PROCEEDINGS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

A latent variable model of developmental instability in relation to men's sexual behaviour

Steven W. Gangestad, Kevin L. Bennett and Randy Thornhill

Proc. R. Soc. Lond. B 2001 **268**, 1677-1684
doi: 10.1098/rspb.2001.1675

THE ROYAL SOCIETY

doi 10.1098/rspb.2001.1675

A latent variable model of developmental instability in relation to men's sexual behaviour

Steven W. Gangestad[1]*, Kevin L. Bennett[1] and Randy Thornhill[2]

[1]*Department of Psychology, and* [2]*Department of Biology, University of New Mexico, Albuquerque, NM 87131, USA*

A single trait's fluctuating asymmetry (FA) is expected to be a poor measure of developmental instability. Hence, studies that examine associations between FA and outcomes expected to covary with developmental instability often have little power in detecting meaningful relationships. One way of increasing the power of detecting relationships between developmental instability and outcomes is through the use of multiple traits' FA. The way multiple traits have typically been used is in trait aggregates. Here, we illustrate another way of examining relationships with developmental instability using multiple traits' FA: through structural equation modelling. Covariances between measures of FA and an outcome variable are interpreted within the context of an explicit model of associations between variables, which is tested for fit and the parameters specified within the model are estimated. We used nine traits' FA as markers of a latent variable of men's developmental instability, which was associated with the number of sexual partners. The results indicate a sizeable correlation between developmental instability and men's sexual history, despite small correlations between individual traits' FA and sexual history.

Keywords: developmental instability; fluctuating asymmetry; mate choice; structural equation modelling; sexual selection

1. INTRODUCTION

One of the most heavily researched topics in sexual selection in recent years is fluctuating asymmetry (FA). FA, the argument goes, reflects developmental instability, i.e. an organism's inability to cope with developmental stresses. In turn, developmental instability can be caused by a paucity of resistance to or resources to deal with stresses (e.g. pathogen resistance, immunocompetence and heterozygosity), the amount of stresses themselves (e.g. mutations, pathogens, toxins and nutritional stress) and growth parameters that affect the expression of noise (Klingenberg & Nijhout 1999). (Some authors (e.g. Leung & Forbes 1997) separate these sources and refer to the first as developmental instability. In practice, they are rarely separable. Our usage follows Houle (2000).) Because coping resources (e.g. resistance against pathogens or toxins) and stresses (e.g. mutations) have heritable components, developmental instability may possess additive genetic variation maintained by mutation–selection balance and coevolutionary processes (e.g. host–parasite coevolution) (see also Klingenberg & Nijhout 1999). Owing either to genetic benefits passed onto offspring or direct material advantages (e.g. superior direct parental care, better protection and lack of disease), individuals may evolve to prefer mates who evidence a lack of developmental instability (either low asymmetry itself or correlated traits). As a result, individuals who have low FA may experience greater mating success.

In the past decade, many studies involving a wide variety of species have examined relationships between FA and mating success or attractiveness. Møller & Thornhill (1998) performed a meta-analysis on these studies yielding several key findings: (i) overall, symmetrical individuals enjoy a mating advantage, (ii) this overall advantage is qualified by the fact that studies have yielded heterogeneous effects, and (iii) the advantages have been particularly large in experiments as opposed to observational studies; the advantages are also larger in males than females and when FA is on sexually selected traits (though this factor is confounded by the type of study). Effects were not observed in all species and, naturally, overall positive findings do not imply effects in all species (see also Thornhill *et al*. 1999).

In general, these associations are small. The studies that most directly address their ecological importance examine mating success in the field. In Møller & Thornhill's (1998) meta-analysis, the weighted mean size of the correlation between a single non-sexually selected trait's FA and mating success in such studies was a mere −0.13. Palmer (1999) questioned the biological significance of such small effects.

Clearly, a single trait's FA is not typically a powerful predictor of mating success. From a theoretical standpoint, however, the important relationships are not those between FA and mating success, but those between developmental instability and mating success. Van Dongen (1998), Whitlock (1998) and Gangestad & Thornhill (1999) have noted that a single trait's FA may be only weakly associated with underlying developmental instability. Based on the kurtosis of the distributions of signed FA observed in large samples in a number of species, Gangestad & Thornhill (1999) estimated that a single trait's FA typically owes only *ca*. 7% of its variance to underlying individual differences in developmental instability (but see also Van Dongen & Lens 2000). This weak association is not due to a lack of individual differences; indeed, the coefficient of variation (CV) of individual differences in developmental instability may be over 20, which is much greater than typical ordinary morphological traits. (Individual differences with a CV typical of ordinary morphological traits (Houle 1992), i.e. 5, would account for less than one-half of 1% of the variance in FA.) Rather, it is due to the fact that the random effects of

*Author for correspondence (sgangest@unm.edu).

developmental error generate amounts of variation in asymmetry much larger than that created by substantial, systematic individual differences in propensity to experience error. (Some studies show larger amounts of variation in asymmetry due to underlying individual differences (e.g. Blanckenhorn *et al*. 1998; Lens & Van Dongen 1999; Van Dongen *et al*. 1999). As Houle (1997, 2000) has argued, some of these variations appear to require too much variance in underlying developmental instability to qualify as FA, i.e. to be due to developmental error; for instance, breaking or damage to a trait may create very large asymmetries not due to FA. However, due to bias in the estimation of repeatabilities of measures of FA on a rough measurement scale some of these estimates are seriously inflated (Gangestad & Thornhill 2001). Nonetheless, in some studies traits have owed 25% or more of their variance to systematic individual differences, reflecting an underlying CV of developmental instability of 50–60 (Gangestad & Thornhill 2001).)

