Intelligence 38 (2010) 393–401

Contents lists available at ScienceDirect

Intelligence

Smarter people are (a bit) more symmetrical: A meta-analysis of the relationship between intelligence and fluctuating asymmetry

George C. Banks *, John H. Batchelor, Michael A. McDaniel

Virginia Commonwealth University, United States

ARTICLE INFO

Article history:
Received 24 February 2010
Received in revised form 12 April 2010
Accepted 16 April 2010
Available online 15 May 2010

Keywords:
Fluctuating asymmetry
Symmetry
Intelligence
IQ
General fitness factor
g
Publication bias

ABSTRACT

Individual differences in general mental ability (*g*) have important implications across multiple disciplines. Research suggests that the variance in *g* may be due to a general fitness factor. If this is the case, a relationship should exist between *g* and other reliable indicators of fitness. Some empirical results indicate a relationship between *g* and fluctuating asymmetry. However, there have been inconsistencies in the results, some of which may be due to random sampling error, and some of which may be due to moderating variables, publication bias, and methodological issues. To help clarify the literature, a meta-analysis was conducted on the relationship between *g* and fluctuating asymmetry. Based on 14 samples across 1871 people, estimates of the population correlation ranged from −.12 to −.20. There was a difference in the magnitude of the correlation between published studies and unpublished studies with published studies showing larger magnitude negative correlations and unpublished studies yielding results closer to zero. The implications for our understanding of *g* and its relationship with fluctuating asymmetry are discussed.

1. Introduction

The source of individual differences in general mental ability (*g*) has been a focus of scientific inquiry since Spearman (1904). The comprehension of individual differences in cognitive ability has broad implications across diverse research areas from personnel selection (McDaniel & Banks, 2010) and health and mortality (Arden, Gottfredson, & Miller, 2009; Arden, Gottfredson, Miller, & Pierce, 2009; Deary, 2008; Gottfredson, 2004) to brain volume (McDaniel, 2005) and understanding how sexual selection might have affected the development of cognitive ability (Miller, 2000a). This latter point will be the focus of this meta-analysis. The objective is to evaluate the cumulative data concerning the relationship between *g* and body symmetry as caused by a general fitness factor. Furthermore, the potential moderating effects of sex, methodological approaches, and publication bias are explored. Finally, we identify methodological and reporting practices which retard the development of advances in the literature. Such limitations are typical of research programs in their early stage of development and we hope that our recommendations serve to guide future research.

Miller (2000a, 2007) argued that human mental evolution is driven not only by natural selection, but also by sexual selection and mate choice. As sexual selection and mate choice are driven by reproductive success, one proposed indicator of general fitness is *g*. In this context, fitness is a latent variable of multiple adaptations that increases reproductive success and general survival (Miller, 2000b). Costly signaling theory (Bradbury & Vehrencamp, 1998; Miller, 2007) supports the notion that *g* may be considered a reliable indicator of fitness because it has high costs and is difficult to fake. Similar to how a male peacock's tail illustrates to females a signal of fitness, *g* may serve the same purpose for humans (Luxen & Buunk, 2006). In fact, research has demonstrated a relationship between semen quality (e.g., sperm concentration and sperm count) and *g* (Arden et al., 2009b). From this perspective, *g* is caused by a general fitness factor. If so, one

* Corresponding author. Virginia Commonwealth University, 301 West Main Street, PO Box 844000, Richmond, VA 23284, United States.
E-mail address: banksgc@vcu.edu (G.C. Banks).

0160-2896/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.intell.2010.04.003

would expect a relationship between *g* and other proposed indicators of general biological fitness, such as body symmetry.

An increasingly researched question is: What is the relationship between *g* and fluctuating asymmetry (FA)? In this instance, FA is characterized by variation from perfect symmetry in the body where there is symmetry in the non-pathological population of humans (Møller & Swaddle, 1997; Van Valen, 1962). Individual differences in body symmetry can result from genetic causes (Møller & Thornhill, 1997) or from adverse environmental conditions (Gangestad & Thornhill, 1999). Subsequently, FA in the lengths and widths of bilateral body parts such as fingers and toes, hands and feet, and other attributes of the human body may be used by researchers as a measure of morpho-developmental stability. Developmental stability refers to how accurately genotypes are turned into phenotypes (Furlow, Armijo-Prewitt, Gangestad, & Thornhill, 1997). Put another way, developmental stability demonstrates how precisely an organism develops "according to plan." Developmental instability has been linked to interference in proper neurological functioning resulting in mental retardation (Thornhill & Møller, 1997). Deviation in cognitive and bilateral trait development can both occur during early fetal stages (Rahman, Wilson, & Abrahams, 2004). Furlow et al. (1997) published the seminal work in a stream of research which explored the hypothesis that there exists a relationship between indicators of developmental instability, such as low levels of *g* and FA. The majority of studies published in this area have indicated a modest correlation between these two factors. However, a review of this literature indicates several issues for concern. These issues include small sample sizes as well as variability in the measurement of FA and *g*. These issues are addressed in turn.

1.1. Challenges in measuring FA and g

The majority of the studies included in the literature have relatively small samples sizes (averaging 134) with the largest published study in this area having a sample size of only 263 participants (Johnson, Segal, & Bouchard, 2008). With these relatively small sample sizes, much of the variability in this literature is likely due to random sampling error. A meta-analysis of results across studies could address this limitation.

There are also causes for concern due to the wide variety of approaches used to measure FA in the samples. Some of the variability in results across studies is likely a function of the approach to the measurement of FA, that is, some FA measures may have greater construct validity for the measurement of FA than others. This will likely cause variance across studies in the correlation of FA and g. Moreover, some of the variation in the results may be due to methodological differences. For instance, due to the limitations of their archival data set, the study by Johnson et al. (2008) was only able to report measurements that were taken once which is a practice inconsistent with the majority of the other studies we reviewed. However, this study did claim to find heritability of FA, which suggests they had at least some reliability in their measure. Also, there is variation in the number of bilateral body traits measured, ranging from 2 to 35. In addition, some of the traits measured were soft tissue body parts which can be subjected to minor fluctuations. Still, most of the studies focus on the traits demonstrated by Thornhill and Gangestad (1994) to be reliably measured, such as finger lengths.

