

COMMENTARY

SPECIAL ISSUE

INTEGRATING BIOMECHANICS, ENERGETICS AND ECOLOGY IN LOCOMOTION

Understanding mechanisms of avian flight by integrating observations with tests of competing hypotheses

Douglas L. Altshuler*, Vikram B. Baliga, Anthony B. Lapsansky, Phillip Lee, Eric R. Press and Jolan S. Theriault

ABSTRACT

A long-standing problem in the study of avian flight is determining how biomechanics and physiology are associated with behaviour, ecological interactions and evolution. In some avian clades, flight mechanisms are strongly linked to ecology. Hummingbirds, for example, exhibit traits that support both hovering flight and nectar foraging. In most avian clades, however, features such as wing shape are highly variable among taxa without clear relationships to biomechanics, energetics or ecology. In this Commentary, we discuss challenges to understanding associations between phenotype and performance in avian flight. A potential pitfall in studies that attempt to link trait specialization with performance is that the most relevant traits and environments are not being considered. Additionally, a large number of studies of the mechanisms of avian flight are highly phenomenological. Although observations are essential for hypothesis development, we argue that for our discipline to make progress, we will need much more integration of the observational phase with developing crucial tests of competing hypotheses. Direct comparison of alternative hypotheses can be accomplished through analytical frameworks as well as through experimentation.

KEY WORDS: Bird flight, Competing hypotheses, Manoeuvrability, Strong inference, Wing morphing

Introduction: challenges to studying phenotype–performance relationships in avian flight

Avian flight is an extraordinary behaviour. Observing birds during feeding, competition, mate displays and predator avoidance naturally elicits hypotheses about how their morphology, physiology and biomechanics support their flight characteristics. Despite decades of research with increasingly sophisticated tools, however, many of these ‘obvious’ hypotheses are either weakly or not at all supported. For example, a perennial research question asks how the morphology of the wing, which varies dramatically in size and shape across species, has been shaped by selection to the ecology and energetics of flight (Greenewalt, 1975; Magnan, 1922; Norberg, 1990; Pennycuik, 2008; Rayner, 1988; Savile, 1957). As flight is a particularly power-intensive form of locomotion, one would expect there to be a tight link between flight and wing traits across clades. Surprisingly, phylogenetically informed analyses have yet to find anything more than a weak association between wing shape and flight styles across avian families (Baliga et al.,

2019; Baumgart et al., 2021; Taylor and Thomas, 2014; Wang and Clarke, 2015).

When there is only a weak association between phenotypic traits and ecology, despite *a priori* predictions, it suggests that either our predictions are based on flawed reasoning or we have failed to measure critical phenotypic or ecological variables. An illustration of this challenge comes from work on cichlid fishes, which exhibit a remarkable variety of feeding morphologies in a hyper-diverse lineage. The morphological diversification suggests that this radiation was driven by intense competition for resources. Studies by Karel Liem and others in the 1970s and 80s, however, indicated that many cichlids retained dietary flexibility, despite possessing specialized morphologies (see Liem, 1980). This phenomenon, wherein species with specialized feeding morphologies consume generic resources, has become known as Liem’s paradox (Greenwood, 1984; Liem, 1980; Mayr, 1984; Robinson and Wilson, 1998). The most common explanation for Liem’s paradox is that feeding specializations can evolve if they allow species to utilize ‘fallback foods’, especially if the costs for maintaining a generalist diet are low (Benkman and Miller, 1995; Liem and Kaufman, 1984; McKaye and Marsh, 1983; Robinson and Wilson, 1998; Smith, 1990). Thus, species’ traits may reflect adaptation not just for general circumstances but also for periods of intense challenge that may not be easily observed.

Another potential confound in comparative studies of avian flight is Simpson’s paradox, in which a trend that is observed within groups disappears or even reverses when the groups are combined (Pearson et al., 1899; Simpson, 1951). For example, the scaling relationships of morphology, wingbeat kinematics and power against body mass in hummingbirds differ when analysed intra- versus inter-specifically (Skandalis et al., 2017) (Fig. 1). Within hummingbird species, as body mass increases, larger hummingbirds beat their wings faster and have higher induced power requirements for hovering. Evolution, however, has scope to compensate for unfavourable scaling. Across hummingbirds, larger species also have larger wings with the consequence that the interspecific relationships of wing velocity and induced power requirements do not scale with body mass. Such differences in scaling within and between groups is possible for multiple aspects of the avian phenotype. Potential targets of selection include wing shape, size and material properties as well as variation in skeletal and muscle morphology.

In this Commentary, we discuss several examples in which long-standing questions about mechanisms of animal flight were resolved either through new methods or through new frameworks for investigation and analysis. Although there is no single solution, we recommend that integrating biomechanics, physiology and ecology will require at least three approaches. First, it is important to measure behavioural and ecological variation, regardless of frequency or ubiquity. Key insights can come from new logging technologies that

Department of Zoology, University of British Columbia, Vancouver, BC, Canada, V6T 1Z4.

*Author for correspondence (doug@zoology.ubc.ca)

 D.L.A., 0000-0002-1364-3617; V.B.B., 0000-0002-9367-8974; A.B.L., 0000-0001-7530-7830; P.L., 0009-0008-6897-430X; E.R.P., 0000-0002-1944-3755; J.S.T., 0000-0002-8673-8406

