The causal influence of brain size on human intelligence: Evidence from within-family phenotypic associations and GWAS modeling

James J. Lee[a*]
Matt McGue[a]
William G. Iacono[a]
Andrew M. Michael[b,c]
Christopher F. Chabris[b]

[a]Department of Psychology
University of Minnesota Twin Cities
75 East River Parkway
Minneapolis, MN 55455, USA

[b]Geisinger Health System
120 Hamm Drive Suite 2A
Lewisburg, PA 17837, USA

[c]Duke Institute for Brain Sciences
Duke University
308 Research Drive, LSRC M051
Durham, NC 27708, USA

[*]To whom correspondence should be addressed;
E-mail: leex2293@umn.edu

Abstract

There exists a moderate correlation between MRI-measured brain size and the general factor of IQ performance (g), but the question of whether the association reflects a theoretically important causal relationship or spurious confounding remains somewhat open. Previous small studies ($n < 100$) looking for the persistence of this correlation within families failed to find a tendency for the sibling with the larger brain to obtain a higher test score. We studied the within-family relationship between brain volume and intelligence in the much larger sample provided by the Human Connectome Project ($n = 1,022$) and found a highly significant correlation (disattenuated $\rho = 0.18$, $p < .001$). We replicated this result in the Minnesota Center for Twin and Family Research ($n = 2,698$), finding a highly significant within-family correlation between head circumference and intelligence (disattenuated $\rho = 0.19$, $p < .001$). We also employed novel methods of causal inference relying on summary statistics from genome-wide association studies (GWAS) of head size ($n \approx 10,000$) and measures of cognition ($257,000 < n < 767,000$). Using bivariate LD Score regression, we found a genetic correlation between intracranial volume (ICV) and years of education (*EduYears*) of 0.41 ($p < .001$). Using the Latent Causal Variable method, we found a genetic causality proportion of 0.72 ($p < .001$); thus the genetic correlation arises from an asymmetric pattern, extending to sub-significant loci, of genetic variants associated with ICV also being associated with *EduYears* but many genetic variants associated with *EduYears* not being associated with ICV. This is the pattern of genetic results expected from a causal effect of brain size on intelligence. These findings give reason to take up the hypothesis that the dramatic increase in brain volume over the course of human evolution has been the result of natural selection favoring general intelligence.

Contents

0 General Introduction

People with bigger brains tend to be smarter. Furthermore, overall brain size is the only proxy for neural complexity that can be reliably obtained from fossils across a wide range of species and times, specifically from skull endocasts (Jerison, 1973). For these reasons overall brain size is perhaps the anatomical correlate of g that most interests researchers contemplating the long-term evolution of human intelligence (Burkart, Schubiger, & van Schaik, 2017). Among all such correlates, none that can be simply measured shows a consistently replicated correlation with IQ exceeding the ∼0.25 shown by brain size (Chabris, 2007; Pietschnig, Penke, Wicherts, Zeiler, & Voracek, 2015; Haier, 2017; Gignac & Bates, 2017). Furthermore, since brain volume has tripled during the most recent 3 million years of our evolutionary history (Holloway, 2008; Herculano-Houzel, 2016), the study of its relationship with g promises to shed light on the nature of the selective advantage that has propelled this remarkable change.

There continues to be some doubt, however, about whether it is justified to interpret this correlation in terms of brain size affecting intelligence (Nisbett et al., 2012). One

powerful method of testing whether a correlation is the result of spurious confounding by other variables affecting the two focal traits is to see whether the correlation persists within reared-together sibling pairs (Jensen & Sinha, 1993; Turkheimer & Waldron, 2000; McGue, Osler, & Christensen, 2010; Beauchamp, Cesarini, Johannesson, Lindqvist, & Apicella, 2011; Lee, 2012). If the correlation does persist, then it cannot be attributable to aspects of the environment affecting both physical and mental development that are fixed for all siblings. Two previous studies taking this approach, each with a sample size of about 35 sibling pairs, failed to find a significant correlation (Schoenemann, Budinger, Sarich, & Wang, 2000; Gignac, Vernon, & Wickett, 2003).

We reexamined this empirical issue in a Human Connectome Project (HCP) (Van Essen et al., 2012; Van Essen et al., 2013) sample of 1,022 young adults, each of whom is a twin or ordinary sibling of another sample member. We also tested the significance of the within-family correlation between head circumference and IQ in an independent Minnesota Center for Twin and Family Research (MCTFR) (Miller et al., 2012) sample of 11-year-old children and 17-year-old adolescents (865 MZ and 484 DZ twin pairs); external measurements of the head provide a noisy proxy of underlying brain volume.

The presence or absence of a within-family association between brain size and intelligence is probably the most intuitively compelling observational data that can be brought to bear on the issue of a causal influence of the former trait on the latter. There are alternative explanations of the population correlation between these two traits, however, that cannot be eliminated by the finding of a significant within-family association. These are listed in Table 1.

To weaken these alternatives to the hypothesis of a causal effect of brain size on intelligence, we turned to recently developed methods for interpreting data from genome-wide association studies (GWAS) (Gondro, van der Werf, & Hayes, 2013; Chabris, Lee, Ce-

sarini, Benjamin, & Laibson, 2015; Lee, Vattikuti, & Chow, 2016). The first of these is bivariate LD Score regression, a method for estimating genetic correlations from GWAS data (Bulik-Sullivan et al., 2015a). Each trait can be regarded as the sum of a genetic and environmental term; if we imagine taking away the environmental part of each trait and calculating the correlation between whatever remains, we are left with a quantity called the *genetic correlation* (r_g) by quantitative geneticists (Lynch & Walsh, 1998). (The remaining genetic term is called the "breeding" or "additive genetic" value. This corresponds to the "A" in the ACE model of behavioral genetics.) Care must be taken to ensure that the estimate of the genetic correlation is not itself biased by environmental confounders affecting both traits, and bivariate LD Score regression contains a powerful safeguard against such confounding (Bulik-Sullivan et al., 2015a; Lee, McGue, Iacono, & Chow, 2018). An estimated genetic correlation that is significant and unbiased rules out the possibility of the total correlation being entirely the result of environmental confounders that can assume different values even within the same sibship. It is also consistent with a causal effect of one trait on the other, since genetic influences on the upstream trait must be propagated to the downstream one (Duffy & Martin, 1994; Minica, Dolan, Boomsma, de Geus, & Neale, 2018).

A significant genetic correlation is also consistent, however, with some of the alternative explanations in Table 1. One of these is a kind of genetic confounding known as *pleiotropy*: the same or correlated polymorphic sites in the genome affecting both traits without either trait affecting the other. Another possibility, perhaps an obvious one, is reverse causation: perhaps it is really the exercise of intelligence that affects brain size, through some form of plasticity.

The method that we employed to address these possibilities that can be understood in the following intuitive manner; see Fig. 1 for an illustration of the logic. Suppose that

Table 1: Strengths and limitations of methods for causal inference.

	Ruled out by . . .		
	WF	LDSC	LCV
Confounding by shared environment	Yes	Yes	No
Confounding by non-shared environment	No	Yes	No
Cross-assortative mating	Yes	No	No
Genetic pleiotropy	No	No	Yes
Reverse causation	No	No	Yes

Each row corresponds to an alternative explanation of the correlation between brain size and intelligence other than the former being a cause of the latter. "Yes" indicates that a positive result obtained with the method given in the column provides evidence against that alternative explanation. A simple yes-or-no answer is always given for the sake of simplicity; consult the text for more nuance. WF, within-family phenotypic association; LDSC, bivariate LD Score regression; LCV, Latent Causal Variable.

the variable X is brain size and Y is a measurement of intelligence. If $X \rightarrow Y$ is the case, then any genetic polymorphisms affecting X must also affect Y. Conversely, there may be many polymorphisms affecting Y through some non-X mechanism and therefore showing no effect on X (Pickrell et al., 2016; Visscher & Yang, 2016; O'Connor & Price, 2018). These predictions also hold, *mutatis mutandis*, if $Y \rightarrow X$ is the case. In contrast, genetic pleiotropy should lead to a rejection of both causal directions. We used the Latent Causal Variable (LCV) method to determine which of these three possibilities— one causal direction, the other causal direction, non-directed pleiotropy—offers the best approximation of the pattern inherent in GWAS summary statistics of head size and cognition.

1 Study 1: Within-Family Association Between IQ and Measures of Brain Size

1.1 Introduction

The common environment shared by all siblings growing up in the same home may conceivably affect both brain size and intelligence, accounting for the correlation between these two traits. A powerful means of ruling out this source of confounding is the calculation of the within-family association: does the sibling with the larger brain also tend to have the higher IQ?

It may be worthwhile to review the two earlier studies that have examined this question. Other studies have obtained relevant data, mostly from twins, but have not analyzed their data in this particular way.

Schoenemann et al. (2000) studied 36 pairs of sisters, administering a number of cognitive tests (including Raven's Progressive Matrices and a vocabulary test) and acquiring the volumes of many different brain regions. The large number of cognitive and brain measures led to many correlations, but it probably suffices to report that they found a within-family correlation between total brain volume and the first principal component of their test battery equaling -0.05. This was despite a nominally significant between-family correlation (i.e., the correlation between sibship average brain volume and test score) of 0.45.

Gignac et al. (2003), performing additional analyses of data first presented by Wickett, Vernon, and Lee (2000), studied 32 pairs of brothers. They found a non-significant within-family correlation of 0.16 between total brain volume and IQ as assessed by the Multidimensional Aptitude Battery. The estimate of the between-family correlation, 0.31, was not significant either.