We note that various methods of assessing FA can be useful for indicating that the results are robust to differences in the measurement. However, little research is available about the comparability of FA measures across studies. To the extent that there are differences across studies in the adequacy of the FA measures, one can expect those differences to cause variability across studies in the FA–*g* relationship.

Additional methodological issues include variation in the types of measures of *g* and differences in the variability of the administration of the measures, even when the measures use the same items. For instance, it was common for the most frequently used measure of cognitive ability in the studies, the Raven's Progressive Matrices (RPM), to be administered as timed in some studies and untimed in others. It was also common to see some studies use the Standard Progressive Matrices (SPM) while other studies use the Advanced Progressive Matrices (APM), the latter being a more challenging test meant for more intelligent participants (Gudjonsson, 1995). Also, the use of the Raven's APM varied from 12 items (Arthur & Day, 1994) to 36 items in Set II. Furthermore, the use of only Set II (without Set I) may affect the correlations of the test with other measures (Raven, Raven, & Court, 1998). Finally, there are limitations to the normative data of the SPM and APM (Gudjonsson, 1995) and we believe that these issues have not yet been addressed in later editions of the tests. We are not arguing that these are not all excellent measures of *g*. Our argument is that the differences between these instruments present serious challenges for estimating the extent of range restriction in these samples and for estimating the magnitude of the FA and *g* correlation in the population.

Range restriction, particularly indirect range restriction (Hunter, Schmidt, & Le, 2006), presents problems in drawing inferences from sample correlations with a limited range of scores on FA and *g* to the population parameter in the unrestricted population. Many studies in this meta-analysis used undergraduate participants which very likely restricted the *g* variance due to cognitive hurdles associated with being admitted to college. To estimate the population correlation in an unrestricted sample, one needs to know the ratio of the standard deviation of *g* in the sample to the standard deviation of *g* in the population. This requires accurate norms on the *g* measure, say the APM, in the population and that the primary studies used the test with the same items and the same administration instructions as was used in the norming study. This also requires that the authors accurately report which test they used (e.g., set 1, set 2, or both item sets of the APM). In addition, to estimate the population correlation without the biasing effect of indirect range restriction, one needs to know the reliability of both the FA measure and the *g* measure (Hunter et al., 2006). In this literature, the choices made concerning the *g* measures and their administration protocols, the failure to calculate and/or report reliability data of both FA and *g*, and the failure to report basic descriptive information such as means, standard deviations, and a correlation matrix, make the accurate estimation of the population correlation quite difficult.

Some authors do better than others in these matters, but as a whole, the authors of this literature, through their methodological and reporting practice decisions, hamper the conclusions that can be drawn across studies. As a result, our estimate of the population correlation (the correlation between *g* and FA both measured without measurement error and without range restriction) is best viewed as a tentative estimate. The estimate of the population parameter will remain tentative until such time as the researchers in this area remedy non-optimal measurement and result reporting practices.

In addition to this literature's methodological limitations, arguments have been made that there is a moderating effect due to sex (Rahman et al., 2004), although there are counter arguments as well (Borgerhoff Mulder, 2009; Hooper & Miller, 2008). There is theoretical and empirical evidence to support the belief that the relationship between *g* and FA is stronger in males than in females. Therefore, some studies use all male samples to examine this relationship (Gangestad, 2010a,b; Prokosch, Yeo, & Miller, 2005; Thoma et al., 2005). A theoretical argument can be made that males are under greater selection pressure to develop reliable indicators of fitness; thus variance in reproductive success would be larger in magnitude in males (Prokosch et al., 2005). Empirical evidence has supported stronger relationships in males than in females between FA and facial attractiveness (Gangestad & Thornhill, 1999; Gangestad, Thornhill, & Yeo, 1994) and levels of depression (Martin, Manning, & Dowrick, 1999).

Counter arguments have been made to this position that in monogamous populations, mutual mate choice puts comparable pressure on both males and females to develop costly signals (Hooper & Miller, 2008). There is also empirical evidence to suggest that variance in reproductive success is not greater for males (Borgerhoff Mulder, 2009). Furthermore, if both FA and *g* do in fact result from a general fitness factor, sex differences would not be likely as both males and females would have equal variance in genetic quality. However, to evaluate the sex moderator accurately, all studies would need to report the FA and *g* correlations separately by sex. This was often not the case.

There is also some empirical and theoretical support for age as a moderating variable of the relationship between *g* and symmetry in facial features as rated by judges (Zebrowitz, Hall, Murphy, & Rhodes, 2002). Zebrowitz et al. (2002) found that the relationship was .27 in childhood, but that the relationship decreased to −.02 in adulthood. The underlying logic is that environmental and genetic stress affect developmental stability which influences *g* and FA (Møller & Swaddle, 1997). In other words, as an individual develops, he or she is vulnerable to environmental stress, but as the individual's development stabilizes the traits are less easily influenced. To evaluate this moderator, the samples would need to have a reasonably wide age range and would need to report correlations separately by some aggregation of age range, for example, 30 years of age and above and under 30 years of age. Data in this format were not reported in any of the studies.

Note that we do not wish to single out FA and *g* researchers for particular scrutiny. Non-optimal decisions in methodology, measurement, and reporting practices are common in psychology and most disciplines. We detail these problems in FA and *g* research because they need to be considered in any meta-analytic synthesis of these data. We also raise these issues because the relationship between FA and *g* is an important topic and we seek to encourage better practices in future studies in this critical research area.

