Fluctuating Asymmetry as a Bioindicator of Stress: Comparing Efficacy of Analyses Involving Multiple Traits

Brian Leung,[1,*] Mark R. Forbes,[2] and David Houle[3]

1. Department of Zoology, University of Cambridge, Cambridge CB2 3EJ, United Kingdom;
2. Biology Department, Carleton University, Ottawa, Ontario K1S 5B6, Canada;
3. Department of Biological Sciences, Florida State University, Tallahassee, Florida 32306-1100

Submitted December 10, 1998; Accepted August 4, 1999

ABSTRACT: Researchers have suggested fluctuating asymmetry (FA) as an indicator of environmental stress and have usually tested this assertion by examining relations between FA of single traits and stress. Fluctuating asymmetry stress relations are real but are typically weak and difficult to detect. Researchers would like to maximize the probability of detecting FA-stress relations when they exist. We assert that analyses based on the FA of multiple traits may provide better methods for detecting stress. In this article, we used computer simulations to compare the ability of six analyses to detect differences in FA between stressed and unstressed populations. We show that the optimal analysis depends upon the underlying form of the FA distributions. We also show that two of the analyses had inflated Type I errors in some situations. Finally, we quantify the advantage of our preferred analysis over those of single-trait FA in detecting stress.

Keywords: bioindicator, composite fluctuating asymmetry, power, stress.

Applied biologists are interested in monitoring environmental stress, preferably before stress irreversibly damages populations (Clarke 1993*a*; Bunn 1995). Researchers have used changes in the biota to indicate stress (i.e., bioindicators). While a variety of bioindicators exist, fluctuating asymmetry (FA) has received increasing attention in the last decade. Most organisms have one or more axes of symmetry about which the body is, basically, a reflection. Most individuals are not exactly symmetrical but differ slightly in the realization of these repeated structures. Fluctuating asymmetry refers to these directionally random, subtle departures from perfect symmetry and is hypothesized to indicate the inability of an organism to maintain precise development (Palmer and Strobeck 1986).

Fluctuating asymmetry variance of populations and absolute FA values of individuals has been found to increase with stress (reviewed in Leung and Forbes 1996). One possible mechanism of relations between FA and stress is that organisms require energy to compensate for stress. This should reduce energy for growth and reproduction (Koehn and Bayne 1989), which may eventually influence populations. Stress may also reduce the energy available to maintain developmental precision (Sommer 1996). Thus, FA should increase with stress. Fluctuating asymmetry could provide advantages over other bioindicators of stress because FA is cost-effective and easy to measure (Clarke 1993*a*). Also, FA has been related to quality of organisms; therefore a change in FA should be biologically relevant (Sommer 1996). However, empirical relations between FA based on single traits and stress are, on average, weak and heterogeneous (Leung and Forbes 1996). Further, we would predict that relations between FA and stress should be weak, based on mathematical algorithms generating FA (Leung and Forbes 1997). Thus, if researchers are serious about using FA as a bioindicator of stress, methods must be found to increase the reliability of FA to detect stress.

Previous analyses and mathematical models have usually considered FA based on single paired traits. Analyses that combine information across traits should prove to be more reliable detectors of stress (Leary and Allendorf 1989; Watson and Thornhill 1994). No study has quantitatively compared the efficacy or the validity of such analyses, although some composite indices of FA (CFAs) have been used in empirical studies (e.g., Whitlock 1993; Dufour and Weatherhead 1996; Manning and Ockenden 1994) and other analyses have been suggested (Palmer 1994; Zhivotovsky 1992). Of those studies that have used a CFA measure, the strengths and directions of FA relations for single traits are not typically reported (with the exception of Manning and Ockenden 1994), and different CFA analyses are not

* To whom correspondence should be addressed; e-mail: bleung@zoo.cam.ac.uk.

Am. Nat. 2000. Vol. 155, pp. 101–115.

typically used on the same data sets. For convenience, we will refer to analyses combining information on FA distributions across traits as CFA analyses.

In this article, we examine properties of CFA analyses using simulations. In a simulation study, we can work with a model for FA whose properties are known and ask how much information about the model can be recovered from various measures of FA. Furthermore, we can isolate and study a number of factors that could affect the relative efficacy of CFA analyses, such as the strength of FA-stress relations, the magnitude of measurement error, the variance in developmental stability of different traits, and the deviation from normality. The objective of this study is to help researchers decide which CFA analysis should be used, given different patterns of FA distributions. We determine which analyses are valid, in terms of the probability of concluding that FA of two samples differ when samples come from identical FA distributions (i.e., Type I error rates), and which analyses provide the greatest probability of detecting relations when FA differs between two populations (i.e., power).

Composite Fluctuating Asymmetry Analyses

We consider six analyses of fluctuating asymmetry across multiple traits, which we refer to as CFA analyses. These are defined below and summarized in appendix A. All of the first three methods involve calculation of a single index of FA across traits, which are then analyzed by a simple *t*-test. The last three methods are multivariate analyses and do not involve calculation of a simple index of FA.

The first composite statistic (CFA 1) is the sum (or mean) of absolute FA values ($|FA|$) for all traits for each individual (e.g., Clarke and McKenzie 1992; Whitlock 1993). Absolute FA values are used because FA of individuals should be directionally random, and it is the magnitude of asymmetry in either direction that may indicate stress. This statistic is simple to calculate, but traits with higher FA magnitudes would be weighted more heavily than more developmentally stable traits. Some researchers have standardized FA by the mean size of the trait in each individual (e.g., Thornhill et al. 1995) or the average trait size of the population (Palmer and Strobeck 1986) before calculating CFA 1. This would be appropriate to control for size differences when FA magnitude across traits scales positively and isometrically to trait size. However, this would not necessarily control for differences in developmental stability across traits. We do not model or consider allometric relations in this manuscript (but see Leung 1998).

Another way to standardize FA magnitudes of traits is to divide each FA value by the average $|FA|$ of a given trait in the populations of interest such that all traits contribute equally to CFA measure and then to sum $|FA|$ values across traits for each individual such that each individual has a composite FA score. This is our second composite index of FA (CFA 2). To our knowledge, CFA 2 has not been applied before. We consider CFA 2 because traits are likely to differ in their developmental stability and because the relationship between the magnitude of developmental stability and the ability to index stress is unclear (Leung and Forbes 1997). Another method of standardizing FA magnitudes across traits is to rank $|FA|$ values of individuals for each trait and then to sum the ranks for each individual such that each individual has a composite FA score (e.g., Zakharov et al. 1991; Dufour and Weatherhead 1996). This summed, ranked measure is our third composite index of FA (CFA 3). Each of these three CFA indices can then be analyzed by a *t*-test, as we do in our simulations, or by an ANOVA if there are more than two samples of individuals to compare.

The fourth analysis (CFA 4), uses standardized, generalized variance to quantify asymmetry (Zhivotovsky 1992). To calculate this, first a variance-covariance matrix consisting of all traits is generated for each population (the main diagonal is the variance associated with each trait; the nondiagonal elements are the covariances between traits). Then the determinant of the matrix is taken, modified according to the number of traits measured, and multiplied by a coefficient (G) to produce an overall variance score for a population (see app. A). The result: variances of two populations can be compared using an *F*-ratio test. Detailed formulas and equations are given by Zhivotovsky (1992). This index, called "generalized FA variance" (GFA) by Zhivotovsky (1992), has the advantage of statistically pooling information across characters, thus taking into account correlations between signed asymmetries. GFA also has the advantage of increasing degrees of freedom, which are a function of both sample size and of number of traits examined. However, GFA indices are not available on many statistical packages, and tables of values are not comprehensive (see tables by Zhivotovsky 1992). Palmer (1994), noting the complexity of GFA, suggested two-way ANOVAs as an alternative, with population (stressed vs. unstressed) and traits as independent variables and $|FA|$ as the dependent variable. This is our fifth composite index of FA (CFA 5). We note that two-way ANOVAs are typically analyzed using two-tailed testing. However, we had a priori predictions with respect to the direction of relations. Thus, we adjusted the critical value and only considered significance if the relation was in the predicted direction (similar to Rice and Gaines [1994] except that we only considered significance in the predicted direction).

MANOVA has also been used to integrate information across traits and to determine differences in $|FA|$ between

samples (e.g., Alados et al. 1993; Clarke 1993*b*). This is our sixth composite index of FA (CFA 6). In this study, we used Wilks's Lambda (Λ) as our test statistic and converted Λ into the *F*-statistic following Lindeman et al. (1980).

All statistics generated by our computer programs were verified by comparing output from several simulations with output generated by STATISTICA statistical package (Statsoft 1993). Fluctuating asymmetry differences between samples for single traits were determined by comparing variances using Levene's test (i.e., *t*-test on absolute FA values as suggested by Palmer and Strobeck [1992]).

