# Of Uberfleas and Krakens: Detecting Trade-offs Using Mixed Models

<table>
<thead>
<tr>
<th>Journal:</th>
<th><em>Integrative and Comparative Biology</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>Manuscript ID</td>
<td>ICB-2017-0004.R1</td>
</tr>
<tr>
<td>Manuscript Type:</td>
<td>Symposium article</td>
</tr>
<tr>
<td>Date Submitted by the Author:</td>
<td>n/a</td>
</tr>
</tbody>
</table>
| Complete List of Authors: | Careau, Vincent; University of Ottawa, Department of Biology  
                          | Wilson, Robbie; UQ, Biological Sciences |
| Keywords:           | mixed model, repeatability, allocation, individual variation, SQuID |

Note: The following files were submitted by the author for peer review, but cannot be converted to PDF. You must view these files (e.g. movies) online.

Careau_et_al_ICB_R_codes.R  
Careau_et_al_ICB_Kraken_population_Figure_3.RData  
Careau_et_al_ICB_Kraken_populations_Figure_4.RData
Of Uberfleas and Krakens:
Detecting Trade-offs Using Mixed Models

Vincent Careau* and Robbie Wilson†

*Canada Research Chair in Functional Ecology, Department of Biology, University of Ottawa, Ottawa, ON, Canada
†School of Biological Sciences, The University of Queensland, St Lucia, QLD, Australia

Running headline: Uberfleas and Krakens

Keywords: mixed model, repeatability, allocation, individual variation, SQuID.

Total number of words in the text (excluding references, tables, and figure legends) = 2,286
Synopsis: All animals experience performance trade-offs as they complete tasks such as capturing prey, defending territories, acquiring mates, and escaping predators. Why then, is it so hard to detect performance trade-offs at the whole-organismal level? Why do we sometimes even obtain positive correlations between two performance traits that are predicted to be negatively associated? Here we explore two plausible explanations. First, most analyses are based on individual maximal values (i.e., personal best), which could introduce a bias in the correlation estimates. Second, phenotypic correlations alone may be poor indicators of a trade-off when contrasting processes occur at the among- vs. within-individual levels. One such scenario is the “big houses big cars” model developed in life-history theory to explain the existence of “uberfleas” that are superior in all regards (because they acquire more resources than others). We highlight that the exact opposite scenario might occur for performance trade-offs, where among-individual trade-offs may be masked by within-individual changes in physical condition. One of the best ways to test among these alternative scenarios is to collect repeated pairs of performance traits and analyse them using multivariate mixed models (MMMs). MMMs allow straightforward and simultaneous examination of trait correlations at the among- and within-individual levels. We use a simple simulation tool (SQuID package in R) to create a population of Krakens, a mythical giant squid-like sea creature whose morphology generates a performance trade-off between swimming speed and strength or ability to sink ships. The simulations showed that using individual maximum values introduces a bias that is particularly severe when individuals differ in the number of repeated samples ($n_{\text{trial}}$). Finally, we show how MMMs can help detect performance (or any other type of) trade-offs and offer additional insights (e.g., help detect plasticity integration). We hope researchers will adopt MMMs when exploring trade-offs in whole-animal performances.
Why are whole-organism performance trade-offs so hard to detect?

The concept of trade-offs is key to many research fields such as evolutionary biology, physiology, behavioral ecology, and functional morphology (Agrawal and others 2010; Garland 2014). Performance trade-offs occur when characteristics that enhance one aspect of performance necessarily decrease another type of performance. For example, greater proportions of slow-twitch oxidative muscle fibers should enhance stamina, whereas more fast-twitch glycolytic fibers would facilitate sprinting abilities (Dohm and others 1996; Garland 1988; Sorci and others 1995). The mechanistic bases of many performance trade-offs are well described and accepted (Esbjornsson and others 1993; Komi 1984; Wilson and James 2004; Wilson and others 2004). Yet evidence for their existence at the whole-animal level is far from convincing (Ford and Shuttlesworth 1986; Garland 1988; Garland and Else 1987; Herrel and Bonneau 2012; Huey and others 1990; Jayne and Bennett 1990; Sorci and others 1995; Tsuji and others 1989; Wilson and others 2002). Many studies exploring trade-offs at the whole-animal level find that high performers in one task are also high performers in other tasks, or find no trade-off between tasks at all (Garland and Else 1987; Huey and others 1990; Jayne and Bennett 1990; Tsuji and others 1989). Given the intuitive physiological basis of many performance trade-offs, the paucity of studies showing them in whole animals is puzzling.

