Appendix B. Univariate vs. vector analysis

The purpose of this Appendix is to demonstrate how univariate testing of vector data can bias non-directed hypothesis testing. To this end we developed and analyzed an arbitrary dataset (Table S2). As in Appendix A, we caution readers that we have constructed these data specifically to demonstrate particular concepts. The reader is therefore left to judge the relevance of this discussion to real (experimental) datasets.

The specific goal of this Appendix is to compare and contrast the (univariate) t test and its (multivariate) vector equivalent: the Hotelling's T^2 test.

Table S2: A simulated dataset exhibiting biased univariate testing. (a) Two-component force vector responses $\boldsymbol{F} = [F_x,\ F_y]^\top$. (b)-(d) Scalar (univariate) testing. (e)-(g) Vector (multivariate) testing. Sources of bias and further details are discussed in the text. Technical overviews of covariance matrices ($\boldsymbol{W}$) and the Hotelling's T^2 statistic are provided in Appendix D and §2.3 (main manuscript), respectively.

		Group A	**Group B**	**Inter-Group**
	(a) Responses	$\boldsymbol{F}_{\mathrm{A1}} = [159,\ 719]^\top$ $\boldsymbol{F}_{\mathrm{A2}} = [115,\ 762]^\top$ $\boldsymbol{F}_{\mathrm{A3}} = [177,\ 681]^\top$ $\boldsymbol{F}_{\mathrm{A4}} = [138,\ 694]^\top$ $\boldsymbol{F}_{\mathrm{A5}} = [98,\ 697]^\top$	$\boldsymbol{F}_{\mathrm{B1}} = [143,\ 759]^\top$ $\boldsymbol{F}_{\mathrm{B2}} = [172,\ 734]^\top$ $\boldsymbol{F}_{\mathrm{B3}} = [161,\ 735]^\top$ $\boldsymbol{F}_{\mathrm{B4}} = [195,\ 733]^\top$ $\boldsymbol{F}_{\mathrm{B5}} = [168,\ 706]^\top$	
Univariate	(b) Means	$(\overline{F_x})_{\mathrm{A}} = 137.4$ $(\overline{F_y})_{\mathrm{A}} = 710.6$	$(\overline{F_x})_{\mathrm{B}} = 167.8$ $(\overline{F_y})_{\mathrm{B}} = 733.4$	$\Delta\overline{F_x} = 30.4$ $\Delta\overline{F_y} = 22.8$
	(c) St.dev.	$(s_x)_{\mathrm{A}} = 28.6$ $(s_y)_{\mathrm{A}} = 28.5$	$(s_x)_{\mathrm{B}} = 16.8$ $(s_y)_{\mathrm{B}} = 16.8$	$s_x = 23.5$ $s_y = 23.4$
	(d) t tests			t_x=1.832; p_x=0.104 t_y=1.380; p_y=0.205
Vector	(e) Means	$\overline{\boldsymbol{F}}_{\mathrm{A}} = [137.4,\ 710.6]^\top$	$\overline{\boldsymbol{F}}_{\mathrm{B}} = [167.8,\ 733.4]^\top$	$\Delta\overline{\boldsymbol{F}} = [30.4,\ 22.8]^\top$
	(f) Covariance	$\boldsymbol{W}_{\mathrm{A}} = \begin{bmatrix} 817.8 & -323.2 \\ -323.2 & 809.8 \end{bmatrix}$	$\boldsymbol{W}_{\mathrm{B}} = \begin{bmatrix} 283.8 & -131.9 \\ -131.9 & 281.8 \end{bmatrix}$	$\boldsymbol{W} = \begin{bmatrix} 550.8 & -227.6 \\ -227.6 & 545.8 \end{bmatrix}$
	(g) T^2 test			T^2=7.113; p=0.028

In Table S2(a) above there are five force vector responses ($\boldsymbol{F} = [F_x,\ F_y]^\top$) for each of two groups: "A" and "B". Their means and standard deviations are shown in Table S2(b)-(c). In Table S2(d) we see that t tests pertaining to both F_x and F_y fail to reach significance; p values are greater than (even an uncorrected) threshold of $p = 0.05$. An adequate interpretation is that the mean force component changes ($\Delta\overline{F_x}$ and $\Delta\overline{F_y}$) are not unexpectedly large given their respective variances (i.e. standard deviations: s_x and s_y).

We next jump ahead to the final results of the vector procedure in Table S2(g): here we see that the Hotelling's T^2 test reached significance ($p = 0.032$). An adequate interpretation is that the mean force vector change ($\Delta\overline{\boldsymbol{F}}$) was unexpectedly large given its (co)variance ($\boldsymbol{W}$). Let us now backtrack and consider why the univariate and vector procedures yield different results.