Model

Baseline Simulations

We began by generating FA values using a baseline model that was defined as follows: the underlying distribution of FA values was normal, centered at 0, with variance ($\sigma^2_{p,t}$), where the subscripts p and t denote population and trait, respectively. Populations $p = 1$ and $p = 2$ refer to the unstressed and stressed populations, respectively. This distribution conforms to typical views of FA (Palmer and Strobeck 1986). Furthermore, FA values were independent between traits, measurement error (ME) was not considered at this point, and the underlying distributions of FA values were identical for all traits. For our baseline simulations, we used 10 traits (k) and a sample size (N) of 40. Subsequently, we modified the baseline model, altering parameters. We used "conditions" as a convenient term to refer to the patterns and parameter sets that we modeled. For each condition, we examined a number of levels; for example, we examined three levels of measurement error.

We analyzed FA difference between populations using the six CFA analyses described above. We used the ratio of variances between populations as a convenient measure of the effect size ($E = \sigma^2_{2,t}/\sigma^2_{1,t}$), where E refers to the effect size and $\sigma^2_{2,t}$ and $\sigma^2_{1,t}$ refer to the population variances of the putatively stressed and unstressed populations, respectively. We considered $E = 1$, 1.2, and 1.5. When $E = 1$, the two FA distributions are identical (i.e., there is no stress). Here, significant differences between stressed and unstressed samples are due to Type I errors. For the other two E values, differences in the percentage of significant analyses reflect differences in power. These E values offer a range of strengths of FA-stress relations. To determine Type I error rates and power, we ran 1,000 simulations for each set of parameter values.

We found that Type I error rates were as predicted ($\alpha = 0.05$) for all CFA analyses when no differences in FA variance existed between populations ($E = 1$; fig. 1*a*; app. B). Thus, all CFA indices could be used validly under these baseline conditions. Power was dramatically improved using CFA analyses compared to single traits; at times it was as much as four times as great (fig. 1*c*; app. B). CFAs 1, 2, and 5 yielded very similar power. These analyses could provide marginally greater power than CFA 3 and CFA 4 analyses (10%–15%). CFA 6 could be as much as 50% less powerful than the other analyses (fig. 1*c*; app. B).

Heterogeneous FA Distributions between Traits

In nature, we might expect the magnitude of FA values to differ between traits. Traits under strong directional selection may have a high FA variance compared to other morphological traits, whereas traits whose asymmetry affects their importance may have lower FA variances. For example, Møller and Höglund (1991) found that sexually selected traits had average FA values an order of magnitude greater than other morphological traits. Gummer and Brigham (1995) found that FA variance in bat wings were about twice as large as variance in tibias. Clearly, FA can differ appreciably between traits. We considered the following five levels of heterogeneity across our 10 traits: In level 1, variances of traits 1–5 were three times the variance of traits 6–10, traits 1–5 were homogeneous, and traits 6–10 were homogeneous (i.e., $\sigma^2_{1,1-5} = 3 \times \sigma^2_{1,6-10}$). In level 2, variance of one trait was three times the variance of the other traits; the other traits were homogeneous. In level 3, variances of traits 4–6 were three times the variance of traits 1–3, variances of traits 7–10 were 10 times the variance of traits 1–3, and traits 1–3, 4–6, and 7–10 were homogeneous. In level 4, variances of traits 1–5 were 10 times the variance of traits 6–10, traits 1–5 were homogeneous, and traits 6–10 were homogeneous. Finally, in level 5, variance of one trait was 10 times the variance of the other traits, the other traits were homogeneous, and other parameters were as for the baseline simulations. Values of E were homogeneous between traits.

As with the baseline simulations, we found that Type I error rates were as predicted (at $\alpha = 0.05$) for all analyses (fig. 1*b*; app. B) and that CFA analyses provided much greater power than analyses using single traits. Once again, CFAs 1, 2, and 5 were the most powerful analyses and CFA 6 the weakest (fig. 1*d*; app. B). In contrast, however, we found that CFA 2 could be marginally more powerful than CFAs 1 and 5, yielding as much as 10% more power depending on the magnitude and nature of the heterogeneity (fig. 1*d*; app. B).

Leptokurtism

Leptokurtic distributions occur when more data points fall in the center and at the tails of a distribution as compared

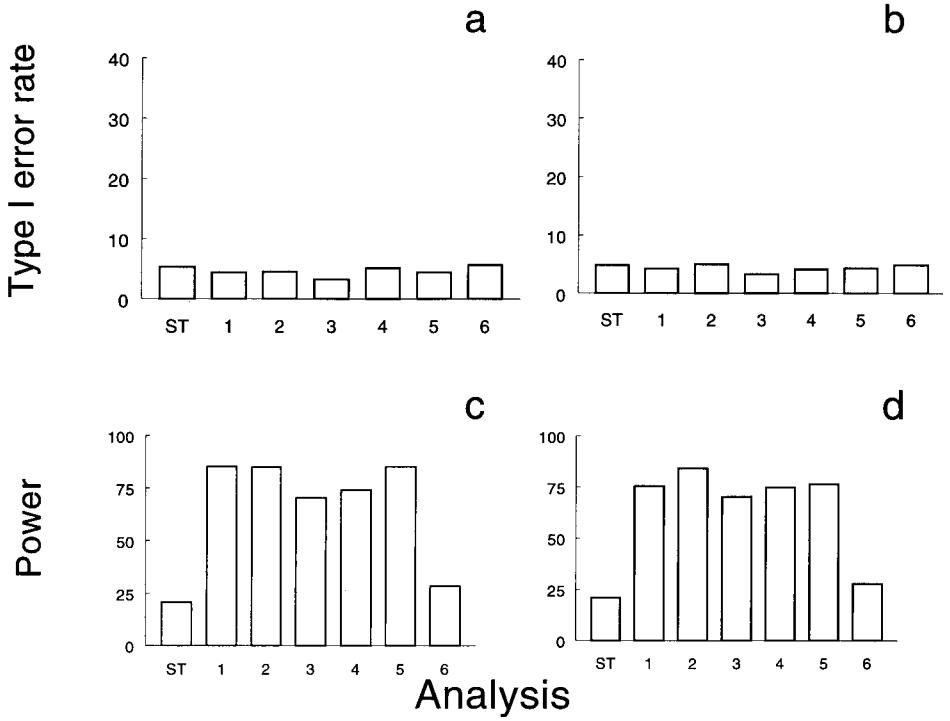

Figure 1: Type I error rates ($E = 1$, panels *a* and *b*) and power ($E = 1.5$, panels *c* and *d*) for six CFA analyses. Analyses using single traits (ST) were shown under different conditions. The conditions were baseline (panels *a* and *c*) and FA distributions heterogeneous between traits (panels *b* and *d*). Baseline conditions were that the underlying FA distributions of all traits were normally distributed, centered at 0, uncorrelated, without measurement error, and homogeneous between traits. We used 40 individuals and 10 traits for the baseline simulations. For the other conditions, we modified one property of the model and maintained baseline properties for the rest of the model. We presented one level of heterogeneity between traits (level 5; see app. B and text). CFA analyses are described in appendix A.

to a normal distribution (Zar 1984). Leptokurtic FA distributions may reflect a composite of many heterogeneous FA distributions (Houle 1997), which can arise due to differences in developmental stability between individuals (Leung and Forbes 1997). Thus, we considered FA variances that were heterogeneous between traits. Again, E values were as for baseline conditions. We considered three levels of kurtosis. The first level, $\sigma^2_{1,t}$, ranged from 1 to 3. The second level, $\sigma^2_{1,t}$, ranged from 1 to 10. And the third level, $\sigma^2_{1,t}$, ranged from 1 to 100. Levels 1–3 produced average kurtosis values (g) of approximately $g = 0.7$, 2.9, and 7.6, respectively. Ranges of $\sigma^2_{2,t}$ could be calculated from $E \times \sigma^2_{1,t}$.

We found that CFA 4 had inflated Type I errors that could be as high as 34% when $g = 7.6$ (top panel of fig. 2; app. B). Thus, we did not consider the power of CFA 4 for leptokurtic FA distributions. The other analyses had predicted levels of Type I errors. It appeared as though the power of both analyses using single traits and CFAs decreased as kurtosis increased. However, it is unknown whether this was a function of kurtosis per se or whether it was due to the additional component of variation. The most powerful CFA analysis depended on the level of kurtosis. When kurtosis was low, CFAs 1, 2, and 5 provided the most powerful analyses. When kurtosis was high, CFA 3 provided the most powerful analysis and could yield up to 25% more power than the other CFA indices (bottom panel of fig. 2; app. B). When kurtosis was high, CFA 2 provided marginally greater power than CFA 1.

