Decreases in Brain Size and Encephalization in Anatomically Modern Humans

Jeff Morgan Stibel
Natural History Museum, Los Angeles, CA, USA

Abstract
Growth in human brain size and encephalization is well documented throughout much of prehistory and believed to be responsible for increasing cognitive faculties. Over the past 50,000 years, however, both body size and brain mass have decreased but little is known about the scaling relationship between the two. Here, changes to the human brain are examined using matched body remains to determine encephalization levels across an evolutionary timespan. The results find decreases to encephalization levels in modern humans as compared to earlier Holocene H. sapiens and Late Pleistocene anatomically modern Homo. When controlled for lean body mass, encephalization changes are isometric, suggesting that much of the declines in encephalization are driven by recent increases in obesity. A meta-review of genome-wide association studies finds some evidence for selective pressures acting on human cognitive ability, which may be an evolutionary consequence of the more than 5% loss in brain mass over the past 50,000 years.
Some evidence that body size in *Homo* has reduced during similar periods has been produced [Ruff et al., 1997], as well as anecdotal support for roughly symmetrical declines in body and brain mass [Henneberg, 1988; Ruff et al., 1997]. There is strong evidence to suggest, however, that body size (mass and stature) has increased during modern times [Henneberg, 1988; Ruff et al., 1997; Bellisari, 2008] despite continued declines in brain size [Ruff et al., 1997]. This has led scholars across various disciplines to propose that brain size declines have resulted in changes to human behavior and cognitive ability [Bailey and Geary, 2009; Crabtree, 2012; Bednarik, 2014; Hare, 2017; Benitez-Burraco et al., 2020; Dor Shilton et al., 2020], a theory that is now prominent in popular literature [McAuliffe, 2011; Ghose, 2012; Stringer, 2014; Alex, 2019]. It is not clear, however, why human brain size has been declining since at least through the Late Pleistocene. Nor do we know the impact, if any, to human cognition.

There are two fundamental questions that must be answered in order to determine whether declines in *Homo* brain size have influenced human cognitive ability. First, why is brain size declining, both from an absolute basis and relative to the body? There is considerable debate as to whether brain size declines are simply a byproduct of evolutionary pressures toward smaller body size or are being directly influenced by natural selection as a result of behavioral changes, metabolic demands, or other pressures that directly act on the brain. Second, regardless of why the brain is declining, is there any evidence to support reductions in human cognition? Studies of both cognitive function and educational achievements have largely supported gains in cognitive ability [Flynn, 1984, 1987, 2009; Barro and Lee, 2013; Pietschnig and Voracek, 2015; Conley and Domingue, 2016; Lee and Lee, 2016], but few studies have examined these results on an evolutionary basis by distinguishing between genetic and environmental factors.

The prevailing theory to explain the trend in *Homo* size changes over the past 50,000 years is that the brain has simply followed a decline in body size [Beals et al., 1984; Henneberg, 1988; Ruff et al., 1997]. To a large extent, brain size is necessarily tied to body size, as more neurons are required to control greater mass in animals [Herculano-Houzel and Kaas, 2011; Dicke and Roth, 2016; Font et al., 2019]. Body height and mass have been shown to account for significant differences in modern brains [Pakkenberg and Voigt, 1964; Dekaban, 1978; Peters et al., 1998; Witelson et al., 2006] and certain physical changes to *Homo* over time correlate with changes to the skull [Pilbeam and Gould, 1974; Frayer, 1981; Ruff et al., 1997], including decreases in size during some of the past 50,000 years [Henneberg, 1988; Ruff et al., 1997]. Body size has been shown to correlate with cranial capacity changes by as much as 45% across much of *Homo* evolution [Pilbeam and Gould, 1974; Beals et al., 1984; Henneberg, 1988; Ruff et al., 1997; McHenry and Caffing, 2000; Rightmire, 2004], so declines in physical size and stature have likely driven proportional declines in brain size.

Despite the correlation between body and brain size, changes to body mass and stature do not uniformly match brain size trends within animal species [Deaner et al., 2007], including hominins [Herculano-Houzel and Kaas, 2011]. While correlations between body and brain size exist for hominins, the ratios do not have a consistent relationship over time [Pilbeam and Gould, 1974; Ruff et al., 1997; McHenry and Caffing, 2000], and across certain periods the correlations appear negative [Henneberg, 1988]. Assuming relative brain size remained evolutionarily advantageous to *Homo*, encephalization would likely have continued to increase, irrespective of body size declines. These discrepancies have caused speculation and led to theories suggesting behavioral reasons for the recent decline in brain size, such as sexual selection [Bednarik, 2014], sociality [Hare, 2017], shifting cognitive demands to tools and technology [Bednarik, 2014], emotional plasticity [Dor Shilton et al., 2020], and self-domestication [Leach, 2003; Theofanopoulou et al., 2017; Benitez-Burraco et al., 2020].

Regardless of whether brain size declines are driven in response to other physical changes to the body versus behavioral adaptations, it is still not clear whether general cognitive function in humans has actually declined. The link between brain size and cognitive ability is spurious at best, but the relationship appears to hold strong validity when looked at with regards to evolutionary changes within species [Bouchard Jr. et al., 1990; Posthuma et al., 2002; Posthuma et al., 2003; Deaner et al., 2007; Pietschnig et al., 2015; Sniekers et al., 2017; Davies et al., 2018; Nave et al., 2018]. While various measures associated with general cognitive ability – particularly with regards to cognitive function and educational attainment – have been found to correlate at the genetic level with brain size [Bouchard Jr. et al., 1990; Posthuma et al., 2002, 2003; Deaner et al., 2007; Pietschnig et al., 2015; Sniekers et al., 2017; Davies et al., 2018; Nave et al., 2018; Hill et al., 2019], cognitive ability is notoriously difficult to test. Cognitive ability is subjective, dependent on interpretation, and measures can be culturally, environmentally, and educationally biased [Reynolds et al., 1984; Richard-
Modern humans (19) 0.032 75.42±3.492 1.73±0.020 – 1,345±30 25.305±1.275 4.616±0.199
Modern humans, low BMI (<25) group (9) 0.032 64.33±4.631 1.716±0.037 – 1,343±47 20.72±1.041 5.175±0.284
Modern humans, high BMI (>25) group (10) 0.032 85.4±2.414 1.708±0.028 – 1,420±30 29.42±1.151 4.693±0.240
Modern humans, low latitude (8) 0.032 71.5±6.668 1.721±0.045 – 1,341±31 24.34±2.574 4.509±.356
Modern humans, high latitude (11) 0.032 78.27±3.653 1.736±0.014 – 1,420±30 26.00±1.252 4.693±0.110
H. sapiens (29) 1 54.98±1.42 1.308±23 1.263±22 5.402±0.109
Holocene AM Homo (1)f 10.2 56.76 – 1.430 1.378 5.703
Late Pleistocene AM Homo (25) 15-38 63.47±1.897 1.46±32 1.40±30 5.386±0.071
Late Middle Pleistocene AM Homo (5) 90-195 67.92±1.927 1.519±29 1.46±27 5.284±0.110
All Hominins, Pliocene and Pleistocene (41) 15-3,180 66.14±1.769 1.439±38 1.38±36 5.122±0.108