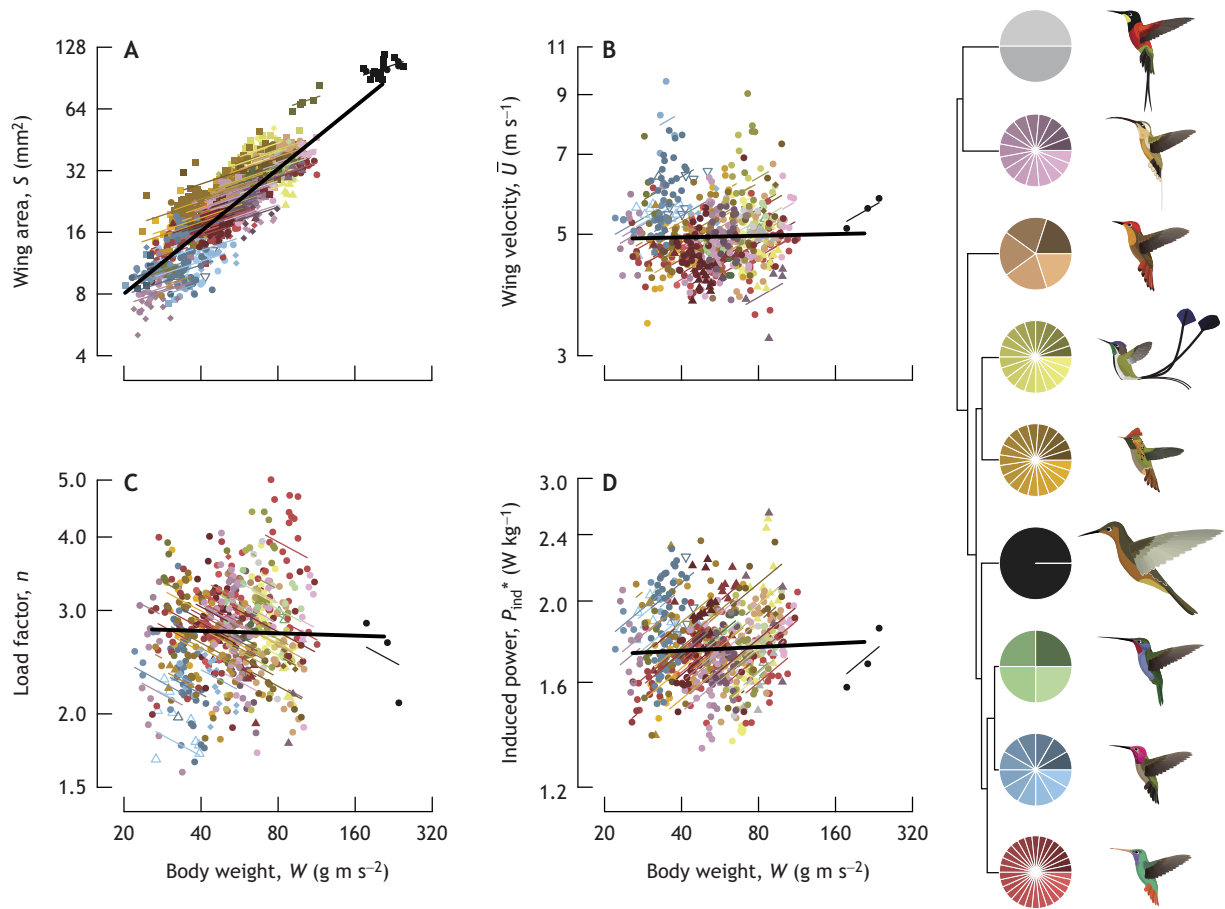


Fig. 1. An example of Simpson's paradox from hummingbirds. (A) Wing area S , (B) wing velocity $\bar{U}$, (C) load factor and (D) induced power P_{ind}^* plotted against body weight (note, data are shown on a log scale). The scaling of morphology, wingbeat kinematics and power requirements for flight was distinct at intra-specific versus inter-specific levels. Individual observations and within-species fits are coloured and shaded by species within clade, according to the cartoon phylogeny on the right. The thick black line in each panel provides the among-species fit, which is distinct from the within-species slopes. Symbols represent different sources of data. Figure reproduced from Skandalis et al. (2017).

allow for measurement of rare behaviours (Tanigaki et al., 2024). Second, this greater understanding of relevant behaviours should be used to develop multiple hypotheses of which phenotypes are most relevant for performance and fitness. Finally, strength of support for these different hypotheses can be distinguished either through a crucial experiment or through competitive analytical frameworks (Fudge, 2014; Platt, 1964). This concept has been introduced before, dating back at least to 1897 (Chamberlin, 1897), but we recommend that greater application of this competitive hypothesis testing is needed to facilitate the next phase of investigations in the biomechanics, physiology and ecology of avian flight.

Observing behavioural variation in space and time for generating hypotheses in avian flight

Prior to delving directly into a manipulative experiment to investigate a new aspect of avian flight, the essential first step is to make observations of bird behaviours across space and time. Breadth is important for the simple reason that restricted observations may cause one to miss a key selective force. Another important issue is that most biological processes are non-linear. The consequence is that as one adds breadth of sampling, it is also important for the analysis to consider Jensen's inequality, where performance at average conditions is seldom equal to average performance across a range of conditions. A discussion of Jensen's inequality and its consequences was presented in an earlier

Commentary in Journal of Experimental Biology by Denny (2017). It is obviously not possible to obtain full resolution of an organism's life history, but we highlight a few examples in which efforts at more fulsome ranges of observations yielded new insights.

The environment, either directly or indirectly through species interactions, is a source of selection in natural populations. Because the environment changes over time, so does the strength of selection it exerts. A classic example of this phenomenon in birds comes from work on Galápagos finches, albeit not from a flight trait. Galápagos finches have experienced rapid morphological diversification in size and feeding morphology. On Daphne Major in the late 1970s, both resident finch species (*Geospiza fortis* and *Geospiza scandens*) exhibited generalist diets with some overlap (Boag and Grant, 1984). Seasonal variation caused shifts in their diets. After the breeding season, the two species fed primarily on seeds, but with different proportions of seed sizes in their diets. A major drought in 1977 caused a reduction in seed abundance and a bias for larger, harder seeds (Boag and Grant, 1981). This shift in seed size distribution drove strong selection in *G. fortis* for larger beaks that could crack the successively harder seeds that remained.

