between 2004 and 2014 and found that among respondents aged 45 or older, the bivariate correlation between WORDSUM scores, and the total number of children is $r = -0.197$ among women and $r = -0.077$ among men. This further confirms that more intelligent women better implement/maintain the desire to remain childless than intelligent men. This increased female selectivity is likely reinforced by sex differences in typical fertility windows, the levels of biological investment necessary to procreate, and cultural norms that place a disproportionate burden on women during child rearing (Trivers, 1972; Trivers & Willard, 1973). These “higher stakes” are compounded by the more “evolutionary novel trade-offs” that women face between work, education and raising a family (Kanazawa, 2010). Matters of sex and reproduction are thus largely driven by female choices.

Second, the issue of when to measure the NoO appears to be a key issue influencing observed effect sizes. Studies with what are referred to as “completed-fertility samples” yielded smaller effect sizes compared to the “incomplete fertility samples.” The extreme groups design (with indeterminate completion removed) also showed a more pronounced difference than when all samples are included. From a purely methodological perspective, these results provide evidence that the size of the estimated relationship will vary depending on the proportion of the sample with “completed fertility.” However, from a substantive perspective, these results also demonstrate that higher g individuals are delaying reproduction longer than their lower g counterparts, even if they eventually go on to have children. That is, more intelligent people appear to be delaying reproduction until later in life compared to less intelligent people who begin reproducing earlier in life. The critical implication is that less intelligent people will have shorter inter-generational intervals than more intelligent people. This implies that the observed effect size at the individual level (e.g., $r = -0.11$) might severely underestimate the total dysgenic effect of intelligence at the macrosocial level.

For example, if more intelligent individuals on average have two children whereas less intelligent individuals on average have four children, it would appear that there would be only twice as many less intelligent individuals in society than more intelligent individuals due to the dysgenic trend. However, at the end of the century, there are actually four times as many less intelligent individuals as more intelligent individuals because there can be five generations in a century among less intelligent individuals (on average reproducing at age 20), whereas there can be two and a half generations among more intelligent individuals (on average reproducing at age 40). Thus, after a century, the population of less intelligent individuals in society is far larger than the negative effect of intelligence on fertility appears to suggest at the individual level (Kanazawa, pers. comm). Consistent with this, Rindermann (2018) estimated a decadal IQ loss of -0.87 points per decade when the interaction between generation length, fertility and IQ is simulated. The study of Kong et al. (2017), which tracked the decline in the frequency of genetic variants promoting educational attainment over seven decades in Iceland estimated an equivalent IQ loss of -0.30 points per decade, which increases to -0.86 points per decade once a latent variable estimate of g ’s heritability is used to correct for ‘missing heritability’ (Woodley of Menie et al., 2017). Kong et al. (2017) note that most of the genetic decline in Iceland is driven by the effect of different generation lengths among those with different levels of educational attainment, rather than different levels of fertility. Predicted losses in g of around one point per decade are close to the g equivalent declines observed in cross-temporal analyses of slowing simple reaction time and decreasing vocabulary knowledge (Woodley of Menie et al., 2017).

Perhaps the most important methodological implication here is that defining “completed fertility” by age is not valid methodology. Men theoretically have boundless fertility windows, whereas women have more biologically finite fertility windows. Thus, the measurement of “completed fertility” should be considered more carefully. We suggest it may be better to assess both intended fertility and actual fertility among both men and women. In addition, future studies should attempt to assess for medically induced infertility. These results reflect the importance of considering the methodological complexity in measuring lifetime reproductive capacities for men versus women.

Of the two historical events coded, only the Roe vs. Wade event (for U.S. based studies) yielded significant differences. The weighted effect for the post Roe vs. Wade period is approximately double the effect for the period prior. This suggests that the *g*-fertility gradient may be stronger when people have more complete and independent control over their reproductive options than when they are restricted by governments or cultural norms (e.g., religious influences). However, it may also be possible that the stronger effects post Roe vs. Wade are merely a reflection of the increasing magnitude of the selection pressure over time. As noted above, our analysis shows a correlation of $r = -0.37$ between the size of the correlations and dates of the studies. This may reflect Kanazawa's (2010) hypothesis that the effect of *g* on cognitive loaded behaviors will become increasingly stronger as the world that we inhabit becomes increasingly complex. Or the secular trend towards strengthening the negative *g*-fertility gradients could be due to the hypothesized effect of reduced ecological stress and concomitantly reduced intergroup competition (Woodley of Menie et al., 2017). Understanding exactly what mechanisms are driving this effect will require further study with more sensitive and detailed measures.