a See Methods and online supplementary Data 1 for details on individual measures and formulas for deriving body mass, height, cranial capacity, brain mass, BMI, and EQ.
b Modern human sample comprised of 11 German and 8 Australian Aboriginal male specimens, of which 4 and 5, respectively, were low BMI.
c H. sapiens sample comprised of 14 female and 15 male Pecos Pueblo specimens.
d Veyrier 1, male.
e Late Pleistocene sample comprised of 9 female, 12 male, and 4 unknown gender specimens originating from high latitudes.
f Late Middle Pleistocene sample comprised of 1 female, 2 male, and 2 unknown gender specimens originating from high latitudes.
g Pliocene and Pleistocene sample comprised of 13 female, 19 male, and 9 unknown gender specimens originating from high latitudes, with the exception of the two earliest specimens which came from low latitudes.

The present study attempts to provide some clarity on why brain size has declined over the past 50,000 years and whether such declines have influenced general cognitive ability in humans. First, the physical substrates for brain size reductions in modern and pre-historic Homo are examined by comparing brain weights of modern and pre-historic specimens. Separately, a meta-review of available GWAS studies is performed specifically with respect to analyzing selective pressures that may be influencing human cognitive abilities, including general cognitive function and educational attainment. The results offer some insight into the selective pressure that has been placed on the brain over the past 50,000 years.
Holocene data were derived from Ruff et al. [1997] and Pliocene and Pleistocene body mass estimates leveraged Ruff et al. [2018], in each case utilizing updated data where available, as provided by the lead authors. Body mass estimates for the Pecos Pueblo dataset and Veyrier 1 skeletal remains were updated based on the latest formulas for femoral head (FH) estimations (body mass = 2.262 × FH – 38.7), which conforms to the pre-Holocene data utilized [Ruff et al., 2018]. No changes were made to samples estimated from bi-iliac breadth. Dates for each of the remains were updated to reflect the latest estimates available in the literature. The final dataset and related notes can be found in online supplementary Data 1.

Prehistoric cranial capacity was measured as a function of endocranial volume in cm³. Cranial measurements were obtained from either original or aggregated sources, and two updated data files from one researcher [Holloway, 1981; Ruff et al., 1997; De Miguel and Henneberg, 2001; Hawks and Wolpoff, 2001; Lee and Wolpoff, 2003; Rightmire, 2004; Mannino et al., 2011; Lordkipanidze et al., 2013]. Endocranial estimates were made through volume measurements using rice, seeds, shot, sand, water, or other fill material. Where multiple estimates were available in the literature for a single skull, the average of all estimates was used, which acted as a control to account for measurement variability. Brain mass (g) was approximated from cranial capacity using a formula derived from a least-squares regression of 27 primate species as provided by Ruff et al. [1997] (brain mass = 1.147 × cranial capacity). The complete dataset and related notes and rationale for inclusion are available in online supplementary Data 1.

For modern humans, methods are straightforward and can be made through direct measurements of recently deceased individuals. However, very few datasets of matched body and brain weights exist, as autopsy studies with the level of granularity required to properly separate body components are infrequent and extremely difficult [Zihlman and Bolter, 2015]. A modern sample was compiled that included matched body (kg) and brain (g) weights along with height (m) measurements. Autopsy data was utilized from 19 deceased individuals without known trauma to the brain, psychological issues, or pathological deficits (Table 1; online suppl. Data 1) [Klekamp et al., 1987]. The sample study measured total body mass, height, and fresh brain weights of 11 male Caucasians of German origin and 8 Australian Aboriginal males. The German specimens originated from the Institute of Forensic Medicine, University of Hamburg, Germany, and the Aboriginal specimens came from the Department of Neuropathology, Royal Perth Hospital, Perth, Australia. Specific dates of death were not known but all subjects were deceased between 1980 and 1982, so 1981 (0.032 kyr BP) was used as the date of death across all specimens.

Encephalization (EQ) was measured using a variant of the method originally derived based on changes to surface area [Jerison, 1973] but updated to account for metabolic turnover [Martin, 1981] as follows: EQ = brain mass/(11.22 × body mass^0.76). Body fat estimates in modern humans were approximated by comparing the relationship of body mass to stature by calculating BMI (weight [kg]/height [m]^2) [Gallagher et al., 2000]. Because of the limited availability of prehistoric fossils and modern autopsy data, the sample was grouped by time period irrespective of sex or geography. While brain and body size differences have been shown to exist across sex and latitudinal clines [Beals et al., 1984; Ruff et al., 1997; McHenry and Coffing, 2000; Ruff, 2002], encephalization levels were not impacted in any time period across any of the samples tested (p = 0.675, t test across latitude for modern humans; p = 0.151, t test across sex for H. sapiens; p = 0.456, t test across sex for Late Pleistocene). This is consistent with prior research that has shown that sex and geography account for less than 2 and 4%, respectively, of encephalization differences within time periods [Ruff et al., 1997]. As a result, encephalization levels appear to be relatively constant within taxon at similar time periods regardless of differences across populations. This permits greater generalization of the encephalization results despite a smaller and less uniform sample.

Specimens were categorized by temporal period, as outlined in the Results section and online supplementary Data 1, and differences in brain mass, body mass, and encephalization were analyzed using independent samples t tests. The modern human and H. sapiens sample set were also grouped by sex and latitude and compared using independent samples t tests to control for sex and geography. Ordinary least squares linear regression (LSR) was used to assess the scaling relationships over time between brain and body mass leveraging the formula for EQ (brain mass/(11.22 × body mass^0.76)). A one-way ANOVA followed by Tukey’s multiple comparisons tests were used to evaluate whether the change in slope of the LSR lines was significant. All analyses were performed in XLSTAT version 2020.5 at alpha = 0.05. Values are shown as the mean ± standard error (SE).

One potential confound deserves special attention: encephalization was used as the primary proxy for cognitive faculty for prehistoric Homo. Relative and absolute brain size have been found to be strong correlates of cognitive function in mammals and humans [Jerison, 1973; Martin, 1981; Bouchard Jr. et al., 1990; Posthumus et al., 2002; Posthumus et al., 2003; Deanner et al., 2007; Pietschnig et al., 2015; Snickers et al., 2017; Davies et al., 2018; Nave et al., 2018]. However, there are limitations to using encephalization to determine cognitive ability without consideration for other factors such as neuron count, neuronal density, interneuron distance, axonal conduction velocity, cortex scaling, and other variables found to be instructive of cognitive ability in humans and other animals [see for instance: Striedter, 2005; Finlay and Workman, 2013; Dicke and Roth, 2016]. Given the nature of prehistoric remains, efforts to determine cognitive ability are necessarily limited to physical remains.