We cannot find any serious fault in either of these studies, other than that their samples of sibling pairs were far too small to draw any definitive conclusions. For example, in the case of a sample size equaling 36 and an estimated correlation -0.05, the upper endpoint of the 95% CI is 0.28. Clearly, even with results favoring the two-sided null hypothesis as closely as one might stipulate, a study of this size is incapable of ruling out theoretically significant values of the within-family correlation. In the current study, we were able to estimate the within-family correlation between MRI-measured brain volume and intelligence with more than 10 times as many families and thereby bring much greater statistical power to ruling out confounding by shared environment as the exclusive contributor to the population correction.

A significant within-family association would also rule out another source of confounding that does not easily fit into the framework of the General Introduction. The definition of the genetic correlation (Lynch & Walsh, 1998) includes contributions from *cross-assortative mating*, the tendency of individuals with a certain attribute to prefer mates with a certain other attribute (Jensen & Sinha, 1993; Redden & Allison, 2006; Keller et al., 2013). An example is the tendency of individuals with higher IQs to prefer mates who are taller. After many generations of such cross-assortative mating, a given genome tending to carry IQ-increasing alleles will also tend to carry height-increasing alleles at distinct sites, an example of long-distance linkage disequilibrium (LD). This mechanism does appear to contribute to the actual height-IQ genetic correlation (Keller et al., 2013).

It does not seem plausible that people find brain size *per se* to be attractive in potential partners, as opposed to the possible effects of brain size on psychological traits. This intuition can be given rigorous support by the finding of a significant within-family correlation between brain size and IQ (Jensen & Sinha, 1993; Beauchamp et al., 2011).

Whereas under cross-assortative mating a given parent may tend to carry alleles increasing both traits, Mendelian independent assortment means that there will be no tendency for an offspring inheriting the allele increasing trait 1 at a site heterozygous in the parent to also inherit the allele increasing trait 2 at an unlinked site.

1.2 Methods

1.2.1 Human Connectome Project

Data from the Human Connectome Project (HCP) (Van Essen et al., 2013) was used as the primary analysis sample. Here, we included only families with at least two siblings. Some descriptive statistics of the HCP sample are given in the first column of Table S1; about 75 percent of the families in the HCP dataset are white. Details relating to participant selection and MRI acquisition have been reported (Van Essen et al., 2012). We chose to use total brain volume as the independent variable in our primary analysis.

We found that the mean and variance of total brain volume in the HCP differs substantially between the two sexes (Table S1). Males have both larger brains on average ($d = 1.67$) and greater variability (VR = 1.31). These differences are consistent with those observed in a recent study of a larger sample (Ritchie et al., 2018). For this reason we decided to standardize brain volume separately in each sex and use the resulting *Z*-score as the independent variable. We employed this standardization procedure in every robustness check.

Family "types" in the HCP are quite heterogeneous. For example, one family may consist of two MZ twins and their younger sibling. We included any family regardless of composition, so long as it contained at least two offspring with the same listed mother and father. In Table 2, the "MZ" subsample consisted solely of MZ twins. If an individual happens to be an MZ twin whose co-twin is absent from the dataset, that individual was

11

not included in the MZ subsample. The "DZ" subsample was constructed similarly. The "non-twin" subsample consisted of all families that did not qualify for either the MZ or DZ subsample; for example, one such family might consists of two siblings—an MZ twin whose co-twin is missing and his older sibling. Thus, by construction, the three HCP subsamples listed in Table 2 below the total sample were non-overlapping.

We selected two of the tests administered to the HCP subjects as our measures of general intelligence (g). The first is called the Penn Progressive Matrices in the HCP Data Dictionary. This is an abbreviated 24-item version of Raven's Standard Progressive Matrices; it is labeled "Non-Verbal IQ" in Table S1. The second is the NIH Toolbox Picture Vocabulary Test (age-adjusted scale score); it is labeled "Verbal IQ" in Table S1. The correlation between these two tests in the HCP data is 0.45, consistent with the g loadings of roughly 0.7 reported for these two types of test in the literature (Jensen, 2001; Johnson, Bouchard, Krueger, McGue, & Gottesman, 2004; Gignac, 2015). To construct our main dependent variable, we standardized each of these two tests separately, added the standardized scores, and finally standardized the sum. We found that this sum exhibits a sex difference of $\sim$0.3 favoring males. We applied another standardization, this time within each sex separately, to produce our final measure of g.

1.2.2   Minnesota Center for Twin and Family Research

The data are derived from the Minnesota Center for Twin and Family Research (MCTFR), including monozygotic (MZ) and same-sex dizygotic (DZ) twin pairs born in Minnesota. Some descriptive statistics of this sample are given in the final two columns of Table S1. Twin births from 1972 to 1984 were identified from Minnesota state birth records. Contact information was found for 91 percent of the families of these twins. Twins and their families were invited to a day-long assessment in a laboratory located in Minneapolis

when the twins were approximately 11 or 17 years of age. Among invited families, 17 percent refused to participate. Over 95 percent of participating families were white. This sample is much simpler in structure than the HCP, in that each family for our purposes consists of two twins (either MZ or DZ).

Head circumference (HC) was measured with a metal tape run around the head immediately above the eyebrows and ears. Males and females had nearly identical means at age 11, but at age 17 males had a roughly 1-SD larger HC (Table S1). The female mean HC at age 11 was 98 percent of the female mean at age 17, and the male mean at age 11 was 95 percent of the male mean at age 17. Males and females did not show a substantial difference in variance at either age. We standardized HC within each combination of sex and age cohort.

Full-Scale IQ, Performance IQ (called "Non-Verbal IQ" in Table S1), and Verbal IQ were measured using abbreviated versions of the Wechsler Adult Intelligence Scale-Revised (WAIS-R) for the 17-year-old twins and the Wechsler Intelligence Scale for Children-Revised (WISC-R) for 11-year-old twins. The short forms consisted of two Verbal subtests (Information, Vocabulary) and two Performance subtests (Block Design, Picture Arrangement). We found the correlation between Verbal and Performance among the 11-year-old twins to be about 0.47, which seems somewhat low; given the loadings on g of the WAIS-R subtests calculated by Gignac (2005), we might expect the correlation to exceed 0.65. Among the 17-year-old twins, the correlation is only 0.28. We do not have an explanation for the small magnitude of this estimated correlation, although it should be kept in mind that its standard error is relatively large. We converted IQ from its conventional scale ($M = 100$, $SD = 15$) to the standardized scale with mean zero and unit variance.

Our dataset largely overlaps with that of Silventoinen, Iacono, Krueger, and McGue

([2012](#)), who reported a significant positive genetic correlation between head circumference and IQ. Here we calculated not the genetic correlation but rather the within-family correlation. It should be understood that these two quantities do not necessarily provide the same information (Beauchamp et al., [2011](#)). For instance, positive correlations between genomic sites affecting two distinct traits as a result of cross-assortative mating contribute to the genetic correlation, but not to the within-family correlation (Table [1](#)).

1.2.3 Data analysis

We introduced dummies for family membership to estimate the within-family association between our measures of brain size and intelligence. All members of a given family were required to share the same mother and father. In the case of two siblings per family, the procedure that we followed—i.e., the regression of intelligence on brain size with family-membership dummies as covariates—is equivalent to the regression of the sibling difference in intelligence on the sibling difference in brain size. In our case the use of dummies is more efficient because many HCP families have more than two siblings.

We included the first three powers of mean-centered age as covariates whenever appropriate. There are no opposite-sex twins in any of our cohorts; in twin-only analyses, we dropped the age covariates. As a robustness check, we re-estimated the regression models with the first three powers of standardized height and weight as additional covariates. See the Supplemental Material for justification of these procedures.

The partial regression coefficients of brain/head size in the prediction of IQ are given in the "$\beta \pm \text{SE}$" column of Table [2](#). Since both independent and dependent variables were standardized, these coefficients should be somewhat comparable to correlations reported in previous studies.

A correlation of 0.45 between the two indicators of g in the HCP leads to an estimated

internal-consistency reliability of the IQ composite equaling 0.62; reliabilities exceeding 0.9 are possible with a higher-quality battery of tests (Brunner, Nagy, & Wilhelm, 2012). Because psychometric quality appears to have a substantial impact on reported estimates of the correlation between brain size and intelligence (Gignac & Bates, 2017), it may be advisable to apply the disattenuation procedure of classical test theory (Lord & Novick, 1968; McDonald, 1999). We corrected the coefficients for unreliability of the IQ measure using the formula $\beta/\sqrt{\omega_h}$, where β is the coefficient to be disattenuated and ω_h is McDonald's hierarchical coefficient omega (Zinbarg, Yovel, Revelle, & McDonald, 2006). In the case of a non-standardized dependent variable, one need not correct regression coefficients for unreliability of the dependent variable because zero-mean noise does not change the conditional expectation of the dependent variable given the independent variable. However, in our case of a standardized dependent variable where both true scores and errors contribute to the variance, such a correction is justified.

The IQ battery used in the MCTFR has excellent norming data. The g loadings of the WAIS-R subtests calculated by Gignac (2005) on the basis of these data suggest that the internal-consistency reliability of the abbreviated Full-Scale IQ in the MCTFR is 0.82, according to the formula for McDonald's coefficient omega. For reasons that are unclear, the correlation between Verbal and Performance observed in the MCTFR does not seem consistent with such high reliability. We nevertheless used a conservative value of 0.82 in calculating the disattenuated associations; the assumption of a lower value would lead to larger apparent effects.

A correction for unreliability using McDonald's hierarchical coefficient omega assumes that it is the common factor, g, rather than group or test-specific factors, that is responsible for the correlation with brain size. Previous results already make this assumption a reasonable one (MacLullich et al., 2002; Kievit et al., 2012; Gignac & Bates, 2017). To

confirm it in our own data, we looked at the within-family correlation between brain/head size and each subtest in our IQ composites.