Long-term evolution studies, especially those capturing seasonal variation like that on the finches of Daphne Major, are rare for flight traits. However, alongside seasonal variation, unpredictable and extreme events (drought, storms) can also shape phenotypic traits (Grant and Grant, 2002; Grant et al., 2017). Bumpus (1898)

documented selection in house sparrows on total length, wingspan, humeral length, femoral length and weight, although these conclusions were based on the survival of immobilized individuals that happened to be brought to the Anatomical Laboratory of Brown University during a severe winter storm. Brown and Brown (1998) reported selection on wing and tail asymmetry of cliff swallows during an extreme cold spell. Taken together with work on Galápagos finches, these results illustrate that environmental selection can vary both within and across years, and that measuring temporal variation in the environment can be key to understanding phenotypic evolution, including traits related to avian flight.

Spatial variation in selective pressures can be another important driver of phenotype. How birds use space is often not well quantified. A recent example comes from a radio telemetry study of two species of thrushes (*Turdus*) in Brazil (Da Silveira et al., 2016). Although both were originally considered habitat generalists, fine-scale tracking revealed a strong preference for foraging in forest edges. Knowledge of habitat use can guide the study of ecological competition and associated changes in morphology. A recent example comes from vultures. Turkey vultures are more likely to forage in forested, canopy-covered areas, whereas black vultures more frequently use open areas. As a result of niche partitioning, turkey vultures spend more time foraging and have a greater wing length to body mass ratio that should allow for more efficient soaring (Byrne et al., 2019). A final point to consider about spatial variation is that co-evolutionary interactions, which can drive changes in phenotype, can also be spatially heterogeneous (Thompson, 2005). This principle is well documented in feeding morphology. Several notable examples include hummingbird bill morphology co-evolving with flower corollas (Soteris et al., 2018; Temeles and Kress, 2003).

Although each of these examples illustrates how careful observations suggest potential mechanisms of avian function, these studies do not span the full range of approaches that include both observations and experimental perturbations that directly test predictions from the hypothesis. Quantifying avian behaviour across space and time is becoming feasible with technologies that include GPS tracking, marker-based tracking and RFID logging. Although there are a number of tools available to quantify biomechanical and physiological performance (e.g. high-speed videography, computational fluid dynamics, respirometry and doubly labelled water), it can be challenging to explicitly link these measurements with fine-scale behaviour. Advances in both behavioural quantification and experimental manipulations continue and, as molecular genetic tools become more accessible in birds, these approaches should connect better. In the meantime, however, observations across the environment provide the raw materials to refine hypotheses and predictions that are then amenable to experimental tests.

Refining hypotheses in avian flight through experimental and natural perturbations

A strategy that can be helpful for matching behavioural and mechanistic measurements in avian flight is to introduce experimental manipulations that challenge biomechanical limits. The goal of these experiments is to magnify differences in biomechanical performance between individuals, developmental stages or species to clarify relationships between ecological forces and phenotype. This strategy has helped to advance several discoveries in flight biomechanics.

An example of a flight trait that has been studied and manipulated in several bird species is tail length. Extreme tail lengths, such as

those in widowbirds, are counterintuitive because these seem to interfere with the birds' day-to-day activities. This paradox led Darwin (1871) and Fisher (1930) to postulate that extreme tail length was due to female choice. To probe this hypothesis, Malte Andersson (1982) and Staffan Andersson (1992) manipulated tail lengths of male widowbirds and examined impacts on reproductive success. They also considered other factors that could influence female choice or tail length selection pressures such as territory quality and intraspecific competition. The studies found a strong positive correlation between male reproductive success and tail length, supporting Darwin's and Fisher's hypothesis that female choice drives exaggeration of ornate tails. This hypothesis has also been addressed through manipulations of tail length in barn swallows, with the similar conclusion that longer tails are derived from sexual selection rather than natural selection (Møller et al., 1998).

The studies of widowbirds and barn swallows did not include direct measurements of how tail length affects flight performance. Norberg (1994) combined observations of barn swallows in flight with wind tunnel measurements from a restrained swallow with a spread tail. This collection of measurements suggested that elongated outer tail feathers can improve manoeuvrability. Measurements of flying Anna's hummingbirds with manipulated tails in a wind tunnel were performed by Clark and Dudley (2009). Several manipulations were made including shortening tails, removing tails and adding the elongated tail of the male red-billed streamertail hummingbird. In each case, metabolic rate was measured through oxygen consumption during forward flight in a wind tunnel. Both elongated and shortened or removed tails negatively impacted hummingbird flight speed, but only elongated tails increased metabolic costs. Overall effects on metabolic costs and forward flight speed were modest. Clark (2011) also used tail manipulations in conjunction with high-speed videography to measure the effects of tail length on low-speed escape manoeuvres in red-billed streamertail hummingbirds. He found that low-speed manoeuvres were not significantly influenced by elongated tails, further supporting the hypothesis that ornate tails are not strongly constrained by flight costs.

The most common type of experimental challenge for studying bird flight is the use of wind tunnels, which has been reviewed elsewhere (Hedenström and Lindström, 2017). Here, we highlight several less common methods for using physical challenges to gain insight into avian flight mechanisms.

Dial and Jackson (2011) asked whether flight traits varied across the lifespan of Australian brush turkeys. They focused on the behaviour of wing-assisted incline running (WAIR), in which birds use their wings to increase foot contact with the ground for walking up steep inclines. Hatchling brush turkeys are unusually adept at WAIR as a predatory avoidance behaviour and consistently outperform adults. Challenging the limits of WAIR for Australian brush turkeys across the lifespan revealed that wing loading and WAIR capacity are inversely correlated, with minimal wing loading predicting maximum performance as hatchlings. For these precocial birds, predatory selection is greatest as hatchlings and declines as birds reach adulthood, coinciding with neuromuscular development, a large change in body mass (0.2–1.8 kg) and a transition from forelimb-propelled locomotion to hindlimb-propelled locomotion.