Our analysis of the effect of differences in the *g*-saturation of measures did not yield any notable effect. However, we caution that due to limitations in the primary literature, we could only analyze a dichotomous proxy variable; studies using factor analytic derived *g*-scores vs. raw test scores. Clearly this is not the sensitive analysis we would have desired and may explain the null finding. However, given that we did not observe meaningful differences, this may serve to somewhat alleviate the concern about measurement differences. This may be due to the fact that raw scores generated by most professionally developed cognitive ability tests, although not psychometrically pure measures of *g*, are by design highly *g*-saturated (clearly some sub-scales are not, but most tend to be). If so, this could also serve to somewhat assuage concerns about measurement error in the independent variable. Reliability estimates were not available in most of the included studies (a notable problem with the extant literature); however, it can be assumed *g*-scores generated from factor analyses are more reliable than observed test scores as they are constructed from only the systematic variance in the observed sigma matrix. And it has been shown (Reeve & Blacksmith, 2009) that when sample sizes are at least 400, the *g*-loadings and *g*-scores generated from different factor analytic methods are highly reliable (i.e., 0.92 to 0.97). Given this, the analysis effectively tested for differences between studies using a (theoretically) perfectly valid and (near perfectly) reliable predictor vs. those that did not. Again, the fact that we did not find a difference may be due to the fact that professionally developed ability tests are quite valid and reliable standard systems of measurement. Thus, there simply may not be enough variance in reliability or validity to create a moderator effect. However, this analysis also points to a key implication for future research. We were not able to conduct a standard correction for unreliability because too many of the studies did not report sufficient psychometric information about the observed scores actually used in the analysis. In many senses, we think this finding (the need for better and more consistent reporting) may be the key implication of this analysis.

Although no evidence from this study emerged to support differences in the other methodological issues examined, this may be due to the limited number of studies available to represent some sub-groups. It

is nonetheless reasonable to expect the use of raw test scores to introduce measurement error with respect to *g*, and in turn introduce measurement error in its relationship with fertility and any other indicators with which correlations may be performed. Resources that can be used to extract *g* from raw test scores are now abundant and should be used when the data permit. Likewise, it is reasonable to expect studies that restrict samples to only adults that have at least one child will attenuate the effects, as we did observe a reduced effect size for these studies.

It was not possible to conduct traditional psychometric comparisons for a variety of variables due to the limited number of studies, or the limited information reported. For example, we located only a single study from a country that may be considered less industrialized than the other samples, most of which come from western, industrialized countries. This highlights the need for more and better data from under-represented regions of the world. Likewise, we could not conduct subgroup analyses for potential moderators that have been noted in previous primary research, such as religiosity, socioeconomic status, and education level.

6. General conclusion

In general, a declining *g* due to selection may lie beneath a range of societal concerns that could impact the course of our future. This possibility is made all the more salient by the aforementioned recent observation that polygenic scores predictive of *g* and educational attainment have now been shown to have actually declined in the population of Iceland over 70 years, in direct proportion to the observed strength of selection acting against them (Kong et al., 2017). In educational systems, this may mean a decline in the abilities of students over time, a dilution of curricula and academic standards, and potentially a slow-down in the progress of science. Although policy decisions and automation are likely the primary drivers, a declining *g* is consistent with the widely-reported income gap, and may combine with these environmental influences to exacerbate the problem. The quality of our decisions as individuals and as a society, particularly how we handle high-stakes events as they unfold on the world stage (e.g., refugee crisis, climate change, overpopulation, shortage of fresh drinking water, etc.), reflect abilities to navigate highly complex problems that are undoubtedly highly dependent on *g*. It is clear that the implications are far-reaching.