If, in fact, a trait's FA owes a mere 7% of its variance to underlying individual differences, it can correlate with other variables at most $\sqrt{0.07}$ or 0.27. A correlation between a single trait's FA and mating success of just −0.13 may thus reflect a correlation between developmental instability and mating success of −0.13/0.27, which is nearly −0.5. An association of this magnitude would seem to be biologically very significant (Thornhill *et al*. 1999). (This argument assumes that FA itself is not used as a cue for affecting mating success directly.)

Another way of approaching this issue is to consider the covariation between multiple traits' FA. Although traits' FA covaries weakly, the size of the correlations appears to be similar to what one could expect if each trait's FA owes only 7% of its variance to developmental instability and much developmental instability is trait general (Gangestad & Thornhill 1999; Van Dongen & Lens 2000; but see also Van Dongen *et al*. 1999). If multiple asymmetries all tap a common source of developmental instability, a composite should be more highly associated with underlying developmental instability than any single trait's FA and, therefore, should covary with mating success more highly as well (e.g. Leung *et al*. 2000).

A series of studies point to a negative association between men's developmental instability and number of sexual partners (age controlled), partly due to female preference. Specifically, (i) men's FA predicts their number of sexual partners (Thornhill & Gangestad 1994; Gangestad & Thornhill 1997*b*), (ii) men's FA predicts their number of extra-pair relationships and the number of times they have been chosen as an extra-pair partner by women (Gangestad & Thornhill 1997*b*), and (iii) women near ovulation prefer the scent of symmetrical men (Gangestad & Thornhill 1998; Rikowski & Grammer 1999; Thornhill & Gangestad 1999*b*). Although most of these studies have been done on US college students using self-report measures of sexual behaviour, more recent research also found an association between men's FA and partner quantity in men from a rural Caribbean village using ratings by other villagers (S. W. Gangestad, R. Thornhill, R. J. Quinlan & M. V. Finn, unpublished data). (Waynforth (1998) found an association in the predicted direction but short of statistical

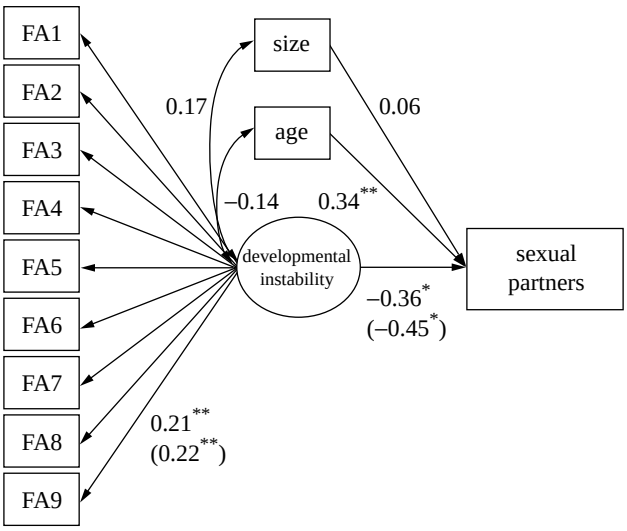

Figure 1. Structural model of relationships between individual asymmetries, developmental instability and the number of sex partners. FA1–FA9, individual asymmetries. Parameter estimates from structural equations modelling are also included. Values not in parentheses are from analysis on 244 cases with missing data imputed. Values in parentheses are key parameter estimates from analysis on 144 cases with no missing data. The statistical significance of paths (where noted) is backed by bootstrapped standard errors. $^{*}p < 0.05$; $^{**}p < 0.01$.

significance in a small sample of Mayan men.) As men with low FA appear to be less attentive partners (being rated by themselves and their partners as more sexualizing of other women and less honest) (Gangestad & Thornhill 1997*a*), one possibility is that female preferences for them have evolved because of indirect genetic benefits. The fact that women prefer the scent of symmetrical men specifically when they are fertile is consistent with this possibility; more evidence is needed in order to test it further. (See Thornhill *et al*. (1995), Furlow *et al*. (1997, 1998), Manning & Wood (1998), Rhodes *et al*. (1998), Mealey *et al*. (1999) and Perrett *et al*. (1999) for additional research on men's FA, attractiveness, sexual relationships and other social traits.)

This research has typically involved a multiple-trait composite measure of FA. In recent studies, up to ten traits have been measured: width of the ears, elbows, wrists, ankles and feet and length of the ears and four digits of the hand. As with other typical applications, we have aggregated the multiple traits' FA into a single FA index (see Leung *et al*. 2000). Another way of using multiple traits' FA for examining the association between developmental instability and theoretically meaningful outcomes is to apply structural equation modelling.

Structural equation modelling is a statistical technique that is used for testing models of associations between observed and latent traits. Figure 1 illustrates one common way of thinking about the associations between developmental instability, individual traits' FA and the number of sexual partners (e.g. Gangestad & Thornhill 1999). Developmental instability is a latent (i.e. not directly observed) trait reflecting perturbations in development due to stresses. It leads to observable variation in

asymmetry on individual traits (effects that are represented by arrows leading from developmental instability to asymmetries). It also affects other traits (not explicitly represented in the diagram) that, in turn, affect the number of sexual partners (through female preference or intrasexual competition, which is represented by an arrow leading from developmental instability to the number of sexual partners). As developmental instability may be associated with age or overall character size, which may also influence the number of sexual partners, these variables are also represented in the model.