A change in elevation is a natural physical challenge to avian flight because this corresponds to systematic decreases in air density and oxygen partial pressure, as well as average decreases in temperature and increases in wind speed. It is well established that

elevation corresponds with shifts in avian flight traits. For example, a study of 105 North American bird species with considerable individual sample size and phylogenetic resolution demonstrated that relative wing length increases with greater elevation among species (Youngflesh et al., 2022). Earlier studies of Andean hummingbirds demonstrated that with increasing elevation, species increase in body mass, wing area and wing stroke amplitude during hovering flight (Altshuler and Dudley, 2003; Altshuler et al., 2004). A surprising result was that the increase in wing area and wing stroke amplitude were sufficient to offset much of the cost of hovering at lower air density such that the mechanical costs of hovering were not significantly affected by elevation among species. In addition to elevation, the birds were also challenged using transient load lifting, in which a hummingbird had a rubber harness attached to a string of weights. As the birds ascended to escape, they lifted progressively more weight until reaching a maximum lift. At all elevations, hummingbirds reached maximum load lifting when the wing stroke amplitude reached ~180 deg. Any further increases in wing stroke amplitude would cause wing interference, both physically and aerodynamically. Taken together, the combined physical challenges of elevation and transient load lifting revealed that although the costs of hovering flight are constant across elevations, the excess mechanical power for other behaviours declines systematically with increased elevation.

Competitive frameworks

The concept of using competing hypotheses was most famously explained by John Platt (1964) in his article titled, 'Strong Inference: Certain systematic methods of scientific thinking may produce much more rapid progress than others'. An updated framing of this seminal paper is available in another Commentary in *Journal of Experimental Biology* by Doug Fudge (2014). As Fudge explains, the idea of having multiple working hypotheses was elucidated as early as 1897 by Thomas Chamberlin (1897).

Platt's insight was that some disciplines of science were advancing rapidly through important and challenging questions by systematically applying the following steps: '(1) devising alternative hypotheses; (2) devising a crucial experiment with alternative possible outcomes, each of which will exclude one or more of the hypotheses; (3) carrying out the experiments so as to get a clean result; (1') recycling the procedure'. Why is this approach different from other approaches? The answer is that it is often not informative to perform an experimental manipulation that simply perturbs a system using whatever technology happens to be available. A more powerful technique for moving beyond simple associations is to devise a framework where the result of the experiment(s) can be interpreted in a context that indicates causality. Here, we review two of the attempts that our laboratory has made to apply this framework.

We applied the Platt approach to improve understanding of how the manoeuvring flight of hummingbirds is influenced by challenges to flight along an elevational gradient (Segre et al., 2016). In the first phase of the study, we transported the same individual Anna's hummingbirds to both high and low elevations and measured their manoeuvring performance. This perturbation revealed that many metrics of manoeuvrability declined with altitude. Because the decline in manoeuvring flight performance at high elevation could be due to either lower air density or lower oxygen partial pressure, we conducted an experiment with a different group of Anna's hummingbirds that would allow us to distinguish between these alternatives. Using a large, sealed chamber, we used gas mixtures to manipulate the air while

measuring manoeuvring performance. Replacing normal air with nitrogen caused a systematic decrease in oxygen availability while maintaining near-constant air density. Replacing normal air with heliox (79% helium, balance oxygen) systematically reduced air density while maintaining constant oxygen availability. This experiment provided a clear result that the decreases in hummingbird manoeuvrability were associated with the decrease in air density rather than oxygen partial pressure.

For many questions in avian flight mechanics, especially evolutionary questions, it may not currently be feasible to design a manipulative experiment to distinguish among alternative hypotheses. In these cases, data analysis can be constructed in a way that is analogous to a crucial experiment in that it allows for competing and comparing alternative hypotheses. We applied this approach to gain insight into the evolution of the avian wing, and specifically to understand why static measurements of wing shapes are not well correlated with flight style or body mass (Wang and Clarke, 2015). We hypothesized that dynamic measurements of wing range of motion may be more constrained by the demands of flight and therefore more closely matched to flight style (Baliga et al., 2019). This type of evolutionary comparative question is not amenable to experimentation, but we were able to measure both static wing shape and dynamic range of motion from a sufficiently large number of species to test alternative hypotheses in a phylogenetic context. With the overall analytical approach in place, we tested several aspects of range of motion: extension, linkage trajectory and out of plane motion (bending and twisting). Dynamic range of motion was more informative than static wing shape in all cases, and it was also more evolutionarily labile. Additional insight included that some features of range of motion (linkage trajectory) were almost exclusively associated with flight style whereas others (out of plane motion) were almost exclusively associated with body mass.

We highlight these two studies because they provide examples in which placing alternative hypotheses in competition helped advance understanding of a long-standing problem. They also illustrate that when a 'crucial' manipulative experiment is not available, alternative hypotheses can also be compared through an analytical framework. One important caveat is that testing alternative hypotheses through competition is only useful if the hypotheses are plausible explanations of observed data, and ideally have some level of previous support. In the next section, we discuss how to design tests that distinguish among plausible alternative hypotheses.

Approaches for testing alternative hypotheses

In another Commentary paper in *Journal of Experimental Biology*, Fudge and Turko (2020) provide detailed guidance for designing a crucial experiment. They point out that a difficult transition in Platt's (1964) approach is to move from step 1 (generating alternative hypotheses from observations) to step 2 (designing a crucial experiment). The bridge, they explain, is to generate high-quality predictions that are both critical and persuasive. A prediction is critical if when found to be false, it would cause one to reject the hypothesis. The concept of a critical prediction is therefore an essential component of falsifiability (Popper, 1959). A prediction is persuasive if when found to be true, it would increase confidence in the hypothesis. Although it has long been recognized that good predictions are critical, unless a prediction is also persuasive, it may yield relatively little insight. Fudge and Turko's (2020) Commentary includes worked examples and a checklist for answering new questions. It is key reading for the experimental

approach. We will focus on analytical approaches that are in many ways parallel to their recommendations for experiments.