The first step of the vector procedure is to compute mean vectors; in Table S2(e) we can see that the vector means have the same component values as the univariate means from Table S2(b). However, there is already one critical discrepancy to note: the vector procedure assesses $\Delta\overline{\boldsymbol{F}}$, which is the **resultant** vector connecting the Group A and Group B means (Fig.S3). From Pythagoras' theorem:

$$|\Delta\overline{\boldsymbol{F}}|^2 = \Delta\overline{F}_x^2 + \Delta\overline{F}_y^2 \tag{B.1}$$

it is clear that the magnitude of the resultant will <u>always</u> be greater than the magnitude of its components — except in the experimentally unlikely cases of $\Delta\overline{F}_x$=0 and/or $\Delta\overline{F}_y$=0. This is non-trivial for two reasons. First, since the vector procedure assesses the maximum difference between the two groups, it is more robust to Type II error than univariate procedures (note: the univariate tests in Table S2 exhibit Type II error by failing to reach significance). Second, the vector technique's assessment of differences is independent of the xy coordinate system definition; whereas the component effects ($\Delta\overline{F}_x$ and $\Delta\overline{F}_y$) can change when the xy coordinate system definition changes, both the resultant and the variance along the resultant direction will always have the same magnitude. This may have non-trivial implications for biomechanical datasets that employ difficult-to-define coordinate systems (e.g. joint rotation axes).

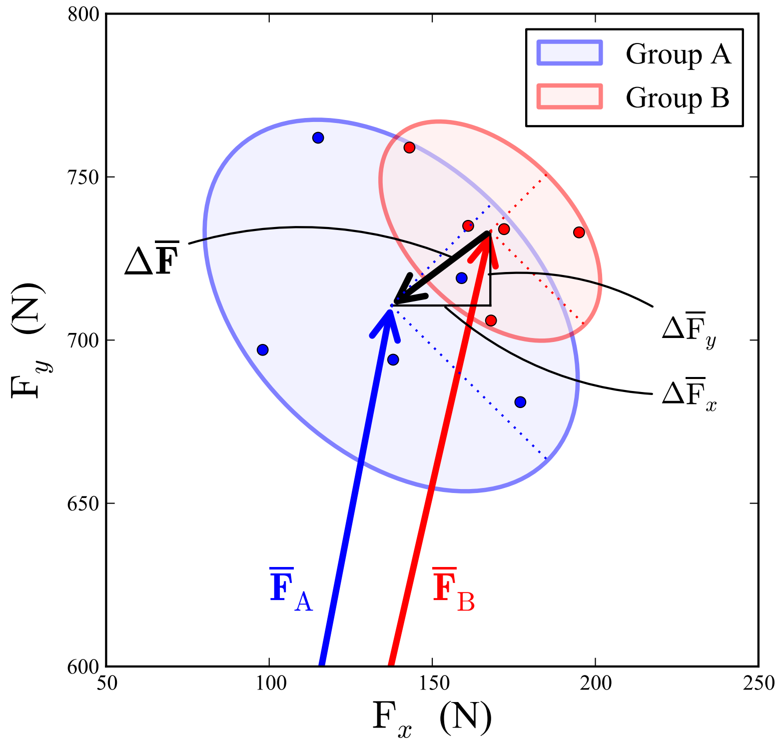

Figure S3: Graphical depiction of the data from Table S2. Small circles depict individual responses. Thick colored arrows depict the mean force vectors for the two groups. The thick black arrow depicts the (vector) difference between the two groups, and thin black lines indicate its x and y components. The ellipses depict within-group (co)variance; their principal axes (thin dotted lines) are the eigenvectors of the covariance matrices in Table S2(f). Here covariance ellipse radii are scaled to two principal axis standard deviations (to encompass all responses).

The next step of the vector procedure is to compute covariance matrices $\boldsymbol{W}$ (Appendix D). The diagonal elements of $\boldsymbol{W}_{\mathrm{A}}$ and $\boldsymbol{W}_{\mathrm{B}}$ in Table S2(f) are simply the variances (i.e. squared standard deviations) s_x^2 and s_y^2 from Table S2(c). The off-diagonal terms are equal and represent the covariance (i.e. correlation) between F_x and F_y. If F_x tends to increase when F_y increases then the off-diagonal terms would be positive, but in this case they are negative, indicating that F_x tends to decrease when F_y increases. This tendency can be seen in the raw data (small circles) in Fig.S3.

The presence of non-zero off-diagonal terms thus has a critical implication: changes in F_x

and F_y are not independent. This is critical because univariate tests implicitly assume that F_x and F_y are independent.

To appreciate this point it is useful to recognize that covariance matrices may be interpreted geometrically as ellipses: the eigenvectors of $\boldsymbol{W}$ represent the ellipse's principal axes, and its eigenvalues represent the variance along each principal direction. This is perfectly analogous to inertia matrices: the eigenvectors of an inertia matrix define a body's principal axes of inertia, and eigenvalues specify the principal moments of inertia.

The importance of this geometric interpretation becomes clear when visualizing covariance ellipses. In Fig.S3 we can see that the principal axes of the covariance matrices are not aligned with the xy coordinate system, implying that changes in F_x and F_y are not independent. Critically, we can also see that the direction of minimum variance is very similar to the direction of $\Delta\overline{\boldsymbol{F}}$. Thus the standard deviations s_x and s_y (used in the univariate analyses) are larger than the standard deviation in the direction of $\Delta\overline{\boldsymbol{F}}$.

In summary, vector statistical testing more objectively detects vector changes because : (a) it is coordinate system-independent, (b) it considers both the maximum difference between groups (i.e. the resultant difference) and the variation along this direction. This Appendix has demonstrated how univariate testing of vector data can lead to Type II error. With a different dataset it would also be possible to demonstrate Type I error, but in interest of space we end here. The most important point, the main paper contends, is that non-directed hypothesis testing mustn't assume vector component independence.