Latent traits are implicitly defined in structural equation modelling by associations between observed measures to which they are linked. Hence, developmental instability is largely defined by associations between individual measures of FA. If these measures are completely uncorrelated, there is no basis for inferring a latent trait of (trait-general) developmental instability. Based on covariations between observed traits, modelling procedures lead to (i) parameter estimates for the paths between traits under a specific model, (ii) overall measures of fit between covariances and the model, which indicate whether the model is plausible, and (iii) tests of the statistical significance of particular paths (i.e. effects) within the model. These procedures thus allow one to assess the strengths of relationships and test model assumptions that are not permitted by use of a simple trait aggregate.

Using these procedures, we asked the following questions. Does the model represented in figure 1 fit the covariances between measures of individual asymmetries and the number of sexual partners? If so, how much variation in specific asymmetries is associated with these individual differences? What is the inferred effect of developmental instability on the number of sexual partners? Although we explored empirical issues specifically pertaining to human developmental instability and its relation to men's sexual history, our study also illustrates the application of structural equation modelling to FA research.

2. METHODS

The participants were 271 men from the University of New Mexico, Albuquerque, USA. Most had participated in studies reported elsewhere, for example Furlow *et al*. (1997) (52 men) (this same sample was also reported in Furlow *et al*. (1998)), Gangestad & Thornhill (1998) (42 men) and Thornhill & Gangestad (1999*b*) (80 men); the former studies examined FA, IQ and fighting and the latter two concerned FA and scent. Some (97) had participated in a third study examining scent (R. Thornhill, S. W. Gangestad, R. Miller, G. Scheyd, J. Knight & M. V. Flinn, unpublished data). The results concerning the number of sexual partners reported here have not been reported elsewhere. All participants received credit towards a psychology class research requirement or extra credit in a psychology or biology class in return for participation and signed an informed consent form prior to participating. As part of all studies, men were asked to fill out a questionnaire asking them about their age, sexual orientation (heterosexual, bisexual or homosexual) and number of sexual partners in their lifetime. A sexual partner was defined as someone with whom they had penile–vaginal intercourse. The participants were asked to exclude all partners paid for sex and all coerced sex. Only heterosexual men who reported their number of sexual partners were included in the analyses. In order to reduce the heterogeneity of age, we excluded six men over the age of 36 years (more than three standard deviations from the mean; their exclusion does not affect the conclusions we report). Our analyses included 244 men whose average age was 21.2 years (s.d. = 3.28 years and range = 17–33 years).

Measurements were taken for ten bilateral traits on all men: width of the ears, elbows, wrists, ankles and feet and length of the ears and the four fingers (excluding the thumb). Measurements were taken by undergraduate research assistants who were trained for taking reliable measurements of FA. Although the research assistants were not aware that the data would be used for the present analyses (they knew that they would be used to examine IQ or scent), they were majoring in biology and psychology and could not be expected to know nothing about FA. Thus, the measurement procedures were designed in order to maximize and check accuracy and thereby minimize bias. The measurers were trained in measurement and practised measuring each other prior to the study. Only measurers who demonstrated high repeatability during practice sessions were used. During the study, the measurers did not write down their own measurements; instead they called them out to a recorder. Hence, their task was simply to measure as accurately as possible and this fact was emphasized to them. Each side of each trait was measured twice. Right and left sides of traits were measured in clusters (e.g. the four digits of one hand were measured consecutively); it is very difficult to remember the measurements of multiple traits on one side while measuring the other side even if one is trying and the measurers were instructed not to try. After the measurers completed one set of measurements on the traits, they obtained second measurements. Because many measurements intervened the repeated measures of the same trait, even if the measurers tried to recall their previous measurements (and, again, they were instructed not to do so), they would have had extreme difficulty in doing so. Hence, the two sets of measurements were independent replicates. Repeatabilities of asymmetry measurements have been reported in previous studies. All traits showed significant repeatability in all studies, with mean intraclass correlations for signed asymmetries of 0.79, 0.83 and 0.85 for Furlow *et al*. (1997), Gangestad & Thornhill (1998) and Thornhill & Gangestad (1999*b*), respectively (all $p < 0.00001$). For unsigned asymmetries, they averaged 0.60, 0.67 and 0.73, respectively (all $p < 0.0005$). For participants in the remaining sample, the repeatabilities were once again highly significant (signed asymmetries, mean $r_{IC} = 0.86$ and range = 0.79 (elbow width) to 0.93 (foot width) and unsigned asymmetries, mean $r_{IC} = 0.73$ and range = 0.63 (elbow width) to 0.87 (foot width)) (all $F_{96,97} \geqslant 8.35$ and $p < 0.00001$ and all $F_{96,97} \geqslant 4.37$ and $p < 0.00001$, respectively). Because signed asymmetries reflect both direction and size whereas unsigned asymmetries reflect only absolute size, the repeatabilities for signed asymmetries should and do exceed those for unsigned asymmetries. The fact that the independent measurements of the signed FA for the same trait correlated on average *ca*. 0.84 suggests high accuracy of measurement.

Each participant was asked whether he had ever broken, sprained or otherwise damaged any character. We excluded measurements on injured traits for all analyses. The percentages of men who had injured a trait were 2.0% for ear length, 8% for ear width, 4.5% for elbows, 15.6% for wrists, 8.6% for index fingers, 13.1% for middle fingers, 12.3% for ring fingers, 14.3%

for small fingers, 37.7% for ankles and 4.9% for feet. With nearly two out of five men missing data on ankle asymmetry (due to injury), we did not include FA of that trait in our analyses. (However, results from analyses that include ankle asymmetry are almost identical to those reported.)

We tested each trait for directional asymmetry. Only one exhibited significant directional asymmetry, namely foot width ($t_{231} = 4.40$ and $p < 0.001$) (right > left). These results replicate analyses on a largely independent sample of over 700 individuals (Furlow *et al*. 1997). The directional foot asymmetry was fairly small, with the mean signed asymmetry displaced 0.29 standard deviations from zero. We performed analyses in two ways, calculating FA in foot width correcting and not correcting for directional asymmetry (these two measures correlated 0.91). The statistical conclusions from the two analyses were identical. We report those not correcting for directional asymmetry, with reference to analyses correcting for it.

