# Appendix 2: a worked example (Corrected Jan. 13, 2009)

This is Appendix 2 from Hansen, T. F., and D. Houle. 2008. Measuring and comparing evolvability and constraint in multivariate characters. J. Evol. Biol. 21:1201-1219. John Stinchcombe brought several errors in this Appendix to our attention, and this version corrects them. Changes are highlighted in yellow. Most importantly, the $\beta_3$ value used for all calculations was given incorrectly, and was 0.01, rather than 0.1. In addition, calculation errors are corrected in the respondabilities and response differences in Table A4, and the conditional evolvabilities in Table A5. Several rounding errors have also been corrected, but not highlighted. Corrections made on Jan. 13, 2009 by David Houle.

To illustrate how our measures of evolvability are calculated, and how they can be interpreted, consider the two hypothetical three-trait **G** matrices in Table A1. We chose this simple example to represent some typical problems in inferring and comparing evolvability from **G** matrices. Traits 1 and 2 represent lengths of morphological features, and trait 3 represents the life-history trait fecundity. Comparison of the matrices themselves does not immediately suggest how each population will respond to selection. The diagonals suggests that traits 1 and 3 might respond more rapidly to selection in population 2 and that trait 2 would respond better in population 1. The degree of correlation among traits seems a bit higher in population 2 than population 1. It is clear, however, that more than a glance is necessary to ascertain which population would evolve more rapidly under particular circumstances. Furthermore, note that the units are not commensurate across all the traits, so the raw numerical values cannot sensibly be compared when the directions of response are not the same. In addition, note that traits and populations vary in the relationships between trait means and variances, so the appropriate standardization for each matrix is different.

Table A2 shows the example matrices standardized by trait means (**G**$_\mu$), trait variances (**G**$_\sigma$), and the square root of the **P** matrix (**G**$_P$). The diagonal of **G**$_\mu$ are $I_A$ values of each trait. The diagonals of **G**$_\sigma$ are the heritabilities of each trait.

Table A3 gives some selection-response statistics when a single gradient is applied to both of the example populations. The selection gradient giving change in relative fitness per unit change in trait is **β'** = [0.005/mm, –0.001/mm, 0.01/egg]. These values were chosen to yield standardized **β** values that are in line with typical standardized strengths of selection (Hereford *et al.,* 2004). Table A4 shows the evolvability statistics developed in this paper, which are based on response to a **β** in the same direction, but standardized to length 1 on whichever scale the parameter estimates are on. Note that this means that the 'standard' strength of selection is different for each standardization.

The first section of each table gives the evolvability statistics for unstandardized data, where the units are a mixture of egg numbers and millimeters. Although the statistics are readily calculated, we see no useful interpretation of any of our statistics on this dog's-breakfast scale. The dimensionless ratios of evolvabilities, *e*, and conditional evolvabilities, *c*, shown in the 'compare' column in A4, however, do have value in expressing the relative progress possible under selection. We do not show the ratio of respondabilities, *r*, as these are measured in different directions and are therefore not comparable. If the gradient **β** had included only one non-zero element, indicating that all the selection was on a single trait, the fact that the *e* and *c* each summarize only the response in the selection direction would give them the units of the single selected trait, and these values could be interpreted. The value of the response difference, *d*, is difficult to interpret, as it is a distance along a different direction in phenotype space from **β** and thus has different units.

The interpretability of these statistics increases on a mean-standardized scale. The individual elements of the response vector shown in Table A3 are in proportions of the mean of each trait. In this coordinate system, the selection gradient results in an average unconstrained change of 1.3% in population 1 and 2.9% in population 2 in the direction of $\boldsymbol{\beta}$. The respondability, $r$, in population 1 is thus 43% of that in population 2. Turning to evolvability, $e$, if the mean-standardized selection gradient had been of unit length in each population, corresponding to a strength of selection equal to that on fitness, and no stabilizing selection occurred, then the response would have been 1.2% in population 1 and 2.4% in population 2. The evolvability in population 1 would have been 50% of that in population 2. The ratio of the projections of $\Delta\bar{\mathbf{z}}$ on $\boldsymbol{\beta}$ does not equal the ratio of the evolvabilities because populations 1 and 2 have different means, so standardizing the selection gradients by their own means changes the relative size of the selection gradients applied. This suggests that it may sometimes be useful to standardize both gradients and **G** by common values, as outlined in the text for the expected response difference, $\bar{d}$.