In summary, this meta-analysis seeks to summarize the relationship between *g* and FA. We also seek to evaluate two potential moderators for which some amount of data exist. Finally, we seek to evaluate the extent to which, if any, publication bias (Rothstein, Sutton, & Borenstein, 2005) may be distorting conclusions drawn from this literature.

2. Methods

2.1. Literature review

A literature search was conducted using *Google Scholar, EBSCOhost* and research databases including *Academic Search Complete, CINAHL, Education Research Complete, ERIC, Health Source, MEDLINE, PsycArticles, Psychology and Behavioral Sciences Collection,* and *PsycINFO*. The key words used were "fluctuating asymmetry", "symmetry" paired with "intelligence," and "asymmetry" paired with "intelligence." Furthermore, a citation index search was conducted of all publications known to include the variables of interest using the *Social Science Citation Index*. The references in these articles were also reviewed. The preliminary findings of this study were then sent to the researchers of the identified studies and the authors were asked for leads to any additional published or unpublished studies. We also solicited suggestions and critiques of the work in progress.

2.2. Decision rules

This meta-analysis evaluated all obtained correlations between *g* and FA. The most commonly used measure of intelligence was the RPM. When studies used more than one cognitive ability test, we coded the relationship between FA and the RPM in an attempt to achieve consistency between studies. One exception to this decision rule was when a primary author used a principal component analysis to estimate *g* (Penke et al., 2009). A few studies provided results for more than one sample. One correlation was coded for each sample.

When possible, the studies were also coded by age and sex. Insufficient information was reported in the studies to analyze results by age. One study by Yeo, Gangestad, Thoma, and Repa (1997) was excluded because it did not include a comparable measure of *g* (e.g., dichotic listening). The schizophrenic group from Euler, Thoma, Gangestad, Canive, and Yeo (2009) was excluded from the analysis to increase the generalizability of the results to the normal population. The correlation between the comprehensive FA index (when controlling for directional asymmetry) and *g* at age 83 was chosen as the effect size from Penke et al. (2009) because *g* was measured at the same time as FA. All the other correlations in the meta-analysis were zero-order correlations. Our efforts to reach Arnold (2005) via personal correspondence were unsuccessful and thus we were not able to obtain the required zero-order correlation between FA and *g* for the study's participants. Some of the data analyzed were based on personal communication from the authors.

2.3. Analysis approach

This study used multiple meta-analytic approaches. The psychometric approach advocated by Hunter and Schmidt (1990, 2004) was used because it permits corrections for measurement error and range restriction. This allows an estimation of the FA and *g* relationship at the construct level, that is, it permits an estimate of the population parameter not downwardly biased by measurement error and range restriction. Although the samples were not subject to direct range restriction (no study stated indicated that its samples were explicitly pre-screened on cognitive ability and FA), the nature of the samples (e.g., all volunteers; mostly college students) make the presence of indirect range restriction almost certain.

Although the psychometric meta-analysis approach permits an estimation of the population correlation unaffected by measurement error and range restriction, it incorporates no empirical methods for the detection of publication bias as is found in meta-analysis in the Hedges and Olkin (1985) meta-analytic tradition. Therefore, analyses were also performed using Comprehensive Meta-analysis (CMA) (Borenstein, Hedges, Higgins, & Rothstein, 2005) which incorporates empirical methods for the evaluation of publication bias. An additional advantage of using multiple perspectives of meta-analysis is that one can evaluate the extent of convergence across analysis approaches. To the extent that results are similar across studies, one can have greater confidence in the results.

We had limited data on measurement reliability and range restriction for both FA and *g*. All the studies for which we were able to obtain standard deviations indicated that there was range restriction on *g*. This is not surprising because the majority of participants in the samples were undergraduate students who typically have higher mean intelligence than the general population. In instances where information regarding the standard deviation were not available, and we did not receive a response back from the author of the corresponding primary article, we used the weighted average standard deviation from the samples in our analysis as the standard deviation for the studies with missing data. We also imputed the standard deviation for those studies which used the RPM developed by Arthur and Day (1994) as we were unaware of any normative data for this measure.

The sample means and standard deviations of *g* that we did have were converted and compared to an IQ score where the normal population mean is 100 and the standard deviation is 15. Peck (1970) demonstrated a conversion of the RPM to an IQ scale and noted that the use of percentiles in these calculations limited the possibilities of an accurate conversion. Furthermore, we point the reader to criticisms that should be noted of the standardized testing used to obtain the normative data of the Raven's APM and SPM (Gudjonsson, 1995). Such norms provide the estimated population standard deviation for the Raven's tests.

Using normative data from Table APM21 in the Raven's manual (Raven et al., 1998) we inferred a Raven's APM standard deviation of 9. We then converted Raven's APM scores for Rahman et al. (2004) and Prokosch et al. (2005). We do not offer these score adjustments as perfect translations of data into the IQ metric (100, 15). However, we do offer these score adjustments as the most reasonable ways of processing the reported data.

Few reliabilities were reported in the primary studies for either FA (Euler et al., 2009; Furlow et al., 1997; Gangestad, 2010a,b; Prokosch et al., 2005) or *g* (Furlow et al., 1997; Luxen & Buunk, 2006). The reported reliabilities ranged from .56 to .92 and from .46 to .79 for FA and *g* respectively. Based on the reliabilities reported and our familiarity with the reliability of *g* measures, we assigned an alpha reliability of .80 to all the FA and *g* measures. Although a reliability of .80 for *g* is slightly higher than that which was reported in the two studies in this meta-analysis that offered reliabilities, it is lower than the reliability of .90 which is reported in the Raven's manual (Raven, Court, & Raven, 1994). We do not offer this imputation as a perfect reproduction of the actual (and either unknown or not reported) reliabilities but as reasonable estimates. Although we assert that imputation of range restriction and reliability estimates are not unreasonable, the cumulative uncertainty of the imputation decisions makes our estimate of the population correlation tentative. We encourage re-estimation of the population correlation as data from studies that do not require imputation become available.