Indeed, the decline of a heritable trait so crucial to health, occupational, and socio-behavioral outcomes yields substantial ramifications at both the individual and national level (Woodley, 2012; Woodley & Figueredo, 2013; Woodley of Menie et al., 2017). For example, childhood intelligence predicts physical, functional, and subjective health in adulthood, as well as hospital visits and sick leave days later in life (Wrulich et al., 2014), and has even been shown to predict social outcomes such as political involvement, liberal attitudes and social progress (Deary, Batty, & Gale, 2008; Rindermann, Sailer, & Thompson, 2009). At the national level, research shows that *g* is a critical social resource that can predict changes in the rates of scientific and technological innovation, which clearly yields implications for current and future societal well-being. For example, in addition to the observed “Woodley Effects” on measures of simple visual and auditory reaction times, working memory, color discrimination, three-dimensional rotation, and vocabulary level (see: Sarraf, 2017), per capita rates of eminent individuals and macro-innovations, or conspicuously novel innovations like the plough and splitting the atom have shown a concomitant decline (e.g., Woodley, 2012; Woodley & Figueredo, 2013; Woodley of Menie et al., 2017). That the phenomenon of declining *g* should be so comprehensively linked to a decline in critical real-world outcomes should be of imminent concern, especially as the countervailing Flynn Effect (the secular increase in narrow and low-heritability cognitive abilities independently of *g* over time; te Nijenhuis & van der Flier, 2013) does not appear to be mitigating these declines. Should this

trend continue unabated, coupled with continuing overpopulation, there is a risk it will eventually trigger economic stagnation and decline, and civil instability (Nyborg, 2012; Woodley, 2012).

Although the range of practical, and ethically proper, policy options may be limited, it seems clear that more and better education regarding family planning could improve the fortunes of prospective parents and their families, and through this potentially attenuate the existing trend. These efforts should consist of not only instruction about the appropriate use of contraceptives, but also a convincing narrative that explains the tradeoffs between having children (particularly early in life), education and long-term financial security. Although there have been many efforts to accomplish this, even more resources devoted to the development and implementation of evidence-based interventions would surely benefit both individuals and society at large.