To account for some of these limitations, a meta-review of studies that directly tested selection against phenotypes associated with general cognitive ability was used to test whether there has been an evolutionary impact to cognitive function in humans. Two phenotypes were used that have been shown to correlate closely with general cognitive ability: tests of cognitive function [Reynolds et al., 1984; Plomin and Deary, 2015] and surveys of educational attainment [Bouchard Jr. et al., 1990; Deary et al., 2007; Rietveld et al., 2013; Rietveld et al., 2014; Plomin and Deary, 2015; Okbay et al., 2016; Snickers et al., 2017]. (Unlike tests of cognitive function, education is an indirect phenotype not directly acted on by natural selection. Nevertheless, both phenotypes correlate closely and have been used as proxies for general cognitive ability in genome studies [Bouchard et al., 1990; Rietveld et al., 2014].) To test for effects of natural selection, various phenotypes associated with reproductive success were used that have been shown to influence evolutionary fitness across successive generations [Zietsch et al., 2014].

Multiple genome datasets were utilized within the different studies, helping to extend the power of the meta-analysis and pro-
vide for a more diverse and robust sample. Datasets were derived from the following sources, all of which provided at least one genotype that has been shown to correlate significantly with general cognitive ability: CHARGE Cognitive Consortium, UK Biobank, Generation Scotland: Scottish Family Health Study, Health and Retirement Study, deCODE Genetics, Snickers GWAS, Barban GWAS, Rietveld GWAS, and Okbay GWAS. It should be noted that the majority of genome datasets are available to western cultures and not representative of the global population.

Participants in the CHARGE Cognitive Consortium (n = 53,949) were of European descent split across 31 geographies, skewed slightly toward women (n = 58,914) and men (n = 53,237). A single test of cognitive ability was used that examined verbal-numerical reasoning. The test contained 13 questions (6 verbal; 7 numerical) and was completed within 2 minutes. Reaction time and memory was also tested but the tests used were not found to correlate strongly enough to general cognitive ability to be used in the results. Of the total number of subjects, 36,035 were genotyped for verbal-numerical reasoning and 111,114 for educational attainment at a college level (binary for college degree). Specific details of the dataset and genotyping are available in the original report [Davies et al., 2015].

Participants in the UK Biobank (n = 112,151) included subjects aged 40–73 years during the time period of 2006 and 2010 and were roughly evenly split between women (n = 58,914) and men (n = 53,237). The majority of participants were questioned after the age of 40 years during the time period of 2006 and 2010 and were roughly evenly split between women (n = 58,914) and men (n = 53,237). A single test of cognitive ability was used that examined verbal-numerical reasoning. The test contained 13 questions (6 verbal; 7 numerical) and was completed within 2 minutes. Reaction time and memory was also tested but the tests used were not found to correlate strongly enough to general cognitive ability to be used in the results. Of the total number of subjects, 36,035 were genotyped for verbal-numerical reasoning and 111,114 for educational attainment at a college level (binary for college degree). Specific details of the dataset and genotyping are available in the original report [Davies et al., 2016].

Participants in the Generation Scotland: Scottish Family Health Study (n = 24,090) included subjects with a median age of 47.6 years that was skewed toward female (female, n = 14,163; male, n = 9,927). Most participants provided blood or saliva samples (n = 23,919) and lived in Glasgow, Tayside, Ayrshire, Arran, and North-East Scotland. After controlling for errors, genotyped data was provided for 20,032 individuals (female = 11,804; male = 8,228). The majority of participants were questioned after the age of 40 years during the time period of 2006 and 2010 and were roughly evenly split between women (n = 58,914) and men (n = 53,237). A single test of cognitive ability was used that examined verbal-numerical reasoning. The test contained 13 questions (6 verbal; 7 numerical) and was completed within 2 minutes. Reaction time and memory was also tested but the tests used were not found to correlate strongly enough to general cognitive ability to be used in the results. Of the total number of subjects, 36,035 were genotyped for verbal-numerical reasoning and 111,114 for educational attainment at a college level (binary for college degree). Specific details of the dataset and genotyping are available in the original report [Davies et al., 2016].

Participants in the Health and Retirement Study were Americans born between 1900 and 1992, with the vast majority of the genotyped individuals (>95%) having been born before 1953. The longitudinal study has been interviewing participants roughly every 2 years since 1992. One of the survey questions for the study relates to educational attainment and another to fitness (lifetime reproductive success). Genotyped data on these phenotypes were available on a subset of individuals in the study and a reduced set, mitigating certain confounds, was used in the primary research [Beauchamp, 2016; Conley et al., 2016], details of which can be found in the Results section herein.

The deCODE Genetics genealogical database contains information on roughly 840,000 individuals from Iceland. deCODE provides information on roughly all of the 317,000 living Iceland residents and broad ancestral data dating back to 1650, with the oldest subject having a recorded birth of 740 AD. Where available, the database includes year of birth, year of death, gender, family linkage, place of birth, primary residence, residence at time of death, and educational attainment. The database is principally composed of 14 Icelandic censuses from 1703 to 1930, parish records dating back to 1780, the Iceland national registry dating back to 1994, available Icelandic annals, genealogical publications, biographical lists of members from professional associations, and available verifiable records. Additional information can be found in the primary study that used this dataset [Kong et al., 2017].

The Snickers GWAS employed thirteen different datasets that sampled 78,308 individuals. The first two datasets come from the UK Biobank described above but a web-based measure of cognitive ability was added that had not been previously published. Six of the remaining cohorts came from the CHIC Consortium, which included the Avon Longitudinal Study of Parents and Children (n = 5,517), the Lothian Birth Cohorts of 1921 (n = 464) and 1936 (n = 947), the Brisbane Adolescent Twin Study subsample of Queensland Institute of Medical Research (n = 1,752), the Western Australian Pregnancy Cohort Study (n = 936), and the Twins Early Development Study for children aged 6–18 years (n = 2,825). The final five cohorts were constructed by using the following datasets: the Erasmus Rucphen Family Study (n = 1,076), the Generation R Study (n = 3,701), the Harvard Union Study (n = 389), the Minnesota Center for Twin and Family Research Study (n = 3,367), and the Swedish Twin Registry Study (n = 3,215). Various tests of cognitive function were performed, and more details can be found in the original supplementary information [Snijders, 2017] and within the research outlined in the Results section [Hill et al., 2019].

The Barban GWAS employed 42 different datasets with a total sample size across the meta-analysis of 251,151 for reproductive onset (women, n = 189,656; men, n = 48,408) and 343,072 for reproductive success (women, n = 225,230; men, n = 103,909). Genotyping was run for each of educational attainment, reproductive onset, and reproductive success (number of children ever born), along with other phenotypes that were not relevant to the current study. A detailed analysis of the design and methods can be found in the original study’s supplementary information [Barban et al., 2016].