3. Results

3.1. Primary analyses

The findings of this meta-analysis were derived from 14 samples obtained from the 12 studies that met the described decision criteria (Bates, 2007; Euler et al., 2009; Furlow et al., 1997; Gangestad, 2010a,b; Johnson et al., 2008; Luxen & Buunk, 2006; Penke & Asendorpf, 2008; Penke et al., 2009; Prokosch et al., 2005; Rahman et al., 2004; Thoma et al., 2005). Our data are listed in Table 1 and our analysis results are in Table 2.

Our first set of analyses was fixed and random effect models using CMA. The fixed model yielded a point estimate of −.12 and the random effects model yielded a point estimate −.14. Whereas the studies varied with respect to the way the FA and *g* were measured and because there may be moderating variables that cause additional heterogeneity (i.e., non-sampling error variance in the distribution of correlation coefficients), the random effects model is the model most suitable for the analysis of these data. We next conducted a sensitivity analysis in which we removed one study at a time and recalculated the meta-analysis. This analysis yielded 14 point estimates (one for each time a study was removed) and helps to determine whether any specific study had an undue influence on the results. The means ranged from −.12 to −.17.

Second, we ran the fixed and random meta-analyses again with the removal of the sample obtained from Johnson et al. (2008). We completed this second step because several of the primary authors with whom we corresponded expressed concerns regarding Johnson et al.'s (2008) method of only measuring bilateral body parts once and the large sample size from this study might downwardly bias the analysis. This sensitivity analysis demonstrated a negligible change in both the CMA fixed and random effect models (fixed −.12 vs. −.15; random −.14 to −.16). Thus, we concluded that the

Table 1
Summary of studies included in the meta-analysis.

Study	Published vs. unpublished sample	N	r	u	% of women in sample	# of bilateral traits measured
Bates (2007) Sample 1 [a]	Published	98	−.43	.90	.67	9
Bates (2007) Sample 2 [a]	Published	164	−.29	.90	.76	4
Euler et al. (2009) [a]	Unpublished	23	.33	.90	.30	10
Furlow et al. (1997) Sample 1	Published	112	−.21	.84	.50	9
Furlow et al. (1997) Sample 2	Published	128	−.24	.85	.60	10
Gangestad (2010a)	Unpublished	67	.18	.90	.00	10
Gangestad (2010b)	Unpublished	98	.00	.90	.00	10
Johnson et al. (2008) [a]	Published	263	.03	.99	.61	10
Luxen & Buunk (2006)	Published	81	−.25	.90	.46	6
Penke & Asendorpf (2008) [a]	Unpublished	284	.02	.90	.50	12
Penke et al. (2009) [a]	Published	216	−.07	.90	.56	35
Prokosch et al. (2005) [a]	Published	78	−.39	.75	.00	10
Rahman et al. (2004)	Published	240	−.13	.90	.50	2
Thoma et al. (2005) [a]	Published	19	−.49	.80	.00	6

u = The ratio of the sample standard deviation of IQ to the population standard deviation of IQ.
[a] Additional information was provided by an author of the published study.

inclusion of the Johnson et al. data did not have a large downward influence on the mean correlation. The details of these specific analyses are available from the first author of this paper. We note that in these analyses there was no correction for the downward bias associated with measurement error and range restriction. To address these biases we reanalyzed the data with the psychometric meta-analysis approach.

One primary difference in methods between Hedges and Olkin (1985) fixed effect model and the Hunter and Schmidt (2004) psychometric approach is the weighting of the mean correlation. The former uses the inverse of the standard error and the latter uses the sample sizes. Although the former is a fixed-effect model and the latter a random-effect model, the two sets of weights are often very similar and in this set of studies are correlated at .82. This high relative similarity in study weights is reflected in the same estimate of the correlation (−.12) between the CMA fixed effect meta-analysis and the uncorrected correlation using the Hunter and Schmidt (1990, 2004) psychometric meta-analysis (commonly referred to as a *bare bones* analysis). When one removes the Johnson et al. (2008) effect size the bare bones analysis yields a correlation of −.14. Once again, we would conclude that the Johnson et al. (2008) correlation does not have an overpowering effect on the results. The strength of the psychometric meta-analysis approach is its estimation of the population correlation in the absence of the downwardly biasing effects of measurement error and range restriction. Following the approach detailed in Hunter et al. (2006), we individually corrected the correlations for measurement error and range restriction. With these corrections, the estimated population correlation was −.16 and with the Johnson et al. (2008) study removed, the correlation was −.20. When deleting the Johnson et al. study, the change in the psychometric meta-analysis (−.16 to −.20) was larger than the random effects meta-analysis using CMA (−.14 to −.16), due to differences in weighting the studies between the two meta-analysis methods. CMA, consistent with the Hedges and Olkin tradition, alters weights in a random effects meta-analysis and the weighting reduces the impact of larger sample studies. Even with the larger difference (−.16 to −.20), the choice of including or excluding the Johnson et al. study had no dramatic impact on the conclusion.

Table 2
Summary of meta-analysis: the correlations between FA and g.

	CMA random	CMA fixed	Bares bones correlation	Psychometric corrected correlation
Mean	−0.14	−0.12	−0.12	−0.16
CI	−.24 to −.05	−.17 to −.08	−.20 to −.03	−.27 to −.04
I^2	75%	75%	73%	75%

Confidence Interval (CI) set at .05 level. The standard error of the psychometric corrected correlation was based on the standard error of ρ reported in Hunter and Schmidt (2004, page 206) for correlations corrected individually. The statistic I^2 (Higgins, Thompson, Deeks & Altman, 2003) is the percent of the variance not due to statistical artifacts. For the first three analyses, the I^2 is the percent of observed variance not due to sampling error. For the last analysis where the correlations were individually corrected for measurement error and indirect range restriction, the I^2 is the percent of the observed variance of the corrected correlations that is not due to sampling error.