Among-individual variation in “quality”

Van Damme and others (2002) and Wilson and others (2014) have suggested that individual variation in overall “quality” can mask within-individual trade-offs in performance. Because individuals vary in health, physical fitness, nutrition, developmental stage or genetics, some individuals perform better or worse across all types of performance tasks than others. This means
that when researchers try to understand *intra*-individual functional trade-offs using *inter*-individual variation in performance, then trade-offs that do occur within individuals can be masked by among-individual variation in overall “quality” (*Lailvaux and Kasumovic 2011*).

Previous studies have shown that correcting for quality can reveal functional trade-offs between motor tasks such as speed, power, and endurance that would otherwise be missed (*Van Damme and others 2002*; *Wilson and others 2014*). However, how to unambiguously define and account for quality remains controversial (*Walker and Caddigan 2015*).

**Uberfleas and the “big houses big cars” scenario**

The pattern described above is essentially the same as the classic Y-model of life-history evolution in which inter-individual variation in acquisition can mask allocation trade-offs (*van Noordwijk and de Jong 1986*). In other words, some individuals can circumvent an apparent trade-off in allocation of resources by acquiring more resources. *Reznick and others (2000)* referred to this situation as the “big houses big cars” scenario: because the resources available for families to spend on both a house and a car are finite, the more a family spends on their car then the less they have available for their house (*vice versa*). However, some families can have both big houses and flashy cars if they have more resources (money) available to them. It is quite easy to see how this situation could apply to animals. For example, limits to energy stores and/or maternal abdominal volume (*e.g., Du and others 2005*) can yield a trade-off between egg size and number (*Roff 1992; Stearns 1992*). However, some female “uberfleas” (superior in all regards) may be able to lay many large eggs because they are able to obtain more energy and nutrients and have larger abdominal volume than “unterfleas” (inferior in all regards) (*Reznick and others 2000*). In such a scenario (Fig. 1A), the relative amount of variance in allocation vs.
acquisition will dictate whether egg size and number are correlated negatively (i.e., trade-off revealed) or positively (inverse of a trade-off) at the phenotypic level (Fig. 1A).

**Trade-offs within energy budget**

One implication of the “big houses big cars” scenario is that studies based at a single level of variation – or those aggregating repeated measures to reduce the hierarchical nature of the data and facilitate statistical analysis – may fail to detect an association between two traits even when they are connected through processes occurring at different levels (Dingemanse and Dochtermann 2013; Downs and Dochtermann 2014). This situation is illustrated by a recent re-analysis of data in humans, where individuals differed in their total energy intake, but nevertheless had to allocate energy to maintenance vs. locomotor activity [Fig. 2, taken from Careau (2017)]. Westerterp and others (1992) collected data on sedentary men and women between the ages of 28–41 years old as they followed a training program to run a half marathon (four sessions per week, increasing running time to 10-30 min, 20-60 min and 30-90 min per training session after 8, 20 and 40 weeks respectively). Westerterp and others (1992) quantified energy intake (self-reported) and basal metabolic rate (respirometry) on four occasions for each individual. The original analysis of this dataset greatly contributed to our understanding of the sex-specific and population-levels effects of exercise training on energy balance and the impacts on body composition (Westerterp and others 1992), but did not consider how traits were correlated at the among- vs. within-individual levels. Reanalysis of this dataset using multivariate mixed models revealed that while surplus energy intake and basal metabolic rate tended to be positively correlated at the among-individual level, there was a negative correlation at the within-individual level (Fig. 2).
Reverse scenario: trade-offs at the among-individual level