As is expected of asymmetry due to developmental error with individual differences in propensity to experience error (Houle 1997; Gangestad & Thornhill 1999), signed asymmetries tended to be weakly leptokurtic (mean $g_2 = 0.58$ and range $= -0.12$ to 1.37).

In order to control for overall skeletal size, we created an overall size variable by summing eight characters excluding the ear measures. This measure weights finger length heavily but, once again, excluding finger lengths did not change the results.

Structural equation modelling was performed on AMOS 4.0 (Arbuckle 1995–1999). Amos imputes full maximum likelihood estimates for missing values. Test statistics based on maximum likelihood procedures assume that variables are normally distributed. Naturally, neither unsigned asymmetry nor the number of sexual partners satisfy this assumption. We addressed this potential problem through bootstrapping, which is a means of estimating parameters and standard errors that are not formally derived but rather are inferred from resampling from observed distributions in the sample itself (e.g. Efron 1982; Dianonis & Efron 1983). Bootstrapped estimates are not based on normal distribution theory and, hence, do not require commitment to any distributional assumptions.

3. RESULTS

Before performing structural equation modelling, we estimated the proportion of total variance in individual traits' FA that was attributable to individual differences in developmental instability (i.e. trait repeatability). Three estimation procedures have been proposed. Whitlock (1998) and Van Dongen (1998) formally derived procedures based on asymmetry variances and CVs. Gangestad & Thornhill (1999) used simulations for estimating the association between an observable statistic, the kurtosis of unsigned asymmetry and trait repeatability. As the sampling distributions of the statistics used in these procedures are variable, no procedure may be highly accurate in estimating the repeatability of individual traits except in very large samples (e.g. Gangestad & Thornhill 1999); their mean across traits should estimate repeatability more accurately.

The mean repeatabilities estimated by the procedures were similar: that of Whitlock (1998) was 0.070 (range $= -0.004$ to 0.145), that of Van Dongen (1998) was 0.075 (range $= -0.005$ to 0.146) and that of Gangestad & Thornhill (1999) was 0.088 (range $= -0.017$ to -0.190). The overall mean, which was 0.078, is also very similar to those estimated by Gangestad & Thornhill (1999) for a number of non-human species on which large samples of observations (> 1000) were available: 0.066–0.083. Additional data on primates and other mammals yield similar values (Gangestad & Thornhill 2001). (Gangestad & Thornhill (1999) estimated a somewhat lower mean for humans (0.045) but those data included several samples with higher measurement error.) Differences across individual traits may largely reflect sampling variability. If, as these results indicate, 7–8% of the variance in the individual traits' FA is due to individual differences, these differences have a CV of *ca*. 23 (26 if measurement error is subtracted out) (Gangestad & Thornhill 1999).

We next turned to structural equation modelling. The full correlation matrix (as well as means and standard deviations) is available in electronic Appendix A, available on The Royal Society's Web site. Individual asymmetries correlated weakly with one another (mean $r = 0.042$). The standard deviation of these correlations is about what would be expected by chance if all correlations were equal in the population (0.076 observed versus 0.071 based on sampling variability of the correlation coefficient; see also the text on structural equations modelling below). Individual asymmetries also correlated weakly with the number of sexual partners, with correlations partialling out age and size ranging from -0.019 to -0.143 and averaging -0.075. Only one correlation reached statistical significance, that involving foot width (-0.143) ($p < 0.05$). (The correlation with foot width FA correcting for directional asymmetry was very similar ($r = -0.138$ and $p < 0.05$).) A test of a sum of the correlations (using Olkin & Finn's (1990) method for testing contrasts on dependent correlations) revealed that the correlations are, on average, significantly negative ($z = -3.26$ and $p < 0.002$). Moreover, the sum of the individual asymmetries (which is expressed as proportions of total character size, i.e. relative FA) (Palmer 1994) covaried significantly with the number of sexual partners (with age partialled out), ($r = -0.267$ and $p < 0.001$), thereby replicating Thornhill & Gangestad (1994) and Gangestad & Thornhill (1997*b*) (partialling size out as well produced identical results).

Structural equation modelling is an appropriate means for testing whether the associations of weak, individual markers are due to a common trait and, if so, the strength of the effect of the latent trait. The model illustrated in figure 1 was fitted to covariances between unsigned asymmetry on the nine individual traits (all standardized through conversion to *z*-scores in order to ensure equal variance), the number of sexual partners, age and size (the parameter estimates in the figure can be ignored at this point). In order to simplify the model, associations between individual traits' FA and developmental instability were constrained to be equal. Hence, this model purports to account for correlations between all FA indicators with a single parameter value, the correlation between an individual indicator and the latent trait of developmental instability. Because age and size did not correlate significantly ($r = -0.083$ and n.s.), we did not include a path between them in the model. We refer to this model as the standard model throughout this paper.

This model fitted the data very well. Hu & Bentler (1999) recommend a two-index evaluation, which includes the standardized root mean squared residual (SRMR) combined with any one of a number of indices, including the Tucker–Lewis index (TLI) and root mean squared error of approximation (RMSEA). Recommended criteria for rejection are SRMR > 0.09, TLI < 0.95 and RMSEA > 0.06. The SRMR could not be calculated with missing values imputed (but see below). The TLI was 1.00 and the RMSEA was < 0.001, both indicating excellent fit of the model to the data. In addition, the χ^2 goodness-of-fit test was non-significant ($\chi^2_{(60)} = 58.64$ and $p = 0.53$.