The Rietveld GWAS employed 62 different datasets that sampled 251,151 individuals for educational attainment and 343,072 individuals for college graduation. Subjects were of European descent. The majority of participants were questioned after the age of 30 years and each study employed independent surveys and methodologies for obtaining each phenotype, all of which were standardized on a 7-point scale for educational attainment and a binary scale for college degree. Genotyping was run for each of educational attainment and college graduation, along with other phenotypes that were not relevant to the current study. A detailed analysis of the study’s design and methods can be found in the supplementary information to the original study [Rietveld et al., 2013].

The Okbay GWAS was an update to the Rietveld GWAS that sampled 293,723 individuals for educational attainment and 280,007 individuals for college graduation. Standardization of...
each phenotype followed the method previously outlined in the Rietveld GWAS [Rietveld et al., 2013]. There was significant overlap between the two GWAS datasets (overlap of 126,413 for educational attainment) and they were significantly correlated as a result. Genotyping was run for each of educational attainment and college graduation, along with other phenotypes that were not relevant to the current study. A detailed analysis of the design and methods can be found in the supplementary information to the original study [Okbay et al., 2016].

Across each GWAS dataset, all phenotypes associated with cognitive ability were included in the meta-review to the extent that they were genotyped and there was a corollary phenotype tied to fitness available. For fitness, the total number of children ever born (reproductive success) was the primary phenotype tested against but age at first birth, average age at childbirth, and age at first sexual encounter were also used as each have been considered strong proxies for fitness [Barban et al., 2016; Day et al., 2016; Kong et al., 2017; Hill et al., 2019].

**Results**

A general trend toward declining encephalization appears to be present throughout the Holocene. The modern human sample was 17% less encephalized than the *H. sapiens* sample (*p* < 0.001, *t* test, entire sample; *p* < 0.01, controlling for sex and latitude; Fig. 1). While there is only one known Early Holocene AM *Homo* (dated 10 kyr BP) with both body and brain mass available, the specimen represented a relatively large encephalization level, roughly 5% larger than the mean for the *H. sapiens* and 19% larger than the modern human mean (Table 1).

It was initially believed that encephalization scaled following a linear trajectory throughout prehistory [Pilbeam and Gould, 1974] but subsequent evidence suggests that there have been a number of punctuated points of en- cephalization changes across the genus *Homo* [Walker, 1993; Ruff et al., 1997; Shultz et al., 2012; Püschel et al., 2021]. Walker initially proposed differences in encephalization between AM *Homo* and earlier species, and it was later found that there were at least three demonstrable gradients [Ruff et al., 1997; Shultz et al., 2012; Püschel et al., 2021], as well as a period of stasis in *Homo* encephalization between 1,800 and 600 kyr BP before renewed growth lasting throughout the Late Pleistocene [Ruff et al., 1997]. The current findings demonstrate a trend reversal in encephalization during the Holocene and into modern times. When placed in the broader context of encephalization over the past 4 million years, the results suggest that encephalization levels continued to climb throughout the Late Pleistocene but have declined at least across the past 1,000 years (Fig. 1a, b).

The modern humans in particular demonstrated lower encephalization than during earlier periods of the Holocene (*p* < 0.001, *t* test), Late Pleistocene (*p* < 0.001, *t* test), and even the late Middle Pleistocene (*p* < 0.01, *t* test). A contributing factor may be changes in diet. Greater availability of food during modern times has increased consumption amongst human populations, which has yielded disproportionate increases in non-functional body mass, such as large fat reserves [Bellisari, 2008]. As a result, body mass as it relates to height (the body mass index or BMI) has risen significantly in modern times, owing in large part to obesity [Bellisari, 2008; Gallagher et al., 2000]. With an average index of 25.3, the modern sample had a BMI that is considered overweight (25–30) [Gallagher et al., 2000]. Given that a significant amount of brain mass is attributed to bodily function and not fatty tissue [Schoenemann, 2004; Font et al., 2019], it is likely that increases in fatty body mass disproportion-
ately reduced encephalization in the modern sample relative to earlier periods.

Whereas body mass has been shown to correlate with brain mass throughout much of Homo history [Von Bonin, 1934; Tobias, 1971; Pilbeam and Gould, 1974; Beals et al., 1984; Henneberg, 1988; Henneberg and Steyn, 1993; Ruff et al., 1997; McHenry and Coffing, 2000; Rightmire, 2004; Burini and William, 2018], the modern human sample did not show any meaningful correlations (Fig. 2a). This is in contrast to each of the other samples used herein, including the Early Holocene group, all of which demonstrated significant correlations between body and brain mass (Fig. 2a). Increases in body fat composition between modern and prehistoric Homo may have contributed to this effect. Given that earlier anatomically modern H. sapiens were leaner than humans today [Bellisari, 2008], body fat differences may have diluted the relationship between body and brain mass in the modern sample. This theory is consistent with animal studies which have shown that gut size [Tsuboi et al., 2015] and body fat [Navarrete et al., 2011] inversely correlate with brain size in certain animals. The theory is reinforced by a highly significant correlation between BMI and encephalization across the modern sample (Fig. 2b).

Comparing EQ between the overweight BMI group (>25) and the remaining modern specimens yields a highly significant difference ($p < 0.004$, $t$ test). When compared to the earlier populations, the lower BMI modern human sample demonstrated no meaningful difference in encephalization across the earlier Holocene sample ($p = 0.125$, $t$ test) or the Late Pleistocene sample ($p = 0.245$, $t$ test), suggesting that the differences in EQ were likely due to higher levels of obesity in modern times (Table 1).

For modern humans, BMI appears to be the primary driver of encephalization differences, indicating that lean body mass could be a better determinate of encephalization than total body mass.

While declines in brain mass appear to be an intermediate response due to selective pressures resulting from body size changes, the reduction in brain mass may still have an impact on cognitive ability. From a genetic standpoint, evolutionary changes in brain mass have been found to correlate strongly with changes in cognitive ability [Bouchard Jr. et al., 1990; Posthuma et al., 2002, 2003; Deaner et al., 2007; Pietschnig et al., 2015; Sniekers et al., 2017; Davies et al., 2018; Nave et al., 2018]. Few genomic studies, however, have examined the genetic variants associated with cognitive abilities or investigated whether these variants are associated with traits that drive evolutionary fitness. To test for genetic trends in cognitive ability, a meta-review of GWAS studies was performed (see Methods). Across seven GWAS studies that directly tested selection against phenotypes associated with general cognitive ability (cognitive function and educational attainment), the results indicated selective pressures acting against cognitive ability and educational attainment (Table 2).