In general, it is intuitive to think of allocation trade-offs as processes occurring within individuals. In the case of performance trade-offs, however, one may expect situations in which trade-offs occur among individuals (Fig. 1B). For example, if individuals consistently differ in many morphological (e.g., gear ratio) and physiological (muscle fiber types) traits that enhance their performance in certain tasks (e.g., speed) while hindering performance in others (e.g., endurance), then the trade-off should be apparent at the among-individual level (Fig. 1B). We may also intuitively expect that within-individual changes in general physical condition will simultaneously affect all performance traits, thus generating positive correlations among performance traits within individuals (Fig. 1B). Although this scenario is the exact opposite as the “big houses big cars” scenario, we argue it is equally likely in the case of performance trade-offs. We refer to this scenario as the “sink or swim” scenario (explained below and in Fig. 1).

How to detect trade-offs?

When contrasting processes occur at the among- vs. within-individual levels, how should one go about verifying the presence of a trade-off? To ensure one does not miss out on information regarding the alternative “big houses big cars” (Fig. 1A) and “sink or swim” (Fig. 1B) scenarios, it is imperative to partition correlations at the among- and within-individual levels. Otherwise, it is likely that the calculated phenotypic correlation ($r_P$) will be a poor indicator of the presence of trade-offs (Fig. 3). We argue that the use of $r_P$s calculated on individual maximum values – rather than properly partitioning correlation across levels – is one of the reasons that whole-animal performance trade-offs are rarely detected, and even spurious positive correlations are sometimes obtained.
Simulating a performance trade-off in a legendary sea monster

We used the SQuID package in R (Allegue and others 2016) to simulate a population of 100 individual organisms that experience a trade-off between two performance traits. We used the legendary giant squids known as Krakens (https://en.wikipedia.org/wiki/Kraken) for our simulation. Krakens are mythical creatures (probably), which live off the coast of Norway, made famous in various fictional works that depict them attacking ships at sea. Let us for one moment imagine that Krakens experience a trade-off between swimming speed and their ability to sink a ship and swallow all the seamen (hence the “sink or swim” scenario). We expect this trade-off is related to an individual’s size and shape; such that slimmer squids can swim faster while bulkier squids can overpower ships. This trade-off occurs at the among-individual level as depicted by the black dots and lines in Fig. 3A. Moreover, we simulated a positive relationship between speed and strength occurring within individuals, as depicted by thin lines in Fig. 3A. This positive relationship occurring within individuals could be due to correlated phenotypic plasticity as function of variation in an unknown factor. Such a factor could be temperature, for example, as it could positively affect both speed and strength. Another factor could be age (Krakens can live up to 100 years and become terrible sea creatures as they grow and get both faster and stronger).

Phenotypic correlations based on a single value per individual

Now imagine we can only catch a Kraken once in its lifetime to obtain a single measure of speed and strength for each individual ($n_{\text{trial}}=1$). Only a single measure per animal would yield a random sample like in Fig. 3B, in which the phenotypic correlation ($r_p$) is not significantly different from zero even at $N=100$. This is because the $r_p$ between speed and strength reflects a
mixture of the processes occurring at the among- and within-individual levels (like in Fig. 1B), as shown by the following equation:

\[ r_P = r_{\text{ind}} \sqrt{R_{\text{speed}} \times R_{\text{strength}}} + r_e \sqrt{(1 - R_{\text{speed}}) \times (1 - R_{\text{strength}})} \]

Where \( r_{\text{ind}} \) and \( r_e \) denote the among- and within-individual correlations and \( R_{\text{speed}} \) and \( R_{\text{strength}} \) denote the repeatability of speed and strength, respectively. This equation (modified from equation (1) in Dingemanse and others 2012) implies that if speed and strength are both repeatable at \( R=0.5 \) (as in our simulations) and linked via a \( r_{\text{ind}} \) and a \( r_e \) of -0.5 and 0.5, respectively, then their \( r_P \) will be 0. By contrast, if speed and strength were repeatable at \( R=0.25 \) then their \( r_P \) will be 0.25 (see for yourself by changing input$Vind[1,1]$ and input$Vind[5,5]$ to 0.25 in the R code provided as supplementary material available at ICB online). Thus, under the “big houses big cars” and “sink or swim” scenarios, variation in repeatability alone can result in a wide range of “context-dependent” \( r_P \). Dingemanse and Dochtermann (2013) modeled a similar scenario (\( r_{\text{ind}}=-0.5 \) and \( r_e=0.5 \)) and showed that \( r_P \) based on a single measure per individual were inaccurate, irrespective of the sample size (number of individuals sampled).