It is possible to use statistical tests to compete alternative hypotheses about avian flight using non-experimental data. There is, however, no one specific test to do this. Rather, there are approaches that can facilitate hypothesis comparison. We begin by making a point that is logically obvious but may seem controversial. When two distributions based on sufficiently sampled data are so distinct that there is no overlap whatsoever, then a statistical test of their difference is not needed. The general rule in biology is that many of the datasets are subject to multiple sources of variation and are therefore inherently messy. This is the case for most avian behavioural data, for example. It is therefore essential to perform at least two checks to ensure that the data are suitable to the question, a repeatability check and a sensitivity check. Repeatability involves a statistical analysis of the consistency of repeated measures. If measurements of the same trait from the same individual have low or zero repeatability, then that trait is not suitable for further study (Segre et al., 2015). Sensitivity analyses determine how the uncertainty of statistical models relates to uncertainty in the underlying variables. These analyses involve recalculating outcomes under reasonable, alternative assumptions (Harvey et al., 2019). If the results are highly dependent on the assumptions, then there is a fundamental problem with the data collection and/or the hypotheses.

Biological data are often not well summarized by the mean and standard deviation. It can therefore be challenging to determine whether measured samples are from distinct distributions (alternative hypothesis) or from the same distribution (null hypothesis). A permutation test is a powerful non-parametric method that requires no assumptions about the distribution of the population(s) (Ernst, 2004). It assesses the significance of an observed effect against repeated random shuffling of the data. For example, if you have 30 measurements of traits from different individuals (or different species), you can create a shuffled distribution by randomly pairing the traits from different individuals (or species). This null distribution is biologically plausible because it is based on the real data. If the strength of the association for the actual data is similar to the association strength of the shuffled data, there is no meaningful association among your variables of interest. This approach can also allow one to distinguish among alternative hypotheses by comparing the results of their respective permutation tests. Permutation tests are commonly used in some disciplines within the life sciences and it is also possible to perform permutation tests within a phylogenetic framework (Adams and Collyer, 2015).

We provide two examples from our own work in which we applied multiple permutation tests to distinguish alternative hypotheses about bird flight (Fig. 2). In a comparative study of manoeuvring performance among species (Dakin et al., 2018), we used two shuffled distributions to help determine whether each of 25 hummingbird species could be distinguished based on either manoeuvring performance or morphology. If we randomly guessed the identities of these species, we would anticipate being correct 4% of the time, but such random guesses would ignore the data entirely. Instead, we created two null distributions, one by shuffling manoeuvring traits and the other by shuffling morphological traits. Models fitted to these null distributions correctly predicted species <20% of the time. The actual manoeuvrability data were, in contrast, able to predict the correct species 34% of the time. Although this was lower accuracy than what was provided by the morphological data (65% accuracy), the

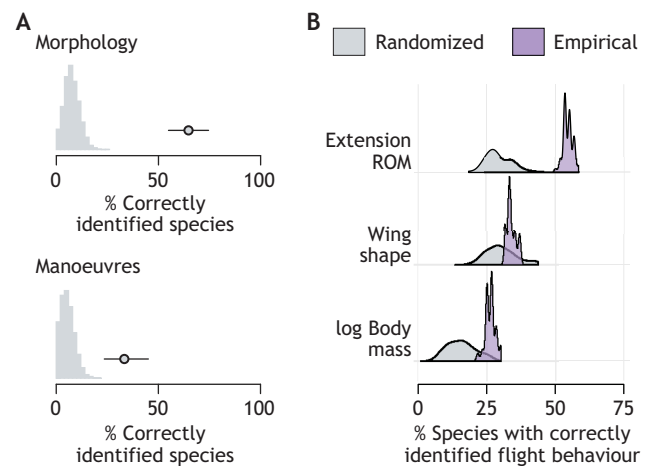


Fig. 2. Examples of permutation tests to compare hypotheses about hummingbird manoeuvrability and avian wing morphing. (A) Both morphology (top) and manoeuvrability are sufficient to correctly identify hummingbird species compared with chance. Shuffled data are shown as grey distributions. Reproduced with permission from Dakin et al. (2018). (B) Range of motion (ROM) correctly identifies avian species' flight behaviours more reliably than wing shape or body mass. Reproduced with permission from Baliga et al. (2019).

highly dynamic measurements of manoeuvrability were much more informative than expected by chance. In another example, from our comparative study of wing range of motion (Baliga et al., 2019), we used three permutation tests to ask how 1 of 9 flight behaviours could be predicted from range of motion, static shape or body mass. This revealed that the range of motion data were the most informative by far. In the permutation tests, these data had the highest accuracy (>50%) and were the only distribution that had no overlap with the shuffled (null) distribution. These examples illustrate how inherently noisy data may contain hidden information that is only revealed when compared with a biologically plausible null distribution.

Another challenge for studies of avian flight performance is that there are often multiple sources of variation, and it is difficult to know *a priori* which sources are meaningful. Statistical models with more parameters will almost always explain more variance than models of the same type with fewer parameters. Models with many parameters may be too tightly tuned to a specific dataset and therefore not be generalizable. This is referred to as overfitting. It is therefore essential to compare the explanatory power among alternative hypotheses by comparing the goodness of fit or predictive power among corresponding statistical models while penalizing for model complexity (Burnham and Anderson, 2010). An important choice is the number of models to compete because as more models are fitted, the chances of obtaining a false positive (type I error) increase. There is no diagnostic test to select the number of models, but best practice is to use as few models as possible to capture the variables that are expected to be most meaningful. Alternative models should be limited to those with predictions that are both critical and persuasive (Fudge and Turko, 2020). A good null model is often an intercept-only model because this assumes that the predictor variables have no value.

Model comparisons were essential features for research in our laboratory on hummingbird manoeuvrability and the comparative biology of avian wing morphing. To identify which variables predicted each metric of hummingbird manoeuvring performance,

we compared models that included the random effect of individual bird and the fixed effects of body mass, wing length, wing aspect ratio and the amount lifted during maximal load lifting (Segre et al., 2015). This analysis revealed the importance of burst muscle power for many metrics of manoeuvring performance in hummingbirds. Including the null model was particularly helpful in our study of wing morphing (Baliga et al., 2019). In the case of wing shape, the intercept-only model had the highest support, which indicated that neither body mass nor flight behaviour had any explanatory power for this metric.