Correlations of FA Values between Traits

If there is a component of developmental stability that is organism-wide, then we would expect weak concordance among FA values of traits, albeit typically nonsignificant (see Leung and Forbes 1997 for reasoning as to why FA concordance should be weak). Further, we would expect $|\mathrm{FA}|$ values but not necessarily signed FA values to be concordant. We modeled this by defining $\sigma^2_{p,t}$ as being composed of two variances ($\sigma^2_{p,t} = V1^2_{p,t} + V2^2_{p,t}$), where $V1^2_{p,t}$ represents the portion of $\sigma^2_{p,t}$ that is uncorrelated between traits within individuals and $V2^2_{p,t}$ represents a

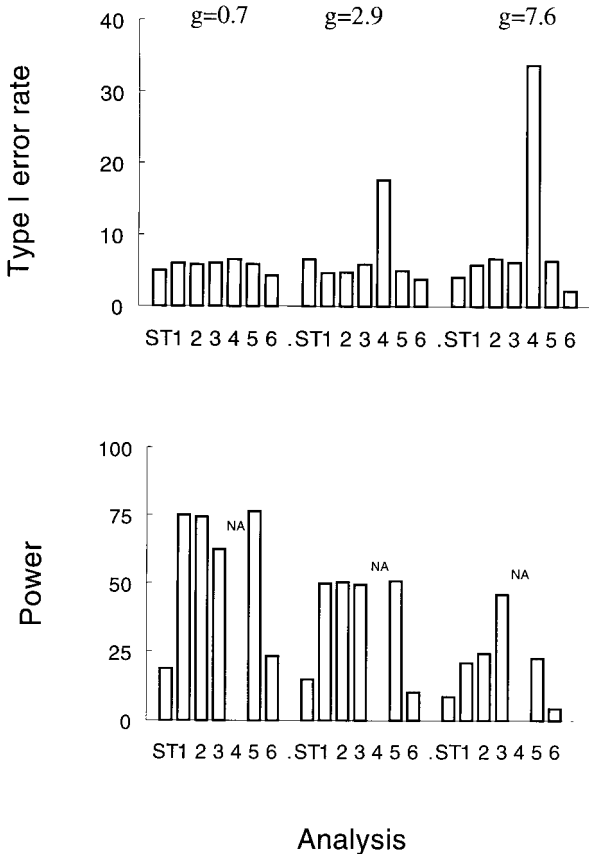

Figure 2: Type I error rates ($E = 1$) and power ($E = 1.5$) for six CFA analyses and analyses using single traits (ST) were shown for three levels of kurtosis (g). Type I error rates were inflated for CFA 4, thus power of CFA 4 was not considered (NA = not applicable). These levels corresponded to levels 1–3 in the text and appendix B. CFA analyses are described in appendix A.

component of $\sigma^2_{p,t}$ that is, common between traits. Absolute values comprising $V2^2_{p,t}$ were correlated but were directionally random such that signed FA values would not be correlated but |FA| would. The common component (C) of $\sigma^2_{p,t}$ could be expressed as a proportion of the total variance (i.e., $C = V2^2_{p,t}/\sigma^2_{p,t}$). The underlying distribution from which FA values were chosen was homogeneous between individuals and traits within populations.

We examined three levels: $C = 0.091$, $C = 0.23$, and $C = 0.33$. All three corresponded to average FA-FA correlations of $r = 0.0052$, $r = 0.039$, and $r = 0.085$, respectively. Thus, we modeled very weak concordance between traits.

We found that CFA 4 and CFA 5 yield inflated Type I error rates when |FAs| were concordant across traits (top panel of fig. 3; app. B). Thus, the power of these CFA analyses was not considered when FA values were concordant between traits. We found that the power of the other CFA analyses declined as the magnitude of FA-FA concordance increased. However, the relative positioning of these CFA analyses did not change from baseline conditions, with CFA 1 and CFA 2 yielding the greatest power (bottom panel of fig. 3; app. B).

Measurement Error

Measurement error reduces the ability of FA to detect stress (Palmer 1994). It is unknown what effect measurement error should have on the relative efficacy of different CFA analyses. Similar to the above simulations, we modeled $\sigma^2_{p,t}$ to be composed of two variances ($\sigma^2_{p,t} = V1^2_{p,t} + V^2_{\mathrm{ME}}$), where V^2_{ME} is the variation due to measurement error. The degree of measurement error (ME) could be expressed as a proportion of $\sigma^2_{1,t}$ (i.e., $\mathrm{ME} = V^2_{\mathrm{ME}}/\sigma^2_{1,t}$).

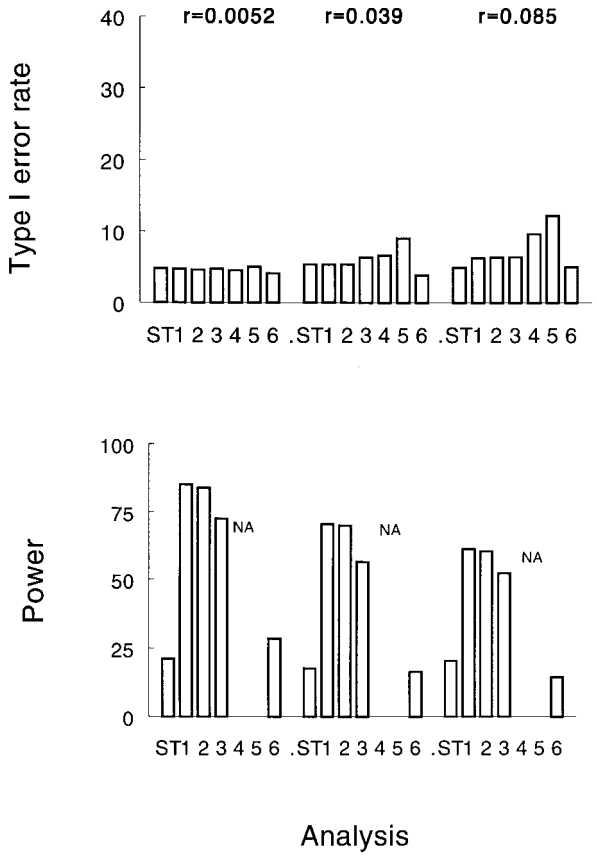

Figure 3: Type I error rates ($E = 1$) and power ($E = 1.5$) for six CFA analyses and analyses using single traits (ST) were shown for three levels of correlations between FA values of traits (r). Type I error rates were inflated for CFA 4 and 5, thus power of CFA 4 and 5 were not considered (NA = not applicable). These levels corresponded to levels 1–3 in the text and in appendix B. CFA analyses are described in appendix A.

We examined three levels: ME = 0.091, ME = 0.23, and ME = 0.33. Stressed and unstressed populations experienced the same value of V_{ME}^2.

Not surprisingly, the power to detect differences between traits decreased with increasing ME. However, Type I errors and relative positioning of CFA analyses in terms of power remained unchanged compared to the baseline simulations. These results are presented only in appendix B.

Number of Traits and Sample Size

The number of traits (k) combined into a CFA measure and the sample size (N) both would affect the power of a CFA analysis to detect stress. However, it is unknown whether these factors influence the relative efficacy of different CFA analyses. Thus, we examined different values of k and N. We used $N = 40$ ($k = 5$), $N = 80$ ($k = 10$), and $N = 20$ ($k = 5$) as three separate sets of simulations. When $N = 20$, we set $k = 5$ because it was not possible to use CFA 4 with $N = 20$ and $k = 10$ (with 10 individuals in each of the stressed and unstressed populations and 10 traits, we would have 0 degrees of freedom).

As for the baseline simulations, Type I errors were as predicted ($\alpha = 0.05$) for all CFA analyses and CFAs 1, 2, and 5 provided the greatest power. These results are presented only in appendix B.

Discussion

Fluctuating asymmetry could potentially offer several advantages as a bioindicator of environmental stress (Clarke 1993*a*), provided FA is used and interpreted correctly. Researchers argue that FA should be higher in a stressed population than in a relatively unstressed one and that elevated FA indicates the presence of stress. Therefore, it is critical that methods of analyzing FA do not falsely indicate differences in FA between populations beyond predicted Type I error rates, while simultaneously maximizing the power or the probability of detecting differences in FA when they exist. In the sections that follow, we present recommendations concerning how FA should be analyzed and reasons for the advantages/disadvantages of different analyses.

Recommendations

CFA analyses typically provided much greater power than analyses based on single traits. For example, CFA 2 could detect differences in FA among populations 85% of the time compared to only 20% using FA of single traits. In fact, to obtain the same power of CFA 2, we would require a sample size of k^*N with analyses using single traits, where N is the sample size using CFA 2 and k is the number of traits combined (reason detailed below). Thus, CFA analyses should be preferred over analyses that use single traits (for qualitative arguments, see Leary and Allendorf 1989 and Watson and Thornhill 1994).

