Comment on "Evolutionary Trade-Offs, Pareto Optimality, and the Geometry of Phenotype Space"

Pim Edelaar[1,2]

Shoval et al. (Reports, 1 June 2012, p. 1157) showed how configurations of phenotypes may identify tasks that trade off with each other, using randomizations assuming independence of data points. I argue that this assumption may not be correct for most and possibly all examples and led to pseudoreplication and inflated significance levels. Improved statistical testing is necessary to assess how the theory applies to empirical data.
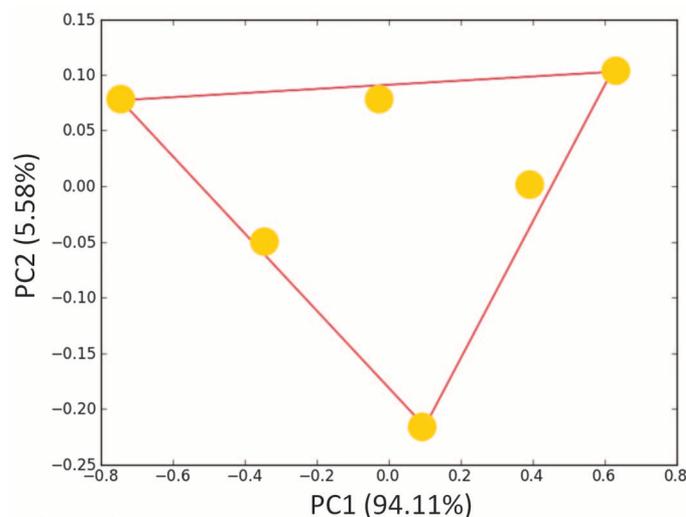
Shoval et al. (1) presented empirical examples to support their interesting argument that the distribution of phenotypes in multivariate space can be used to identify key tasks that trade off with each other. For example, a trade-off between two tasks is identified by phenotypes falling along a line in x-y space, whereas phenotypes trading off three functions would fall within a triangle. Statistical support was presented by comparing the degree to which actual data points form a line or a triangle with that of random configurations obtained by drawing x and y values independently from the same cumulative distribution as the actual data (see the supplementary materials of Shoval et al.).

In the example of the Darwin's finches, the randomization is performed over 120 data points under a null hypothesis that x and y values are independent. However, these data were composed of separate mean values for males and females from each of 60 populations. Subsequently randomizing male and female values independently is incorrect, because it is unlikely that male and female phenotypes will ever evolve fully independently within a population, even in the absence of performance trade-offs. In Shoval et al. (figure S6B), not surprisingly, populations of the same species resemble each other more than populations from different species. Hence, randomizing these values as if they are fully independent is incorrect because this assumes that populations were completely free to evolve (i.e., species have on average identical phenotypes), whereas the data strongly suggest that populations of the same species resemble each other—e.g., due to gene flow, limited time since divergence, or interspecific competition.

Thus, the data have a hierarchical structure (sexes within populations, populations within species), and any randomization test used should focus on the level relevant for the hypothesis, in this case species differences. This effectively reduces the sample size from 120 to 6 (or even less; see below). This is important, because the probability of finding a high degree of triangularity increases as sample size decreases: Three points will always form a triangle (except in the unlikely case that they fall exactly on a line). Using the data set and software provided by Shoval et al., I calculated the six species mean values and randomized these 10,000 times to try to estimate in how many cases a random configuration of six data points could reach degrees of triangularity greater than or equal to the observed data (2). Unfortunately and erroneously, the software provided by Shoval et al. only counts the number of random configurations that have a degree of triangularity greater than the observed data (which already have the maximum degree of triangularity; see Fig. 1) and not those that have equal degrees. The probability that chance alone can result in the observed degree of triangularity of the six Darwin's finches (Fig. 1) therefore remains to be determined, but it is clear that the reported $P < 10^{-4}$ is incorrect and far too small. Furthermore, the first principal component axis already explains >90% of the variance in the data (Fig. 1) and the second axis is not significant (2), so one could even argue that the test tries to explain variability that a priori does not need any explanation.