Leveraging data from the Sniekers GWAS and the UK Biobank, cognitive ability ($n = 199,242$) and total years of education ($n = 329,417$) were found to be under negative
selection pressure [Hill et al., 2019]. Multiple tests of cognitive ability were used for the Sneakers GWAS and those sets of tests were combined into a single phenotype. The verbal-numerical reasoning test was used as a proxy phenotype for cognitive ability in the UK Biobank. Each phenotype was genotyped and compared against two reproductive traits that have been shown to drive selection: “age at first birth” and total “number of children” [Barban et al., 2016; Day et al., 2016; Hill et al., 2019]. In all cases, general cognitive ability and education were found to be under negative selection.

Using data from the UK Biobank, a study found that both cognitive ability and educational attainment negatively correlated with reproductive success and age at first sexual intercourse [Day et al., 2016]. Surveys were used to assess both educational attainment (1 question) and cognitive ability (13 questions). Of the relevant correlations, significant enrichment was found with early reproductive onset (negative genetic correlation) and lifetime reproductive success (negative genetic correlation), for both educational attainment and cognitive function, indicating a selection bias against educational attainment and cognitive function.

Using data from the deCODE genetics database, the Oakbay GWAS, and the UK Biobank, a study found that a negative selection bias was present for educational attainment [Kong et al., 2017]. Across 109,120 individuals (female, n = 58,560; male, n = 50,560) born between 1910 and 1975, a negative correlation was found between educational attainment and three measures of fitness: reproductive success, early reproductive onset, and average age at first childbirth. In addition, the study found a correlation between educational attainment and cognitive ability (as measured by the Weschler Abbreviated Scale of Intelligence, recorded between 2009 and 2016) across a smaller sample within the deCODE data (n = 1,577). Leveraging the correlation, the study estimated a decline of between 0.038 and 0.30 intelligence (IQ) points per decade depending on which polygenic score was used for educational attainment.

### Table 2. Negative selection found in cognitive function and educational attainment across seven GWAS

| Sample | Cognitive trait (genotype unless noted) | Selection trait | Genetic correlation rg ± SE (p) |
|--------|----------------------------------------|-----------------|--------------------------------|
| UK Biobank | Cognitive function (UK Biobank) | Age at first birth | 0.56±0.0229 (2.25E–133) |
| UK Biobank | Cognitive function (UK Biobank) | Number of children | –0.23±0.027 (8.03E–18) |
| Sneakers | Cognitive function (Sneakers) | Age at first birth | 0.45±0.0342 (3.82E–39) |
| Sneakers | Cognitive function (Sneakers) | Number of children | –0.17±0.039 (8.75E–06) |
| UK Biobank | Cognitive function | Age at first sex | Reported as “significant” |
| UK Biobank and Sneakers | Education | Age at first birth | 0.72±0.0221 (9.29E–235) |
| UK Biobank and Sneakers | Education | Number of children | –0.26±0.0309 (1.46E–17) |
| US Health and Retirement | Education | Number of children | –0.041±0.009 (<0.01) |
| US Health and Retirement | Education female | Number of children | –0.033±0.010 (<0.01) |
| US Health and Retirement | Education male | Number of children | –0.031±0.012 (<0.05) |
| US Health and Retirement and Rietveld | Education | Number of children | –0.037 (9.91E–04) |
| Barban | Education | Age at first birth | 0.712±0.022 (2.72E–239) |
| Barban | Education | Number of children | –0.236±0.031 (6.63E–14) |
| US Health and Retirement | Education | Number of children | –0.131±0.047 (<0.001) |
| UKBiobank | Education | Age at first sex | 0.305±0.074 (3.70E–05) |
| deCODE, Oakbay, and UK Biobank | Education female | Number of children | –0.097 (1.7E–23) |
| deCODE, Oakbay, and UK Biobank | Education female | Age at first birth | 0.59 (1.3E–70) |
| deCODE, Oakbay, and UK Biobank | Education male | Average age at childbirth | 0.42 (6.4E–76) |
| deCODE, Oakbay, and UK Biobank | Education male | Number of children | –0.053 (3.50E–06) |
| deCODE, Oakbay, and UK Biobank | Education male | Age at first birth | 0.43 (3E–22) |
| US Health and Retirement | Education female | Average age at childbirth | 0.36 (1.4E–19) |
| US Health and Retirement | Education male (phenotype) | Number of children | –0.057±0.003 (<0.001) |
| US Health and Retirement | Education male (phenotype) | Number of children | –0.022±0.003 (<0.001) |
| US Health and Retiremen and Rietveld | Education (phenotype) | Number of children | –0.166 (0) |

Adapted from: a Hill et al. [2019]; b Beauchamp [2016]; c Conley et al. [2016]; d Barban et al. [2016]; e Conley and Domingue [2016]; f Day et al. [2016]; g Kong et al. [2017].

1 Regression (rg), standard error (SE), and significance (p) were provided where available.
2 Noted in the original study as significant but specifics were not provided in the main article or supplementary materials.
In a study using data from the US Health and Retirement genome study, education was found to be under negative selection pressure [Beauchamp, 2016]. Total years of education was genotyped and compared against lifetime reproductive success for both male (phenotypic $n = 5,436$; genotypic $n = 2,571$) and female (phenotypic $n = 6,414$; genotypic $n = 3,416$) Americans of European decent born between 1931 and 1953. The results imply that the genetic impact of the selection bias is a drop of roughly 1.5 months of education per generation.

A second study utilizing data from the Health and Retirement study found a similar selection bias against educational attainment using lifetime reproductive success as a proxy for fitness [Conley et al., 2016]. The study was primarily looking at assertive mating and therefore limited its dataset to 2,343 married couples ($n = 4,686$), which was further reduced by 871 respondents that did not have answers to the reproductive success question. The sample was increased by adding additional subjects which yielded a total of 8,855 participants. Polygenic scores were computed for each of the phenotypes and results demonstrated a small but statistically significant bias toward lower educational attainment.

Another negative selection bias was found using data from the Health and Retirement study and the Rietveld GWAS [Conley and Domingue, 2016]. The results found were similar to the other studies using the Health and Retirement study, but the power was enhanced by including the Rietveld GWAS ($n = 8,865$). Consistent with the other Health and Retirement Studies, a small but significant inverse correlation was found between education and reproductive success, indicating negative selection. The study noted that the effect is such that for each standard deviation of educational attainment achieved, there is a corresponding drop of roughly 1/15th in reproductive success.

In a genome study looking for phenotypes that influence reproductive behavior using the Barban GWAS [Barban et al., 2016], a negative selection bias was found between educational attainment and both reproductive success ($n = 343,072$) and early reproductive onset ($n = 251,151$).

Two additional studies provide further support of selection operating against cognitive ability and educational attainment using alternative or indirect methods. In one case, as part of a series of studies investigating genetic ties between general cognitive function and psychological disease, it was found that regions of the genome that have been under negative selection drive a large proportion of the heritability of cognitive ability [Hill et al., 2016]. Using data from the CHARGE Cognitive Consortium and the UK Biobank, both general cognitive function (CHARGE, $n = 53,949$) and verbal-numerical reasoning (UK Biobank, $n = 36,035$) were found to be under selective pressure.