However, not all CFA analyses were equally useful. We suggest that researchers analyze data using standardized, summed FA values across traits (i.e., CFA 2 and CFA 3, depending on the level of kurtosis). Both parametric analyses (CFA 2) and nonparametric, ranked analyses (CFA 3) have predicted Type I error rates and yield the greatest power. For normal FA distributions that were heterogeneous between traits and for slightly leptokurtic FA distributions, CFA 2 provided the greatest power (although the difference in power between CFAs 1, 2, and 5 was < 10%). The power of CFA 2 was essentially identical to CFA 1 (unstandardized, summed FA values across traits) and CFA 5 (two-way ANOVAs) when FA distributions were homogeneous between traits. We expect, however, that FA distributions would not typically be identical between traits in nature.

If FA distributions are leptokurtic, we suggest either CFA 2 or CFA 3. For low kurtosis values, CFA 2 was approximately 15% more powerful than CFA 3, whereas, for high kurtosis values, CFA 3 was approximately 25% more powerful than CFA 2. Our results are consistent with those of Gangestad and Thornhill (1998) in that CFA 2 yielded predicted Type I error rates regardless of the degree of kurtosis. Further, CFA 2 was more powerful than CFA 3 when the signed FA distribution was normal, even though the $|\text{FA}|$ distribution on which analysis was conducted theoretically would be nonnormal. However, our results differ from Gangestad and Thornhill (1998) in that CFA 3 (a nonparametric measure of FA) provided much greater power as the degree of nonnormality of the signed FA distribution increased.

The obvious question, of course, is, What constitutes high versus low kurtosis? Unfortunately, the answer is not as simple as whether or not the FA distributions are significantly leptokurtic. The point at which CFA 3 becomes more powerful than CFA 2 is a function of sample size and of effect size; thus, it is impossible to give a precise kurtosis value at which CFA 3 should be used. Nevertheless, we present tables of kurtosis values based on computer simulations at which CFA 3 becomes more powerful that examine a range of sample sizes (N), effect sizes (E), and number of traits combined (k; app. C). We note that there is a great deal of overlap in terms of which index should be most powerful at intermediate levels of kurtosis (fig. *C*1 in app. C).

The other CFA analyses (4–6) were flawed. Specifically, CFA 4 (GFA) yielded inflated Type I errors when signed FA distributions were leptokurtic (cf. Monte Carlo sim-

ulations by Zhivotovsky [1988]) and when |FA| values were related between traits. Further, CFA 4 was much more difficult to calculate and was not amenable to most statistical packages. Similarly, CFA 5 (two-way ANOVAs) yielded inflated Type I errors when |FA| values were related between traits. For example, there was a 12% Type I error rate using CFA 5 when the average correlation was only $r = 0.085$ and typically nonsignificant (corresponding to $r^2 = 0.007$ or only 0.7% of the variation explained). Thus, researchers could not use significance as an indication as to when relations between |FA| values of different traits would be a problem. While CFA 6 (MANOVA) had predicted Type I error rates, its power was considerably less than the other CFA indices (we could detect differences in FA between populations 50% more often by using CFA 2 than by using CFA 6).

Mechanisms of Relative Benefits of CFA Indices Compared to Single Traits

The mechanisms underlying relative benefits of GFA, two-way ANOVAs, and MANOVA are clear. Both GFA and two-way ANOVAs reduce the standard error by increasing the effective sample size and the degrees of freedom (Zhivotovsky 1992 and Palmer 1994, respectively). MANOVA estimates coefficients associated with each trait FA such that the maximal amount of variation is explained (Lindeman et al. 1980). The reasons why analyses using CFAs 1–3 are more powerful are perhaps less clear.

As above, CFAs 1–3 also act to reduce the standard error. The |FA| of each trait of each individual comes from a distribution of |FA| values from which one value is chosen. As more traits are measured, the standard error around the mean |FA| across all traits for a given individual would decrease relative to mean |FA|, following the central limit theorem (Zar 1984). This resulted in reduced standard error in a population. Thus, we would have an increased probability of detecting differences between populations if |FA| values of individuals in one population were generally greater than those in another population (as we would predict to occur due to stress).

Specifically, in the absence of FA-FA correlations, the standard error would be reduced by approximately a factor of $(k)^{0.5}$, where k is the number of traits combined. Thus, using single traits, we would require an average sample size of k^*N to obtain the same power as analyses using CFA 2. Of course, the advantage of combining traits decreases as the degree of correlation between traits increases; for perfect correlation between traits, analyses using CFA 2 would provide no advantage over single traits.

Comparison of CFA Indices: Mechanisms of Differences in Power and Validity

All five CFAs had similar power under baseline conditions because all of them acted to reduce the standard error (although not with the same efficiency). The difference in power between CFA 2 and CFA 3 reflected the difference between parametric versus nonparametric CFA measures; these differences in power depended on the degree of nonnormality (app. C). When FAs were heterogeneous between traits, a higher power of CFA 2 than of CFAs 1 and 5 occurred because CFAs 1 and 5 would implicitly weight asymmetric traits more heavily than less asymmetric traits. To illustrate the potential consequences of heterogeneity, consider a trait whose FA variance was an order of magnitude larger than other traits (e.g., Møller and Höglund 1991). CFAs 1 and 5 would primarily reflect this one trait. Thus, the standard error would primarily reflect a single trait rather than being optimally reduced by taking the standardized average of many traits. Not surprisingly, CFAs 1, 2, and 5 were essentially identical when FA distributions were homogeneous between traits. In such cases, traits would be weighted approximately equally using these analyses. We note that CFAs 1 and 5 could potentially be equal to or stronger than CFA 2 if highly asymmetric traits were consistently more strongly affected by stress.

As noted above, CFA 5 treated traits as independent data points and therefore had inflated Type I errors when |FA| values of traits were correlated. GFA was also sensitive to correlations of unsigned asymmetries. Although GFA takes into account signed asymmetries and hence may not be biased by developmentally linked traits, unsigned asymmetries may be correlated without signed asymmetries being correlated. Hence, GFA could be biased by slight concordance among traits (albeit typically nonsignificant) due to some organism-wide component of developmental stability or quality (Leung and Forbes 1997).

MANOVA was typically the weakest index. This was because a separate parameter needed to be estimated for each trait. Thus, more variability needed to be explained to find significance. In contrast, CFA 2 increased in power following the central limit theorem but did not concurrently require additional variation to be explained in order to find significance. MANOVA also had to be two tailed, since each trait had a parameter estimate that could be in either direction, whereas the other analyses could be analyzed using one-tailed testing. Although there are situations in which two-tailed testing would be appropriate, MANOVA would still be the weakest index.

Researchers may also ask whether single traits can sometimes be more reliable than CFA indices. If the effect sizes of FA-stress relations are heterogeneous, this is certainly possible. CFA 2 would be more powerful, typically, than

the average of single traits. However, CFA 2 may be a combination of both strong and weak FA relations. Weak FA relations would act like random noise and could result in a weaker CFA index compared to reliable single traits. However, two problems would be associated with using reliable single traits. First, consideration of multiple individual traits would require some correction for multiple tests and would therefore result in reduced power (e.g., sequential Bonferonni, Palmer 1994). Second, it is unknown which trait would be most reliable. Thus, a priori testing would be required to find such traits without causing inflated Type I error rates. Further, if there were a priori reasons for expecting one trait to be more reliable than another, it could be possible to weight CFA 2 values by the estimated reliability of individual traits (e.g., taking into account measurement error). However, one cautionary note regarding such weighting procedures is that inappropriate weighting would weaken CFA 2 indices. In such instances, CFA 2 would reflect some traits more than others; hence reduction in standard error due to taking averages would not be as great. Thus, while weighting may provide advantages even over CFA 2, optimal methods of weighting have yet to be determined.

Finally, we note that the CFA indices discussed here are not the only methods possible for combining information across traits. Repeated measures of ANOVA treating traits as the repeated measure or a nested ANOVA with individuals nested within groups should yield similar results to CFA 1. In fact, these tests turn out to be mathematically similar, and the degrees of freedom reflect the number of individuals rather than the number of traits by individuals (i.e., no pseudoreplication). These tests may offer the advantage of allowing for tests of interactions between stress and traits. However, consideration of such issues as compound symmetry (or sphericity) may be required. Further, if the main consideration is to determine whether a putatively stressed sample has higher FA than an unstressed sample (as typically appears to be the case), then CFAs 1–3 offer the most powerful and simplest tests, whereas analyses using repeated measures or nested designs could quickly become complicated.

FA-Quality Relations within Samples

This study dealt exclusively with detection of FA differences between samples. However, many studies examine relations between FA and some measure of quality (e.g., mating success) within studies. The logic underlying the findings of this article extends to tests of FA-quality relations. Further, although single traits sometimes relate significantly to quality, FA-quality distributions within samples should typically show a triangular-bivariate distribution, and FA of single traits may only be predictive of quality at high FA values (see arguments by Leung and Forbes [1997]; but cf. Rowe et al. 1997 for antisymmetry-quality distributions). However, with composite measures, symmetry need not be the common phenotype, due to the central limit theorem, and FA can potentially be predictive of quality over the entire range of FA values.