On the mean scale, autonomies, $a(\beta)$, the ratio of conditional evolvability to evolvability, are 39% in population 1 but only 1.2% in population 2. The integration values, $i(\beta)$, are 1- $a(\beta)$, so population 1 is 61% integrated and population 2 is 98.8% integrated in these directions. Despite the larger unconstrained evolvability of population 2 in direction $\boldsymbol{\beta}$, evolution would therefore be much more constrained by stabilizing selection on the remaining traits in population 2 than in 1. The conditional response is nearly 16 times as large in population 1 as in population 2. This is reflected by the difference in the angle of the unconstrained response relative to $\boldsymbol{\beta}$, 18° in population 1 and 36° in population 2. When the other traits are under stabilizing selection, this increased deflection will be counteracted and the constrained response reduced. The angle between the response vectors in the two populations is 18°. When we standardize by the average of the mean vectors response difference, $d(\beta)$, is 1.0%, which is a substantial proportion of the direct responses.

On the variance-standardized scale, elementwise standardization places the individual elements of the response vectors in standard-deviation units. Respondability, $r$, is 21% of a standard deviation in this coordinate system in population 1 and 22% of a standard deviation in population 2 in direction $\boldsymbol{\beta}$. The ratio of the projections of $\boldsymbol{\Delta \bar{z}}$ on $\boldsymbol{\beta}$ does not equal the ratio of the evolvabilities because of the different standardizations employed in the two populations. The autonomies, $a(\beta)$, on the standard-deviation scale are higher in both populations than those on the mean-standardized scale, particularly in population 2, with the result that conditional evolvabilities, $c(\beta)$, are very similar. The angles between the responses and $\boldsymbol{\beta}$ are higher in population 1 than on the mean-standardized scale, as is the angle between response vectors, $\theta_d$. When we use elementwise variance standardization, the response difference, $d(\beta)$ , is 17% of a standard deviation, about as large as the direct responses.

Standardization with the square root of the **P** matrix places the lengths of response vectors and evolvabilities in standard deviation units appropriate to their direction. For example, the evolvability, $e(\beta)$, in population 1 is 9.5% of the phenotypic standard deviation in direction $\boldsymbol{\beta}$. The oblique transformation of the coordinate system makes the elements of the response vectors difficult to interpret. In this case, **P** standardization results in similar vectors and scalar measures of evolvability to $\boldsymbol{\sigma}$ standardization.

The many differences between the statistics calculated on different scales make clear that the choice of scale can strongly influence the results. Each standardization gives a unique weighting of the traits that stretches or compresses each of the bases of phenotype space to a different degree. In addition, the square-root-of-**P** transformation also performs an oblique rotation of the bases.

Finally, we can compare the evolvability statistics over the entire phenotype space. Table A5 shows the mean evolvability, $\overline{e}$ ,conditional evolvability, $\overline{c}$ , respondability, $\overline{r}$ , and autonomy, $\overline{a}$ , values for the two hypothetical populations. The average unconditional evolvability, $\overline{e}$ , on a mean-standardized scale is 0.5% in population 1 and 1.4% in population 2. The average conditional evolvability, $\overline{c}$ , is 0.15% of the mean in population 1 but just 0.03% in population 2. This difference is reflected in the lower autonomies, $\overline{a}$ , in population 2. The raw and standard-deviation scales show a similar pattern, in which population 2 is more unconditionally evolvable but also more constrained than population 1. The key cause of this result is that the eigenvalues of **G** matrix 2 are more uneven than those in **G** matrix 1.