A second study found that rare genetic variants contribute to negative selection for cognitive function but not educational attainment. Various phenotypes associated with general cognitive ability were compared against hereditary data to determine possible correlations and selection bias for rare variants associated with each phenotype across the Generation Scotland dataset [Hill et al., 2018]. This study provides the closest link between the recent genetic work and the historical work performed by psychologists, demographers, and sociologists. The study looked at kinship, sibling pairs, and familial relationships over time and correlated them at the genotype level to limit environmental confounds. In addition to comparing common variants, the study also included rare genetic variants for education and general cognitive ability. Tests of heredity and fitness were derived leveraging data on kinship and familial relationship. Educational attainment, four individual tests of cognitive function (the Mill Hill Vocabulary Scale, the Wechsler Digit Symbol Substitution Task, the Wechsler Logical Memory, and a test of executive function) and a combined general measure of cognitive ability derived from those four tests were used as measures associated with general cognitive function. Both education and general cognitive ability correlated closely to one another, as did each of the four independent tests. General cognitive ability was found to be under negative selection; however, in contrast to every other study reviewed, education did not appear to be under evolutionary pressure. This is an important finding given how closely education and general cognitive function typically correlate with one another. The authors suggested that because the results relied on rare genetic variants which may have separated the causal connection between education and cognitive function, it may be that general cognitive ability is an intermediate variable acting on educational attainment. Thus, when the independent variants across each phenotype are separated, education may not be under negative selection independent of cognitive ability.

**Discussion**

The current research supports a growing body of evidence demonstrating a decline in human brain size since at least 50 kyr BP [Henneberg, 1988; Henneberg and Steyn, 1993; Ruff et al., 1997]. As compared to the Upper
### Table 3. General intelligence gains (Flynn effect) across multiple cognitive performance tests for 14 nations from 1932 to 2006

| Location | Psychological test | Dates | IQ gain | IQ gain annualized |
|----------|--------------------|-------|---------|--------------------|
| Leipzig⁵ | Ravens             | 1968–1978 | 10–15  | 1.250              |
| France⁶  | Ravens             | 1949–1974 | 25.12  | 1.005              |
| Belgium⁵ | Ravens             | 1958–1967 | 7.15   | 0.794              |
| Belgium⁵ | Shapes             | 1958–1967 | 6.45   | 0.716              |
| Netherlands⁵ | Ravens | 1952–1982 | 20     | 0.667              |
| Norway⁴  | Matrices           | 1954–1968 | 8.8    | 0.629              |
| West Germany⁴ | Horn-Ravens (altered) | 1961–1978 | 10     | 0.588              |
| Australia⁴ | Jenkins         | 1949–1981 | 15.67  | 0.490              |
| Edmonton⁴ | Ravens             | 1956–1977 | 8.44   | 0.402              |
| Australia⁴ | Ravens             | 1950–1976 | 8.76   | 0.337              |
| Norway⁴  | Matrices           | 1968–1980 | 2.6    | 0.217              |
| Great Britain⁴ | Ravens       | 1938–1979 | 7.75   | 0.189              |
| Great Britain⁴ | Ravens       | 1940–1979 | 7.07   | 0.181              |
| Japan⁵   | Wechsler (altered) | 1951–1975 | 20.03  | 0.835              |
| Vienna⁵  | Wechsler           | 1962–1979 | 12–16  | 0.824              |
| West Germany⁴ | Wechsler       | 1954–1981 | 20     | 0.741              |
| Zurich⁵  | Wechsler           | 1954–1977 | 10–20  | 0.652              |
| Edmonton⁴ | CTMM              | 1956–1977 | 11.03  | 0.525              |
| France⁶  | Wechsler (altered) | 1955–1979 | 9.12   | 0.380              |
| USA⁶     | Wechsler-Binet (altered) | 1932–1972 | 12–17  | 0.300              |
| USA⁶     | Wechsler (altered) | 1954–1978 | 8.76   | 0.337              |
| Solothurn⁵ | Wechsler       | 1977–1984 | 1.3    | 0.186              |
| Saskatchewan⁶ | Otis (altered) | 1958–1978 | 12.55  | 0.628              |
| Norway⁴  | Verbal-Math        | 1954–1968 | 8.15   | 0.582              |
| Belgium⁵ | Verbal-Math        | 1958–1967 | 3.67   | 0.408              |
| France⁶  | Verbal-Math        | 1949–1974 | 9.35   | 0.374              |
| Saskatchewan⁶ | Otis (altered) | 1958–1978 | 6.95   | 0.348              |
| New Zealand⁶ | Otis            | 1936–1968 | 7.73   | 0.242              |
| Norway⁴  | Verbal-Math        | 1968–1980 | –1.6   | –0.133             |
| USA⁶     | Stanford-Binet, 1932 and Wechsler, 1947 1/2 | 1932–1947.5 | 5.76   | 0.372              |
| USA⁶     | Stanford-Binet, 1932 and Wechsler, 1953 1/2 | 1932–1953.5 | 7.49   | 0.348              |
| USA⁶     | Stanford-Binet, 1932 and Wechsler, 1964 1/2 | 1932–1964.5 | 8.96   | 0.276              |
| USA⁶     | Stanford-Binet, 1932 and Stanford-Binet, 1971 1/2 | 1932–1971.5 | 9.89   | 0.250              |
| USA⁶     | Wechsler, 1936 1/2 and Wechsler, 1947 1/2 | 1936.5–1947.5 | –2.03  | –0.185             |
| USA⁶     | Wechsler, 1936 1/2 and Wechsler, 1953 1/2 | 1936.5–1953.5 | 4.69   | 0.276              |
| USA⁶     | Wechsler, 1947 1/2 and Wechsler, 1953 1/2 | 1947.5–1953.5 | 2.64   | 0.440              |
| USA⁶     | Wechsler, 1947 1/2 and Wechsler, 1964 1/2 | 1947.5–1964.5 | 2.7    | 0.159              |
| USA⁶     | Wechsler, 1947 1/2 and Wechsler, 1971 1/2 | 1947.5–1971.5 | 11.98  | 0.499              |
| USA⁶     | Wechsler, 1947 1/2 and Wechsler, 1972 | 1947.5–1972 | 8.41   | 0.343              |
| USA⁶     | Wechsler, 1953 1/2 and Wechsler, 1972 | 1953.5–1972 | 6.65   | 0.359              |
| USA⁶     | Wechsler, 1953 1/2 and Wechsler, 1978 | 1953.5–1978 | 8.04   | 0.328              |
| USA⁶     | Wechsler, 1964 1/2 and Stanford-Binet, 1971 1/2 | 1964.5–1971.5 | 4.41   | 0.630              |
| USA⁶     | Wechsler, 1964 1/2 and Wechsler, 1972 | 1964.5–1972 | 4.26   | 0.568              |
| USA⁶     | Wechsler, 1972 and Wechsler, 1978 | 1972–1978 | 0.96   | 0.160              |
| USA⁶     | WAIS-III (1995) and Stanford-Binet-S (2001) | 1995–2001 | 5.5    | 0.917              |
| USA⁶     | WAIS-R (1978) and Stanford-Binet-4 (1985) | 1978–1985 | 3.42   | 0.489              |
| USA⁶     | WAIS-III (1995) and WISC-IV (2001.75) | 1995–2001.75 | 3.1    | 0.459              |
| USA⁶     | WISC-III (1989) and Stanford-Binet-S (2001) | 1989–2001 | 5      | 0.417              |
| USA⁶     | WISC-III (1989) and WISC-IV (2001.75) | 1989–2001.75 | 4.23   | 0.332              |
| USA⁶     | WISC-R (1972) and WISC-III (1989) | 1972–1989 | 5.3    | 0.312              |
| USA⁶     | WISC-R (1972) and Stanford-Binet-4 (1985) | 1972–1985 | 2.95   | 0.227              |
| USA⁶     | Stanford-Binet-4 (1985) and Stanford-Binet-S (2001) | 1985–2001 | 2.77   | 0.173              |
| USA⁶     | WAIS-R (1978) and WAIS-III (1995) | 1978–1995 | 4.2    | 0.247              |
| USA⁶     | Stanford-Binet-LM (1972) and Stanford-Binet-4 (1985) | 1972–1985 | 2.16   | 0.166              |
| USA⁶     | WISC-R (1972) and WAIS-R (1978) | 1972–1978 | 0.9    | 0.15               |
| USA⁶     | WISC-III (1989) and WAIS-III (1995) | 1989–1995 | –0.7   | –0.117             |
| USA⁶     | WAIS-III (1995) and WAIS-IV (2006) | 1995–2006 | 3.37   | 0.306              |
| USA⁶     | WISC-IV (2001.75) and WAIS-IV (2006) | 2001.75–2006 | 1.2    | 0.282              |