Summary

It may be possible to use FA to monitor stress. However, current methods of analyses usually employ FA of single traits and appear to be weak and unreliable. If FA is to be used, it is essential to maximize the reliability of FA as a monitor of stress. Composite measures of FA can provide vast advantages over single traits in terms of the probability of detecting FA differences between populations. However, different methods of CFA analyses differ both in their validity (Type I error rates) and in their power, and it is critical to determine which CFA analyses should be used. In this study, we demonstrated that the most powerful, simplest, and robust CFA analyses are simply *t*-tests using standardized, summed FA values. When kurtosis is low, parametric, summed FA values provide the greatest power (CFA 2). When kurtosis is high, nonparametric, summed, ranked FA values provide the greatest power (CFA 3).

Acknowledgments

We thank M. Driedger, S. Findlay, M. Gunness, and A. Woodside for help with multivariate analysis; K. Dufour for discussions on FA; and J. M. Emlen, R. Palmer, J. Travis, and an anonymous reviewer for helpful comments. This work was supported by a Natural Sciences and Engineering Research Council of Canada (NSERCC) scholarship to B. Leung and by an NSERCC grant to M. R. F and D. H.

APPENDIX A

Table A1: Composite FA analyses (CFA)

Index	Equation	Description
CFA 1	$CFA_i = \lvert FA_{ij} \rvert \; j = 1 \text{ to } k$	Summation of absolute FA values across traits (j) for each individual (i). CFA and FA denote composite asymmetry values for an individual and asymmetry for each trait, respectively.
CFA 2	$CFA_i = \lvert FA_{ij} \rvert / \text{avg} \lvert FA_j \rvert \; j = 1 \text{ to } k$	Summation of standardized absolute FA values. FA values of a given trait are first divided by the average absolute FA magnitude for that trait. Thus, FA values of all traits are given equal weight.
CFA 3	$CFA_i = \text{ranked} \lvert FA_{ij} \rvert \; j = 1 \text{ to } k$	FA values of individuals are ranked for each trait. Thus, FA is standardized across traits. The ranked values of each trait of an individual are summed to give a CFA value for that individual (i).
CFA 4	$GFA = G^*(\det \mathbf{S})^{1/k}$	Generalized FA variance, where G is a coefficient that eliminates statistical bias, det S is the determinant of the variance-covariance matrix of FAs of different traits. GFAs of stressed versus unstressed populations can be compared using F-ratio test. Zhivotovsky (1992) provides necessary matrix notation and formulae to calculate G and a brief table of G values and degrees of freedom.
CFA 5	$FA_{ijg} = \mu + \alpha_j + \beta_g + (\alpha\beta)_{jg} + \epsilon_{ijg}$	Two-way ANOVA model (Type I) using trait and sample (stressed vs. unstressed) as independent factors and FA as the dependent factors. The expressions μ, α_j, β_g, $(\alpha\beta)_{jg}$, ϵ_{ijg}, and the subscript g denote the grand mean FA across populations and traits, the effect of trait, the effect of stress, the interaction term, the error term, and population (grouping variable), respectively (Palmer 1994). Here, we are interested in whether the effect of stress (β_g) is significant. Equation following Lindeman et al. (1980, p. 137).
CFA 6	$D_{ig} = u_j(\lvert FA_{ijg} \rvert)$	MANOVA examining FA differences between samples. Coefficients (u) are calculated to maximize the differences between D values of stressed and unstressed populations. Equation following Lindeman et al. (1980, p. 171).

Note: The CFA numbers correspond to numbers used in the text. Subscripts: i = individual, j = trait, and g = population. The variable k represents the number of traits per individual.

APPENDIX B

Ranges of Type I errors and power were presented for analyses using single traits and CFA measures (see app. A for CFA measures). Baseline conditions were that underlying FA distributions of all traits were normally distributed, centered at 0, uncorrelated, without measurement error, and homogeneous between traits. We used 40 individuals and 10 traits for the baseline simulations. We then modified parameters to examine five levels of heterogeneity between traits, three levels of leptokurtosis, three levels of correlations of FA values between traits, three levels of measurement error, two other sample sizes, and a different number of traits. Levels 1–3 for kurtosis corresponded to average kurtosis values (g) of approximately $g = 0.7$, $g = 2.9$, and $g = 7.6$, respectively. Levels 1–3 for correlations of FA values corresponded to average FA-FA correlations of $r = 0.0052$, $r = 0.039$, and $r = 0.085$, respectively. Each level of each condition was examined at three different effect sizes ($E = 1$, $E = 1.2$, and $E = 1.5$). Effect size (E) was measured as the ratio of FA variances between stressed ($\sigma^2_{2,t}$) and unstressed populations ($\sigma^2_{1,t}$; i.e., $E = \sigma^2_{2,t}/\sigma^2_{1,t}$), where t denotes a trait. When $E = 1$, there was no difference between populations and significance was due to Type I errors. Conditions are discussed in more depth in the text.

Table B1: Type I error rates at $\alpha = 0.05$ ($E = 1$, see text) of CFA analyses at different parameter values

	Type I errors (% significant)						
Conditions/levels	Single trait	CFA 1	CFA 2	CFA 3	CFA 4	CFA 5	CFA 6
Baseline	5.3	4.4	4.5	3.2	5.1	4.4	5.6
Heterogeneous FA distributions:							
1. $\sigma^2_{1,1-5} = 3 \times \sigma^2_{1,6-10}$	4.6	5.3	5.2	3.7	4.1	4.9	5.2
2. $\sigma^2_{1,1} = 3 \times \sigma^2_{1,2}\sigma^2_{1,2} = \sigma^2_{1,3} = \ldots = \sigma^2_{1,10}$	4.7	3.9	4.4	3.0	4.0	4.1	5.3
3. $\sigma^2_{1,4-6} = 3 \times \sigma^2_{1,1-3}$ and $\sigma^2_{1,7-10} = 10 \times \sigma^2_{1,3}$	4.2	5.2	4.5	2.9	4.0	5.2	5.2
4. $\sigma^2_{1,1-5} = 10 \times \sigma^2_{1,6-10}$	4.8	5.1	7.1	4.2	5.8	5.8	5.8
5. $\sigma^2_{1,1} = 10 \times \sigma^2_{1,2,}\sigma^2_{1,2} = \sigma^2_{1,3} = \ldots = \sigma^2_{1,10}$	4.8	4.2	5.0	3.3	4.1	4.3	4.8
Leptokurtism:							
1. $\sigma^2_{p,t}$ ranged from 1 to 3	5.0	6.0	5.8	6.1	6.6	5.9	4.3
2. $\sigma^2_{p,t}$ ranged from 1 to 10	6.6	4.7	4.8	5.9	17.6	5.0	3.8
3. $\sigma^2_{p,t}$ ranged from 1 to 100	4.1	5.8	6.7	6.2	33.7	6.4	2.2
Correlations of FA values:							
1. $C = 0.091$	4.8	4.7	4.6	4.7	4.5	5.0	4.1
2. $C = 0.23$	5.3	5.3	5.3	6.3	6.6	9.0	3.8
3. $C = 0.33$	4.9	6.2	6.3	6.4	9.6	12.2	5.0
Measurement error:							
1. ME = 0.091	4.6	5.3	5.5	5.6	3.7	5.1	4.2
2. ME = 0.23	4.3	5.1	5.2	4.8	4.2	4.9	4.4
3. ME = 0.33	5.2	5.2	5.2	5.2	4.5	5.2	4.8
Number of traits ($k = 5$)	6.3	6.0	5.8	4.7	5.1	5.4	6.1
Sample size:							
1. $N = 80$	5.5	4.5	4.6	3.1	3.9	4.4	4.4
2. $N = 20$ ($k = 5$)	5.4	4.2	4.0	4.3	4.1	3.3	4.6