Therefore, even studies based on \( r_P \) estimated with a large sample size one could erroneously conclude that speed and strength are not linked in Kraken, while in fact they are connected through different processes occurring at the among- and within-individual levels and cancelling each other out at the phenotypic level.

**Phenotypic correlations based on average values**

If one has access to all 20 repeated measures of performances for each Kraken, how should one go about verifying if there is a trade-off as simulated? One easy solution is to calculate \( r_P \) based on the individual mean values, which has the advantage of facilitating statistical analysis by
reducing the hierarchical nature of the data. This approach proves effective at detecting the trade-off under the parameter settings used for simulating the Kraken dataset (Fig. 3C). However, the $r_p$ based on individual mean values are typically biased towards $r_e$ (see Appendix in Dingemanse and others 2012). This bias depends on the level of repeatability in the traits involved (see equation above) and the number of repeated measures ($n_{trial}$) within individuals. With 20 repeated measures, the estimated $r_p$ based on mean values (-0.37±0.09) was not too far from the simulated value (-0.5), but the bias can rapidly increase with decreasing $n_{trial}$ (i.e., keeping the first 5 repeated measures per individual yields a non-significant $r_p$ of -0.03±0.10). Clearly, under the “big houses big cars” and “sink or swim” scenarios, the $r_p$ based individual mean values may not reveal much about the presence of trade-offs because it depends on the geometric mean repeatability of traits involved and the sample size per individual ($n_{trial}$).

**Phenotypic correlations based on maximum values**

An alternative approach to analysing the data would be to select the best performance for each Kraken among the 20 repeated measures and then calculate $r_p$ on the resulting 100 “personal best” samples, which yielded a non-significant correlation (Fig. 3D). This is perhaps not surprising, given that correlations based on maximum values have the same problems as those based on mean values (see above), in addition to other problems (Adolph and Hardin 2007). While the expected value of a mean is independent of $n_{trial}$, the expected value of a maximum increases with $n_{trial}$. Moreover, individuals may differ in their underlying distribution of repeated performances, which should affect maximum values more than mean values. Therefore, the undesirable properties of the sampling distribution of extreme values can further conceal
performance trade-offs ([Head and others 2012](#)) and in some cases even generate spurious counterintuitive results (see below).

**Multivariate mixed models to the rescue**

To obtain an unbiased estimate of the strength of the “sink or swim” trade-off in Krakens, we need to quantify the among-individual correlation ($r_{ind}$) separately from the within-individual (or residual) correlation ($r_e$). Whenever two traits are repeatedly measured for a set of individuals, we can estimate $r_{ind}$ and $r_e$ in one step using multivariate mixed models (MMMs) ([Dingemanse and Dochtermann 2013](#)). MMMs are now relatively easy to learn thanks to several “how to” papers ([Dingemanse and Dochtermann 2013](#); [Houslay and Wilson 2017](#); [Roche and others 2016](#)) and widely available software like ASReml ([Butler and others 2009](#)), SAS, and the R package MCMCglmm ([Hadfield 2010](#)). Using a MMM we successfully detected the trade-off as indicated by the negative $r_{ind}$ (-0.41±0.09), which is close to the simulated value of -0.5.