References*

- *Bajema, C. J. (1968). Relation of fertility to occupational status, IQ, educational attainment, and size of family of origin: A follow-up study of a male Kalamazoo public school population. *Eugenics Quarterly*, 15, 198–203.
- *Chen, H., Chen, Y., Liao, Y., & Chen, H. (2013). Relationship of fertility with intelligence and education in Taiwan: A brief report. *Journal of Biosocial Science*, 45(4), 567–571. <http://dx.doi.org/10.1017/S0021932012000545>.
- *Conrad, H. S., & Jones, H. E. (1932). A field study of the differential birth rate. *Journal of the American Statistical Association*, 27, 153–159.
- *Higgins, J. V., Reed, E. W., & Reed, S. C. (1962). Intelligence and family size: A paradox resolved. *Eugenics Quarterly*, 9(2), 84–90. <http://dx.doi.org/10.1080/19485565.1962.9987508>.
- *Kanazawa, S. (2014). Intelligence and childlessness. *Social Science Research*, 48, 157–170.
- *Lynn, R., & Van Court, M. (2004). New evidence of dysgenic fertility for intelligence in the United States. *Intelligence*, 32(2), 193–201.
- *Meisenberg, G., & Kaul, A. (2010). Effects of sex, race, ethnicity, and marital status on the relationship between intelligence and fertility. *Mankind Quarterly*, 50, 151–187.
- *Meisenberg, G., Lawless, E., Lambert, E., & Newton, A. (2006). The social ecology of intelligence on a Caribbean island. *Mankind Quarterly*, 46(4), 395–433.
- *Reeve, C. L., Lyerly, J. E., & Peach, H. (2013). Adolescent intelligence and socio-economic wealth independently predict adult marital and reproductive behavior. *Intelligence*, 41(5), 358–365. <http://dx.doi.org/10.1016/j.intell.2013.05.010>.
- *Retherford, R. D., & Sewell, W. H. (1989). How intelligence affects fertility. *Intelligence*, 13(2), 169–185. [http://dx.doi.org/10.1016/0160-2896\(89\)90015-9](http://dx.doi.org/10.1016/0160-2896(89)90015-9).
- *Vining, D. R. (1982). On the possibility of the reemergence of a dysgenic trend with respect to intelligence in American fertility differentials. *Intelligence*, 6, 241–264. [http://dx.doi.org/10.1016/0160-2896\(82\)90002-2](http://dx.doi.org/10.1016/0160-2896(82)90002-2).
- *Vining, D. R. (1995). On the possibility of the reemergence of a dysgenic trend with respect to intelligence in American fertility differentials: An update. *Personality and Individual Differences*, 19(2), 259–263. [http://dx.doi.org/10.1016/0191-8869\(95\)00038-8](http://dx.doi.org/10.1016/0191-8869(95)00038-8).
- *von Stumm, S., Batty, G. D., & Deary, I. J. (2011). Marital status and reproduction: Associations with childhood intelligence and adult social class in the Aberdeen children of the 1950s study. *Intelligence*, 39, 161–167.
- *Wang, M., Fuerst, J., & Ren, J. (2016). Evidence of dysgenic fertility in China. *Intelligence*, 57, 15–24. <http://dx.doi.org/10.1016/j.intell.2016.04.001>.
- *Willoughby, R. R. (1928). Fertility and parental intelligence. *The American Journal of Psychology*, 40, 671–672. <http://dx.doi.org/10.2307/1414387>.
- *Willoughby, R. R., & Coogan, M. (1940). The correlation between intelligence and fertility. *Human Biology*, 12(1), 14–119.
- *Woodley of Menie, M. A., Figueredo, A. J., Dunkel, C. S., & Madison, G. (2015). Estimating the strength of genetic selection against heritable g in a sample of 3520 Americans, sourced from MIDUS II. *Personality and Individual Differences*, 86, 266–270. <http://dx.doi.org/10.1016/j.paid.2015.05.032>.
- *Woodley of Menie, M. A., Schwartz, J. A., & Beaver, K. M. (2016). How cognitive genetic factors influence fertility outcomes: A mediational SEM analysis. *Twin Research and Human Genetics*, 19(6), 628–637. <http://dx.doi.org/10.1017/thg.2016.82>.
- Abelson, R. P. (1985). A variance explanation paradox: When a little is a lot. *Psychological Bulletin*, 97, 128–132.
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. John Wiley & Sons.
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2015). *Comprehensive meta-analysis, Version 3*. Englewood, NJ: Biostat.
- Chiappe, D., & MacDonald, K. (2005). The evolution of domain-general mechanisms in intelligence and learning. *Journal of General Psychology*, 132, 5–40. <http://dx.doi.org/10.3200/GENP.132.1.5-40>.
- Cohen, J. (1969). *Statistical power analysis for the behavioral sciences*. New York: Academic Press.
- Cooper, H. (2010). *Research synthesis and meta-analysis: A step-by-step approach* (4th ed.). Thousand Oaks, CA, US: Sage Publications, Inc.
- Deary, I. J., Batty, G. D., & Gale, C. R. (2008). Bright children become enlightened adults. *Psychological Science*, 19(1), 1–6. <http://dx.doi.org/10.1111/j.1467-9280.2008.02036.x>.
- Fieder, M., & Huber, S. (2007). Parental age difference and offspring count in humans. *Biology Letters*, 3(6), 689–691. <http://dx.doi.org/10.1098/rsbl.2007.0324>.
- Gordon, R. A. (1997). Everyday life as an intelligence test: Effects of intelligence and intelligence context. *Intelligence*, 24(1), 203–320. [http://dx.doi.org/10.1016/S0160-2896\(97\)90017-9](http://dx.doi.org/10.1016/S0160-2896(97)90017-9).
- Gottfredson, L. S. (1997). Why g matters: The complexity of everyday life. *Intelligence*, 24, 79–132.
- Gottfredson, L. S. (2004). Intelligence: Is it the epidemiologists' elusive fundamental cause of social class inequalities in health? *Journal of Personality and Social Psychology*, 86, 174–199.