Recommendations for implementing a competitive hypothesis framework

The framework of using alternative hypotheses to study avian flight requires that all of the alternative hypotheses are biologically plausible. Aside from a reasonable null hypothesis, there is no value to including alternative hypotheses for which there is no support, either from new observations or from previous studies. Although this instruction may seem obvious at first glance, it needs to be kept in mind throughout because it can be challenging to develop a single hypothesis with critical and persuasive predictions, and much more so for multiple hypotheses. Strong inference is usually described as progressing directly from observation to experimental design to measurement and analysis. In reality, developing the correct set of questions, experiments and analyses is a non-linear process. For example, we have found that our questions often improve considerably as we begin an experiment.

The central goal of competing alternative hypotheses against each other is to break one or more of them. Hypotheses that withstand multiple attempts to disprove them are by definition strongly supported. New technologies typically provide new mechanisms to perform manipulative experiments. Our understanding of the biomechanics and physiology of bird flight has clearly advanced through the increased availability of biologists (Hedenström and Hedh, 2024), flow measurements (Usherwood et al., 2020; Warrick et al., 2005) and force measurements (Lentink et al., 2007; Tobalske et al., 2003). Breaking hypotheses is a regular feature of research with genetic model organisms such as *Drosophila melanogaster*, zebrafish and mouse. Genetic tools provide powerful techniques for testing both necessity and sufficiency through a combination of imaging, activation, inactivation and reporting. Tools from model organisms are becoming more available in non-genetic models, including birds (Roberts et al., 2012; Xiao et al., 2018). Implementing such techniques can be challenging and time consuming but they have considerable potential to test predictions that are both critical and persuasive.

What does it mean when a hypothesis that was supported by observations or previous research is broken? The phenomenon is a common feature of complex systems because of confounding or latent variables. Bird flight is a complex behaviour that is not a simple product of discrete traits. Often, multiple traits determine the performance of behaviour (so-called many-to-one mapping). These traits may trade-off with one another, contribute additively or exhibit more complex interactions to determine performance and fitness. Similarly, most traits influence performance across multiple behaviours (one-to-many mapping). When a hypothesis is disproven through a manipulative, crucial experiment, the result provides the foundation for the next round of study. As explained so powerfully by Platt (1964), this framework can greatly accelerate research when applied consistently and we believe the time is ripe for committing to this approach in the study of avian flight.

Acknowledgements

We thank Journal of Experimental Biology and the meeting organizers for a stimulating symposium and the opportunity to write this Commentary.

Competing interests

The authors declare no competing or financial interests.

Funding

Our research on the biomechanics and physiology of avian flight is funded by a Discovery Grant from the Natural Sciences and Engineering Research Council of Canada (NSERC, RGPIN-2021-02977). Some authors have also been supported through graduate and postdoctoral fellowships from the Michael Smith Foundation for Health Research, the Natural Sciences and Engineering Research Council of Canada, the US National Science Foundation, and the University of British Columbia.

Special Issue

This article is part of the special issue 'Integrating Biomechanics, Energetics and Ecology in Locomotion', guest edited by Andrew A. Biewener and Alan M. Wilson. See related articles at https://journals.biologists.com/jeb/issue/228/Suppl_1.

References

- Adams, D. C. and Collyer, M. L. (2015). Permutation tests for phylogenetic comparative analyses of high-dimensional shape data: what you shuffle matters: brief communication. *Evolution* **69**, 823–829. doi:10.1111/evo.12596
- Altshuler, D. L. and Dudley, R. (2003). Kinematics of hovering hummingbird flight along simulated and natural elevational gradients. *J. Exp. Biol.* **206**, 3139–3147. doi:10.1242/jeb.00540
- Altshuler, D. L., Dudley, R. and McGuire, J. A. (2004). Resolution of a paradox: hummingbird flight at high elevation does not come without a cost. *Proc. Natl. Acad. Sci. USA* **101**, 17731. doi:10.1073/pnas.0405260101
- Andersson, M. (1982). Female choice selects for extreme tail length in a widowbird. *Nature* **299**, 818–820. doi:10.1038/299818a0
- Andersson, S. (1992). Female preference for long tails in lekking Jackson's widowbirds: experimental evidence. *Anim. Behav.* **43**, 379–388. doi:10.1016/S0003-3472(05)80098-3
- Baliga, V. B., Szabo, I. and Altshuler, D. L. (2019). Range of motion in the avian wing is strongly associated with flight behavior and body mass. *Sci. Adv.* **5**, eaaw6670. doi:10.1126/sciadv.aaw6670
- Baumgart, S. L., Sereno, P. C. and Westneat, M. W. (2021). Wing shape in waterbirds: morphometric patterns associated with behavior, habitat, migration, and phylogenetic convergence. *Integr. Org. Biol.* **3**, obab011. doi:10.1093/iob/obab011
- Benkman, C. W. and Miller, R. E. (1995). Morphological evolution in response to fluctuating selection. *Evolution* **50**, 2499–2504. doi:10.2307/2410716
- Boag, P. T. and Grant, P. R. (1981). Intense natural selection in a population of Darwin's finches (Geospizinae) in the Galápagos. *Science* **214**, 82–85. doi:10.1126/science.214.4516.82
- Boag, P. T. and Grant, P. R. (1984). Darwin's Finches (*Geospiza*) on Isla Daphne Major, Galapagos: breeding and feeding ecology in a climatically variable environment. *Ecol. Monogr.* **54**, 463–489. doi:10.2307/1942596
- Brown, C. R. and Brown, M. B. (1998). Intense natural selection on body size and wing and tail asymmetry in cliff swallows during severe weather. *Evolution* **52**, 1461–1475. doi:10.2307/2411315
- Bumpus, H. (1898). *The Elimination of the Unfit as Illustrated by the Introduced Sparrow. Biological lectures delivered at the Marine Biological Laboratory of Wood's Hole.* Gin.
- Burnham, K. P. and Anderson, D. R. (2010). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach.* New York: Springer.
- Byrne, M. E., Holland, A. E., Turner, K. L., Bryan, A. L. and Beasley, J. C. (2019). Using multiple data sources to investigate foraging niche partitioning in sympatric obligate avian scavengers. *Ecosphere* **10**, e02548. doi:10.1002/ecs2.2548
- Chamberlin, T. C. (1897). Studies for students: the method of multiple working hypotheses. *J. Geol.* **5**, 837–848. doi:10.1086/607980
- Clark, C. J. (2011). Effects of tail length on an escape maneuver of the red-billed streamertail. *J. Ornithol.* **152**, 397–408. doi:10.1007/s10336-010-0550-8
- Clark, C. J. and Dudley, R. (2009). Flight costs of long, sexually selected tails in hummingbirds. *Proc. R. Soc. B Biol. Sci.* **276**, 2109–2115. doi:10.1098/rspb.2009.0090
- Da Silveira, N. S., Niebuhr, B. B. S., Muylaert, R. L., Ribeiro, M. C. and Pizo, M. A. (2016). Effects of land cover on the movement of frugivorous birds in a heterogeneous landscape. *PLoS One* **11**, e0156688. doi:10.1371/journal.pone.0156688
- Dakin, R., Segre, P. S., Straw, A. D. and Altshuler, D. L. (2018). Morphology, muscle capacity, skill, and maneuvering ability in hummingbirds. *Science* **359**, 653–657. doi:10.1126/science.aao7104
- Darwin, C. R. (1871). *The Descent of Man, and Selection in Relation to Sex.* London: John Murray.