The equality of the within-family coefficient across sibship types (MZ, DZ, non-twin full sibling) is not only compatible with a causal relationship between brain volume and intelligence but in fact is predicted by such a relationship in the absence of confounding by genetic pleiotropy or non-shared environment (McGue et al., 2010; Frisell, Öberg, Kuja-Halkola, & Sjölander, 2012). That is, if brain size affecting intelligence is the only reason for the correlation between these variables, then any given difference between siblings in brain size—even if induced environmentally (which is the only possibility in an MZ pair)—is expected to produce the same difference between them in g. To test whether the equality of the relationship holds across sibship types, we re-estimated the regression model with family fixed effects in the non-overlapping samples of MZ twins only, DZ twins only, and non-twin siblings only. We then tested whether any difference between estimates is significantly different from zero.

1.3 Results

1.3.1 Human Connectome Project

The overall correlation between MRI-measured brain volume and IQ is 0.26 (0.33 upon correction for unreliability), in line with previous research (Pietschnig et al., 2015; Haier, 2017; Gignac & Bates, 2017). In the case of brain volume, the MZ and DZ intraclass correlations are 0.96 (CI95%: 0.94, 0.97) and 0.74 (CI95%: 0.62, 0.82) respectively; IQ, 0.74 (CI95%: 0.65, 0.80) and 0.57 (CI95% CI: 0.39, 0.70). Both MZ correlations are larger than their DZ counterparts, as expected in the case of a heritable trait.

The results of the within-family analysis are given in Table 2. In a model controlling for family background, the partial regression coefficient of brain volume is 0.142 standardized

Table 2: Within-family associations between brain/head size and IQ.

Dataset	Families	Measure	$\beta \pm \text{SE}$	ρ	p-value
HCP					
All sibs	381	Brain volume	0.142 ± 0.040	0.18	4×10^{-4}
MZ	138	Brain volume	0.175 ± 0.168	0.22	0.30
DZ	79	Brain volume	0.055 ± 0.109	0.07	0.61
Non-twins	161	Brain volume	0.162 ± 0.064	0.21	0.01
MCTFR					
All sibs	1,349	Head circumference	0.172 ± 0.031	0.19	3×10^{-8}
MZ 11 yr	642	Head circumference	0.150 ± 0.049	0.17	0.002
DZ 11 yr	363	Head circumference	0.219 ± 0.053	0.24	5×10^{-5}
MZ 17 yr	223	Head circumference	-0.050 ± 0.086	-0.06	0.56
DZ 17 yr	121	Head circumference	0.209 ± 0.104	0.23	0.05

β is the estimated partial regression coefficient of brain/head size in a model predicting IQ with family fixed effects. Both size and IQ were standardized. "All sibs" means all families in the dataset, regardless of type (e.g., zygosity). Within each dataset, the rows beneath "All sibs" correspond to subsamples of the total sample. ρ, the partial regression coefficient (β) divided by the square root of the IQ test's internal-consistency reliability.

units of IQ per standardized unit of brain volume (CI95%: 0.062, 0.222; $p < .001$). A correction for unreliability brings this to 0.18. The upper endpoint of the CI95% becomes 0.28 upon disattenuation, and this may be consistent with the equality of the within-family and population correlations. It should be kept in mind, however, that the exact upper endpoint bounding the disattenuated correlation is uncertain as a result of quantities such as the internal-consistency reliability being estimated rather than known.

As can be seen in Table S2, the within-family correlation is also significant when the dependent variable is either the vocabulary or matrix-reasoning component of the IQ composite (both $p < .05$). This supports the correlation being with the general factor g rather than any test-specific factor.

No difference between subsample coefficients in the HCP is significant (Table 2), although the statistical power to find any difference is poor. If a given difference between

siblings in brain volume were truly associated with the same difference in IQ, regardless of how the siblings are related, this would be powerful evidence for our postulated causal relation.

A number of robustness checks affirmed the conclusions above. Upon excluding the non-whites and re-estimating the regression model with the same specification, we obtained a very similar result ($\beta = 0.130$, $p < .005$). We also re-estimated the regression model with the first three powers of standardized height and weight as additional covariates (Table S3). The result is nearly unchanged ($\beta = 0.124$, $p < .005$).

1.3.2   Minnesota Center for Twin and Family Research

As can be seen in Table 2, we found a significant within-family correlation between HC and IQ ($p < 10^{-7}$). This replicates two earlier studies (Jensen & Johnson, 1994; Jensen, 1994) and improves upon them, in that one of these studies was small (143 pairs) and the other only examined children no older than 7 years. In the age-17 cohort, the DZ within-family coefficient is almost significantly larger than the MZ ($p = .055$). If this were a reliable difference, it would provide evidence for the presence of confounding in addition to our hypothesized causal relation, perhaps attributable to pleiotropic influences on HC and IQ (Silventoinen et al., 2012). Our inclination, however, is to treat this difference as a chance fluctuation. By contrast, in the age-11 cohort, the MZ within-family correlation between HC and IQ is very significant ($p < .005$) and not distinguishable in magnitude from the DZ within-family correlation.

Overall, then, the body of results in Table 2 is roughly consistent with the hypothesis of brain size (measured by its proxy HC) affecting intelligence. The disattenuated coefficient of 0.19 is again consistent with the equality of the within-family and population correlations. It happens that our sample individual-level correlation is 0.15.

Table S2 shows that the within-family correlations of HC with both Verbal and Performance IQ are also significant ($p < .001$), supporting a correlation of head size with g rather than any group or test-specific factor.

Upon inclusion of the first three powers of standardized and weight as covariates, the coefficient of HC becomes slightly smaller but remains highly significant ($\beta = 0.148$, $p < 10^{-4}$).

1.4 Discussion

In two datasets we found positive and very robustly significant estimates of the within-family association between brain size and intelligence. Regardless of how brain size was measured, the type of IQ test, the composition of the sample, and choice of covariates, we did not discern any meaningful difference in the within-family association. These results provide strong evidence against the population correlation between brain size and intelligence being entirely the result of cross-assortative mating, confounding by the common environment shared by children in the same home, or some combination of these two mechanisms. Note that there remain theoretically viable alternative explanations of the population correlation (Table 1), and it is the purpose of Study 2 to address these.

The more specific finding of a significant within-MZ correlation would further bolster the case for brain size affecting IQ. Additionally, the within-MZ coefficient must be equal to the other within-family coefficients to affirm that pleiotropic genetic confounding and non-shared environment make no contribution. As can be seen in Table 2, the evidence for these propositions is tentative. This must be regarded as a limitation of Study 1. It is evident that there is some value in collecting samples of twins and siblings even larger than those we have relied upon here. Such samples would be able to estimate the overall within-family correlation much more precisely. Our disattenuated estimate of 0.18 on the

basis of the HCP data qualifies as a relatively small effect size according to the guidelines of Gignac and Szodorai (2016), but the confidence interval extends upward to values that would qualify as typical. The reduction of this uncertainty may be worthwhile. Note that since many within-family correlations are smaller than their population counterparts, the guidelines of Gignac and Szodorai (2016) should be considered conservative.

2 Study 2: Causal Inference Based on GWAS Data

2.1 Introduction

The calculations of the within-family correlation between IQ and brain/head size in Study 1 required phenotypic data only. In Study 2 we tested the causal status of the relationship by using recently developed methods for interpreting data from genome-wide association studies (GWAS), which are capable of ruling out the additional alternative explanations of reverse causation, a non-directed genetic overlap (pleiotropy), and confounding by non-shared environment (Table 1).

There have been many twin studies reporting a positive genetic correlation between brain size and IQ (Pennington et al., 2000; Posthuma et al., 2002; van Leeuwen et al., 2009; Betjemann et al., 2010; Silventoinen et al., 2012). There is nevertheless great value in estimates based on the alternative data source provided by GWAS, especially since the GWAS-based technique of bivariate LD Score regression deals with confounding in a manner that appears quite different from the twin method and its equal-environment assumption. Recall that a significant and unbiased estimate of the genetic correlation rules out environmental confounding, whether by shared or non-shared factors, as the sole explanation of the phenotypic correlation.

A genetic correlation by itself is consistent with both reverse causation and pleiotropic confounding. We therefore turned to another method, also based on GWAS data, that

bears on the issue of causality. If genetic variation affecting X is propagated to Y in the simple causal chain $G \rightarrow X \rightarrow Y$, then any parts of G identified in GWAS of X must also show concordant effects on Y. Furthermore, the effects of these GWAS-identified single-nucleotide polymorphisms (SNPs) on Y must be proportional to their effects on X. In contrast, there may be many SNPs affecting Y through some non-X mechanism, and GWAS of Y should therefore detect many SNPs with no discernible effects on X. To follow this logic, it may be helpful to consult Fig. 1.

Actual empirical results are not likely to conform exactly to these simple predictions even under the most favorable conditions. It is easy to imagine that there are parts of the brain, perhaps subcortical, whose sizes could hypothetically be altered in a manner increasing the brain's overall volume without affecting higher cognition (Jerison, 1973; Jung & Haier, 2007). The portion of the brain serving as the substrate of cognition would then act effectively as a confounder, its size sending causal arrows to both overall brain size and intelligence. Nevertheless, we attempted to determine the extent to which the predictions from the simple $G \rightarrow X \rightarrow Y$ model are borne out by GWAS data.