Table A1 Example **G** and the residual matrix **E = P – G** matrices and trait means.

| Population | Trait (Units) | **G** | **E** | $\bar{z}$ |
|---|---|---|---|---|
| 1 | 1 (mm)<br>2 (mm)<br>3 (eggs) | $\begin{bmatrix}10 & 10 & 20\\ & 30 & 20\\ & & 80\end{bmatrix}$ | $\begin{bmatrix}10 & 13 & 50\\ & 30 & 40\\ & & 890\end{bmatrix}$ | $\begin{bmatrix}73\\ 138\\ 82\end{bmatrix}$ |
| 2 | 1 (mm)<br>2 (mm)<br>3 (eggs) | $\begin{bmatrix}20 & 16 & -10\\ & 20 & 20\\ & & 150\end{bmatrix}$ | $\begin{bmatrix}20 & 20 & 20\\ & 50 & 100\\ & & 600\end{bmatrix}$ | $\begin{bmatrix}80\\ 152\\ 64\end{bmatrix}$ |

Table A2 Example **G** matrices from Table A1 standardized by trait means and variances.

| Population | $\mathbf{G}_{\mu} \times 100$ | $\mathbf{G}_{\sigma}$ | $\mathbf{G}_{P}$ |
|---|---|---|---|
| 1 | $\begin{bmatrix} 0.188 & 0.099 & 0.334 \\ & 0.158 & 0.177 \\ & & 1.190 \end{bmatrix}$ | $\begin{bmatrix} 0.500 & 0.289 & 0.144 \\ & 0.500 & 0.083 \\ & & 0.082 \end{bmatrix}$ | $\begin{bmatrix} 0.602 & -0.086 & 0.084 \\ & 0.536 & 0.023 \\ & & 0.067 \end{bmatrix}$ |
| 2 | $\begin{bmatrix} 0.313 & 0.132 & -0.195 \\ & 0.087 & 0.206 \\ & & 3.662 \end{bmatrix}$ | $\begin{bmatrix} 0.500 & 0.302 & -0.058 \\ & 0.286 & 0.087 \\ & & 0.200 \end{bmatrix}$ | $\begin{bmatrix} 0.424 & 0.154 & -0.105 \\ & 0.188 & 0.016 \\ & & 0.196 \end{bmatrix}$ |

Table A3 Standardized selection gradients and response vectors for example populations in Table A1 on the raw and three standardized scales.

| Pop. | | $\Delta\bar{\mathbf{z}}$ | $\boldsymbol{\beta}_{\mu}$ | $\Delta\bar{\mathbf{z}}_{\mu}$ | $\boldsymbol{\beta}_{\sigma}$ | $\Delta\bar{\mathbf{z}}_{\sigma}$ | $\boldsymbol{\beta}_{P}$ | $\Delta\bar{\mathbf{z}}_{P}$ |
|---|---|---|---|---|---|---|---|---|
| 1 | vector | $\begin{bmatrix}0.24\text{ mm}\\0.22\text{ mm}\\0.88\text{ eggs}\end{bmatrix}$ | $\begin{bmatrix}0.37\\-0.14\\0.82\end{bmatrix}$ | $\begin{bmatrix}0.0033\\0.0016\\0.0107\end{bmatrix}$ | $\begin{bmatrix}0.022\\-0.008\\0.311\end{bmatrix}$ | $\begin{bmatrix}0.054\\0.028\\0.028\end{bmatrix}$ | $\begin{bmatrix}0.036\\0.016\\0.319\end{bmatrix}$ | $\begin{bmatrix}0.047\\0.013\\0.025\end{bmatrix}$ |
| | length | 0.938 | 0.908 | 0.011 | 0.312 | 0.067 | 0.321 | 0.055 |
| 2 | vector | $\begin{bmatrix}-0.02\text{ mm}\\0.26\text{ mm}\\1.43\text{ eggs}\end{bmatrix}$ | $\begin{bmatrix}0.40\\-0.15\\0.64\end{bmatrix}$ | $\begin{bmatrix}-0.0002\\0.0017\\0.0223\end{bmatrix}$ | $\begin{bmatrix}0.032\\-0.008\\0.274\end{bmatrix}$ | $\begin{bmatrix}-0.003\\0.031\\0.052\end{bmatrix}$ | $\begin{bmatrix}0.025\\0.042\\0.268\end{bmatrix}$ | $\begin{bmatrix}-0.011\\0.016\\0.051\end{bmatrix}$ |
| | length | 1.454 | 0.770 | 0.022 | 0.276 | 0.061 | 0.273 | 0.054 |

The selection gradient is $\boldsymbol{\beta}'$ = [0.005/mm, –0.001/mm, 0.01/egg], and the length (norm) of this vector is 0.011 in a combination of egg and mm units.