Table B2: Power (% significant) of CFA analyses at different parameter values

	Power (% significant)						
Conditions/levels	Single trait	CFA 1	CFA 2	CFA 3	CFA 4	CFA 5	CFA 6
Baseline:							
$E = 1.2$	10.2	34.9	35.3	27.4	25.8	35.1	9.4
$E = 1.5$	20.6	85.2	84.8	70.3	74.1	85.1	28.3
Heterogeneous FA distributions:							
$E = 1.2$:							
1	8.6	31.4	31.5	26.8	26.8	32.6	8.5
2	9.7	33.0	33.3	26.7	27.5	33.4	8.7
3	8.9	29.4	31.6	26.4	25.8	30.2	8.5
4	10.8	28.0	31.6	25.8	25.1	27.8	9.0
5	10.7	29.2	34.7	28.3	26.7	29.4	8.3
$E = 1.5$:							
1	19.1	83.5	83.7	72.9	74.6	84.3	29.5
2	19.6	82.8	83.5	69.3	75.0	83.8	27.0
3	19.2	79.0	84.3	71.5	77.4	80.1	29.0
4	21.7	77.4	84.1	72.5	75.3	78.9	28.8
5	20.9	75.4	84.0	70.2	74.7	76.3	27.8
Leptokurtism:							
$E = 1.2$:							
1	8.5	28.0	26.5	21.1	30.5	28.1	7.3
2	5.9	18.6	17.6	18.3	34.2	18.8	5.0
3	4.5	10.2	10.7	16.0	43.3	10.0	2.3
$E = 1.5$:							
1	18.8	75.1	74.4	62.5	73.0	76.4	23.3
2	14.8	49.9	50.5	49.6	60.6	50.9	10.1
3	8.5	20.9	24.4	46.0	59.4	22.6	4.3
Correlations of FA values:							
$E = 1.2$:							
1	10.3	32.6	31.3	26.1	26.4	35.2	8.3
2	11.1	26.8	26.3	21.5	29.0	36.0	7.4
3	8.2	21.0	20.9	18.1	29.4	35.2	5.1
$E = 1.5$:							
1	21.1	85.2	84.0	72.6	74.4	86.9	28.4
2	17.7	70.7	70.1	56.6	70.7	81.2	16.4
3	20.5	61.5	60.7	52.4	70.5	78.5	14.4
Measurement error:							
$E = 1.2$:							
1	9.8	28.3	28.0	22.4	22.6	29.0	6.2
2	9.7	22.1	22.8	17.9	19.8	24.1	6.2
3	8.1	18.5	18.1	14.8	14.8	18.9	6.6
$E = 1.5$:							
1	18.4	78.3	78.0	63.8	69.3	80.4	22.0
2	16.4	69.9	68.2	53.8	58.4	70.6	17.3
3	13.9	58.6	57.3	44.2	47.5	60.0	12.6
Number of traits:							
$E = 1.2$:							
1	8.1	20.9	20.9	16.7	17.3	21.3	7.4
$E = 1.5$:							
1	19.9	59.2	59.2	46.9	55.8	59.7	21.4
Sample size:							
$E = 1.2$:							
1	13.1	51.4	51.5	40.9	49.9	51.3	13.7
2	8.8	14.4	14.1	12.2	11.8	15.4	6.2
$E = 1.5$:							
1	32.7	98.6	98.1	90.7	98.3	98.7	64.6
2	13.2	33.7	32.6	26.8	27.1	36.1	9.2

Note: Condition numbers correspond to condition numbers and descriptions in table B1.

APPENDIX C

Estimated kurtosis values at which analyses using CFA 3 become stronger than CFA 2. We calculated *t*-values using CFA 3 and CFA 2 for 100 simulations. Between simulations, we increased the degree of kurtosis of the FA distribution such that a range of kurtosis values was produced. We subtracted the *t*-values of CFA 2 from the *t*-values of CFA 3 as a measure of the differences in power of these two analyses ($D = t_{\text{CFA 3}} - t_{\text{CFA 2}}$). When D was positive, analyses using CFA 3 were stronger than CFA 2. We regressed D against kurtosis and determined a best-fit line. The point at which the best-fit line intercepted the X-axis (i.e., $D = 0$) indicated the point at which CFA 3 became stronger than CFA 2. However, this was only a rough estimate, as there was a great deal of overlap at intermediate levels of kurtosis. This is illustrated in figure C1, which uses a sample size $N = 40$, number of traits $k = 10$, and effect size $E = 1.5$. The point at which CFA 3 became stronger than CFA 2 was at a kurtosis value of approximately $g = 3.5$. Below, we present a series of tables of kurtosis values at which CFA 3 became stronger than CFA 2. Each value presented is based on 10 X-intercepts (i.e., 1,000 simulations), such that we could determine the variation in our estimates (standard deviation of 10 estimates shown in parentheses). The percent of significant relations (using CFA 2) is also given in square brackets to show the range of powers that we examined. We examined a range of sample sizes (N), effect sizes (E), and number of traits (k; 500,000 simulations in total). Note that kurtosis values are based on signed FA values of single traits, whereas analyses using CFA 2 and CFA 3 are based on $|\text{FA}|$.

Table C1: Kurtosis values based on number of traits $k = 2$

N	$E = 1.1$	$E = 1.2$	$E = 1.3$	$E = 1.4$	$E = 1.5$	$E = 1.6$	$E = 1.7$	$E = 1.8$	$E = 1.9$	$E = 2.0$
10	.85(.31)[5]	.70(.26)[3]	1.21(.41)[7]	.95(.34)[6]	1.00(.33)[7]	.72(.24)[9]	.78(.26)[8]	.94(.31)[10]	.69(.34)[10]	.39(.22)[11]
20	2.62(.80)[5]	2.46(.76)[6]	2.58(.79)[9]	2.37(.73)[9]	1.92(.61)[10]	2.01(.63)[9]	1.66(.52)[13]	1.97(.60)[13]	1.88(.59)[16]	1.74(.54)[14]
30	4.11(1.26)[7]	3.42(1.04)[7]	3.58(1.09)[9]	3.48(1.06)[10]	2.68(.87)[12]	2.94(.91)[14]	3.02(.93)[15]	2.61(.79)[19]	2.54(.78)[17]	2.51(.77)[21]
40	5.46(1.69)[4]	4.40(1.34)[7]	4.47(1.37)[9]	4.22(1.29)[11]	3.59(1.13)[14]	3.37(1.03)[15]	3.70(1.14)[17]	3.33(1.04)[20]	2.91(.90)[21]	2.53(.80)[25]
50	6.18(2.10)[7]	5.45(1.69)[7]	4.97(1.58)[10]	3.74(1.19)[14]	3.95(1.31)[15]	4.20(1.27)[16]	4.00(1.23)[20]	3.40(1.10)[24]	3.35(1.05)[25]	3.39(1.04)[26]
60	7.48(2.28)[7]	5.28(1.66)[7]	4.36(1.45)[10]	5.40(1.67)[12]	4.89(1.49)[17]	4.48(1.37)[20]	3.82(1.26)[20]	3.63(1.13)[28]	3.63(1.12)[29]	3.53(1.13)[31]
70	7.28(2.27)[7]	7.39(2.25)[8]	5.65(1.85)[11]	5.56(1.72)[13]	5.57(1.70)[18]	5.16(1.60)[22]	4.38(1.37)[26]	4.15(1.29)[27]	3.07(.96)[31]	2.44(.83)[34]
80	9.03(2.78)[5]	7.86(2.39)[7]	6.61(2.05)[12]	4.85(1.56)[15]	5.44(1.68)[19]	4.58(1.49)[20]	4.14(1.30)[29]	3.55(1.13)[28]	3.62(1.19)[32]	2.87(1.10)[36]
90	8.45(2.68)[6]	8.90(2.73)[10]	5.86(1.98)[13]	5.51(1.78)[16]	3.85(1.57)[22]	4.64(1.50)[26]	3.31(1.05)[27]	2.99(.99)[30]	2.66(1.23)[38]	3.26(1.02)[39]
100	11.84(3.71)[6]	8.54(2.65)[10]	6.70(2.08)[13]	5.37(1.88)[18]	5.20(1.65)[24]	4.11(1.34)[25]	4.34(1.53)[31]	3.42(1.26)[37]	2.62(1.11)[37]	1.82(1.17)[41]

Table C2: Kurtosis values based on number of traits $k = 3$

N	$E = 1.1$	$E = 1.2$	$E = 1.3$	$E = 1.4$	$E = 1.5$	$E = 1.6$	$E = 1.7$	$E = 1.8$	$E = 1.9$	$E = 2.0$
10	1.00(.36)[5]	1.26(.41)[6]	1.11(.34)[6]	.80(.28)[8]	.76(.26)[8]	.77(.24)[10]	.66(.23)[12]	.60(.24)[12]	.57(.22)[13]	.76(.28)[15]
20	2.76(.85)[6]	2.84(.86)[7]	2.46(.75)[10]	2.08(.66)[9]	2.04(.63)[10]	2.17(.67)[14]	1.84(.56)[16]	2.01(.62)[18]	1.67(.53)[21]	1.68(.51)[20]
30	4.15(1.26)[7]	3.62(1.09)[8]	3.75(1.14)[9]	3.13(.96)[11]	3.09(.95)[15]	2.96(.91)[14]	2.85(.88)[18]	2.41(.76)[21]	2.65(.81)[25]	2.24(.69)[25]
40	5.26(1.60)[5]	4.58(1.40)[9]	4.02(1.24)[10]	3.86(1.18)[14]	3.64(1.11)[18]	3.48(1.08)[17]	3.39(1.05)[21]	3.43(1.05)[23]	3.01(.94)[31]	3.12(.95)[30]
50	5.84(1.78)[6]	5.76(1.75)[9]	5.30(1.64)[12]	5.08(1.56)[16]	4.10(1.28)[17]	4.15(1.31)[21]	3.45(1.08)[25]	3.15(.99)[30]	3.22(1.00)[33]	2.71(.85)[35]
60	6.59(2.02)[7]	6.95(2.12)[10]	4.74(1.50)[13]	5.18(1.58)[16]	4.47(1.41)[20]	4.09(1.30)[23]	4.35(1.36)[31]	3.45(1.06)[31]	3.46(1.06)[36]	2.91(.94)[37]
70	7.84(2.38)[7]	6.80(2.18)[10]	6.16(1.88)[18]	4.72(1.47)[18]	5.18(1.59)[24]	3.71(1.33)[24]	3.60(1.13)[31]	3.47(1.12)[37]	4.16(1.28)[38]	3.79(1.17)[43]
80	7.83(2.44)[8]	6.79(2.11)[11]	6.80(2.08)[16]	5.61(1.76)[21]	4.37(1.53)[24]	3.66(1.23)[28]	3.92(1.25)[35]	2.90(.93)[40]	3.35(1.27)[44]	3.16(1.02)[46]
90	8.48(2.66)[6]	7.13(2.26)[11]	6.16(1.91)[15]	6.24(1.91)[21]	5.08(1.56)[25]	5.43(1.70)[33]	4.76(1.53)[34]	3.91(1.37)[43]	3.34(1.15)[47]	2.42(.88)[49]
100	10.07(3.07)[9]	7.88(2.43)[12]	6.35(2.11)[16]	6.44(1.98)[23]	4.09(1.41)[26]	3.15(1.19)[31]	3.56(1.18)[36]	2.28(.89)[43]	3.66(1.16)[50]	2.15(.96)[52]