3.2. Meta regression analyses

To evaluate the potential moderating effect of sex, we would ideally have FA and g correlations for each sample by sex. Because that was not the case, we conducted a meta-regression analysis to evaluate whether percent of women in the sample moderated the FA and g relationship. The moderating effect of percent of women explained near zero variance in the FA and g correlations and the slope of the regression was not statistically significant. We note that the number of observations for the analysis is the number of effects. Given the analysis was based on only 14 observations, we suggest that conclusions concerning the strength and direction of the effect await additional data.

The number of bilateral traits measured in the FA composites was also tested as a moderator. The slope of the regression was not statistically significant and the magnitude of the moderating effect was near zero. Given that this analysis was based on only 14 observations, we suggest that conclusions concerning the strength and direction of the moderating effect, if any, await additional data.

3.3. Publication bias analyses

Publication bias exists to the extent that the studies available to be reviewed are not a representative sample of all studies. Typically, publication bias results in point estimates (i.e., the weighted mean correlation) that overestimate the magnitude of the population effect. We first examined differences between published and unpublished studies. The

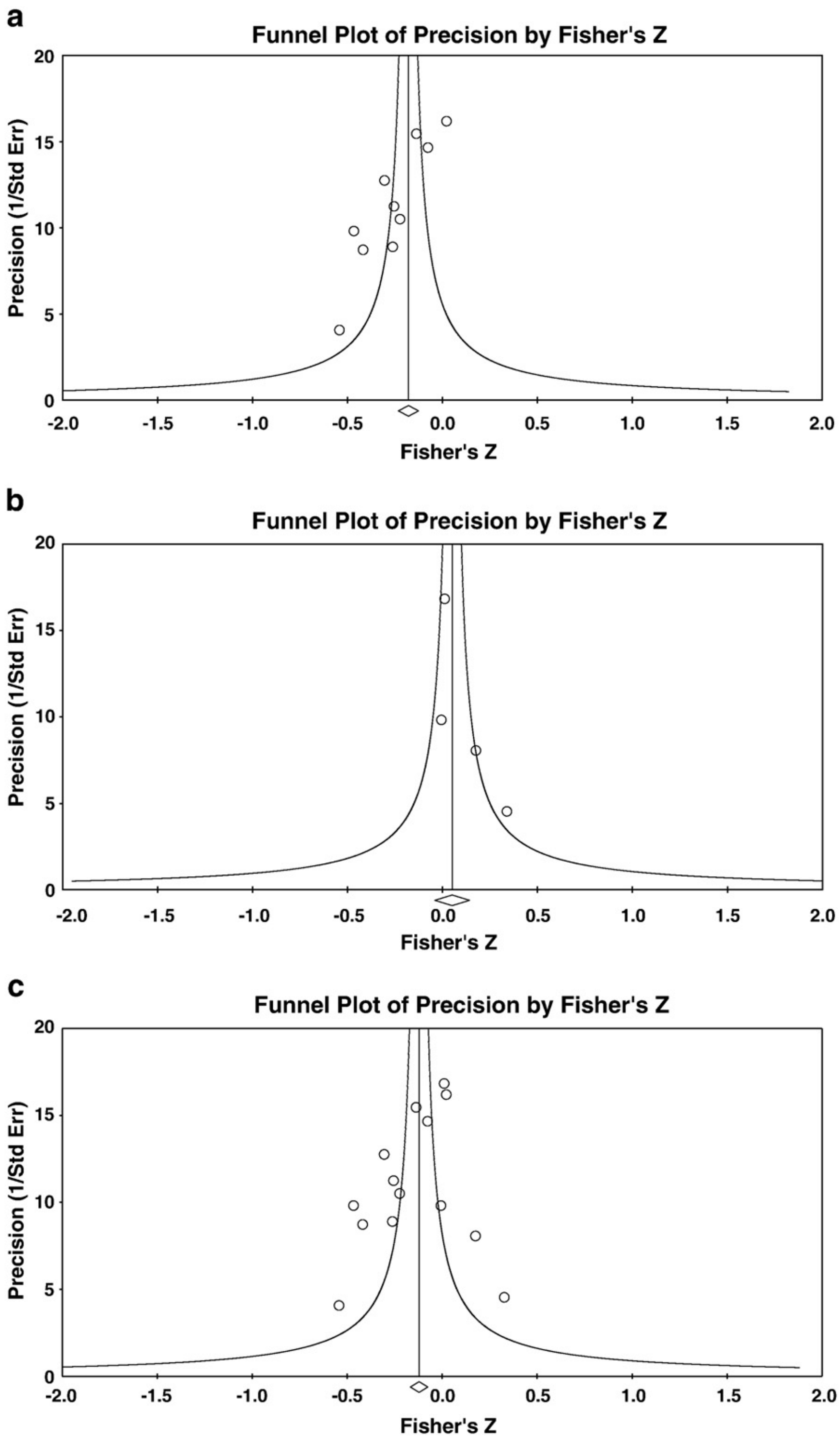

Fig. 1. a. Funnel plot of published studies ($n = 10$). b. Funnel plot of unpublished studies ($n = 4$). c. Funnel plot of published and unpublished studies combined ($n = 14$).