2.2 Methods

2.2.1 Bivariate LD Score regression

Here we present estimates of relevant genetic correlations between measures of head size and cognition, updated from previous GWAS reports to reflect increases in sample size (Okbay et al., 2016; Hagenaars et al., 2016; Sniekers et al., 2017). We used bivariate LD Score regression (Bulik-Sullivan et al., 2015b; Bulik-Sullivan et al., 2015a; Lee et al., 2018), as implemented in the LCV software tool (O'Connor & Price, 2018). If there is a tendency for a SNP with a relatively large effect on one trait to also show a relatively large effect of the same sign on the other, then this method will return a positive esti-

mate of the genetic correlation. An attractive feature of LD Score regression is that its means of controlling the influence of confounding variables extends not just to the evidence of association at individual SNPs but also to its genome-wide estimates of genetic correlations; ideally, the regression intercepts absorb all confounding, including the effects of environmental variables correlated with genetic variation that happen to affect both traits. The Supplemental Material discusses this point in more detail.

Whereas a positive result obtained by LD Score regression does provide evidence against environmental confounding, such a result does not bear on environmental mediators of genetic influence, including mediators conveying a causal influence in the direction opposite of what we are hypothesizing. For example, there may be genetic influences on intelligence or intellectual curiosity that lead to more educational attainment, which may then increase brain size through a plasticity-based mechanism. Such possibilities can be addressed not by LD Score regression itself but rather the LCV method, which will be described shortly in 2.2.2.

We obtained the GWAS summary statistics of *EduYears*, IQ, and infant head circumference from the websites of the consortia that conducted these GWAS (Taal et al., 2012; Okbay et al., 2016; Lee et al., 2018). The GWAS summary statistics of adult intracranial volume can be obtained by contacting the ENIGMA Consortium (Hibar et al., 2015). We ran bivariate LD Score regression, as implemented in the LCV software (downloaded July 3, 2018) (O'Connor & Price, 2018). We used all recommended default settings. One recommendation is to remove all SNPs in the MHC region, but to our knowledge the LCV developers do not provide a definition of this region. We adopted the definition used by the DEPICT bioinformatics tool of all SNPs on chromosome 6 between positions 25,000,000 and 35,000,000 (Pers et al., 2015).

As measures of intelligence, we examined both years of education (*EduYears*) and IQ.

We designated *EduYears* as our primary measure because the attainable GWAS sample size for *EduYears* is currently larger than that for IQ. Previous studies have found a high genetic correlation (∼0.7) between IQ and *EduYears*, and the top SNPs identified in the respective GWAS of these traits show very strong concordance. Our two measures of brain size are adult intracranial volume (ICV) (Hibar et al., 2015) and infant head circumference (HC) (Taal et al., 2012). ICV is an excellent proxy for brain volume with the advantage that it does not decline sharply with age (Davis & Wright, 1977); thus, any correlation with intelligence cannot be explained by the effects of neural atrophy on cognitive decline. According to bivariate LD Score regression, adult ICV and infant HC show a genetic correlation of 0.75 (Adams et al., 2016).

The samples for IQ and *EduYears* consist of the cohorts studied in the largest published GWAS of *EduYears* at the time of writing with the exception of 23andMe (Lee et al., 2018). The *EduYears* sample size is ∼760,000, reflecting a net increase of ∼465,000 since the last reported calculation of the ICV-*EduYears* genetic correlation. The sample size for IQ is ∼260,000. The sample sizes for adult ICV and infant HC, 11,373 and 10,678 respectively, are unchanged since their GWAS were published. There has been a more recent and larger GWAS of ICV (Adams et al., 2016), but its full summary statistics are not publicly available.

It is possible to estimate the genetic correlations between measures of brain size and cognition using twin data, of the kind that we employed in Study 1, rather than GWAS summary statistics. Our HCP dataset contains too few twins, however, to produce a reliable estimate of the genetic correlation. As we stated earlier, the twins in the MCTFR have already been used to estimate the genetic correlation between IQ and HC (Silventoinen et al., 2012).

Another method for estimating genetic correlations is the genomic relatedness re-

stricted maximum-likelihood (GREML) method (Lee, Yang, Goddard, Visscher, & Wray, 2012), which has been used in many applications, including a study showing that both twin analyses and GREML produce similarly large estimates of the genetic correlations between different mental tests (Trzaskowski et al., 2013). We discuss the data requirements and assumptions of LD Score regression and GREML respectively in the Supplemental Material.

2.2.2   Latent Causal Variable

The most recent GWAS of intracranial volume reported seven genome-wide significant SNPs in its Table 1 (Adams et al., 2016). We included the four additional SNPs in Supplementary Table S9 of this paper that were found to reach genome-wide significance upon a meta-analysis of both adult intracranial volume and infant head circumference. These are the 11 SNPs plotted in the left panel of our Fig. 1. One of these SNPs, rs138074335, has changed its identifier and is not present in the summary statistics of Lee et al. (2018). We took its association statistics from the summary file of Okbay et al. (2016).

Eleven genome-wide significant SNPs does not provide enough power to make use of the testing framework proposed by Pickrell et al. (2016) for the data plotted in Fig. 1. We therefore turned to LCV, a recently developed method that overcomes the limitation of insufficient significant SNPs by using genome-wide summary statistics and additionally addresses the problem of the candidate causal variable X being a proxy affected by residual genetic variation (O'Connor & Price, 2018). The Supplemental Material contains an extensive discussion of the LCV method, providing intuitive explanations on top of those here in the main text and comparing it in great depth with a distinct class of methods (Mendelian randomization) for drawing causal inferences from GWAS data. The output

24

of LCV is the *genetic causality proportion* (GCP), which ranges in possible value from -1 to $+1$. A value of zero means that the genetic correlation is entirely attributable to undirected pleiotropy; a value of $+1$ means that all genetic variation affecting X is propagated to Y in a perfect causal chain; a value of -1 means that it is actually Y affecting X. The developers suggest $|\text{GCP}| > 0.6$ as a cutoff for considering the relationship between the two traits to approximate a causal one reasonably well.

To be clear, Fig. 1 does not depict the results of our main analysis relying on LCV. It depicts the data that would be used in the method of Pickrell et al. (2016), which has been generalized and made more robust by LCV. Fig. 1 is nevertheless useful for illustrating the intuition behind the LCV method and suggesting the pattern in the genome-wide summary statistics that drives our positive results.

The LD Score regression slope Z-statistic of ICV is roughly 4, which falls short of the threshold of 7 recommended by the LCV developers. The Supplemental Material explains the rationale behind this recommendation. Therefore, as a robustness check, we meta-analyzed ICV and HC with PLINK 1.9 (Chang et al., 2015) and used the resulting ICV + HC trait with an effectively twofold sample size as the candidate causal variable in LCV. Recall that the genetic correlation between ICV and HC is high; both measures of brain size are in turn genetically correlated with IQ (Sniekers et al., 2017; Savage et al., 2018).

2.3 Results

2.3.1 Genetic correlations between head size and measures of cognition estimated by bivariate LD Score regression

As can be seen in Table 3, IQ shows a significant genetic correlation with adult ICV (r_g = 0.29, $SE = 0.11$), as does *EduYears* ($r_g = 0.41$, $SE = 0.08$). These estimates are fully consistent with previous DNA-level studies, one of which used largely overlapping data

Table 3: Genetic correlations and genetic causality proportions.

Traits	$r_g \pm \mathrm{SE}$	$\mathrm{GCP} \pm \mathrm{SE}$
Intracranial volume		
IQ	0.291 ± 0.107	0.746 ± 0.146
EduYears	0.411 ± 0.077	0.721 ± 0.174
ICV + HC		
IQ	0.306 ± 0.083	0.495 ± 0.137
EduYears	0.353 ± 0.064	0.450 ± 0.136

r_g, the genetic correlation estimated by bivariate LD Score regression between the measures of head size (Taal et al., 2012; Hibar et al., 2015) and cognition (Lee et al., 2018); GCP, the genetic causality proportion, which ranges between zero (genetic variation acts purely as a confounder $X \leftarrow G \rightarrow Y$) and one (joint genetic influences act purely through the causal chain $G \rightarrow X \rightarrow Y$); ICV + HC, GWAS meta-analysis of adult intracranial volume and infant head circumference.

to study IQ (Savage et al., 2018). The estimates are also consistent with previous twin studies. To give just one example, the HC-IQ genetic correlations based on the MCTFR (one of the datasets in Study 1) are 0.32 in the 11-year-old boys (CI95%: 0.14, 0.50), 0.28 in the 11-year-old girls (CI95%: 013, 0.46), 0.25 in the 17-year-old boys (CI95%: 0.10, 0.42), and 0.17 in the 17-year-old girls (CI95%: 0.04, 0.30) (Silventoinen et al., 2012).