**Magnitude of bias using individual maximum performance**

To make a more convincing case against using maximum values – and a more convincing case for using MMMs – we simulated 100 populations of Krakens and tested the 3 methods on each population (MMMs, using individual means, and using individual maxima). Moreover, we tested each method with all 20 repeated measures per individual (equal $n_{trial}$) and once again after we deleted a randomly varying number of observations for each individual (unequal $n_{trial}$). As expected, MMMs yielded on average a $r_{ind}$ close to the simulated value and were not sensitive to variable $n_{trial}$ (Fig. 4). Although mean values yield acceptable $r_P$s in most populations when $n_{trial}$ was equal, there was a clear bias when $n_{trial}$ was unequal among individuals (Fig. 4). Finally, there was a clear bias when using individual maximum values, which was especially severe.
when \( n_{\text{trial}} \) is variable among individuals (Fig. 4). Retaining individual maximum values seems to remain a common practice in studies of organismal performance even though the problems with this practice have been voiced before (Adolph and Hardin 2007; Adolph and Pickering 2008; Head and others 2012). Adolph and Hardin (2007) developed a method to obtain an unbiased estimation of the correlation by using mean values and correcting the obtained \( r_P \) for the repeatability of the traits of interest. Still, this correction factor may yield a biased \( r_{\text{ind}} \) because it assumes that \( r_e \) is zero (Dingemanse and others 2012). As shown by Downs and Dochtermann (2014), this correction factor of Adolph and Hardin (2007) can result in an underestimation of \( r_{\text{ind}} \) when there is a positive \( r_e \) (as in the case in our simulation of speed and strength in Krakens).

**Implications of a within-individual correlation**

A significant \( r_e \) can be generated by correlated measurement error or correlated phenotypic plasticity at the population or individual levels. At the population level, correlated plasticity could give rise to a \( r_e \) if both traits are simultaneously affected by a factor (e.g., temperature, age) that is unaccounted for. In this case, simply adding the relevant factor as a fixed effect fitted to both traits would eliminate the \( r_e \). At the individual level, correlated plasticity could give rise to a \( r_e \) if individuals differ in how their performances in multiple tasks change similarly across an environmental factor. This is how we generated a positive \( r_e \) in our Kraken population, but so far assumed individuals did not vary in how they responded to the environmental factor (i.e., we did not include random slopes). If we take temperature as an example, some Krakens may be more sensitive to temperature than others, such that individuals whose speed increase more steeply with temperature also increase more in strength with temperature. By modelling random intercepts only, individual (co)variation in plasticity ended up in the \( r_e \) (Brommer 2013a), which was 0.43±0.02. To fully capture the among-individual correlation in phenotypic plasticity of
speed and strength, we used a bivariate random regression \cite{Henderson1982} to model correlated plasticity separately from the residuals. This model yielded a correlation of $0.49\pm0.08$ among the slopes of the individual reaction norms (see R codes in the supplementary material available at ICB online). Accordingly, the bivariate random regression yielded a $r_e$ close to zero ($0.01\pm0.02$), which is expected given that we did not include correlated measurement error in our simulation.

**Plasticity integration**

Correlated plasticity may indicate “plasticity integration”, a concept that has long been recognised and intermittently investigated over the past ~30 years \cite{Husby2010, Newman1994, Scheiner1991, Schlichting1986, Spitze1996}, but remains an understudied phenomenon \cite{Pigliucci2003}. Using multivariate random regression models to study plasticity integration has the potential to greatly enhance our understanding of how performance traits (or any morphological, behavioural, and physiological traits) are co-adapted across environmental gradients. For example, the evolution of thermal reaction norms is a highly coadaptive process \cite{Angilletta2006} involving phenotypic plasticity in behaviour \cite{Biro2010}, metabolism \cite{Careau2014b}, and performance \cite{Artacho2013, Careau2014a}. Thus, correlated plasticity in response to temperature, when left unaccounted for, could very likely generate significant $r_e$ among traits.