10 published studies yielded a mean effect size of −.22 whereas the 4 unpublished studies yielded a mean closer to zero (.06). We also examined funnel plots separately for the published and unpublished studies. Asymmetry in a funnel plot is an indication of potential publication bias because in the absence of heterogeneity (i.e., moderators and other non-sampling error sources of variance), Fisher-*z* transformed correlations will approach symmetry (Duval, 2005). Fig. 1a is the funnel plot for the published samples (Bates, 2007; Furlow et al., 1997; Johnson et al., 2008; Luxen & Buunk, 2006; Penke et al., 2009; Prokosch et al., 2005; Rahman et al., 2004; Thoma et al., 2005). The plot is substantially asymmetric. Fig. 1b is the funnel plot for the unpublished samples (Euler et al., 2009; Gangestad, 2010a,b; Penke & Asendorpf, 2008). This plot is also asymmetric but in a direction counter to the published studies. Collectively, the mean differences between published and unpublished studies and the complementary asymmetry in the funnel graphs indicates the FA and *g* studies are more likely to be published when they show a larger negative correlation than if they show a near zero or positive correlation. This is a cause for concern. If one only relies on published results, the magnitude of the typical FA and *g* correlation will likely overestimate the relationship between FA and *g*. We note that when one examines a funnel plot of the published and unpublished correlations combined (Fig. 1c), the funnel plot is largely symmetrical. Thus, we conclude that there is a possibility of publication bias in the published studies which is substantially diminished with the inclusion of the unpublished studies. Based on all 14 studies, our results are consistent with a conclusion of a small negative correlation between FA and g. Across analyses, about 72 to 75% of the variance across studies remained unexplained (see Table 2). Some of this variance may be due to differences across studies in the manner in which FA and *g* were measured. It is also possible that one or more moderators influence the relationship between FA and *g*. With this much unexplained variance, the mean correlation estimates (−.12 to −.16) should be interpreted with caution until the source(s) of the unexplained variance can be determined.

4. Discussion

Using multiple meta-analytic methods, we demonstrated that the correlation between FA and *g* is negative with estimates of the population correlation ranging from −.12 to −.20. The relationship between FA and *g* may be explained by their association with a general fitness factor which influences sexual selection and mate choice (Miller, 2000a, 2007). Here the general fitness factor is said to increase both the chance of survival and reproductive success (Miller, 2000b). Because of its high resource cost and resistance to faking, *g* is considered a reliable proxy for fitness (Bradbury & Vehrencamp, 1998; Miller, 2007). Our results support prior evidence that *g* may serve the same purpose in humans.

The relationship between *g* and FA is not large which indicates that common and distinct genetic and environmental factors likely contribute to variance in both. Common genetic factors such as genes or mutations and environmental factors such as prenatal infection, malnutrition, and environmental toxins have been suggested to affect the relationship between *g* and physical health (Arden, Gottfredson, & Miller, 2009). Evolutionary genetic models provide valuable guidance in explaining these potential causes of variance in the relationship between intelligence, symmetry, health, and other indicators of Darwinian fitness (for a review see Keller & Miller, 2006). It's quite likely that there exists a genetic correlation which underlies the phenotypic correlation that is observed. This finding is in need of further investigation as preliminary findings have not provided support (Johnson et al., 2008). Even if the association is largely genetic, it may not reflect a general fitness factor if other alleged fitness indicators don't load on the same genetic factor. If there is a general fitness factor, it may arise from basic effects of pleiotropic mutation load that would affect all species, and may not have any relation to sexual selection for intelligence as a fitness indicator.

In meta-analysis, the credibility of moderator analyses is a function of the number of studies. Our analyses of sex and number of FA measurements were based on only 14 observations, one for each effect size. We argue that it is best to reserve judgment about the presence of moderators until more studies become available.

Our analyses comparing published to unpublished studies were intriguing in that correlations from published studies tended to be larger than correlations from unpublished studies. The funnel plots suggested that there are forces at work that cause data consistent with a negative correlation to be published and for correlations near zero or positive to not be published. Author decisions may explain this pattern of findings. If an author has a result that is consistent with most published work, in this case a negative correlation, the author may be more motivated to submit it for publication. An author with a result inconsistent with most published work may be less motivated to publish the work or may wish to work on another project first before turning to the study that is counter to the trend in the literature. It would be beneficial to this research area if all competently conducted studies were published irrespective of the magnitude and direction of the effect.

4.1. Limitations and suggestions for future research

This research literature, not unlike most research literatures, could be improved. We offer several suggestions concerning data reporting and areas for greater research attention.

4.1.1. Data reporting issues

This research area is relatively new and, as is common with new research areas, there is room for improvement regarding reporting practices. A common contribution of a meta-analysis is the identification of issues that impede the culmination of results. The majority of studies did not report zero-order correlations separately by sex which limits the ability of a study such as this one to evaluate sex as a moderator. Also, many studies did not report means and standard deviations for the cognitive measures and did not report reliabilities for measures used. The research literature is also hampered when authors do not report the specific cognitive ability measure that was used and when the measure is shortened or the administration protocol is altered. Studies seldom reported the reliability of the FA measure. We note that measures of agreement are not necessarily measures of reliability.

We recommend that FA measures be based on a minimum of two assessors and that an FA composite be calculated separately for each assessor. One can then correlate the FA composites from the two assessors as a preliminary estimate of the reliability of the FA composite. We argue that this is a preliminary estimate because most researchers would want to average the two FA composites. The reliability estimate of the average of the two composites is the correlation boosted using the Spearman–Brown formula (Nunnaly & Bernstein, 1994). Additionally, each rater should follow the recommendations of Palmer and Strobeck (1986) as well as Gangestad et al. (1994) and use repeated measurements of the same bilateral traits.

4.1.2. The ideal study[1]

In order to continue to grow, the nature of the research studies in this area can be improved. An ideal study might be one that is taken from a community sample so as to better represent a normal population. This would serve to limit issues related to range restriction. Ideally, the study could include all MZ and DZ twins (or an extended-twin family design). This would allow researchers to test for both genetic and phenotypic correlations between FA and *g* (Johnson et al., 2008). Furthermore, the study should include both sexes (and report correlations separately by sex) as well as a wide range of ages. This would allow researchers to test for both sex and age moderators. Other demographic characteristics such as race and ethnicity should be reported. Again, this would allow researchers to test hypotheses related to different sexual selection histories and different degrees of genetic outbreeding and admixture.

It would also be valuable for researchers to look at the association in anthropological samples from small-scale societies of hunter–gatherer–horticulturalists, and from countries where average nutrition and health care is low, such that variance in genetic quality is more apparent phenotypically which could result in a stronger relationship between symmetry and intelligence. If the fitness factor is supported, it would be difficult to find single-nucleotide polymorphisms (SNPs) that predict either intelligence or symmetry (or both) in genome-wide association studies (Keller & Miller, 2006).