Nonindependence at a higher taxonomic level has likely inflated the significance of the test for triangularity in morphology of bats. Here, Shoval et al. report that body mass and wing aspect ratio of 108 species form a triangle associated with a $P < 0.03$. However, this P value is based on a null hypothesis that species evolved their morphologies independently with respect to those of other species. This is often not the case, because more closely related species typically resemble each other more than more distantly related species do (3, 4). The degree of phylogenetic signal present in the data limits the degree of independence among data points, and neglecting this aspect can lead to spurious results and inflated significance (3, 4). The bat data suggests that phylogenetic signal is present: Fig. 2 (5) shows how bat species of the same family are not randomly distributed in morphospace. Incorporating nonindependence reduces the degree to which data points should be allowed to move independently when randomized and reduces the effective sample size (3), which increases P values and likely would result in a nonsignificant result here. The lack of control for phylogenetic nonindependence likely affects the comparisons of bats, Darwin's finches, and mice. In comparative studies, it is customary to take phylogenetic nonindependence into account or otherwise to show that it is absent (3, 4).

Other sources of nonindependence may also occur. The bacterial expression levels may show temporal autocorrelation that would have to be taken into account in randomizations (2). Nonindependence may also have arisen from multiple tests of the same hypothesis within the same system (2). For example, relative poison sac length and head width are reported to form a

Fig. 1. Position of six Darwin's finches in principal component space. The red line indicates the "minimum area triangle" as determined by the software provided by Shoval et al. Note that the minimum area triangle and the convex hull enclosing the data coincide, and hence no alternative configuration could have a higher degree of triangularity, but it could have the same degree of triangularity. Also note that the axes have different scales, such that untransformed data fall almost perfectly on a line.

[1]University Pablo de Olavide, Carretera Utrera Km 1, ES-41013, Sevilla, Spain. [2]Estación Biológica de Doñana (EBD-CSIC), Avenida Américo Vespucio s/n, ES-41092, Sevilla, Spain. E-mail: edelaar@upo.es