2.3.2 Causal inference from asymmetry of GWAS results

The effects of the top SNPs identified in GWAS of ICV and *EduYears* respectively are displayed in Fig. 1. Note that we use the term "effect" loosely, because the SNPs plotted in Fig. 1 may not be causal themselves but in LD with the actual causal sites (Lee & Chow, 2013). The left panel displays the 11 genome-wide significant SNPs identified in the largest GWAS of ICV (Adams et al., 2016). Since we can be highly confident that these SNPs are associated with brain size, a causal effect of brain size on intelligence implies
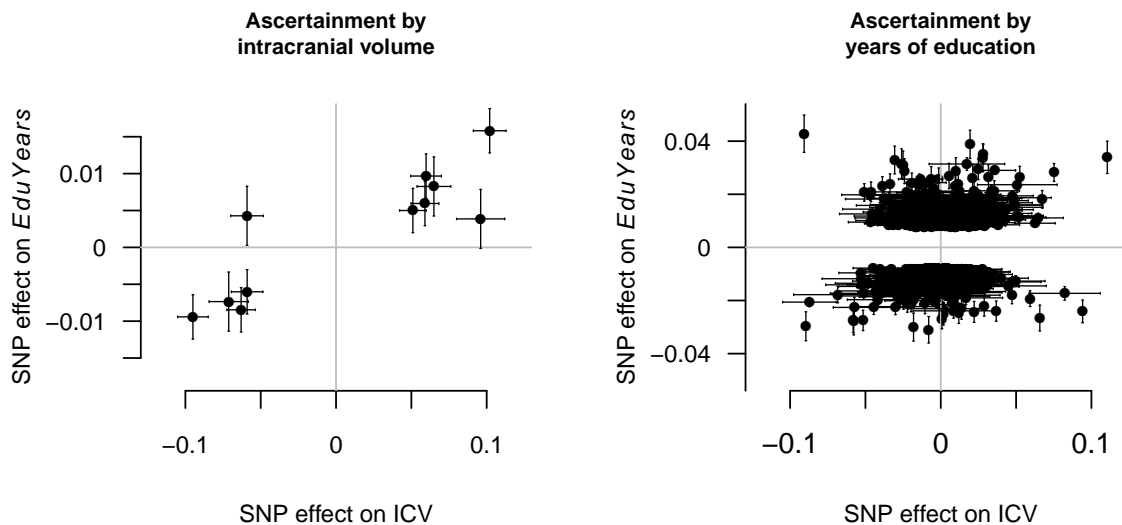
Figure 1: The nature of the causal relationship between brain volume and years of education inferred from the asymmetry of GWAS results. Each data point represents the effects of a single-nucleotide polymorphism (SNP) on intracranial volume (ICV) and years of education (*EduYears*) respectively, as estimated in genome-wide association studies (GWAS) of these two traits. The left panel displays 11 SNPs reaching genome-wide significance in the latest GWAS of ICV (Adams et al., 2016), whereas the right panel displays 1,044 SNPs reaching genome-wide significance in the latest *EduYears* GWAS (Lee et al., 2018) that are also found in the publicly available summary statistics of the ICV GWAS (Hibar et al., 2015). Both traits have been standardized. The error bars represent plus/minus 1 SE. In the right panel, ICV errors larger than 0.03 (about 5 percent of all SNPs) have been omitted for clarity.

.

the tight clustering of the data points around a straight line. The right panel displays the 1,044 genome-wide significant SNPs identified in the largest GWAS of *EduYears* (Lee et al., 2018) that are also contained in the ICV summary statistics. We can be highly confident that these SNPs are associated with *EduYears*, but the complement of the subset acting through brain size should show no association with ICV.

Ten of the 11 SNPs associated with ICV show concordant effects on *EduYears* (sign-test one-sided $p < .01$). Of the 10 concordant SNPs, the associations of 9 with *EduYears*

27

are nominally significant (all $p < .05$). In contrast, only 592 of the 1,044 SNPs ascertained through the GWAS of *EduYears* show concordant associations with ICV (57 percent, sign-test one-sided $p < 10^{-5}$). The Spearman correlation between effect sizes is 0.83 when ascertainment is by ICV, whereas it is only 0.17 when ascertainment is by *EduYears*. Qualitatively, then, this pattern is consistent with a causal effect of brain size on intelligence.

To determine whether the pattern evident in Fig. 1 extents to the entire genome, including sub-significant SNPs, we applied LCV to the complete GWAS summary statistics of our traits. The results can be found in Table 3. When the candidate causal variable is ICV, the estimated values of GCP exceeding 0.7 are very significantly different from zero (both $p < 10^{-4}$ on the assumption of a normal sampling distribution). The LCV developers consider GCP > 0.6 to be indicative of a good approximation to a causal relationship between the two variables. In fact, for neither *EduYears* nor IQ can we reject the hypothesis GCP = 1 (both $p > 0.15$).

In our setting, the LCV model is consistent with an appealing interpretation. Suppose that only a portion of the brain affects intelligence. Upon standardization of all breeding (additive genetic) values, the breeding value of the intelligence-supporting part of the brain has an effect on the ICV breeding value equaling 0.884 and an effect on the *EduYears* breeding value equaling 0.467. These two path coefficients reproduce the values of r_g and GCP given in Table 3 for ICV and *EduYears*. Thus it may be reasonable to treat a measure like ICV as the actual causal variable affecting intelligence—despite the likely existence of brain regions unrelated to higher cognition function (Jerison, 1973; Jung & Haier, 2007)—because genetic influences uniquely affecting such regions are comparatively slight.

As a robustness check, we meta-analyzed adult ICV and infant HC and used the

resulting GWAS summary statistics as input to the LCV method. This use of ICV + HC as the outcome variable led to GCP estimates between 0.45 and 0.50 (Table 3). An effect of the intelligence-supporting portion of the brain on ICV + HC equaling 0.753 and on *EduYears* equaling 0.472 reproduces the relevant values of r_g and GCP, although the interpretation of the variable ICV + HC is admittedly unclear.

A potential limitation of Study 2 is that the relevant intercepts in bivariate LD Score regression may not adequately reflect the degree of confounding in the GWAS if the cohorts were sampled from different latitudes in Europe along a cline in average intelligence; the Supplemental Material further discusses this possibility. As another robustness check, we reran the LCV software with the GWAS summary statistics of *EduYears* taken this time, not from Lee et al. (2018), but from a single large cohort (the UK Biobank) analyzed with a linear mixed model to address confounding and other biases (Loh, Kichaev, Gazal, Schoech, & Price, 2018). We obtained very similar results ($r_g = 0.43$, $SE = 0.08$; GCP = 0.78, $SE = 0.15$), indicating that confounding is indeed unlikely to be seriously biasing our inferences.

2.4   Discussion

Applying bivariate LD Score regression to GWAS summary statistics, we found significant genetic correlations between measures of head size and cognition whose magnitudes agree well with those obtained in twin studies. This agreement across methods suggests that each provides a reasonably accurate estimate of this key quantity and thus can be relied upon to rule out confounding by non-heritable environmental factors as the sole contributor to the total phenotypic correlation.

A nonzero genetic correlation by itself is silent about whether the relationship between the two traits is causal and the direction of the causal influence. To make the further

causal inference, we turned to the LCV method, which generalizes the logic illustrated in Fig. 1: if $G \to X \to Y$ is the case, then those parts of G identified in GWAS of X must show concordant and proportional associations in GWAS of Y, whereas the top signals in GWAS of Y may often affect Y through non-X mechanisms and thus show no association with X. We obtained estimates of the genetic causality proportion exceeding 0.7, which can be interpreted as ICV being an effective proxy for those parts of the brain where size affects intelligence. We obtained a smaller estimate when meta-analyzing ICV and HC as a robustness check, which perhaps implies that HC measured in infancy is a noisier proxy for the sizes of the relevant brain regions. Further precision will come with increasing the GWAS sample size for three-dimensional measures of brain size, which certainly seems likely to occur; there are currently plans to acquire high-quality imaging data of many different organs in the UK Biobank.

The novelty of the LCV method should be kept in mind. Other ingenious methods in statistical genetics have often disclosed non-obvious intricacies only years after their introduction (Speed, Hemani, Johnson, & Balding, 2012; Lee & Chow, 2014; Lee et al., 2018; Berg et al., 2018), although it is notable how robust these methods remain even after they have become better understood. The Supplemental Material summarizes at some length the simulation studies conducted by the LCV developers, but it is hard to know whether a given set of simulations covers all plausible states of nature. This is an issue worth following as research unfolds. Right now, what we can say is that the LCV method is a generalization of the notions associated with Fig. 1 and thus appears to rest on a solid basis (Pickrell et al., 2016; Visscher & Yang, 2016).

3 General Discussion

Altogether, the results of Studies 1 and 2 provide a compelling case for brain volume (or the volumes of certain brain regions) affecting intelligence within the human species. Study 1 found that the sibling with a larger MRI-measured brain volume (or externally measured head circumference) tends to have the higher level of *g*. Study 2 found that SNPs in the human genome that are associated with intracranial volume also tend to be associated with both years of education and IQ; furthermore, whereas the effect of a SNP on intracranial volume is strongly predictive of its effect on intelligence, there appear to be many SNPs associated with intelligence that have no appreciable effect on intracranial volume.

We now discuss the relevance of these empirical findings to the notion of brain size affecting intelligence.

3.1 Significant within-family correlation is evidence against confounding by shared environment and cross-assortative mating

A detailed consideration of the within-family correlation leads to many theoretical arcana, but the basic notion is simple and compelling: if there is a significant within-family correlation, then the correlation observed in the population at large cannot be wholly ascribed to confounding variables that assume the same value for siblings growing up together. Despite the abundance and variety of the existing evidence, the failure to observe a significant correlation between MRI-measured brain volume and IQ in the few within-family studies reporting this estimate has been perhaps the weightiest point against accepting a nontrivial causal relationship between these two variables (Nisbett et al., 2012).

Here we have examined the within-family correlation in a sample consisting of more than 10 times as many families as those used in either of the two previous studies and obtained a significant result. We supplemented this finding with a calculation of the within-family correlation between IQ and head circumference, which is a noisier proxy for brain volume but is available to us in a much larger sample of twin pairs. We found this correlation to be significant as well.

3.2 Significant genetic correlation estimated by bivariate LD Score regression is evidence against exclusive confounding by non-shared environment

A within-family correlation might still be attributed to environmental confounders that can vary even within the same household (Frisell et al., 2012). We ruled out this kind of confounding as the sole explanation of the correlation between brain size and intelligence by replicating the genetic correlation between head size and two distinct measures of cognition. For this purpose we used bivariate LD Score regression, a powerful method that guards against environmental confounders affecting the two traits whose genetic correlation is to be calculated.

The sensitivity of this method to cross-assortative mating and other evolutionary forces inducing long-range LD between a site affecting trait 1 and a distant site affecting trait 2 is an interesting question that we did not consider here (although for the sake of simplicity Table 1 states that a positive result from bivariate LD Score regression does not provide evidence against this possibility). A within-family correlation between two highly polygenic traits, however, is very unlikely to be the result of long-range LD.