In addition to the ideal sample, there are a variety of methodological changes that could be completed to improve the research studies. Studies should create composites of symmetry which are based on a large number of bilateral traits (10–20) which span a variety of parts of the body. A minimum of 2–3 raters should be used to take measurements and the reliability of these measurements should be reported. Also, we encourage a better understanding of the intercorrelations of various FA measures (Miller, 2000b). Currently, the research implicitly assumes that different FA measurement composites are highly correlated. In other words, the current literature reports FA measurement composites that are derived from different bilateral traits. Future research could attempt to evaluate the correlation between FA measurement composites derived from some bilateral traits (e.g., the breadth and width of fingers) with FA measurement composites from other bilateral traits (e.g., the breadth and width of feet and ankles). Intelligence measures should be ones that have high *g*-loadings (e.g., WAIS scores) and should have well-established normative data and reliability.

Finally, we suggest that researchers of primary studies analyze the relationships between symmetry and intelligence as opposed to fluctuating asymmetry and intelligence. The term “symmetry” is more intuitive to the readers than is “fluctuating asymmetry” as it is a positive correlation between two good things such as symmetry and intelligence (the converse being a correlation between a positive and negative resulting in a negative correlation).

Acknowledgments

The authors appreciate the suggestions and additional data provided by many of the authors who conducted primary research in the area of fluctuating asymmetry and *g*. Reviewer suggestions by Jan te Nijenhus and Geoffrey Miller were valuable in the improvement of the paper and were greatly appreciated.

References[2]

Arden, R., Gottfredson, L. S., & Miller, G. F. (2009). Does a fitness factor contribute to the association between intelligence and health outcomes? Evidence from medical abnormality counts among 3654 US Veterans. *Intelligence*, *37*, 581−591.

Arden, R., Gottfredson, L. S., Miller, G. F., & Pierce, A. (2009). Intelligence and semen quality are positively correlated. *Intelligence*, *37*, 277−282.

Arnold, R. D. (2005). *Behavioral and psychological correlates of fluctuating asymmetry: A within-families study. Unpublished doctoral dissertation.* University of Texas at Austin, Austin, TX.

Arthur, W., & Day, D. V. (1994). Development of a short form for the Raven Advanced Progressive Matrices test. *Educational & Psychological Measurement*, *54*, 394−403.

*Bates, T. C. (2007). Fluctuating asymmetry and intelligence. *Intelligence*, *35*, 41−46. Supplemented via personal communication on December 21, 2009 from Tim Bates to Michael A. McDaniel.

Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2005). *Comprehensive Meta-analysis version 2.* Englewood, N.J.: Biostat.

Borgerhoff Mulder, M. (2009). Serial ponogamy as polygyny or polyandry? Marriage in the Tanzanian Pimbwe. *Human Nature*, *20*, 130−150.

Bradbury, J. W., & Vehrencamp, S. L. (1998). *Principles of animal communication.* Sunderland, MA: Sinauer Associates.

Deary, I. (2008). Why do intelligent people live longer? *Nature*, *456*, 175−176.

Duval, S. J. (2005). The trim and fill method. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), *Publication bias in meta analysis: Prevention, assessment and adjustments*. Chichester, UK: Wiley.

*Euler, M., Thoma, R. J., Gangestad, S. W., Canive, J. M., & Yeo, R. A. (2009). The Impact of developmental instability on voxel-based morphometry analyses of neuroanatomical abnormalities in Schizophrenia. *Schizophrenia Research*, *115*, 1−7. Supplemented via personal communication on January 30, 2010 from Matt Euler to George C. Banks.

*Furlow, F. B., Armijo-Prewitt, T., Gangestad, S. W., & Thornhill, R. (1997). Fluctuating asymmetry and psychometric intelligence. *Proceedings: Biological Sciences*, *264*, 823−829.

*Gangestad, S. W. (2010a). [Sample data and results regarding the relationship between fluctuating asymmetry and *g*]. Unpublished raw data. Supplemented via personal communication on January 2 and 22, 2010 from Steve Gangestad to George C. Banks and Michael McDaniel.

*Gangestad, S. W. (2010b). [Sample data and results regarding the relationship between fluctuating asymmetry and *g*]. Unpublished raw

[1] Our presentation of the ideal study is largely based on reviewer feedback provided by Geoffrey Miller.

[2] References marked with an asterisk indicate studies included in the meta-analysis.