Table A4 Evolvability statistics for the trait $\boldsymbol{\beta}'$ = [0.005/mm, –0.001/mm, 0.01/egg].

| | Standardization | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | None[*] | | | Mean | | | Standard deviation | | | Square root of **P** | | |
| | Population | | | Population | | | Population | | | Population | | |
| Statistic | 1 | 2 | compare | 1 | 2 | compare | 1 | 2 | compare | 1 | 2 | compare |
| $r(\boldsymbol{\beta})$ | 84 | 129 | na[†] | 0.0125 | 0.0291 | 0.43 | 0.214 | 0.221 | 0.97 | 0.170 | 0.199 | 0.85 |
| $e(\boldsymbol{\beta})$ | 78 | 111 | 0.70 | 0.0119 | 0.0236 | 0.50 | 0.100 | 0.184 | 0.55 | 0.095 | 0.188 | 0.50 |
| $c(\boldsymbol{\beta})$ | 29 | 3.8 | 7.67 | 0.0046 | 0.0003 | 15.91 | 0.043 | 0.038 | 1.14 | 0.056 | 0.158 | 0.35 |
| $a(\boldsymbol{\beta})$ | 0.38 | 0.03 | | 0.39 | 0.01 | | 0.43 | 0.21 | | 0.59 | 0.84 | |
| $\theta$[‡] | 22 | 31 | 16 | 18 | 36 | 15 | 62 | 34 | 51 | 56 | 19 | 78 |
| $d(\boldsymbol{\beta})$[§] | | | 54 | | | 0.010 | | | 0.174 | | | 0.234 |

The ‘compare’ column compares the responses in the two populations. For the respondabilities, $r(\boldsymbol{\beta})$, and evolvabilities, $e(\boldsymbol{\beta})$ and $c(\boldsymbol{\beta})$, the comparison is the ratio of the value in population 1 to that in population 2, when each population is standardized with its own vector or matrix. For $\theta$ and $d(\boldsymbol{\beta})$, both populations are standardized by the average of the standardization vectors or matrices in the two populations.

[*] The units for the responses of each population are a mixture of mm and eggs, and therefore most of these statistics have no clear interpretation.

[†] The ratio of respondabilities is meaningless on the raw scale.

[‡] In the columns labeled 1 and 2, this is the angle between $\boldsymbol{\beta}$ and $\boldsymbol{\Delta\bar{z}}$. In the 'compare' column it is the angle between the response vectors in the two populations, $\theta_d$.

[§] Response differences were calculated from a standard length $\boldsymbol{\beta}$ under each standardization. In the original paper we calculated response differences using the unstandardized $\boldsymbol{\beta}$**.**

Table A5  Expectations of evolvability statistics over a uniform distribution of selection gradients in the entire phenotype space for the hypothetical populations.

| | Standardization | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | None[a] | | Mean | | Variance | | **P** | |
| Statistic | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 |
| $\overline{e}$ | 40.00 | 63.33 | 0.0051 | 0.0135 | 0.361 | 0.329 | 0.402 | 0.269 |
| $\overline{r}$ | 51.08 | 83.66 | 0.0070 | 0.0194 | 0.456 | 0.410 | 0.463 | 0.310 |
| $\overline{c}$ | 13.20 | 4.64 | 0.0015 | 0.0003 | 0.131 | 0.060 | 0.189 | 0.180 |
| $\overline{a}$ | 0.454 | 0.101 | 0.460 | 0.034 | 0.495 | 0.243 | 0.575 | 0.832 |

[a] The units for the responses of each population are a mixture of millimeters and eggs, and these statistics therefore have no clear interpretation.