Another likely situation of plasticity integration among performance traits involves training, as it is widely known that training can improve performance in multiple dimensions (e.g., speed and endurance). Such correlated changes in multiple performances in response to training is a form of plasticity integration, which would link performance traits through a positive $r_e$. 
Sampling designs to study performance trade-offs

Researchers wanting to test the “big houses big cars” or “sink or swim” scenarios should adopt sampling designs and sample sizes that allow the proper quantification of the correlation of interests (i.e., $r_{\text{ind}}$ and $r_e$). If two performance traits are assayed repeatedly, but never at the same time, then it is possible to estimate $r_{\text{ind}}$ but not $r_e$. To quantify both $r_{\text{ind}}$ and $r_e$, traits must be measured repeatedly at the same time. Dingemanse and Dochtermann (2013) suggested that when both $r_{\text{ind}}$ and $r_e$ are of interest – as required to detect scenarios in Fig. 1 – sample sizes should be optimized with respect to $r_{\text{ind}}$. This is because $r_{\text{ind}}$ is generally estimated with less power and accuracy than $r_e$, which means that researchers should favour measuring more individuals instead of taking more repeated measures per individual. When traits have a repeatability of 0.5, the power to detect a $r_{\text{ind}} > 0.5$ rapidly increases above acceptable levels (0.8) past sample sizes of >25 individuals, but much larger sample sizes (i.e., >125 individuals) are needed to detect $r_{\text{ind}} < 0.3$ (Dingemanse and Dochtermann 2013).

Researchers interested in studying plasticity integration among performance traits must take multiple measures per trait per individual along a quantified environmental factor (Martin and others 2011). For example, one could measure both speed and strength in multiple individuals at 10, 15, 20, and 25°C. With such a sampling design, it is possible to model the among-individual variation in the slopes of the reaction norms of speed and strength, as well as their correlation (e.g., Careau and others 2014b). Correlated slopes at the individual (or genetic) level can be interpreted as plasticity integration. Otherwise a significant $r_e$ cannot be unambiguously interpreted as plasticity integration, since a $r_e$ can arise from correlated measurement error and population-level plasticity.
In general, researchers studying performance use carefully designed protocols to ensure an animal’s physical capabilities are measured instead of their behaviour (Roche and others 2016). For example, to force animals to run at their maximal speed, researchers can use various techniques of making noises to prompt an animal into an escape response. Moreover, researchers usually measure performance over a series of successive trials on a given individual and retain the maximal value as the performance of that individual on that day. All other measurements are discarded because their lower values may indicate the individual was not maximally motivated. However, further insights might be gained from retaining measurements across repeated trials within each day (Roche and others 2016), which adds another level of variance that can be easily captured using multilevel mixed models (Araya-Ajoy and others 2015). In this case, the residual variance represents the variation among successive trials within a day. Interestingly, there might be some interesting biology hidden at this level of variance (Cleasby and Nakagawa 2011). For example, individuals may differ in the consistency at which they perform over successive trials, which may be related to variation in motivation. Finally, retaining all repeated measures across successive trials would allow researchers to test whether individuals differ in habituation and/or fatigue resistance (Roche and others 2016).

Conclusions

1. Partitioning performance trade-offs at the within- and among-individual levels is required to discern between scenarios like those occurring in mythical organisms like uberfleas (“big houses big cars”; Fig. 1A) and Krakens (“sink or swim”; Fig. 1B).
2. Our simulation exemplifies the magnitude of the bias introduced by using individual maximum values. Under a “sink or swim” scenario, the bias can be severe enough to generate a spurious positive correlation where a trade-off is expected.

2.3. Researchers should pay greater attention to processes occurring at the within-individual level, such as correlated phenotypic plasticity and correlated measurement error. In the presence of a significant $r_v$, then random regression models can be used to separate potential individual differences in plasticity from the residuals (Brommer 2013a; 2013b).

3.4. Finally, we note that while using MMMs to estimate $r_{ind}$ and $r_v$ is informative in its own right, it is only the first step towards identifying genetic and environmental associations among performance traits through larger-scale quantitative genetics studies (e.g., Dohm and others 1996; Garland 1988).

Acknowledgment

We thank S. Lailvaux & J. Husak for their invitation to participate in this symposium.

Funding

This project was supported by grants from Natural Science and Engineering Research Council to V.C. R.S.W is supported by an Australian Research Council Future Fellowship.

Supplementary Data

All of the R codes necessary to reproduce these results were made available to editors and reviewers upon initial submission and are available as supplementary material available at ICB online.
References


Brommer JE. 2013a. On between-individual and residual (co)variances in the study of animal personality: are you willing to take the “individual gambit”? Behavioral Ecology and Sociobiology 67:1027-1032.