We compared the standard model with a model that did not constrain the effects of developmental instability on the nine individual asymmetries to be equal. A likelihood ratio test addresses whether removing constraints significantly improves the fit. The test was non-significant ($\chi^2_{(8)} = 2.96$ and $p = 0.94$). These results show that allowing the nine individual asymmetries to tap developmental instability differentially does not significantly improve the fit of the model.

The parameter estimates of the model are given in figure 1. Individual asymmetries were estimated to correlate 0.21 with developmental instability. This value corresponds to *ca.* 4.4% shared variance. As we estimated that individual differences account for *ca.* 7–8% of the variance in individual asymmetries, these data suggest that slightly more than half of the systematic variance in individual asymmetries is shared and due to trait-general developmental instability.

The standardized path coefficient of the effect of developmental instability and the number of sexual partners was -0.36. The likelihood ratio test against a model setting this effect to zero was significant ($\chi^2_{(1)} = 9.49$ and $p < 0.005$). This effect generates a correlation between developmental instability and the number of sexual partners (with age and size partialled out) of -0.38. Naturally, this model assumes that all of the associations between individual asymmetries and the number of sexual partners are due to shared developmental instability. In order to test whether the individual asymmetries predicted the number of sexual partners above and beyond the effect of shared developmental instability, we compared the standard model with nine other models, each of which added a direct path between one of the individual asymmetries and number of sexual partners. None yielded a significant improvement of fit ($\chi^2_{(1)} \leqslant 1.63$ and n.s.). These results are consistent with the view that these asymmetries are associated with the number of sexual partners only through a trait-general developmental instability.

Under the standard model, all asymmetries correlate with the number of sexual partners equally: *ca.* -0.08. Naturally, this value is almost precisely the mean correlation between the individual asymmetries and number of sexual partners in the sample. These values are small (smaller than the mean of -0.13 in Møller & Thornhill's (1998) meta-analysis) and, again, only one of them is statistically significant. While one might be led to think that these data suggest that foot width asymmetry in particular predicts the number of sexual partners, the data we just presented provide no evidence for this claim. As noted, we found no significant associations between individual asymmetries and the number of sexual partners independent of a shared source of variance; variations in size of the correlations can be attributed to sampling variability. The fact that we can detect a sizeable and significant association between a common latent trait of developmental instability and the number of sexual partners despite weak and generally non-significant associations involving individual asymmetries attests to the power of having multiple indicators.

We also fitted the data to a very different model, one in which (i) there exists no common source of developmental error (i.e. no latent trait-general developmental instability), but instead all asymmetries are due to trait-specific error and (ii) individual asymmetries individually predict the number of sexual partners. This model fitted reasonably well too (probably because the largest correlations between variables involve the number of sexual partners, which this model captures), but not as well as the standard model (TLI $= 0.991$ and RMSEA $= 0.035$). The χ^2 likelihood ratio test yielded a marginally significant lack of fit ($\chi^2_{(55)} = 71.23$ and $p = 0.069$). Models can be compared using the Akaike information criterion (AIC) and the Browne–Cudeck criterion (BCC) (which penalizes complex models slightly more harshly). Better fit is revealed by smaller values. The standard models had AIC and BCC values lower than those of the alternative model (AIC, 118.26 versus 141.23 and BCC, 122.03 versus 145.23). The standard model was superior on all fit indices.

In order to estimate standard errors using bootstrapping, AMOS does not allow missing values on any variables and, hence, these procedures were performed on the subsample ($n = 144$) that had complete information (no broken or sprained characters). The maximum likelihood results paralleled those from the full sample. Here, the SRMR could be calculated. The model fit was excellent (SRMR $= 0.062$, TLI $= 1.01$ and RMSEA < 0.001) ($\chi^2_{(60)} = 46.26$ and $p = 0.90$). The effect of developmental instability on the number of sexual partners was slightly larger ($\beta = -0.45$ and $p < 0.02$), corresponding to a correlation (with age and size partialled) of -0.46.

We bootstrapped standard errors on the basis of 5000 samples. The standard error for the standardized effect of developmental instability was 0.190 (with an s.e. of 0.002). The maximum likelihood estimate over this standard error is 2.34 ($p < 0.02$), indicating that the effect is statistically reliable. The mean standardized effect in the bootstrapped samples was actually slightly larger than the maximum likelihood estimate ($\beta = -0.48$) (corresponding to a partial correlation of *ca.* -0.50).

The mean standardized effect of developmental instability on an individual asymmetry was 0.22. The mean bootstrapped standard error of this estimate is 0.045. All ratios of estimated effects over their standard errors were 4.58 ($p < 0.00001$).

4. DISCUSSION

Two apparently contrasting findings emerged from the current study. First, individual traits' FA correlated weakly with men's number of sexual partners. Across nine traits, the mean correlation was a mere -0.08. Only a

single correlation was statistically significant (and would not be after Bonferroni adjustment for multiple comparisons). Second, structural equation modelling revealed a statistically significant and substantial correlation between trait-general developmental instability underlying these nine specific trait asymmetries and developmental instability and men's number of sexual partners. Controlling for age and size, this correlation was −0.38 to −0.50. In other words, in this population developmental instability may account for 14–25% of the variance in men's self-reported partner number, independent of age and size.

Though these two findings may appear to contradict each other, they are highly compatible if, in fact, an individual trait's asymmetry is only weakly associated with underlying developmental instability. Evidence and theory indicate that, indeed, this situation is both expected and observed. Procedures that estimate the proportion of variance in an individual trait's asymmetry due to individual differences yielded a mean figure of 7–8%. As one critical observer of FA research has noted (Houle 1997, 2000), in theory one should not expect FA to have a substantial amount of variance due to individual differences; instead individual traits' FA should be weak markers of underlying developmental instability.