3.3 Significant LCV estimate is evidence against genetic pleiotropy and reverse causation

After ruling out the ascription of the entire correlation between brain size and intelligence to the combination of family background, non-shared environment, and cross-assortative mating, we are still left with the possibility of pleiotropic confounding—the same genetic polymorphisms affecting the two traits, without one trait affecting the other. We acknowledge that pleiotropy must, strictly speaking, be favored *a priori* as the explanation of the correlation over a directed causal relationship, if for no other reason than the plausible existence of brain regions that do not subserve higher cognitive functions (Jerison, 1973; Jung & Haier, 2007).

On the other hand, there is good reason to think that overall brain size acting as a cause of intelligence could prove to be an extremely useful approximation. Nearly 80 percent of the human brain by volume is taken up by the cerebral cortex (Swanson, 1995), the seat of memory, attention, thought, and language. Furthermore, there are strong individual-level correlations (both genetic and overall phenotypic) between the volumes of different regions within our own species (Rentería et al., 2014). The means available to nature of adjusting one brain region often act on functionally distinct regions as well. If overall brain size is as good a proxy for the sizes of the cognition-supporting regions as suggested by these considerations, then we are well justified in testing the (near) validity of strong hypotheses that fix measures of overall size as causally upstream of intelligence.

Examples of such strong hypotheses include the equality of the within-family regression coefficient across sibship types and a perfectly linear relationship between GWAS effect sizes upon ascertainment of SNPs whose associations with the measure of brain size reach genome-wide statistical significance. Although the first hypothesis was not clearly borne out in the sibling analysis of Study 1, both the data plotted in Fig. 1 and the

results of applying the LCV method do support the approximate validity of the second hypothesis. Furthermore, they provide a nearly definitive refutation of reverse causation. The right panel of Fig. 1 seems to reveal hundreds of SNPs with significant effects on *EduYears* having no effect on intracranial volume, and the formal LCV results back up this impression by providing overwhelming statistical evidence against GCP $= -1$ (perfect causal chain) and the developer-recommended cutoff of GCP $= -0.6$ (reasonable approximation of causal chain). This conclusion agrees with that of van Leeuwen et al. (2009), who argued against reverse causation on the grounds that the high heritability of brain volume is evidence against the propagation of environmental influences on intelligence to brain volume.

3.4 Limitations

We have pointed to limitations of each study in its own Discussion section. Here we mention a limitation common to both studies—the low reliability with which g was often measured (Gignac & Bates, 2017).

In Study 1's HCP dataset, we chose only two subtests as indicators of g for reasons of content validity and consequently suffered from an internal-consistency reliability of roughly 0.6. Similarly, the largest cohort contributing to the GWAS meta-analysis of IQ in Study 2 was assessed with an IQ test containing only a dozen items (Lee et al., 2018); the hierarchical coefficient omega of this test may well be less than 0.6. The relatively low reliability of the HCP test probably led to a diminution of the within-family correlation, although we tried to account for this with corrections based on classical test theory. It is tempting to think that the genetic correlation is not susceptible to unreliability because the greater error variance is simply part of the environmental term wiped away in the definition of this quantity. But care must be taken with such reasoning in genetic studies

of behavioral traits (Chabris et al., 2013). Psychometric errors of measurement may in fact represent stable and heritable attributes that are not affected by brain size, and thus even the estimates in Table 3 may increase with greater reliability. Highly reliable measurements of psychological traits are typically difficult to obtain in large studies where any given research goal is incidental, but their acquisition should remain a priority of differential psychologists involved in such projects.

3.5 Wider implications of brain size affecting intelligence in humans

We think that alternative explanations of the correlation between brain size and intelligence have been sufficiently weakened to justify some tentative remarks about what deeper meaning a causal relationship might hold.

We can imagine a brain becoming larger by either increasing the sizes of its neurons or adding more of them. The number of neurons is probably the more promising focus of theoretical and empirical inquiry, for reasons that we now explain.

It is a well-known paradox that humans are certainly the most cognitively advanced species but nevertheless show a smaller average brain size, by weight or volume, than elephants and whales. One attempt at a resolution of this paradox invoked the encephalization quotient (EQ), a measure of brain mass normalized essentially by body mass (Jerison, 1973). Many researchers found EQ appealing because it served well to distinguish humans from other species, but this measure turns not to provide an acceptable rank order of other species (Deaner, Isler, Burkart, & van Schaik, 2007; MacLean et al., 2014); for example, monkeys are more encephalized than the undoubtedly intelligent great apes (Deaner, van Schaik, & Johnson, 2006). Furthermore, the notion of an EQ may not be applicable to individual differences within the human species, for both birds and mam-

mals exhibit within-species brain-body scaling that is weaker and more variable than in other vertebrates (Tsuboi et al., 2018).

The more likely resolution of this paradox is that whereas average neuron size increases along with average neuron number in most taxa, this relationship has been abolished in primates, where neuron size remains fairly constant across a wide range of neuron number (Herculano-Houzel, Manger, & Kaas, 2014). Some large-bodied representatives of non-primate taxa thus have more voluminous brains than humans merely because their neurons are larger. The relevance of a neuron's size to its function is unclear; the number of synapses per neuron, a critical computational attribute, varies very weakly with neuron size across species (Wang, Ambrosini, & Wittenberg, 2016). When the average number of cortical neurons is the measure of size, then humans are by far the top-ranked species (Herculano-Houzel, Catania, Manger, & Kaas, 2015), followed in an intuitive order by other species (Herculano-Houzel, 2016). It therefore seems reasonable to infer that our intraspecific variation in neuron number is causally antecedent to g. This inference is compatible with the implication of genes and pathways associated with the regulation of neurogenesis in the GWAS of *EduYears* (Okbay et al., 2016; Lee et al., 2018), and it may be possible to add to the coherence of these findings by partitioning the genetic covariance between brain size and intelligence between subsets of the genome with different functional annotations (Lu et al., 2017; Ni, Moser, Schizophrenia Working Group of the Psychiatric Genomics Consortium, Wray, & Lee, 2018).

By positing that neuron number is a cause of g because the human brain has evolved more neurons than that of any other species, we have relied on the assumption that the average level of g in our lineage has increased over the last several million years as a result of evolution by natural selection (Burkart et al., 2017; Lee & Chabris, 2017). Genetic data may soon show whether this assumption is empirically well founded.

References

Adams, H. H. H., Hibar, D. P., Chouraki, V., Stein, J. L., Nyquist, P. A., Rentería, M. E., … Thompson, P. M. (2016). Novel genetic loci underlying human intracranial volume identified through genome-wide association. *Nature Neuroscience*, *19*(12), 1569–1582. doi:10.1038/nn.4398

Beauchamp, J. P., Cesarini, D., Johannesson, M., Lindqvist, E., & Apicella, C. (2011). On the sources of the height-intelligence correlation: New insights from a bivariate ACE model with assortative mating. *Behavior Genetics*, *41*(2), 242–252. doi:10.1007/s10519-010-9376-7

Berg, J. J., Harpak, A., Sinnott-Armstrong, N., Joergensen, A. M., Mostafavi, H., Field, Y., … Coop, G. (2018). Reduced signal for polygenic adaptation of height in UK Biobank. *bioRxiv*. doi:10.1101/354951

Betjemann, R. S., Johnson, E. P., Barnard, H., Boada, R., Filley, C. M., Filipek, P. A., … Pennington, B. F. (2010). Genetic covariation between brain volumes and IQ, reading performance, and processing speed. *Behavior Genetics*, *40*(2), 135–145. doi:10.1007/s10519-009-9328-2

Brunner, M., Nagy, G., & Wilhelm, O. (2012). A tutorial on hierarchically structured constructs. *Journal of Personality*, *80*(4), 796–846. doi:10.1111/j.1467-6494.2011.00749.x

Bulik-Sullivan, B., Finucane, H. K., Anttila, V., Gusev, A., Day, F. R., Loh, P.-R., … Neale, B. M. (2015a). An atlas of genetic correlations across human diseases and traits. *Nature Genetics*, *47*(11), 1236–1241. doi:10.1038/ng.3406

Bulik-Sullivan, B., Loh, P.-R., Finucane, H. K., Ripke, S., Yang, J., Schizophrenia Working Group of the Psychiatric Genomics Consortium, … Neale, B. M. (2015b). LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature Genetics*, *47*(3), 291–295. doi:10.1038/ng.3211

Burkart, J. M., Schubiger, M. N., & van Schaik, C. P. (2017). The evolution of general intelligence (with discussion). *Behavioral and Brain Sciences*, *40*, e195. doi:10.1017/S0140525X16000959

Chabris, C. F. (2007). Cognitive and neurobiological mechanisms of the law of general intelligence. In M. J. Roberts (Ed.), *Integrating the mind: Domain general versus domain specific processes in higher cognition* (pp. 449–491). Hove, UK: Psychology Press.