Fig. 1. Schematic representations of trade-offs and facilitation processes simultaneously occurring at the among- vs within-individual levels. (A) The classic “big houses big cars” scenario (Reznick and others 2000; van Noordwijk and de Jong 1986) where individuals differ in the amount of resources (energy, time, or space) that they allocate to competing demands (egg size vs. quantity). Repeated pairs of measurements on each individual are represented by dots connected by a line. Thus, some female uberfleas appear to always have more resources than others, but on any given reproduction event they nevertheless face the same quantity vs. quality trade-off. Whenever acquisition and allocation processes respectively occur at the among- and within-individual levels, the relative amount of among- vs. within-individual variance in the two traits (i.e., their geometric mean repeatability) will dictate whether the resulting phenotypic correlation will be positive (panel a₁) or negative (panel a₂). (B) Another scenario where individuals consistently differ in morphology (e.g., gear ratio) or physiology (e.g., muscle fiber type) such that, on average, strong individuals are slower. However, whenever individuals train to increase their overall physical condition, their performance is enhanced for both strength and speed. Such a situation would yield a negative correlation at the among-individual level (trade-off), but a positive correlation at the within-individual level (training effect), as in the simulated population of Krakens (see Fig. 3 and 4). Because Krakens face a trade-off between swimming speed and the ability to sink ships, we refer to this scenario as the “sink or swim” scenario. Note that Wilson and others (2014) applied the “big houses big cars” scenario (A) to performance trade-offs, suggesting that among-individual differences in “quality” can mask within-individual trade-offs. Here we suggest that the “sink or swim” scenario (B) is equally likely to apply to performance trade-offs (i.e., that among-individual differences in trade-offs can be masked within-individual changes in condition). Multivariate mixed models can be used to test these
alternative scenarios and quantify the relative importance of among- vs. within-individual
(co)variances driving phenotypic correlations (see Fig. 2 for an example).

Fig. 2. Trade-offs among and within energy budgets. (A) Representation of the total energy
intake of most wild animals as the sum of the energy invested in locomotor activity,
reproduction, growth, thermoregulation, digestion (thermic effect of food; TEF), and basal
metabolic rate (BMR). BMR represent the minimum amount of energy required for the
functioning (e.g., breathing) and the maintenance (e.g., tissue turnover) of maintain vital
systems. Proportions are arbitrary and can change depending on the age, season, and life-history
stage of animals. Any source of energy expenditure above BMR can be grouped into the “surplus
energy”. (B) Energy budget in most modern non-reproductive adult humans, in which there are
no cost of growth and reproduction. The cost of thermoregulation is assumed to be negligible
(people wearing appropriate clothing at room temperature do not have any extra energy
expenditure to maintain body temperature). In this simplified energy budget, a large proportion
(~60%) is taken by BMR and most of the surplus energy is devoted to activity. (C) Among- and
within-individual correlations between surplus energy and BMR in 12 adult men (triangles) and
11 adult women (circles) measured 4 times each during a long-term training program in
preparation for running a half marathon [data from Westerterp and others (1992) and figure taken
from Careau (2017)]. BMR and surplus energy are shown as residuals (open symbols) from
multiple regression models including several fixed effects (testing sequence, age, sex, body
mass, and fat mass). Solid symbols indicate the mean residual values for each individual, and the
thick black line illustrates the positive among-individual relationship. Thin lines show separate
linear regression for each individual, thus illustrating the within-individual relationship. A
bivariate mixed model applied to these data revealed a positive among-individual correlation $(0.49\pm0.27; \ P = 0.113)$ and a negative within-individual correlation $(-0.32\pm0.11; \ P = 0.007)$, providing support for a “big houses big cars” scenario depicted in Fig 1A.