An appropriate analogy can be drawn with psychometric intelligence. Psychologists agree that psychometric intelligence predicts meaningful outcomes. However, a study that examined the correlates of psychometric intelligence with these outcomes using single test items might yield nothing but null findings. The same study that aggregated multiple items (or modelled their associations using a latent trait with structural equations) might reveal very strong associations (see Leung *et al.* (2000) for a discussion of aggregation of multiple traits' FA and power).

The current findings have a number of implications. First, they have implications for understanding the relationship between developmental instability and sexual behaviour in men. Previous studies have reported that composite measures of FA predict men's number of sexual partners (Thornhill & Gangestad 1994; Gangestad & Thornhill 1997*b*). The weighted mean of these correlations is just over −0.2. Palmer (1999) has questioned the biological significance of correlations this small. Based on estimates of the potential validity of FA for measuring developmental instability, Gangestad & Thornhill (1999) estimated that the weighted mean correlation between underlying trait-general developmental instability and men's number of sexual partners in these studies is −0.53. Using a different statistical methodology on an independent sample, the current study yielded a similar (though slightly lower) estimate. These studies were of college men and used self-report measures; the results in one very different population using reputational ratings by peers were similar (S. W. Gangestad, R. Thornhill, R. J. Quinlan & M. V. Flinn, unpublished data), though work on additional populations is needed. Because female preferences may vary across human populations and ecologies, so too may the correlates of developmental instability (Gangestad & Simpson 2000).

However, perhaps even more importantly, the current study may illustrate the difficulties of interpreting the results of studies of FA in non-human species. Specifically, this study demonstrates how underpowered much if not most research on FA probably is. In order to detect the mean correlation between single traits' FA and the number of sexual partners in the current study (a weak −0.08 despite powerful effects of developmental instability) with 80% power (using a two-tailed test), we would need an n of over 1200. Yet most research continues to examine relationships with FA on single traits (even if multiple traits are measured). A sample size of 100 has just 16% power for detecting a correlation of −0.1. When power is so horribly low, very little can be said of null effects. Most heritability studies of FA are similarly underpowered (Fuller & Houle 2001).

Fortunately, researchers are increasingly using aggregate measures of FA. For example, Badyaev *et al.* (2000) summed the FA of eight mandibular traits in order to predict individual condition. Whereas individual traits probably showed little or inconsistent relations (the results were not presented), the composite did predict condition (see also Blanckenhorn *et al.* 1998; Dufour & Weatherhead 1998; Bjorksten *et al.* 2000). Some work using aggregate measures has found few significant relationships of FA with mating success or other fitness traits. These null results may indicate that, in these species, developmental instability bears little if any association with fitness. They may also be due to low power because of a small n. In order to ensure close to 80% power in detecting correlations of ±0.4 with trait-general developmental instability (and 5% variance in single-trait FA due to trait-general developmental instability, as found here), researchers need a sample size of *ca.* 125 with a ten-trait composite of FA and *ca.* 200 if a five-trait composite is used; in order to detect correlations of ±0.3 with 80% power, the necessary sample sizes are approximately double these figures.

The current study also demonstrates the effectiveness of FA as a cue for developmental instability. As noted above, weak effects of single traits' FA may reflect powerful effects of developmental instability, as revealed in the current study through structural equation modelling. However, two different claims about the effects of developmental instability on mating success must be distinguished. On the one hand, the effect of developmental instability may be mediated through preference for symmetry itself. On the other hand, developmental instability may be mediated through a preference for honest indicators other than symmetry. Asymmetry itself is a poor cue for developmental instability (Gangestad & Thornhill 1999; Houle 2000). Indeed, given that a single trait's FA can be expected to have at most a few per cent heritable variance, it can hardly function as a useful indicator of 'good genes' (Woods *et al.* 1999; Bjorksten *et al.* 2000; Chapman & Goulson 2000). The current study underscores this fact. The nine asymmetries and number of sexual partners in figure 1 are all effects of developmental instability. The largest effect by far is on the number of sexual partners, not individual asymmetries. Because the effect on the number of sexual partners is mediated by other traits (see Gangestad & Thornhill 1997*a*), it stands to reason that one or more of these mediators (e.g. informal peer status) (Gangestad & Thornhill 1997*a*) is actually a better marker of

developmental instability (and, hence, a better cue for developmental instability) than is any individual trait's FA. Ironically, of course, researchers can only identify good cues by measuring known (even if weaker) indicators of developmental instability, i.e. single traits' FA and, hence, cannot discard FA in studies of developmental instability. (Naturally, the current study provides no evidence that the asymmetries we measure actually function as cues. The single best-measured predictor of the number of sexual partners, i.e. foot width FA, is probably the least visible during social interaction (though, in fact, we suspect that none of the traits' FA is readily perceptible during ordinary social interaction).)

A number of experimental studies have shown that FA itself does indeed affect mate choice in some species (e.g. see Møller & Thornhill 1998; Thornhill & Gangestad 1999*a* for reviews). Several different scenarios may account for the preferences. First, individuals may pay attention to multiple asymmetries (as might be evident in an elaborate bird tail or a human face) and, hence, FA that is aggregate in nature, which is perhaps a reasonably valid marker of developmental instability. Second, individuals may combine asymmetry information with other more valid markers of developmental health (e.g. trait size). If individuals weight single trait asymmetry heavily, we suspect that it is not because the preferences have largely evolved in order to detect developmental health or good genes; again, traits that have just a few per cent of variance due to genetic differences cannot effectively function as indicators of good genes. Enquist & Arak (1994) and Johnstone (1994) have suggested that preference for symmetry may be due to exploitation of sensory biases. Ironically, this model may account for some preferences for symmetry *per se* even when developmental instability itself powerfully affects mate preference, as mediated by cues other than symmetry. More modelling and research are needed on this issue.