- Denny, M. (2017). The fallacy of the average: on the ubiquity, utility and continuing novelty of Jensen's inequality. *J. Exp. Biol.* **220**, 139–146. doi:10.1242/jeb.140368
- Dial, K. P. and Jackson, B. E. (2011). When hatchlings outperform adults: locomotor development in Australian brush turkeys (*Alectura lathami*, Galliformes). *Proc. R. Soc. B Biol. Sci.* **278**, 1610–1616. doi:10.1098/rspb.2010.1984
- Ernst, M. D. (2004). Permutation methods: a basis for exact inference. *Stat. Sci.* **19**, 676–685. doi:10.1214/088342304000000396
- Fisher, R. A. (1930). *The Genetical Theory of Natural Selection: by R.A. Fisher; Edited With a Foreword and Notes by J.H. Bennett*. Oxford: Oxford University Press.
- Fudge, D. S. (2014). Fifty years of J. R. Platt's strong inference. *J. Exp. Biol.* **217**, 1202–1204. doi:10.1242/jeb.104976
- Fudge, D. S. and Turko, A. J. (2020). The best predictions in experimental biology are critical and persuasive. *J. Exp. Biol.* **223**, jeb231894. doi:10.1242/jeb.231894
- Grant, P. R. and Grant, B. R. (2002). Unpredictable evolution in a 30-year study of Darwin's finches. *Science* **296**, 707–711. doi:10.1126/science.1070315
- Grant, P. R., Grant, B. R., Huey, R. B., Johnson, M. T. J., Knoll, A. H. and Schmitt, J. (2017). Evolution caused by extreme events. *Philos. Trans. R. Soc. B Biol. Sci.* **372**, 20160146. doi:10.1098/rstb.2016.0146
- Greenewalt, C. H. (1975). The flight of birds: the significant dimensions, their departure from the requirements for dimensional similarity, and the effect on flight aerodynamics of that departure. *Trans. Am. Philos. Soc.* **65**, 1. doi:10.2307/1006161
- Greenwood, P. H. (1984). African cichlids and evolutionary theories. In *Evolution of Fish Species Flocks* (ed. A. A. Echelle and I. Kornfield), pp. 141–154. Orono, Maine: University of Maine at Orono Press.
- Harvey, C., Baliga, V. B., Lavoie, P. and Altshuler, D. L. (2019). Wing morphing allows gulls to modulate static pitch stability during gliding. *J. R. Soc. Interface* **16**, 20180641. doi:10.1098/rsif.2018.0641
- Hedenström, A. and Hedh, L. (2024). Seasonal patterns and processes of migration in a long-distance migratory bird: energy or time minimization? *Proc. R. Soc. B Biol. Sci.* **291**, 20240624. doi:10.1098/rspb.2024.0624
- Hedenström, A. and Lindström, Å. (2017). Wind tunnel as a tool in bird migration research. *J. Avian Biol.* **48**, 37–48. doi:10.1111/jav.01363
- Lentink, D., Müller, U. K., Stamhuis, E. J., de Kat, R., van Gestel, W., Veldhuis, L. L. M., Henningson, P., Hedenström, A., Videler, J. J. and van Leeuwen, J. L. (2007). How swifts control their glide performance with morphing wings. *Nature* **446**, 1082–1085. doi:10.1038/nature05733
- Liem, K. F. (1980). Adaptive significance of intra- and interspecific differences in the feeding repertoires of cichlid fishes. *Am. Zool.* **20**, 295–314. doi:10.1093/icb/20.1.295
- Liem, K. F. and Kaufman, L. S. (1984). Intraspecific macroevolution: functional biology of the polymorphic cichlid species *Cichlasoma minckleyi*. In *Evolution of fish species flocks* (ed. A. A. Echelle and I. Kornfield), pp. 203–216. Orono, Maine: University of Maine at Orono Press.
- Magnan, A. (1922). *Les Caractéristiques des Oiseaux Suivant le Mode de Vol. Leur Application à la Construction des Avions*. Masson et Cie.
- Mayr, E. (1984). Evolution of fish species flocks: a commentary. In *Evolution of fish species flocks* (ed. A. A. Echelle and I. Kornfield), pp. 3–12. Orono, Maine: University of Maine at Orono Press.
- McKaye, K. R. and Marsh, A. (1983). Food switching by two specialized algae-scraping cichlid fishes in Lake Malawi, Africa. *Oecologia* **56**, 245–248. doi:10.1007/BF00379697
- Møller, A. P., Barbosa, A., Cuervo, J. J., Lope, F. D., Merino, S. and Saino, N. (1998). Sexual selection and tail streamers in the barn swallow. *Proc. R. Soc. Lond. B Biol. Sci.* **265**, 409–414. doi:10.1098/rspb.1998.0309
- Norberg, U. M. (1990). *Vertebrate Flight*. Berlin: Springer.
- Norberg, R. A. (1994). Swallow tail streamer is a mechanical device for self deflection of tail leading edge, enhancing aerodynamic efficiency and flight manoeuvrability. *Proc. R. Soc. Lond. B Biol. Sci.* **257**, 227–233. doi:10.1098/rspb.1994.0119
- Pearson, K., Lee, A. and Bramley-Moore, L. (1899). VI. Mathematical contributions to the theory of evolution. —VI. Genetic (reproductive) selection: inheritance of fertility in man, and of fecundity in thoroughbred racehorses. *Philos. Trans. R. Soc. Lond. Ser. Contain. Pap. Math. Phys. Character* **192**, 257–330.