Table C3: Kurtosis values based on number of traits $k = 4$

N	*E* = 1.1	*E* = 1.2	*E* = 1.3	*E* = 1.4	*E* = 1.5	*E* = 1.6	*E* = 1.7	*E* = 1.8	*E* = 1.9	*E* = 2.0
10	1.35(.45)[6]	1.32(.41)[5]	.92(.29)[7]	.62(.31)[7]	.95(.32)[10]	.73(.23)[11]	.75(.23)[12]	.69(.23)[11]	.59(.21)[14]	.67(.28)[17]
20	2.79(.85)[5]	2.61(.79)[7]	2.53(.77)[10]	2.35(.71)[12]	2.26(.69)[15]	2.16(.66)[15]	1.90(.58)[19]	1.75(.55)[20]	1.60(.49)[23]	1.64(.51)[25]
30	4.28(1.31)[7]	3.94(1.21)[10]	3.23(1.00)[12]	3.48(1.06)[13]	2.89(.89)[17]	3.29(1.00)[19]	2.87(.87)[23]	2.82(.85)[29]	2.33(.72)[28]	2.27(.70)[30]
40	5.30(1.61)[5]	4.63(1.41)[10]	4.29(1.30)[12]	4.08(1.24)[17]	3.83(1.16)[19]	3.67(1.13)[24]	3.52(1.08)[26]	3.25(.98)[30]	2.91(.90)[34]	2.46(.79)[38]
50	6.54(2.00)[7]	5.64(1.72)[10]	5.10(1.59)[12]	4.42(1.37)[20]	4.25(1.33)[20]	4.35(1.34)[25]	4.13(1.26)[32]	3.94(1.19)[34]	3.27(1.01)[38]	3.02(.92)[42]
60	6.77(2.07)[7]	6.09(1.87)[11]	5.62(1.71)[15]	5.37(1.66)[19]	4.57(1.41)[24]	4.22(1.31)[29]	4.49(1.37)[35]	3.84(1.19)[38]	3.87(1.18)[44]	3.34(1.04)[48]
70	7.98(2.46)[9]	6.16(1.91)[10]	6.41(1.99)[18]	4.84(1.49)[20]	5.11(1.57)[27]	4.25(1.36)[31]	3.93(1.24)[36]	4.17(1.29)[43]	3.22(1.06)[47]	3.69(1.16)[50]
80	8.79(2.68)[9]	6.26(1.98)[12]	5.57(1.89)[18]	5.98(1.86)[22]	5.52(1.70)[29]	4.54(1.47)[36]	4.29(1.35)[41]	3.92(1.26)[44]	2.73(.87)[51]	3.10(1.00)[55]
90	9.30(2.86)[9]	6.92(2.16)[14]	5.11(1.82)[19]	6.44(1.99)[24]	4.91(1.59)[32]	4.60(1.44)[39]	4.51(1.44)[44]	4.14(1.32)[50]	3.09(1.00)[55]	2.53(.86)[59]
100	8.62(2.64)[8]	6.67(2.13)[15]	6.94(2.17)[22]	5.03(1.55)[26]	4.85(1.63)[32]	4.02(1.29)[38]	4.26(1.34)[46]	3.38(1.05)[52]	3.27(1.05)[53]	2.13(.81)[61]

Table C4: Kurtosis values based on number of traits $k = 5$

N	*E* = 1.1	*E* = 1.2	*E* = 1.3	*E* = 1.4	*E* = 1.5	*E* = 1.6	*E* = 1.7	*E* = 1.8	*E* = 1.9	*E* = 2.0
10	1.14(.35)[4]	1.07(.39)[7]	.95(.33)[10]	.98(.30)[10]	.80(.30)[8]	.79(.27)[13]	.64(.22)[14]	.80(.26)[16]	.67(.23)[15]	.58(.19)[20]
20	2.51(.77)[6]	2.39(.74)[7]	2.38(.72)[10]	2.36(.72)[13]	1.97(.60)[13]	2.05(.62)[16]	1.92(.58)[23]	1.56(.49)[23]	1.87(.58)[25]	1.74(.53)[30]
30	4.36(1.33)[7]	3.71(1.12)[8]	3.71(1.12)[14]	3.12(.96)[15]	3.15(.97)[18]	3.01(.93)[23]	3.03(.92)[25]	2.66(.81)[30]	2.45(.75)[32]	2.25(.70)[37]
40	5.45(1.66)[6]	4.88(1.49)[13]	4.38(1.32)[14]	4.25(1.29)[16]	3.95(1.21)[22]	3.48(1.06)[26]	3.34(1.02)[32]	2.76(.85)[34]	2.77(.86)[38]	2.74(.86)[45]
50	6.01(1.83)[7]	5.59(1.70)[11]	5.10(1.56)[17]	4.60(1.40)[20]	4.07(1.26)[24]	3.68(1.14)[32]	3.64(1.11)[34]	3.63(1.11)[39]	3.09(.95)[43]	2.95(.90)[48]
60	6.70(2.07)[7]	6.51(2.00)[11]	6.40(1.95)[19]	4.61(1.43)[22]	4.17(1.31)[26]	4.11(1.25)[34]	3.93(1.20)[38]	3.62(1.12)[44]	3.19(.98)[47]	3.52(1.08)[54]
70	7.71(2.55)[8]	6.45(1.96)[11]	6.26(1.92)[20]	5.31(1.64)[22]	4.52(1.43)[31]	4.35(1.37)[35]	4.03(1.25)[41]	3.28(1.02)[49]	3.50(1.13)[49]	2.84(.99)[57]
80	6.67(2.11)[8]	6.78(2.09)[13]	5.86(1.80)[19]	5.51(1.72)[27]	5.03(1.53)[32]	4.60(1.42)[41]	3.58(1.26)[47]	3.56(1.12)[52]	3.69(1.15)[60]	3.50(1.10)[61]
90	7.74(2.45)[8]	7.76(2.40)[15]	6.18(1.92)[21]	5.05(1.60)[28]	5.39(1.67)[35]	5.59(1.73)[44]	4.78(1.46)[50]	3.46(1.12)[55]	3.44(1.09)[64]	2.44(.86)[63]
100	8.38(2.58)[9]	6.42(2.11)[14]	5.99(1.84)[22]	5.55(1.72)[30]	5.51(1.70)[36]	4.61(1.41)[43]	4.11(1.37)[51]	2.90(.98)[56]	2.57(.98)[64]	3.21(1.02)[70]