- Heeney, M. D., & Reeve, C. L. (2017). *Effect sizes for dysgenic fertility studies*. *Mendeley Data*, v1. <http://dx.doi.org/10.17632/66d7v3vm24.1>.
- Herrnstein, R. J., & Murray, C. (1994). *The bell curve: Intelligence and class structure in American life*. New York, NY: The Free Press.
- Higgins, J. P., & Green, S. (2008). Identifying and measuring heterogeneity. *Cochrane handbook for systematic reviews of interventions version 5.0* (pp. 1). .
- Jensen, A. R. (1998). *The g factor: The science of mental ability*. Westport, CT: Praeger.
- Kan, K. J., Wicherts, J. M., Dolan, C. V., & van der Maas, H. L. J. (2013). On the nature and nurture of intelligence and specific cognitive abilities: The more heritable, the more culture dependent. *Psychological Science*, 24, 2420–2428. <http://dx.doi.org/10.1177/0956797613493292>.
- Kanazawa, S. (2004). General intelligence as a domain-specific adaptation. *Psychological Review*, 111(2), 512–523. <http://dx.doi.org/10.1037/0033-295X.111.2.512>.
- Kanazawa, S. (2008). Temperature and evolutionary novelty as forces behind the evolution of general intelligence. *Intelligence*, 36, 99–108.
- Kanazawa, S. (2010). Evolutionary psychology and intelligence research. *American Psychologist*, 65, 279–289. <http://dx.doi.org/10.1037/a0019378>.
- Kong, A., Frigge, M. L., Thorleifsson, G., Stefansson, H., Young, A. I., Zink, F., ... Stefansson, K. (2017). Selection against variants in the genome associated with educational attainment. *Proceedings of the National Academy of Sciences of the United States of America*, 114. *Proceedings of the National Academy of Sciences of the United States of America* (pp. E727–E732). <http://dx.doi.org/10.1073/pnas.1612113114>.
- Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA, US: Sage Publications, Inc.
- Lubinski, D., & Benbow, C. P. (2000). States of excellence. *American Psychologist*, 55, 137–150.
- Lynn, R. (1991). The evolution of race differences in intelligence. *Mankind Quarterly*, 32, 99–173.
- Lynn, R. (1996). *Dysgenics: Genetic deterioration in modern populations*. Westport: CT, Praeger.
- Lynn, R. (1999). New evidence for dysgenic fertility for intelligence in the United States. *Social Biology*, 46(1–2), 146–153. <http://dx.doi.org/10.1080/19485565.1999.9988992>.
- Lynn, R. (2011). *Dysgenics: Genetic deterioration in modern populations*. London, United Kingdom: Ulster Institute for Social Research.
- Lynn, R., & Harvey, J. (2008). The decline of the world's IQ. *Intelligence*, 36(2), 112–120. <http://dx.doi.org/10.1016/j.intell.2007.03.004>.
- Madison, G., Woodley of Menie, M. A., & Sanger, J. (2016). Secular slowing of auditory simple reaction time in Sweden (1959–1985). *Frontiers in Human Neuroscience*, 10, 407. <http://dx.doi.org/10.3389/fnhum.2016.00407>.
- Meisenberg, G. (2010). The reproduction of intelligence. *Intelligence*, 38(2), 220–230. <http://dx.doi.org/10.1016/j.intell.2010.01.003>.
- Neiss, M., Rowe, D. C., & Rodgers, J. L. (2002). Does education mediate the relationship between IQ and age of first birth? A behavioural genetic analysis. *Journal of Biosocial Science*, 34, 259–275.
- Neyeloff, J. L., Fuchs, S. C., & Moreira, L. B. (2012). Meta-analyses and forest plots using a microsoft excel spreadsheet: Step-by-step guide focusing on descriptive data analysis. *BMC Research Notes*. <http://dx.doi.org/10.1186/1756-0500-5-52>.
- Nyborg, H. (2005). Sex related differences in general intelligence g, brain size and social status. *Personality and Individual Differences*, 39, 497–509.
- Nyborg, H. (2012). The decay of western civilization: Double relaxed Darwinian selection. *Personality and Individual Differences*, 53, 118–125.
- Peach, H., Lyerly, J. E., & Reeve, C. L. (2014). Replication of the Jensen effect on dysgenic fertility: An analysis using a large sample of American youth. *Personality and Individual Differences*, 71, 56–59. <http://dx.doi.org/10.1016/j.paid.2014.07.017>.
- Peterson, R. A., & Brown, S. P. (2005). On the use of beta coefficients in meta-analysis. *Journal of Applied Psychology*, 90(1), 175–181. <http://dx.doi.org/10.1037/0021-9010.90.1.175>.
- Razmyar, S., & Reeve, C. L. (2013). Individual differences in religiosity as a function of cognitive ability and cognitive style. *Intelligence*, 41(5), 667–673. <http://dx.doi.org/10.1016/j.intell.2013.09.003>.
- Reeve, C. L. (2009). Expanding the g-nexus: Further evidence regarding the relations among national IQ, religiosity and national health outcomes. *Intelligence*, 37(5), 495–505. <http://dx.doi.org/10.1016/j.intell.2009.06.003>.
- Reeve, C. L., & Blacksmith, N. (2009). Equivalency and reliability of vectors of g-loadings across different methods of estimation and sample sizes. *Personality and Individual Differences*, 47, 968–972.
- Reeve, C. L., & Bonaccio, S. (2009). Measurement reliability, the Spearman–Jensen effect and the revised Thorndike model of test Bias. *International Journal of Selection and Assessment*, 17(1), 61–68. <http://dx.doi.org/10.1111/j.1468-2389.2009.00451.x>.
- Reeve, C. L., & Bonaccio, S. (2011). The nature and structure of 'intelligence'. In T. Chamorro-Premuzic, S. von Stumm, A. Furnham, T. Chamorro-Premuzic, S. von