- Pennycuik, C. J. (2008). *Modelling the Flying Bird*. Amsterdam: Academic Press.
- Platt, J. R. (1964). Strong Inference: Certain systematic methods of scientific thinking may produce much more rapid progress than others. *Science* **146**, 347–353. doi:10.1126/science.146.3642.347
- Popper, K. (1959). *The Logic of Scientific Discovery*. London: Julius Springer, Hutchinson & Co.
- Rayner, J. M. V. (1988). Form and function in avian flight. In *Current Ornithology* (ed. R. F. Johnston), pp. 1–66. Boston, MA: Springer US.
- Roberts, T. F., Gobes, S. M. H., Murugan, M., Óveczky, B. P. and Mooney, R. (2012). Motor circuits are required to encode a sensory model for imitative learning. *Nat. Neurosci.* **15**, 1454–1459. doi:10.1038/nn.3206
- Robinson, B. W. and Wilson, D. S. (1998). Optimal foraging, specialization, and a solution to Liem's paradox. *Am. Nat.* **151**, 223–235. doi:10.1086/286113
- Savile, O. B. O. (1957). Adaptive evolution in the avian wing. *Evolution* **11**, 212–224. doi:10.2307/2406051
- Segre, P. S., Dakin, R., Zordan, V. B., Dickinson, M. H., Straw, A. D. and Altshuler, D. L. (2015). Burst muscle performance predicts the speed, acceleration, and turning performance of Anna's hummingbirds. *Elife* **4**, e11159. doi:10.7554/eLife.11159
- Segre, P. S., Dakin, R., Read, T. J. G., Straw, A. D. and Altshuler, D. L. (2016). Mechanical constraints on flight at high elevation decrease maneuvering performance of hummingbirds. *Curr. Biol.* **26**, 3368–3374. doi:10.1016/j.cub.2016.10.028
- Simpson, E. H. (1951). The interpretation of interaction in contingency tables. *J. R. Stat. Soc. Ser. B Methodol.* **13**, 238–241. doi:10.1111/j.2517-6161.1951.tb00088.x
- Skandalis, D. A., Segre, P. S., Bahlman, J. W., Groom, D. J. E., Welch, Jr, K. C., Witt, C. C., McGuire, J. A., Dudley, R., Lentink, D. and Altshuler, D. L. (2017). The biomechanical origin of extreme wing allometry in hummingbirds. *Nat. Commun.* **8**, 1047. doi:10.1038/s41467-017-01223-x
- Smith, T. B. (1990). Resource use by bill morphs of an African finch: evidence for intraspecific competition. *Ecology* **71**, 1246–1257. doi:10.2307/1938261
- Soteras, F., Moré, M., Ibañez, A. C., Iglesias, M. D. R. and Cocucci, A. A. (2018). Range overlap between the sword-billed hummingbird and its guild of long-flowered species: an approach to the study of a coevolutionary mosaic. *PLoS One* **13**, e0209742. doi:10.1371/journal.pone.0209742
- Tanigaki, K., Otsuka, R., Li, A., Hatano, Y., Wei, Y., Koyama, S., Yoda, K. and Maekawa, T. (2024). Automatic recording of rare behaviors of wild animals using video bio-loggers with on-board light-weight outlier detector. *PNAS Nexus* **3**, pgad447. doi:10.1093/pnasnexus/pgad447
- Taylor, G. K. and Thomas, A. (2014). *Evolutionary Biomechanics: Selection, Phylogeny, and Constraint*. Oxford: Oxford University Press.
- Temeles, E. J. and Kress, W. J. (2003). Adaptation in a plant-hummingbird association. *Science* **300**, 630–633. doi:10.1126/science.1080003
- Thompson, J. N. (2005). *The Geographic Mosaic of Coevolution*. Chicago: University of Chicago Press.
- Tobalske, B. W., Hedrick, T. L., Dial, K. P. and Biewener, A. A. (2003). Comparative power curves in bird flight. *Nature* **421**, 363–366. doi:10.1038/nature01284
- Usherwood, J. R., Cheney, J. A., Song, J., Windsor, S. P., Stevenson, J. P. J., Dierksheide, U., Nila, A. and Bomphey, R. J. (2020). High aerodynamic lift from the tail reduces drag in gliding raptors. *J. Exp. Biol.* **223**, jeb214809. doi:10.1242/jeb.214809
- Wang, X. and Clarke, J. A. (2015). The evolution of avian wing shape and previously unrecognized trends in covert feathering. *Proc. R. Soc. B Biol. Sci.* **282**, 20151935. doi:10.1098/rspb.2015.1935
- Warrick, D. R., Tobalske, B. W. and Powers, D. R. (2005). Aerodynamics of the hovering hummingbird. *Nature* **435**, 1094–1097. doi:10.1038/nature03647
- Xiao, L., Chattree, G., Oscos, F. G., Cao, M., Wanat, M. J. and Roberts, T. F. (2018). A basal ganglia circuit sufficient to guide birdsong learning. *Neuron* **98**, 208–221.e5. doi:10.1016/j.neuron.2018.02.020
- Youngflesh, C., Saracco, J. F., Siegel, R. B. and Tingley, M. W. (2022). Abiotic conditions shape spatial and temporal morphological variation in North American birds. *Nat. Ecol. Evol.* **6**, 1860–1870. doi:10.1038/s41559-022-01893-x