Average annualized gain: 0.410

Adapted from: ⁵ Flynn [1987]; ⁶ Flynn [1984]; ⁷ Flynn [2009]. IQ annualized gains were derived by comparing data within each study across successive periods and dividing by the number of years per period.
Paleolithic (approx. 50 kyr BP to 15 kyr BP), brain size has declined by 5.415% ($p < 0.001$, t test) in modern humans. In addition to declines in absolute brain size, Homo encephalization has also declined significantly during modern periods.

Body size changes appear to explain most of the recent changes to brain size. With the exception of the modern sample, encephalization levels remained relatively stable across the past 50,000 years. While the modern sample demonstrated a relatively low level of encephalization, increases in BMI appear to have driven much of the change. There is strong evidence that encephalization in mammals is best understood in terms of lean body mass [Schoenemann, 2004] and the present results suggest that lean body mass may be a better measure at least with respect to comparing within species over time. The modern sample, adjusted for BMI, showed no significant differences in encephalization as compared to AM Homo. After controlling for obesity, modern brain and body mass appear to scale isometrically relative to the prehistoric AM Homo sample. The results herein suggest that recent reductions in brain size are an adaptive response to changing physiology, particularly as it relates to body mass changes.

Nevertheless, there is strong evidence that brain mass is highly correlated with cognitive function evolutionarily [Bouchard Jr. et al., 1990; Posthuma et al., 2002, 2003; Deaner et al., 2007; Pietschnig et al., 2015; Sniekers et al., 2017; Davies et al., 2018; Nave et al., 2018]. Absent structural changes that have made the brain more efficient and significant decreases in brain mass could lead to reductions in cognitive function irrespective of encephalization. To some extent, it is possible that the overall makeup of the brain could have evolved toward greater functionality within a smaller cavity. The skull appears to have evolved from an elongated to a more globular shape roughly at the same time of the slowdown in cranial capacity growth (between 100 and 35 kyr BP), indicative of structural changes to the brain [Neubauer et al., 2018]. However, fossil evidence supports relatively distributed brain size reductions [Henneberg and Steyn, 1993] or inconsistent variations [Balzeau et al., 2012; Liu et al., 2014]. One study reported significantly smaller frontal lobes in modern humans as compared to some but not all early Homo and Neanderthal specimens [Balzeau et al., 2012], despite this brain region being attributed to higher levels of cognition. In contrast, another study found that modern brains appear to have larger frontal lobes as compared to early Homo [Liu et al., 2014].

While more work is needed, the overall results of the various GWAS studies that have examined evolutionary changes to cognitive ability suggest that both general cognitive function and educational attainment are under negative selection pressure. While the genetic correlations and underlying relationships are still not fully understood, the data support a genetic decrease in cognitive ability consistent with an evolutionary decline in brain size.

There is a paradox to the genetic data, however: despite the selective pressures on cognitive ability noted in the GWAS studies, measures of general intelligence and educational attainment have all risen during much of the past century [Flynn, 1984, 1987, 2009; Barro and Lee, 2013; Pietschnig and Voracek, 2015; Conley and Domingue, 2016; Lee and Lee, 2016]. Intelligence, as with most phenotypes, is determined by genetic and environmental causes. Short-term changes in general intelligence are largely driven by environmental factors such as health, education, and technology – that can offset or enhance long-term genetic trends [Pietschnig and Voracek, 2015; Bratsberg and Rogeberg, 2018]. Genetic intelligence, in contrast, is driven by heredity. In this way, neither brain size nor genetic intelligence is a pre-determinate of general intelligence at an individual, group, or species level.
Aggregated data from 14 countries over nearly a century demonstrate the long-term positive impact of environmental factors on human intelligence [Flynn, 1984, 1987, 2009], a phenomenon known as the Flynn effect. Gains in IQ scores across all countries averaged 0.410 points per year, with the majority of countries showing significant increases (Table 3) between 1932 and 2006. Similar results have been found for educational attainment, with average gains of roughly 0.068 years of growth annually between 1870 and 2010 across more than 100 countries [Barro and Lee, 2013; Lee and Lee, 2016] (Table 4).

The incongruity between genetic and environmental effects was highlighted in one of the Health and Retirement GWAS studies [Conley and Domingue, 2016], which directly tested whether the effects of negative selection found in polygenic scores of educational attainment manifested themselves in actual decreases in educational attainment. The authors found, consistent with other studies, that educational attainment is increasing in the population despite evolutionary pressures on the phenotype.

Environmental factors are often more transient than genetics so it is not clear whether physical changes to the brain or genetic predispositions will ultimately produce a negative impact on human cognitive ability. There are, however, signs of a possible reversal in the Flynn effect. A significant decrease in IQ has been noted over the past 30 years in many parts of the globe, with the largest declines occurring across industrialized nations [Shayer et al., 2007; Pietschnig and Voracek, 2015; Bratsberg and Rogeberg, 2018; Flynn and Shayer, 2018]. On an evolutionary timescale, environmental improvements may not be able to offset the long-term impact of genetic and physical changes to the brain. This places into question the ability for natural selection in general to drive species level intelligence beyond an upper bound of fitness.