**Fig. 3. Trade-off between speed vs. strength in a simulated population of $N=100$ Krakens** (sea monsters of giant size dwelling off the Norwegian sea). (A) Each legendary giant squid was sampled 20 times ($n_{\text{trial}}=20$) for speed and strength, which were part of trade-off at the among-individual level because of morphological differences (a slender body is good for swimming speed, but a wider body confers strength needed to sink ships). Black dots indicate mean values for each individual and the thick line illustrates the negative among-individual correlation ($r_{\text{ind}}$; simulated value = -0.5). Thin lines show separate linear regression for each individual, thus illustrating the facilitation effects occurring through within-individual correlated phenotypic plasticity in response to an unknow factor (e.g., temperature, age, or training). (B) After randomly sampling each Kraken once ($N=100$; $n_{\text{trial}}=1$), the phenotypic correlation ($r_P$) shows no relationship at all and can vary wildly depending on the random sampling of the original dataset in A. Phenotypic correlations ($r_P$) calculated using (C) individual mean (same as black dots in A) and (D) individual maximum values. This figure shows how interesting contrasting processes can go undetected when $n_{\text{trial}}$ is low (B) and using individual maximum values (D). The data was simulated using SQuID package in R (Allegue and others 2016). See supplementary material available at ICB online for a description of the parameter settings used for simulating and analysing the dataset.
Fig. 4. Attenuated and biased correlations, or not. We used the SQuID package in R (Allegue and others 2016) to simulate 100 populations of Krakens like in Fig. 3 in which there is a negative among-individual correlation of -0.5 (dashed line). For each population, we calculated correlations using the full dataset (equal $n_{\text{trial}}$) and again after deleting a randomly different number of observations for each individual (variable $n_{\text{trial}}$). We first calculated the among-individual correlation ($r_{\text{ind}}$) using a multivariate random mixed model (MMM) in which the among- and within-individual (co)variances were properly modelled. Although the MMM is not sensitive to variable $n_{\text{trial}}$, it is not the case for phenotypic correlations ($r_{\text{P}}$) estimated using individual mean and maximum values. Indeed, individual mean values yield relatively unbiased $r_{\text{P}}$s when $n_{\text{trial}}$ is equal, but the $r_{\text{P}}$s become slightly biased when $n_{\text{trial}}$ is variable among individuals. The bias is clearly worst when using individual maximum values. For equal $n_{\text{trial}}$, we successfully detected the trade-off in only $\sim$50% of populations (with $N=100$, $P<0.05$ for all $|r_{\text{P}}|>0.197$, dotted line). The bias with individual maxima is much more severe when $n_{\text{trial}}$ is variable among individuals – only a few of the negative $r_{\text{P}}$s were significant and even positive $r_{\text{P}}$s are obtained. A complete description of the parameter settings used for simulating the datasets using SQuID and analysing them is available as supplementary material available at ICB online.
Fig. 1.

A "big houses big cars" scenario:

"Mythical animal that follow this pattern
Uberfleas"

B "sink or swim" scenario:

"Krakens"
A Energy budget of many wild animals (arbitrary proportions)

- Activity
- Reproduction
- Growth
- Thermoregulation
- TEF
- Immune syst.
- Cardiovasc. syst.
- Nervous syst.
- Digestive syst.
- Respiratory syst.

B Energy budget in most modern adult non-reproductive humans

- Activity (~30%)
- TEF (~10%)
- BMR (~60%)

C Among- and within-individual correlations between surplus energy (mostly activity) and BMR (data from Westerterp et al. 1992)

\[ \text{Surplus (MJ\text{-day}^{-1}, \text{residuals})} \]

\[ \text{BMR (MJ\text{-day}^{-1}; residuals)} \]
Fig. 3.
**Fig. 4.**

Simulated value

\[ r_{\text{ind}} = -0.5 \quad P = 0.05 \]

- **average replicate pop.**
- Individual maximum (variable \( n_{\text{trial}} \))
- Individual maximum (equal \( n_{\text{trial}} \))
- Individual average (variable \( n_{\text{trial}} \))
- Individual average (equal \( n_{\text{trial}} \))
- MMM (variable \( n_{\text{trial}} \))
- MMM (equal \( n_{\text{trial}} \))

Correlation estimate

\(-0.8 \quad -0.6 \quad -0.4 \quad -0.2 \quad 0.0 \quad 0.2\)