Chabris, C. F., Lee, J. J., Benjamin, D. J., Beauchamp, J. P., Glaeser, E. L., Borst, G., … Laibson, D. I. (2013). Why it is hard to find genes that are associated with social

science traits: Theoretical and empirical considerations. *American Journal of Public Health*, *103*(S1), S152–S166. doi:10.2105/AJPH.2013.301327

Chabris, C. F., Lee, J. J., Cesarini, D., Benjamin, D. J., & Laibson, D. I. (2015). The fourth law of behavior genetics. *Current Directions in Psychological Science*, *24*(4), 304–312. doi:10.1177/0963721415580430

Chang, C. C., Chow, C. C., Tellier, L. C. A. M., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. *GigaScience*, *4*, 7. doi:10.1186/s13742-015-0047-8

Davis, P. J. M. & Wright, E. A. (1977). A new method for measuring cranial cavity volume and its application to the assessment of cerebral atrophy at autopsy. *Neuropathology and Applied Neurobiology*, *3*(5), 341–358. doi:10.1111/j.1365-2990.1977.tb00595.x

Deaner, R. O., Isler, K., Burkart, J., & van Schaik, C. P. (2007). Overall brain size, and not encephalization quotient, best predicts cognitive ability across non-human primates. *Brain, Behavior and Evolution*, *70*(2), 115–124. doi:10.1159/000102973

Deaner, R. O., van Schaik, C. P., & Johnson, V. (2006). Do some taxa have better domain-general cognition than others? a meta-analysis of nonhuman primate studies. *Evolutionary Psychology*, *4*, 149–196. doi:10.1177/147470490600400114

Duffy, D. L. & Martin, N. G. (1994). Inferring the direction of causation in cross-sectional twin data: Theoretical and empirical considerations. *Genetic Epidemiology*, *11*(6), 483–502. doi:10.1002/gepi.1370110606

Frisell, T., Öberg, S., Kuja-Halkola, R., & Sjölander, A. (2012). Sibling comparison designs: Bias from non-shared confounders and measurement error. *Epidemiology*, *23*(5), 713–720. doi:10.1097/EDE.0b013e31825fa230

Gignac, G. E. (2005). Revisiting the factor structure of the WAIS-R: Insights through nested factor modeling. *Assessment*, *12*(3), 320–329. doi:10.1177/1073191105278118

Gignac, G. E. (2015). Raven's is not a pure measure of general intelligence: Implications for *g* factor theory and the brief measurement of *g*. *Intelligence*, *52*, 71–79. doi:10.1016/j.intell.2015.07.006

Gignac, G. E. & Bates, T. C. (2017). Brain volume and intelligence: The moderating role of intelligence measurement quality. *Intelligence*, *64*, 18–29. doi:10.1016/j.intell.2017.06.004

Gignac, G. E. & Szodorai, E. T. (2016). Effect size guidelines for individual differences researchers. *Personality and Individual Differences*, *102*, 74–78. doi:10.1016/j.paid.2016.06.069

Gignac, G. E., Vernon, P. A., & Wickett, J. C. (2003). Factors influencing the relationship between brain size and intelligence. In H. Nyborg (Ed.), *The scientific study of*

general intelligence: Tribute to Arthur R. Jensen (pp. 93–106). Amsterdam, The Netherlands: Pergamon.

Gondro, C., van der Werf, J. H. J., & Hayes, B. J. (Eds.). (2013). *Genome-wide association studies and genomic prediction.* New York, NY: Springer. doi:10.1007/978-1-62703-447-0

Hagenaars, S. P., Harris, S. E., Davies, G., Hill, W. D., Liewald, D. C. M., Ritchie, S. J., … Deary, I. J. (2016). Shared genetic aetiology between cognitive functions and physical and mental health in UK Biobank ($n = 112151$ and 24 GWAS consortia. *Molecular Psychiatry, 21*(11), 1624–1632. doi:10.1038/mp.2015.225

Haier, R. J. (2017). *The neuroscience of intelligence.* New York, NY: Cambridge University Press.

Herculano-Houzel, S. (2016). *The human advantage: A new understanding of how our brain became remarkable.* Cambridge, MA: MIT Press.

Herculano-Houzel, S., Catania, K., Manger, P. R., & Kaas, J. H. (2015). Mammalian brains are made of these: A dataset of the numbers and densities of neuronal and nonneuronal cells in the brain of glires, primates, scandentia, eulipotyphlans, afrotherians and artiodactyls, and their relationship with body mass. *Brain, Behavior and Evolution, 86*(3–4), 145–163. doi:10.1159/000437413

Herculano-Houzel, S., Manger, P. R., & Kaas, J. H. (2014). Brain scaling in mammalian evolution as a consequence of concerted and mosaic changes in numbers of neurons and average neuronal cell size. *Frontiers in Neuroanatomy, 8*(4), 77. doi:10.3389/fnana.2014.00077

Hibar, D. P., Stein, J. L., Rentería, M. E., Arias Vasquez, A., Desrivières, S., Jahanshad, N., … Medland, S. E. (2015). Common genetic variants influence human subcortical brain structures. *Nature, 520*(7546), 224–229. doi:10.1038/nature14101

Holloway, R. L. (2008). The human brain evolving: A personal retrospective. *Annual Review of Anthropology, 37*(1), 1–19. doi:10.1146/annurev.anthro.37.081407.085211

Jensen, A. R. (1994). Psychometric *g* related to differences in head size. *Personality and Individual Differences, 17*(5), 597–606. doi:10.1016/0191-8869(94)90132-5

Jensen, A. R. (2001). Vocabulary and general intelligence. *Behavioral and Brain Sciences, 24*(6), 1109–1110. doi:10.1017/S0140525X01280133

Jensen, A. R. & Johnson, F. W. (1994). Race and sex differences in head size and IQ. *Intelligence, 18*(3), 309–333. doi:10.1016/0160-2896(94)90032-9

Jensen, A. R. & Sinha, S. N. (1993). Physical correlates of human intelligence. In P. A. Vernon (Ed.), *Biological approaches to the study of human intelligence* (pp. 139–242). Norwood, NJ: Ablex.

Jerison, H. J. (1973). *Evolution of the brain and intelligence.* New York, NY: Academic Press.

Johnson, W., Bouchard, T. J., Jr., Krueger, R. F., McGue, M., & Gottesman, I. I. (2004). Just one *g*: Consistent results from three test batteries. *Intelligence, 32*(1), 95–107. doi:10.1016/S0160-2896(03)00062-X

Jung, R. E. & Haier, R. J. (2007). The Parieto-Frontal Integration Theory (P-FIT) of intelligence: Converging neuroimaging evidence (with discussion). *Behavioral and Brain Sciences, 30*(2), 135–187. doi:10.1017/S0140525X07001185

Keller, M. C., Garver-Apgar, C. E., Wright, M. J., Martin, N. G., Corley, R. P., Stallings, M. C., … Zietsch, B. P. (2013). The genetic correlation between height and IQ: Shared genes or assortative mating? *PLoS Genetics, 9*(4), e1003451. doi:10.1371/journal.pgen.1003451

Kievit, R. A., van Rooijen, H., Wicherts, J. M., Waldorp, L. J., Kan, K.-J., Scholte, H. S., & Borsboom, D. (2012). Intelligence and the brain: A model-based approach. *Cognitive Neuroscience, 3*(2), 89–97. doi:10.1080/17588928.2011.628383

Lee, J. J. (2012). Correlation and causation in the study of personality (with discussion). *European Journal of Personality, 26*(4), 372–412. doi:10.1002/per.1863

Lee, J. J. & Chabris, C. F. (2017). Genomic data can illuminate the architecture and evolution of cognitive abilities. *Behavioral and Brain Sciences, 40*, e209. doi:10.1017/S0140525X16002004

Lee, J. J. & Chow, C. C. (2013). The causal meaning of Fisher's average effect. *Genetics Research, 95*(2–3), 89–109. doi:10.1017/S0016672313000074

Lee, J. J. & Chow, C. C. (2014). Conditions for the validity of SNP-based heritability estimation. *Human Genetics, 133*(8), 1011–1022. doi:10.1007/s00439-014-1441-5

Lee, J. J., McGue, M., Iacono, W. G., & Chow, C. C. (2018). The accuracy of LD Score regression as an estimator of confounding and genetic correlations in genome-wide association studies. *Genetic Epidemiology, 42*(8), 783–795. doi:10.1002/gepi.22161

Lee, J. J., Vattikuti, S., & Chow, C. C. (2016). Uncovering the genetic architectures of quantitative traits. *Computational and Structural Biotechnology Journal, 14*, 28–34. doi:10.1016/j.csbj.2015.10.002

Lee, J. J., Wedow, R., Okbay, A., Kong, E., Maghzian, O., Zacher, M., … Cesarini, D. (2018). Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nature Genetics, 50*(8), 1112–1121. doi:10.1038/s41588-018-0147-3

Lee, S. H., Yang, J., Goddard, M. E., Visscher, P. M., & Wray, N. R. (2012). Estimation of pleiotropy between complex diseases using single-nucleotide polymorphism-derived

genomic relationships and restricted maximum likelihood. *Bioinformatics*, *28*(19), 2540–2542. doi:10.1093/bioinformatics/bts474

Loh, P.-R., Kichaev, G., Gazal, S., Schoech, A. P., & Price, A. L. (2018). Mixed-model association for biobank-scale datasets. *Nature Genetics*, *50*(7), 906–908. doi:10.1038/s41588-018-0144-6

Lord, F. M. & Novick, M. R. (1968). *Statistical theories of mental test scores.* Reading, MA: Addison-Wesley.

Lu, Q., Li, B., Ou, D., Erlendsdottir, M., Powles, R. L., Jiang, T., … Zhao, H. (2017). A powerful approach to estimating annotation-stratified genetic covariance via GWAS summary statistics. *American Journal of Human Genetics*, *101*(6), 939–964. doi:10.1016/j.ajhg.2017.11.001

Lynch, M. & Walsh, B. (1998). *Genetics and the analysis of quantitative traits.* Sunderland, MA: Sinauer.

MacLean, E. L., Hare, B., Nunn, C. L., Addessi, E., Amici, F., Anderson, R. C., … Zhao, Y. (2014). The evolution of self-control. *Proceedings of the National Academy of Sciences USA*, *111*(20), E2140–E2148. doi:10.1073/pnas.1323533111

MacLullich, A., Ferguson, K. J., Deary, I. J., Seckl, J. R., Starr, J. M., & Wardlaw, J. M. (2002). Intracranial capacity and brain volumes are associated with cognition in healthy elderly men. *Neurology*, *59*(2), 169–174. doi:10.1212/WNL.59.2.169

McDonald, R. P. (1999). *Test theory: A unified treatment.* Mahwah, NJ: Erlbaum.