Table C5: Kurtosis values based on number of traits $k = 10$

N	*E* = 1.1	*E* = 1.2	*E* = 1.3	*E* = 1.4	*E* = 1.5	*E* = 1.6	*E* = 1.7	*E* = 1.8	*E* = 1.9	*E* = 2.0
10	1.21(.38)[7]	1.06(.33)[8]	.94(.29)[11]	.93(.29)[14]	.89(.28)[16]	.88(.28)[18]	.46(.38)[20]	.85(.26)[20]	.67(.21)[26]	.54(.19)[28]
20	2.66(.81)[6]	2.51(.76)[10]	2.25(.69)[14]	2.31(.70)[18]	2.09(.64)[19]	1.79(.55)[26]	2.04(.62)[30]	1.84(.56)[34]	1.84(.56)[41]	1.80(.55)[42]
30	3.75(1.14)[7]	3.57(1.10)[12]	3.29(1.00)[16]	2.71(.84)[21]	3.34(1.01)[29]	2.84(.86)[32]	2.60(.79)[38]	2.59(.79)[46]	2.40(.74)[49]	2.06(.63)[54]
40	4.89(1.49)[8]	4.49(1.36)[15]	4.12(1.25)[20]	4.01(1.21)[25]	3.43(1.06)[34]	3.35(1.02)[41]	3.12(.98)[47]	2.98(.91)[52]	3.00(.91)[60]	2.82(.87)[63]
50	5.90(1.79)[8]	5.79(1.77)[13]	4.42(1.35)[21]	3.95(1.20)[30]	4.20(1.27)[37]	3.91(1.19)[47]	3.52(1.07)[52]	3.62(1.10)[61]	3.11(.96)[65]	3.05(.94)[69]
60	6.47(1.98)[10]	5.72(1.74)[17]	5.04(1.55)[22]	4.90(1.49)[34]	4.33(1.33)[41]	4.07(1.25)[50]	3.89(1.19)[60]	3.59(1.12)[62]	3.37(1.03)[71]	3.60(1.09)[77]
70	7.14(2.17)[9]	6.36(1.95)[18]	5.35(1.64)[26]	4.98(1.52)[40]	4.39(1.37)[48]	3.93(1.22)[56]	3.53(1.10)[63]	3.58(1.09)[69]	3.16(.98)[76]	3.64(1.13)[80]
80	6.55(2.14)[10]	5.77(1.78)[20]	5.09(1.56)[28]	4.70(1.44)[40]	4.45(1.36)[49]	4.04(1.24)[56]	4.05(1.25)[67]	3.75(1.17)[74]	3.13(.97)[78]	2.81(.92)[83]
90	8.78(2.68)[12]	6.16(1.96)[20]	5.71(1.74)[32]	4.70(1.54)[43]	4.88(1.50)[53]	4.26(1.33)[63]	3.85(1.19)[69]	3.37(1.05)[77]	3.21(.99)[82]	3.41(1.08)[86]
100	8.67(2.67)[12]	5.55(1.94)[22]	4.89(1.58)[35]	5.31(1.62)[46]	4.67(1.44)[56]	4.42(1.36)[67]	4.05(1.25)[73]	3.94(1.22)[77]	3.59(1.13)[85]	3.20(1.01)[89]

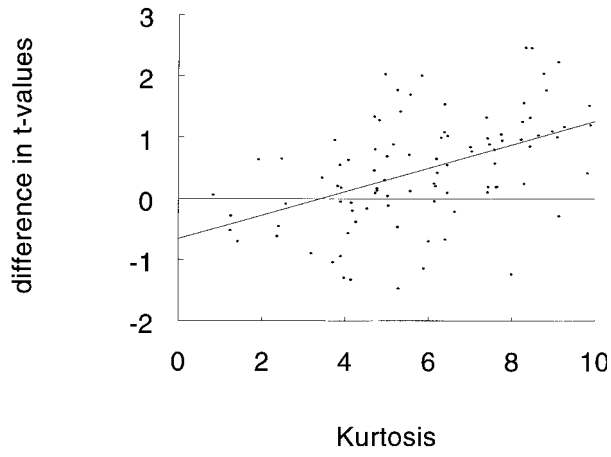

Figure C1: The difference in power of CFA 2 and CFA 3. We subtracted the *t*-values of CFA 2 from *t*-values of CFA 3 as a measure of the differences in power of these two analyses ($D = t_{CFA3} - t_{CFA2}$). We regressed D against kurtosis and determined a best-fit line. The point at which the best-fit line intercepted the X-axis (i.e., $D = 0$) indicated the point at which CFA 3 became stronger than CFA 2. However, note the high degree of overlap

Literature Cited

Alados, C. L., J. Escos, and J. M. Emlen. 1993. Developmental instability as an indicator of environmental stress in the Pacific hake (*Merluccius productus*). U.S. National Marine Fisheries Service: Fishery Bulletin 91:587–593.

Bunn, S. E. 1995. Biological monitoring of water quality in Australia: workshop summary and future directions. Australian Journal of Ecology 20:220–227.

Clarke, G. M. 1993*a*. Fluctuating asymmetry of invertebrate populations as a biological indicator of environmental quality. Environmental Pollution 82:207–211.

———. 1993*b*. Patterns of developmental stability of *Chrysopa perla L.* in response to environmental pollution. Environmental Entomology 22:1362–1366.

Clarke, G. M., and L. J. McKenzie. 1992. Fluctuating asymmetry as a quality control indicator for insect mass rearing processes. Journal of Economic Entomology 85: 2045–2050.

Dufour, K. J., and P. J. Weatherhead. 1996. Estimation of organism-wide asymmetry in red-winged blackbirds and its relation to studies of mate selection. Proceedings of the Royal Society of London B, Biological Sciences 263:769–775.

Gangestad, S. W., and R. Thornhill. 1998. The analysis of fluctuating asymmetry redux—the robustness of parametric statistics. Animal Behaviour 55:497–501.

Gummer, D. L., and R. M. Brigham. 1995. Does fluctuating asymmetry reflect the importance of traits in little brown bats (*Myotis lucifugus*)? Canadian Journal of Zoology 73:990–992.

Houle, D. 1997. Comment on "A meta-analysis of the heritability of developmental stability" by Møller and Thornhill. Journal of Evolutionary Biolology 10:17–20.

Koehn, R. K., and B. L. Bayne. 1989. Towards a physiological and genetical understanding of the energetics of the stress response. Biological Journal of the Linnean Society 37:157–171.

Leary, R. F., and F. W. Allendorf. 1989. Fluctuating asymmetry as an indicator of stress: implications for conservation biology. Trends in Ecology & Evolution 4: 214–216.

Leung, B. 1998. Controlling for allometry in studies of fluctuating asymmetry and quality within samples. Proceedings of the Royal Society of London B, Biological Sciences 265:1623–1629.

Leung, B., and M. R. Forbes. 1996. Fluctuating asymmetry in relation to stress and fitness: effects of trait type as revealed by meta-analysis. Ecoscience 3:400–413.

———. 1997. Modeling fluctuating asymmetry in relation to stress and fitness. Oikos 78:397–405.

Lindeman, R. H., P. F. Merenda, and R. Z. Gold. 1980. Introduction to bivariate and multivariate analysis. Scott Foresman, Glenview, Ill.

Manning, J. T., and L. Ockenden. 1994. Fluctuating asymmetry in racehorses. Nature (London) 370:185–186.

Møller, A. P., and J. Höglund. 1991. Patterns of fluctuating asymmetry in avian feather ornaments: implications for models of sexual selection. Proceedings of the Royal Society of London B, Biological Sciences 245:1–5.

Palmer, A. R. 1994. Fluctuating asymmetry analysis: a primer. Pages 335–364 *in* T. A. Markow, ed. Develop-

mental instability: its origins and evolutionary implications. Kluwer, Dordrecht.

Palmer, A. R., and C. Strobeck. 1986. Fluctuating asymmetry: measurement, analysis and patterns. Annual Review of Ecology and Systematics 17:391–421.

———. 1992. Fluctuating asymmetry: implications of non-normality. Acta Zoologica Fennica 191:57–72.

Rice, W. R., and S. D. Gaines. 1994. Heads I win, tails you lose—testing directional alternative hypotheses in ecological and evolutionary research. Trends in Ecology & Evolution 9:235–237.

Rowe, L., R. R. Repasky, and A. R. Palmer. 1997. Size-dependent asymmetry: fluctuating asymmetry versus antisymmetry and its relevance to condition dependent signaling. Evolution 51:1401–1408.

Sommer, C. 1996. Ecotoxicology and developmental stability as an in situ monitor of adaptation. Ambio 25: 374–376.

StatSoft. 1993. Statistica for Windows 4.3D. StatSoft, Tulsa, Okla.

Thornhill, R., S. W. Gangestad, and R. Comer. 1995. Human female orgasm and mate fluctuating asymmetry. Animal Behaviour 50:1601–1615.

Watson, P. J., and R. Thornhill. 1994. Fluctuating asymmetry and sexual selection. Trends in Ecology & Evolution 9:21–25.

Whitlock, M. 1993. Lack of correlation between heterozygosity and fitness in forked fungus beetles. Heredity 70:574–581.

Zakharov, V. M., E. Pankakoski, B. I. Sheftel, A. Peltonen, and I. Hanski. 1991. Developmental stability and population dynamics in the common shrew, *Sorex araneus*. American Naturalist 138:797–810.

Zar, J. H. 1984. Biostatistical analysis. 2d ed. Prentice Hall, Upper Saddle River, N.J.

Zhivotovsky, L. A. 1988. Some methods of analysis of correlated characters. Pages 423–432 *in* B. S. Weir, E. J. Eisen, M. M. Godman, and G. Namkoong, eds. Proceedings of the Second International Conference on Quantitative Genetics. Sinauer, Sunderland, Mass.

———. 1992. A measure of fluctuating asymmetry for a set of characters. Acta Zoologica Fennica 191:73–77.

Associate Editor: Gregory A. Wray