data. Supplemented via personal communication on January 2 and 22, 2010 from Steve Gangestad to George C. Banks and Michael McDaniel.
Gangestad, S. W., & Thornhill, R. (1999). Individual differences in developmental precision and fluctuating asymmetry: A model and its implications. *Journal of Evolutionary Biology*, *12*, 402−416.
Gangestad, S. W., Thornhill, R., & Yeo, R. A. (1994). Facial attractiveness and fluctuating anatomic asymmetries. *Ethology and Sociobiology*, *15*, 73−85.
Gottfredson, L. S. (2004). Intelligence: Is it the epidemiologists' elusive "fundamental cause" of social class inequalities in health? *Journal of Personality and Social Psychology*, *86*, 174−199.
Gudjonsson, G. H. (1995). Raven's norms on the SPM revisited: A reply to Raven. *Personality and Individual Differences*, *18*, 447.
Hedges, L., & Olkin, I. (1985). *Statistical methods for meta-analysis.* San Diego: Academic Press.
Higgins, J. P., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal*, *327*, 557−560.
Hooper, P. L., & Miller, G. F. (2008). Mutual mate choice can drive costly significance even under perfect monogamy. *Adaptive Behavior*, *16*, 53−70.
Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis: Correcting error and bias in research findings.* Newbury Park, CA: Sage.
Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings.* Newbury Park, CA: Sage.
Hunter, J. E., Schmidt, F. L., & Le, H. (2006). Implications of direct and indirect range restriction for meta-analysis methods and findings. *Journal of Applied Psychology*, *91*, 594−612.
*Johnson, W., Segal, N. L., & Bouchard, T. J., Jr (2008). Fluctuating asymmetry and general intelligence: No genetic or phenotypic association. *Intelligence*, *36*, 279−288. Supplemented via personal communication on December 22, 2009 from Wendy Johnson to Michael A. McDaniel.
Keller, M. C., & Miller, G. (2006). Resolving the paradox of common, harmful, heritable mental disorders: Which evolutionary genetic models work best? *Behavioral and Brain Sciences*, *29*, 385−404.
*Luxen, M. F., & Buunk, B. P. (2006). Human intelligence, fluctuating asymmetry and the peacock's tail: General intelligence (g) as an honest signal of fitness. *Personality and Individual Differences*, *41*, 897−902.
Martin, S. M., Manning, J. T., & Dowrick, C. F. (1999). Fluctuating asymmetry, relative digit length, and depression in men. *Evolution and Human Behavior*, *20*, 203−214.
McDaniel, M. A. (2005). Big-brained people are smarter: A meta-analysis of the relationship between in vivo brain volume and intelligence. *Intelligence*, *33*, 337−346.
McDaniel, M. A., & Banks, G. C. (2010). Cognitive ability. In D. H. Reynolds & J. C. Scott (Eds.), *The Handbook of Workplace Assessment: Selecting and Developing Organizational Talent.* CA: Pfeiffer.
Miller, G. F. (2000). *The mating mind: How sexual choice shaped the evolution of human nature.* New York, NY: Doubleday.
Miller, G. F. (2000). Sexual selection for indicators of intelligence. *Novartis Foundation Symposium*, *233*, 260.
Miller, G. F. (2007). Sexual selection for moral virtues. *The Quarterly Review of Biology*, *82*, 97−125.
Møller, A. P., & Swaddle, J. P. (1997). *Asymmetry, developmental stability, and evolution.* Oxford: Oxford University Press.
Møller, A. P., & Thornhill, R. (1997). A meta-analysis of heritability of developmental stability. *Journal of Evolutionary Biology*, *10*, 1−6.
Nunnaly, J. C., & Bernstein, I. H. (1994). *Psychometric Theory.* New York: McGraw-Hill.
Palmer, A. R., & Strobeck, C. (1986). Fluctuating asymmetry: Measurement, analysis, patterns. *Annual Review of Ecology & Systematics*, *17*, 391−421.
Peck, D. F. (1970). The conversion of progressive matrices and Mill Hill vocabulary raw scores into deviation IQ's. *Journal of Clinical Psychology*, *26*, 67−70.
*Penke, L., & Asendorpf, J. B. (2008). Beyond global sociosexual orientations: A more differentiated look at sociosexuality and its effects on courtship and romantic relationships. *Journal of Personality and Social Psychology*, *95*, 1113−1135. Supplemented via personal communication on January 12, 2010 from Lars Penke to Michael A. McDaniel.
*Penke, L., Bates, T. C., Gowa, A. J., Pattiea, A., Starra, J. M., Jones, B. C., et al. (2009). Symmetric faces are a sign of successful cognitive aging. *Evolution and Human Behavior*, *30*, 429−437. Supplemented via personal communication on January 22, 2010 from Lars Penke to George C. Banks.
*Prokosch, M. D., Yeo, R. A., & Miller, G. F. (2005). Intelligence tests with higher *g*-loadings show higher correlations with body symmetry: Evidence for a general fitness factor mediated by development stability. *Intelligence*, *33*, 203−213. Supplemented via personal communication on December 20, 2009 from Geoffrey Miller to Michael A. McDaniel.
*Rahman, Q., Wilson, G. D., & Abrahams, S. (2004). Developmental instability is associated with neurocognitive performance in heterosexual and homosexual men, but not in women. *Behavioral Neuroscience*, *118*, 243−247.
Raven, J., Raven, J. C., & Court, J. H. (1998). *Manual for Raven's Progressive Matrices.* Oxford: Oxford Psychologists Press Ltd.
Raven, J. C., Court, J. H., & Raven, J. (1994). *Advanced progressive matrices: Sets I and II. Manual for Raven's progressive matrices and vocabulary scales.* Oxford, England: Oxford Psychologists Press.
Rothstein, H. R., Sutton, A. J., & Borenstein, M. (2005). Publication bias as a threat to validity. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), *Publication bias in meta analysis: Prevention, assessment and adjustments Chichester*. UK: Wiley.
Spearman, C. (1904). "General Intelligence", objectively determined and measured. *American Journal of Psychology*, *15*, 201−293.
*Thoma, R. J., Yeo, R. A., Gangestad, S. W., Halgren, E., Sanchez, N. M., & Lewine, J. D. (2005). Cortical volume and developmental instability are independent predictors of general intellectual ability. *Intelligence*, *33*, 27−38. Supplemented via personal communication on December 21, 2009 from Robert Thoma to John H. Batchelor.
Thornhill, R., & Gangestad, S. W. (1994). Human fluctuating asymmetry and sexual behavior. *Psychological Science*, *5*, 297−302.
Thornhill, R., & Møller, A. P. (1997). Developmental instability, disease, and medicine. *Biological Review*, *72*, 497−548.
Van Valen, L. (1962). A study of fluctuating asymmetry. *Evolution*, *16*, 125−142.
Yeo, R. A., Gangestad, S. W., Thoma, R. J., & Repa, K. (1997). Developmental instability and cerebral lateralization. *Neuropsychology*, *11*, 552−561.
Zebrowitz, L. A., Hall, J. A., Murphy, N. A., & Rhodes, G. (2002). Looking smart and looking good: Facial cues to intelligence and their origins. *Personality and Social Psychology Bulletin*, *28*, 238−249.