Finally, the current study illustrates the use and benefits of using structural equation modelling in research on developmental instability. First, the procedures yield an estimate of the effect of developmental instability on the variable. Because individual measures of FA are generally very weak indicators of developmental instability, this effect will often be much greater than the strength of the associations between individual measures and the variable. In the present instance, the individual asymmetries possessed small (albeit on average reliable) associations with the number of sexual partners, but these associations reflected a very substantial effect of developmental instability on the number of sexual partners. Second, they also allow one to estimate the extent to which individual asymmetries tap trait-general developmental instability and, in conjunction with estimated hypothetical repeatabilities, the extent to which developmental instability is due to trait-general or trait-specific influences. The issue of how much developmental instability is due to trait-general influences has been a matter of concern as individual traits' asymmetries typically covary minimally (e.g. Van Valen 1962; Palmer & Strobeck 1986). However, because so little variance in FA is due to any individual differences, even small correlations between individual asymmetries are consistent with developmental instability being largely trait-general rather than trait-specific (Gangestad & Thornhill 1999). In the current example, correlations of only 0.042 were consistent with trait-general developmental instability accounting for over half the total repeatable variance. Structural equation modelling puts the matter on empirical grounds. Third, one need not merely accept a trait-general model without critical reflection. Model fit can be evaluated and compared with the fit of alternative models. In the present instance, the trait-general developmental instability model fitted very well and we suspect that it will in many others. However, in some circumstances it may not. Unfortunately, in some instances researchers have concluded that effects are trait-specific without actually testing whether the effects of traits differ significantly (e.g. Woods *et al*. 1999). Again, structural equation modelling puts the matter on empirical grounds. In the present instance, foot width FA was the only FA that significantly predicted the number of partners by itself. However, it did not outperform the FA of other traits significantly and, had we inferred that it did by comparing significance levels, we would have committed an error in statistical reasoning. One potential drawback of these analyses is that they require a substantial sample size. A minimum sample size of 100 is needed and 200 or more is recommended. However, adequate power in studies of FA, even using trait aggregates, may require samples of this size.

REFERENCES

Arbuckle, J. L. 1995–1999 *Amos 4.0*. Chicago, IL: Smallwaters Corp.

Badyaev, A. V., Foresman, K. R. & Fernandes, M. V. 2000 Stress and developmental stability: vegetation removal causes increased fluctuating asymmetry in shrews. *Ecology* **81**, 336–345.

Bjorksten, T., David, P., Pomiankowski, A. & Fowler, K. 2000 Fluctuating asymmetry of sexual and nonsexual traits in stalk-eyed flies: a poor indicator of developmental stress and genetic quality. *J. Evol. Biol*. **13**, 89–97.

Blanckenhorn, W. U., Reusch, T. & Mühlhäuser, C. 1998 Fluctuating asymmetry, body size and sexual selection in the dung fly *Sepsis cynipsea*—testing the good genes assumptions and predictions. *J. Evol. Biol*. **11**, 735–753.

Chapman, J. W. & Goulson, D. 2000 Environmental versus genetic influences on fluctuating asymmetry in the housefly, *Musca domestica*. *Biol. J. Linn. Soc*. **70**, 403–413.

Dianonis, P. & Efron, B. 1983 Computer-intensive methods in statistics. *Sci. Am*. **248**, 116–120.

Dufour, K. W. & Weatherhead, P. J. 1998 Reproductive consequences of bilateral asymmetry for individual male red-winged blackbirds. *Behav. Ecol*. **9**, 232–242.

Efron, B. 1982 *The jackknife, the bootstrap, and other resampling plans*. Philadelphia, PA: Society for Industrial and Applied Mathemathics.

Enquist, M. & Arak, A. 1994 Symmetry, beauty and evolution. *Nature* **372**, 169–172.

Fuller, R. & Houle, D. 2001 The heritability of developmental instability. In *Developmental instability* (ed. M. Polak). Cambridge University Press. (In the press.)

Furlow, B., Armijo-Prewitt, T., Gangestad, S. W. & Thornhill, R. 1997 Fluctuating asymmetry and psychometric intelligence. *Proc. R. Soc. Lond*. B **264**, 823–829.

Furlow, B., Gangestad, S. W. & Armijo-Prewitt, T. 1998 Developmental stability and human violence. *Proc. R. Soc. Lond*. B **265**, 1–6.

Gangestad, S. W. & Simpson, J. A. 2000 The evolution of human mating: trade-offs and strategic pluralism. *Behav. Brain Sci.* **23**, 675–687.

Gangestad, S. W. & Thornhill, R. 1997*a* The evolutionary psychology of extrapair sex: the role of fluctuating asymmetry. *Evol. Hum. Behav.* **18**, 69–88.

Gangestad, S. W. & Thornhill, R. 1997*b* Human sexual selection and developmental stability. In *Evolutionary personality and social psychology* (ed. J. A. Simpson & D. T. Kenrick), pp. 169–195. Hillsdale, NJ: Erlbaum.

Gangestad, S. W. & Thornhill, R. 1998 Menstrual cycle variation in women's preference for the scent of symmetrical men. *Proc. R. Soc. Lond.* B **265**, 927–933.

Gangestad, S. W. & Thornhill, R. 1999 Individual differences in developmental precision and fluctuating asymmetry: a model and its implications. *J. Evol. Biol.* **12**, 402–416.