McGue, M., Osler, M., & Christensen, K. (2010). Causal inference and observational research: The utility of twins. *Perspectives on Psychological Science*, *5*(5), 546–556. doi:10.1177/1745691610383511

Miller, M. B., Basu, S., Cunningham, J., Eskin, E., Malone, S. M., Oetting, W. S., … McGue, M. (2012). The Minnesota Center for Twin and Family Research genome-wide association study. *Twin Research and Human Genetics*, *15*(06), 767–774. doi:10.1017/thg.2012.62

Minica, C. C., Dolan, C. V., Boomsma, D. I., de Geus, E. J. C., & Neale, M. C. (2018). Extending causality tests with genetic instruments: An integration of Mendelian Randomization with the classical twin design. *Behavior Genetics*, *48*(4), 337–349. doi:10.1007/s10519-018-9904-4

Ni, G., Moser, G., Schizophrenia Working Group of the Psychiatric Genomics Consortium, Wray, N. R., & Lee, S. H. (2018). Estimation of genetic correlation via linkage disequilibrium score regression and genomic restricted maximum likelihood. *American Journal of Human Genetics*, *102*(6), 1185–1194. doi:10.1016/j.ajhg.2018.03.021

Nisbett, R. E., Aronson, J., Blair, C., Dickens, W., Flynn, J. R., Halpern, D. F., & Turkheimer, E. (2012). Intelligence: New findings and theoretical developments. *American Psychologist*, *67*(2), 130–159. doi:10.1037/a0026699

O'Connor, L. J. & Price, A. L. (2018). Distinguishing genetic correlation from causation across 52 diseases and complex traits. *Nature Genetics*, *50*(12), 1728–1734. doi:10.1038/s41588-018-0255-0

Okbay, A., Beauchamp, J. P., Fontana, M. A., Lee, J. J., Pers, T. H., Rietveld, C. A., … Benjamin, D. J. (2016). Genome-wide association study identifies 74 loci associated with educational attainment. *Nature*, *533*(7604), 539–542. doi:10.1038/nature17671

Pennington, B. F., Filipek, P. A., Lefly, D., Chhabildas, N., Kennedy, D. N., Simon, J. H., … DeFries, J. C. (2000). A twin MRI study of size variations in human brain. *Journal of Cognitive Neuroscience*, *12*(1), 223–232. doi:10.1162/089892900561850

Pers, T. H., Karjalainen, J., Chan, Y., Westra, H.-J., Wood, A. R., Yang, J., … Franke, L. (2015). Biological interpretation of genome-wide association studies using predicted gene functions. *Nature Communications*, *6*, 5890. doi:10.1038/ncomms6890

Pickrell, J. K., Berisa, T., Liu, J. Z., Ségurel, L., Tung, J. Y., & Hinds, D. A. (2016). Detection and interpretation of shared genetic influences on 42 human traits. *Nature Genetics*, *48*(7), 709–717. doi:10.1038/ng.3570

Pietschnig, J., Penke, L., Wicherts, J. M., Zeiler, M., & Voracek, M. (2015). Meta-analysis of associations between human brain volume and intelligence differences: How strong are they and what do they mean? *Neuroscience and Biobehavioral Reviews*, *57*, 411–432. doi:10.1016/j.neubiorev.2015.09.017

Posthuma, D., de Geus, E. J. C., Baaré, W. F. C., Hulshoff Pol, H. E., Kahn, R. S., & Boomsma, D. I. (2002). The association between brain volume and intelligence is of genetic origin. *Nature Neuroscience*, *5*(2), 83–84. doi:10.1038/nn0202-83

Redden, D. T. & Allison, D. B. (2006). The effect of assortative mating upon genetic association studies: Spurious associations and population substructure in the absence of admixture. *Behavior Genetics*, *36*(5), 678–686. doi:10.1007/s10519-006-9060-0

Rentería, M. E., Hansell, N. K., Strike, L. T., McMahon, K. L., de Zubicaray, G. I., Hickie, I. B., … Wright, M. J. (2014). Genetic architecture of subcortical brain regions: Common and region-specific genetic contributions. *Genes, Brain, and Behavior*, *13*(8), 821–830. doi:10.1111/gbb.12177

Ritchie, S. J., Cox, S. R., Shen, X., Lombardo, M. V., Reus, L. M., Alloza, C., … Deary, I. J. (2018). Sex differences in the adult human brain: Evidence from 5216 UK Biobank participants. *Cerebral Cortex*, *28*(8), 2959–2975. doi:10.1093/cercor/bhy109

Savage, J. E., Jansen, P. R., Stringer, S., Watanabe, K., Bryois, J., de Leeuw, C. A., … Posthuma, D. (2018). Genome-wide association meta-analysis in 269,867 individuals

identifies new genetic and functional links to intelligence. *Nature Genetics*, *50*(7), 912–919. doi:10.1038/s41588-018-0152-6

Schoenemann, P. T., Budinger, T. F., Sarich, V. M., & Wang, W. S.-Y. (2000). Brain size does not predict general cognitive ability within families. *Proceedings of the National Academy of Sciences USA*, *97*(9), 4932–4937. doi:10.1073/pnas.97.9.4932

Silventoinen, K., Iacono, W. G., Krueger, R., & McGue, M. (2012). Genetic and environmental contributions to the association between anthropometric measures and IQ: A study of Minnesota twins at age 11 and 17. *Behavior Genetics*, *42*(3), 393–401. doi:10.1007/s10519-011-9521-y

Sniekers, S., Stringer, S., Watanabe, K., Jansen, P. R., Coleman, J. R. I., Krapohl, E., … Posthuma, D. (2017). Genome-wide association meta-analysis of 78,308 individuals identifies new loci and genes influencing human intelligence. *Nature Genetics*, *49*(7), 1107–1112. doi:10.1038/ng.3869

Speed, D., Hemani, G., Johnson, M. R., & Balding, D. J. (2012). Improved heritability estimation from genome-wide SNPs. *American Journal of Human Genetics*, *91*(6), 1011–1021. doi:10.1016/j.ajhg.2012.10.010

Swanson, L. W. (1995). Mapping the human brain: Past, present, and future. *Trends in Neurosciences*, *18*(11), 471–474. doi:10.1016/0166-2236(95)92766-J

Taal, H. R., St Pourcain, B., Thiering, E., Das, S., Mook-Kanamori, D. O., Warrington, N. M., … Early Growth Genetics Consortium. (2012). Common variants at 12q15 and 12q24 are associated with infant head circumference. *Nature Genetics*, *44*(5), 532–538. doi:10.1038/ng.2238

Trzaskowski, M., Davis, O. S. P., DeFries, J. C., Yang, J., Visscher, P. M., & Plomin, R. (2013). DNA evidence for strong genome-wide pleiotropy of cognitive and learning abilities. *Behavior Genetics*, *43*(4), 267–273. doi:10.1007/s10519-013-9594-x

Tsuboi, M., van der Bijl, W., Kopperud, B. T., Erritzøe, J., Voje, K. L., Kotrschal, A., … Kolm, N. (2018). Breakdown of brain-body allometry and the encephalization of birds and mammals. *Nature Ecology and Evolution*, *2*(9), 1492–1500. doi:10.1038/s41559-018-0632-1

Turkheimer, E. & Waldron, M. (2000). Nonshared environment: A theoretical, methodological, and quantitative review. *Psychological Bulletin*, *126*(1), 78.

Van Essen, D. C., Ugurbil, K., Auerbach, E., Barch, D., Behrens, T. E. J., Bucholz, R., … WU-Minn HCP Consortium. (2012). The Human Connectome Project: A data acquisition perspective. *NeuroImage*, *62*(4), 2222–2231. doi:10.1016/j.neuroimage.2012.02.018

Van Essen, D. C., Smith, S. M., Barch, D. M., Behrens, T. E. J., Yacoub, E., Ugurbil, K., & WU-Minn HCP Consortium. (2013). The WU-Minn Human Connectome Project: An overview. *NeuroImage*, *80*, 62–79. doi:10.1016/j.neuroimage.2013.05.041

van Leeuwen, M., Peper, J. S., van den Berg, S. M., Brouwer, R. M., Pol, H. E. H., Kahn, R. S., & Boomsma, D. I. (2009). A genetic analysis of brain volumes and IQ in children. *Intelligence*, *37*(2), 181–191. doi:10.1016/j.intell.2008.10.005

Visscher, P. M. & Yang, J. (2016). A plethora of pleiotropy across complex traits. *Nature Genetics*, *48*(7), 707–708. doi:10.1038/ng.3604

Wang, S. S.-H., Ambrosini, A. E., & Wittenberg, G. M. (2016). Evolution and scaling of dendrites. In G. Stuart, N. Spruston, & M. H ausser (Eds.), *Dendrites* (3rd ed., pp. 47–75). New York, NY: Oxford University Press. doi:10.1093/acprof:oso/9780198745273.003.0002

Wickett, J. C., Vernon, P. A., & Lee, D. H. (2000). Relationships between factors of intelligence and brain volume. *Personality and Individual Differences*, *29*(6), 1095–1122. doi:10.1016/S0191-8869(99)00258-5

Zinbarg, R. E., Yovel, I., Revelle, W., & McDonald, R. P. (2006). Estimating generalizability to a latent variable common to all of a scale's indicators: A comparison of estimators for ω_h. *Applied Psychological Measurement*, *30*(2), 121–144. doi:10.1177/0146621605278814