Acknowledgements

Thanks to Christopher Ruff, John Perry, David Hill, Daniel Dennett, and Philip Lieberman for helpful commentary on specific areas of the manuscript; Christopher Ruff for providing expanded datasets; Lindsey Long and Caitlin Mason for assistance with data collection, tabulation, and standardization; Zack Stokes for assistance with mathematical equations and statistical modeling; and Kobe Bryant for not accepting “I don’t know” for an answer.

Statement of Ethics

The author has no ethical conflicts to disclose. No experiments on living animals were performed, and no animals were euthanized purposefully for the work reported herein.

Conflict of Interest Statement

The author declares no conflict of interest.

Funding Sources

No funding was acquired for this work.

Data Availability Statement

The data supporting the findings of this study are available within the paper and its online supplementary information files (online suppl. Data 1).

References

Alex B. The Human brain has been getting smaller since the stone age. Discover Magazine. 2019. https://www.discovermagazine.com/planet-earth/the-human-brain-has-been-getting-smaller-since-the-stone-age.

Bailey DH, Geary DC. Hominid brain evolution. Hum Nat. 2009;20(1):67–79.

Balzeau A, Holloway RL, Grimaud-Hervé D. Variations and asymmetries in regional brain surface in the genus Homo. J Hum Evol. 2012 Jun;62(6):696–706.

Barban N, Jansen R, de Vlaming R, Vaez A, Mandemakers JJ, Tropf FC, et al.; BIOS Consortium; LifeLines Cohort Study. Genome-wide analysis identifies 12 loci influencing human reproductive behavior. Nat Genet. 2016 Dec;48(12):1462–72.

Barro RJ, Lee J. A new data set of educational attainment in the world, 1950–2010. J Dev Econ. 2013;104:184–98.

Beals KL, Smith CL, Dodd SM, Angel JL, Armstrong E, Blumenberg B, et al. Brain size, cranial morphology, climate, and time machines. Curr Anthropol. 1984;25(3):301–30.

Beauchamp JP. Genetic evidence for natural selection in humans in the contemporary United States. Proc Natl Acad Sci USA. 2016 Jul;113(28):7774–9.

Bednarik RG. Doing with less: hominin brain atrophy. Homo. 2014 Dec;65(6):433–49.

Bellissari A. Evolutionary origins of obesity. Obes Rev. 2008 Mar;9(2):165–80.

Benitez-Burraco A, Clay Z, Kempe V. Editorial: self-domestication and human evolution. Front Psychol. 2020 Aug;11:207.

Bouchard TJ Jr, Lykken DT, McGue M, Segal NL, Tellegen A. Sources of human psychological differences: the Minnesota Study of Twins Reared Apart. Science. 1990 Oct;250(4978):223–8.

Bratsberg B, Rogeberg O. Flynn effect and its reversal are both environmentally caused. Proc Natl Acad Sci USA. 2018 Jun;115(26):6674–8.

Burini RC, William RL. The evolutionary roles of nutrition selection and dietary quality in the human brain size and encephalization. Nutrire. 2018;43(1):43.
Decreasing Brain Size in Humans

Pietzchnig J, Voracek M. One Century of Global IQ Gains: A Formal Meta-Analysis of the Flynn Effect (1909–2013). Perspect Psychol Sci. 2015 May;10(3):282–306.

Pilbeam D, Gould SJ. Size and scaling in human evolution. Science. 1974 Dec;186(4167):892–901.

Plomin R, Deary IJ. Genetics and intelligence differences: five special findings. Mol Psychiatry. 2015 Feb;20(1):98–108.

Posthuma D, Baaré WF, Hulshoff Pol HE, Kahn RS, Boomsma DI, De Geus EJ. Genetic correlations between brain volumes and the WAIS-III dimensions of verbal comprehension, working memory, perceptual organization, and processing speed. Twin Res. 2003 Apr;6(2):131–9.

Posthuma D, De Geus EJ, Baaré WF, Hulshoff Pol HE, Kahn RS, Boomsma DI. The association between brain volume and intelligence is of genetic origin. Nat Neurosci. 2002 Feb;5(2):83–4.

Püschel HP, Bertrand OC, O’Reilly JE, Bobe R, Püschel TA. Divergence-time estimates for hominins provide insight into encephalization and body mass trends in human evolution. Nat Ecol Evol. 2021 Jun;5(6):808–19.

Reynolds CR, Brown RT, editors. Perspectives on IQ composition in mammals. Brain Behav Evol. 2004;63(1):47–60.

Shayer M, Ginsburg D, Coe R. Thirty years on - a large anti-Flynn effect? The Piagetian test and encephalization in Pleistocene Homo. Nature. 1997 May;387(6629):173–6.

Schoenemann PT. Brain size scaling and body composition in mammals. Brain Behav Evol. 2004;63(1):47–60.

Shultz S, Nelson E, Dunbar RI. Hominin cognitive evolution: identifying patterns and processes in the fossil and archaeological record. Philos Trans R Soc Lond B Biol Sci. 2012 Aug;367(1599):2130–40.

Snieder S, Stringer S, Watanabe K, Jansen PR, Coleman JR, Krapolh E, et al. Genome-wide association meta-analysis of 78,308 individuals identifies new loci and genes influencing human intelligence. Nat Genet. 2017 Jul;49(7):1107–12.

Striedter GF. Principles of brain evolution. Sinauer Associates; 2005.

Stringer C. Brain size has increased for most of our existence, so why has it started to diminish for the past few thousand years? Sci Am Mind. 2014;25(6):74.

Tobias PV. The Brain in Hominid Evolution. New York: Columbia University Press; 1971.

Tsai M, Husby A, Kotschral A, Hayward A, Buechel SD, Zidar J, et al. Comparative support for the expensive tissue hypothesis: big brains are correlated with smaller gut and greater parental investment in Lake Tanganı́yika cichlids. Evolution. 2015 Jan;69(1):190–200.

Von Bonin G. On the size of man’s brain as indicated by skull capacity. J Comp Neurol. 1934;59(1):1–28.

Walker A. The Nariokotome Homo Erectus Skeleton. In: Walker A, Leakey R, editors. Cambridge: Harvard University Press; 1993. pp. 411–30.

Whiten A, van de Waal E. Social learning, culture and the ‘socio-cultural brain’ of human and non-human primates. Neurosci Biobehav Rev. 2017 Nov;82:58–75.

Witelson SF, Beresh H, Kigal DR. Intelligence and brain size in 100 postmortem brains: sex, lateralization and age factors. Brain. 2006 Feb;129(Pt 2):386–98.

Zietsch BP, Kuja-Halkola R, Walum H, Verweij KJ. Perfect genetic correlation between number of offspring and grandoffspring in an industrialized human population. Proc Natl Acad Sci USA. 2014 Jan;111(3):1032–6.