Gangestad, S. W. & Thornhill, R. 2001 Fluctuating asymmetry, developmental instability, and fitness: toward model-based interpretation. In *Developmental instability* (ed. M. Polak). Cambridge University Press. (In the press.)

Houle, D. 1992 Comparing evolvability and variability of traits. *Genetics* **130**, 195–204.

Houle, D. 1997 Comment on 'A Meta-analysis of the Heritability of Developmental Stability' by Møller and Thornhill. *J. Evol. Biol.* **10**, 17–20.

Houle, D. 2000 A simple model of the relationship between asymmetry and developmental stability. *J. Evol. Biol.* **13**, 720–730.

Hu, L. & Bentler, P. M. 1999 Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct. Equation Model.* **6**, 1–55.

Johnstone, R. A. 1994 Female preference for symmetrical males as a by-product of selection for mate recognition. *Nature* **372**, 172–175.

Klingenberg, C. P. & Nijhout, H. F. 1999 Genetics of fluctuating asymmetry: a developmental model of developmental instability. *Evolution* **53**, 358–375.

Lens, L. & Van Dongen, S. 1999 Evidence for organism-wide asymmetry in five bird species of a fragmented Afrotropical forest. *Proc. R. Soc. Lond.* B **266**, 1055–1060.

Leung, B. & Forbes, M. R. 1997 Modelling fluctuating asymmetry in relation to stress and fitness. *Oikos* **78**, 397–405.

Leung, B., Forbes, M. R. & Houle, D. 2000 Fluctuating asymmetry as a bioindicator of stress: comparing efficacy of analyses involving multiple traits. *Am. Nat.* **155**, 101–115.

Manning, J. T. & Wood, D. 1998 Fluctuating asymmetry and aggression in boys. *Hum. Nat.* **9**, 53–65.

Mealey, L., Bridgestock, R. & Townsend, G. C. 1999 Symmetry and perceived facial attractiveness: a monozygotic cotwin comparison. *J. Pers. Soc. Psychol.* **76**, 151–158.

Møller, A. P. & Thornhill, R. 1998 Bilateral symmetry and sexual selection: a meta-analysis. *Am. Nat.* **151**, 174–192.

Olkin, I. & Finn, J. 1990 Testing correlated correlations. *Psychol. Bull.* **108**, 330–333.

Palmer, A. R. 1994 Fluctuating asymmetry analysis: a primer. In *Developmental instability: its origins and evolutionary implications* (ed. T. Markow), pp. 355–364. Kluwer: Dordrecht.

Palmer, A. R. 1999 Detecting publication bias in meta-analysis: a case study of fluctuating asymmetry and sexual selection. *Am. Nat.* **154**, 220–233.

Palmer, A. R. & Strobeck, C. 1986 Fluctuating asymmetry: measurement, analysis, patterns. *A. Rev. Ecol. Syst.* **17**, 391–421.

Perrett, D. I., Burt, D. M., Penton-Voak, I. S., Lee, K. J., Rowland, D. A. & Edwards, R. 1999 Symmetry and human facial attractiveness. *Evol. Hum. Behav.* **20**, 295–307.

Rhodes, G., Proffitt, F., Grady, J. & Sumich, A. 1998 Facial symmetry and the perception of beauty. *Psychomet. Bull. Rev.* **5**, 659–669.

Rikowski, A. & Grammer, K. 1999 Human body odour, symmetry, and attractiveness. *Proc. R. Soc. Lond.* B **266**, 869–874.

Thornhill, R. & Gangestad, S. W. 1994 Fluctuating asymmetry and human sexual behavior. *Psychol. Sci.* **5**, 297–302.

Thornhill, R. & Gangestad, S. W. 1999*a* Facial attractiveness. *Trends Cognit. Sci.* **3**, 452–460.

Thornhill, R. & Gangestad, S. W. 1999*b* The scent of symmetry: a human sex pheromone that signals fitness? *Evol. Hum. Behav.* **20**, 175–201.

Thornhill, R., Gangestad, S. W. & Comer, R. 1995 Human female orgasm and mate fluctuating asymmetry. *Anim. Behav.* **50**, 1601–1615.

Thornhill, R., Møller, A. P. & Gangestad, S. W. 1999 The biological significance of fluctuating asymmetry and sexual selection: a reply to Palmer. *Am. Nat.* **154**, 234–241.

Van Dongen, S. 1998 How repeatable is the estimation of developmental stability by fluctuating asymmetry? *Proc. R. Soc. Lond.* B **265**, 1423–1427.

Van Dongen, S. & Lens, L. 2000 The evolutionary potential of developmental instability. *J. Evol. Biol.* **13**, 326–335.

Van Dongen, S., Sprengers, E. & Löfstedt, C. 1999 Correlated development, organism-wide asymmetry and patterns of asymmetry in two moth species. *Genetica* **105**, 81–91.

Van Valen, L. 1992 A study of fluctuating asymmetry. *Evolution* **10**, 139–146.

Waynforth, D. 1998 Fluctuating asymmetry and human male life history traits in rural Belize. *Proc. R. Soc. Lond.* B **265**, 1497–1501.

Whitlock, M. 1998 The repeatability of fluctuating asymmetry: a revision and extension. *Proc. R. Soc. Lond.* B **265**, 1429–1431.

Woods, R. E., Srgó, C., Hercus, M. J. & Hoffmann, A. A. 1999 The association between fluctuating asymmetry, trait variability, trait heritability and stress: a multiply replicated experiment on combined stresses in *Drosophila melanogaster*. *Evolution* **53**, 493–505.

As this paper exceeds the maximum length normally permitted, the authors have agreed to contribute to production costs.

An electronic appendix to this paper can be found at (http://www.pubs.royalsoc.ac.uk).