Fig. 2. Relationship between wing aspect ratio and body mass in bats. Species belonging to the same family (as indicated by symbol type) are clustered; e.g., the family Molossidae (indicated by filled dots) dominates and is restricted to the top of the distribution. Note that megabats (indicated by *x*) are not included in the analysis by Shoval *et al.* [Figure reproduced from (*5*), with permission from the Royal Society]
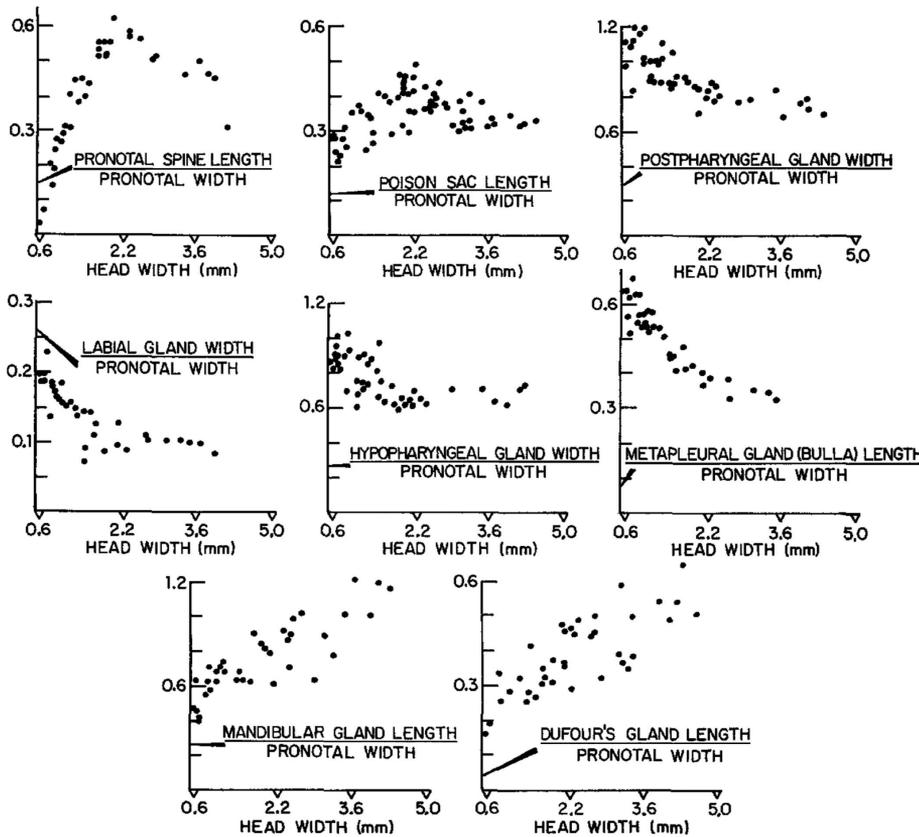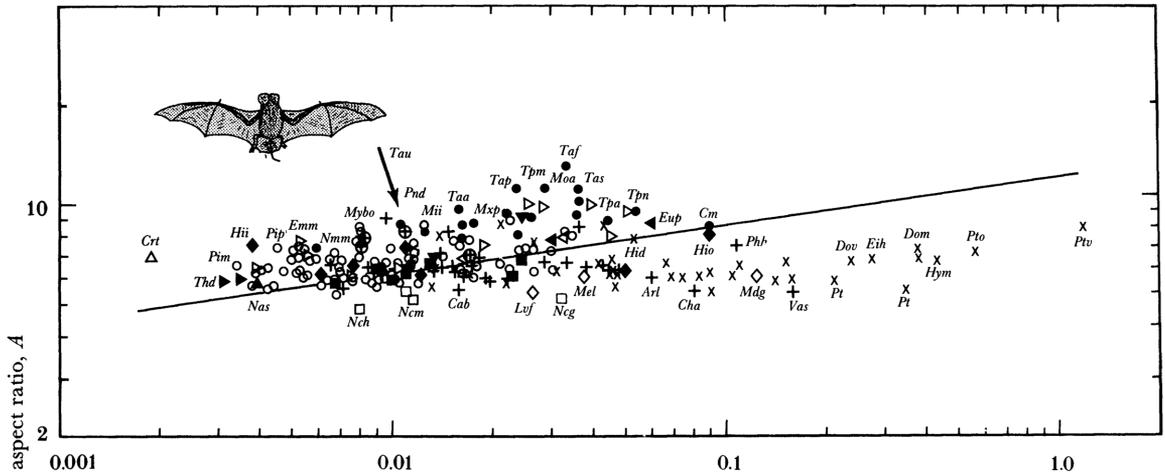

Fig. 3. Relative organ size versus head width in the ant *Atta sexdens*. Phenotypes fall within a triangle for relative poison sac length, but not for seven other organs. [Figure reproduced from (*6*), with permission from Springer Science+Business Media]

triangle among individual ants, but in the same ants none of the other seven traits do so (Fig. 3); instead, it may be more parsimonious to conclude that relative poison sac length and head width are distributed along a curved line, just as relative pronotal spine length and head width are. Similarly, wing aspect ratio and body mass of bats form a (probably nonsignificant; see above) triangle, but wing loading and body mass do not [figure 7 in (*5*)]. Because there is no formal theory to predict which traits will be involved in strong trade-offs, it is possible that multiple tests on data sets from the same system have identified some coincidental cases of data falling significantly on a line or in a triangle, resulting in inflated type 1 error (*2*). Reporting negative results and appropriate correction of *P* values for multiple testing (*2*) would be necessary in that case.

Whether the approach presented by Shoval *et al.* can help us to deduce which tasks are important for fitness and form the basis of trade-offs thus remains unconvincingly demonstrated. Future demonstrations would need to take statistical nonindependence among data points and tests properly into account.

References and Notes

1. O. Shoval *et al.*, *Science* **336**, 1157 (2012).
2. G. P. Quinn, M. J. Keough, *Experimental Design and Data Analysis for Biologists* (Cambridge Univ. Press, Cambridge, 2002).
3. R. P. Freckleton, *J. Evol. Biol.* **22**, 1367 (2009).
4. P. H. Harvey, M. D. Pagel, *The Comparative Method in Evolutionary Biology* (Oxford Univ. Press, Oxford, 1991).
5. U. M. Norberg, J. M. V. Rayner, *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **316**, 335 (1987).
6. E. O. Wilson, *Behav. Ecol. Sociobiol.* **7**, 143 (1980).