* References marked with an asterisk indicate sources that were included in the meta-analysis.

- Stumm, & A. Furnham (Eds.). *The Wiley-Blackwell handbook of individual differences* (pp. 187–216). Wiley-Blackwell.
- Reeve, C. L., & Lam, H. (2007). The relation between practice effects, test-taker characteristics and degree of g-saturation. *International Journal of Testing*, 7, 225–242.
- Rindermann, H. (2012). Intellectual classes, technological progress, and economic development: The rise of cognitive capitalism. *Personality and Individual Differences*, 53(2), 108–113. <http://dx.doi.org/10.1016/j.paid.2011.07.001>.
- Rindermann, H. (2018). *Cognitive capitalism: Human capital and the wealth of nations*. Cambridge, UK: Cambridge University Press.
- Rindermann, H., Sailer, M., & Thompson, J. (2009). The impact of smart fractions, cognitive ability of politicians and average competence of peoples on social development. *Talent Development and Excellence*, 1, 3–25.
- Rosenthal, R., & Rosnow, R. L. (1991). *Essentials of behavioral research: Methods and data analyses*. New York, NY: McGraw Hill.
- Rushton, J. P. (1995). *Race, evolution, and behavior: A life history perspective*. New Brunswick: Transaction.
- Sarraf, M. (2017). Review of historical variability in heritable general intelligence: Its evolutionary origins and socio-cultural consequences. *Personality and Individual Differences*, 109, 238–241.
- SAS Institute Inc (2017). *SAS Version 9.4*. (Cary, NC).
- Schmidt, F. L., & Hunter, J. E. (2015). *Methods of meta-analysis: Correcting error and bias in research findings* (3rd ed.). (Thousand Oaks, CA; London and New Delhi).
- te Nijenhuis, J., & van den Hoek, M. (2016). Spearman's hypothesis tested on black adults: A meta-analysis. *Journal of Intelligence*, 4, 6. <http://dx.doi.org/10.3390/jintelligence4020006>.
- te Nijenhuis, J., & van der Flier, H. (2013). Is the Flynn effect on g? A meta-analysis. *Intelligence*, 41, 802–807.
- Trivers, R. (1972). Parental investment and sexual selection. In B. Campbell (Ed.). *Sexual selection and the descent of man, 1871–1971* (pp. 136–179). Chicago: Aldine.
- Trivers, R., & Willard, D. (1973). Natural selection of parental ability to vary the sex ratio of offspring. *Science*, 179, 90–92.
- van Bloois, R. M., Geutjes, L.-L., te Nijenhuis, J., & de Pater, I. E. (2009, December). *g loadings and their true score correlations with heritability coefficients, giftedness, and mental retardation: Three psychometric meta-analyses. Paper presented at the 10th annual meeting of the international society for intelligence research* (Madrid, Spain).
- 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 of Menie, M. A. (2015). How fragile is our intellect? Estimating losses in general intelligence due to both selection and mutation accumulation. *Personality and Individual Differences*, 75, 80–84.
- Woodley of Menie, M. A., Figueredo, A. J., Fernandes, H. B. F., Peñaherrera-Aguirre, M., Sarraf, M. A., & Hertler, S. (2017). The rhythm of the west: A biohistory of the modern era AD 1600 to the present. *Journal of Social Political and Economic Studies*, 37 (Monograph Series).
- Wrulich, M., Brunner, M., Stadler, G., Schalke, D., Keller, U., & Martin, R. (2014). Forty years on: Childhood intelligence predicts health in middle adulthood. *Health Psychology*, 33(3), 292–296. <http://dx.doi.org/10.1037/a0030727>.