# Replication and generalization in evolutionary ecology

Russell Bonduriansky

Evolution and Ecology Research Centre, School of Biological, Earth and Environmental Sciences, University of New South Wales, Sydney NSW 2052 Australia

r.bonduriansky@unsw.edu.au

**Abstract**

Several authors have expressed concern about replicability in evolutionary ecology. But what does replicability mean in this field, and how can replication enable generalization? Evolutionary ecologists investigate probabilistic processes across a vast diversity of unique historical entities. I argue that the aims and subject matter of evolutionary ecology complicate inference and necessitate distinct concepts of replication and replicability grounded in biological diversity. I define narrow replicability as the potential to obtain consistent results within a species, and broad replicability as the potential to observe consistent patterns across species possessing relevant attributes. I argue that broad replicability is essential to progress in evolutionary ecology because broad replicability enables generalization.Furthermore, many outcomes in evolutionary ecology can result from more than one underlying mechanism, and tests of a single, simple prediction against a statistical null hypothesis can therefore rarely enable robust inference. Complex predictions that discriminate alternative biological models can enable "severe tests" and provide a better measure of replicability. Likewise, biological diversity and complexity often preclude quantitative predictions, necessitating qualitative tests of hypotheses and qualitative measures of replicability.

**Key words**:

replicability, severe tests, scientific method, hypothesis testing, predictions

## How is evolutionary ecology different?

Replicability, generally defined as the consistency of results across studies addressing the same question (Table 1), is an essential foundation for progress in science. The meaning of replicability is fairly straightforward in physics, chemistry, and some areas of the biological sciences, such as molecular biology and physiology, because research in these fields explores largely deterministic, consistent entities and processes. A molecule's behaviour can be predicted quantitatively from its structure, weight and charge, irrespective of its individual history. Likewise, many basic aspects of organelles, cells, tissues and physiological processes are relatively consistent and deterministic. Failure of replication can therefore be interpreted as a problem with the materials, equipment, experimental procedures, analysis, or reporting.

Evolutionary ecology's subject matter is very different. Evolutionary ecologists seek to understand how probabilistic processes, such as variation, selection and adaptation, shape biological diversity. Each species consists of distinct populations, genotypes, epigenomes and individual phenotypes. At each of these levels, variation is shaped by contingent historical processes involving numerous interacting environmental and genetic factors. Behaviour is especially challenging to study because it is shaped by the properties of neural networks whose development is partly stochastic and partly shaped by individual experience. Consequently, a replicated study can yield different results if done using a different species or genotype. And even for the same species and genotype, the results could vary depending on individuals' environment, age, or social experience.

Evolutionary ecology's subject matter also complicates prediction and analysis. The complexity and diversity of life means that evolutionary ecologists often lack sufficient knowledge to make highly specific predictions, such as which traits will respond to experimental treatment. For example, a treatment effect on fitness could manifest in multiple ways, and it is often not possible to predict which fitness components (e.g. longevity, fecundity, offspring quality, etc) will

be affected. This uncertainty often necessitates measuring multiple traits and responses, and basing inference on a complex set of results. Quantifying replicability across studies therefore also requires integrating multiple outcomes. For similar reasons, evolutionary ecologists are rarely able to make quantitative predictions. Effect sizes are likely to vary considerably among species, genotypes, environments, and age groups. Consequently, evolutionary ecologists are usually limited to testing qualitative predictions such as effect direction, and qualitative measures of replicability might be most meaningful as well.

There has been much recent discussion and concern about replicability in evolutionary ecology (Palmer 2000; Nakagawa and Parker 2015; Fidler et al. 2017; Kelly 2019; Lind 2019; Powers and Hampton 2019; Fraser et al. 2020; Brecht et al. 2021; Filazzola and Cahill Jr 2021; Yang et al. 2024; Mundinger et al. 2025; Godoy et al. 2026). Yet, it is clear that null expectations for replicability in evolutionary ecology cannot be the same as in the physical sciences, molecular biology or physiology. If evolutionary ecologists work just as carefully as do scientists in those disciplines, their ability to obtain consistent results in replicated studies will still be substantially lower. Even psychologists might have an easier time replicating studies because, despite the complexity and plasticity of the human mind, human behaviour research focuses on just one species. Thus, while there is undoubtedly scope to improve replicability in evolutionary ecology, it is not clear whether the field faces a replicability crisis. To determine whether a problem exists and identify ways to solve it, it is first essential to clarify what is involved in replicating studies in evolutionary ecology, and what kind of replicability evolutionary ecologists should strive for.

## Replicability and biological diversity

Replicability is key to scientific progress because it enables generalization: if a result is replicable across studies then it is indicative of a genuine regularity in nature. Replicable results are therefore key to testing theory. In ecological studies, generalizability can be defined in terms of

many different contexts, such as location, time of the season, experimental methods, or the properties of sampled populations (Spake et al. 2022). Here I focus on an especially important form of generalizability in evolutionary ecology: phylogenetic generalizability. To achieve such generalizability, I argue that evolutionary ecology research necessitates a distinct concept of replicability grounded in biological diversity.

Consider the question of how adult dietary protein intake affects fitness, addressed through a hypothetical experiment where *Drosophila melanogaster* is subjected to diets differing in protein content. This study could be replicated in *D. melanogaster* using a similar experimental set-up. If consistent outcomes are obtained, the results could be regarded as replicable in the "direct" or "exact" sense (Table 1). However, in the context of evolutionary ecology, a key feature of such replication is that it is within a particular species. I will call this 'narrow replicability'.

As Nakagawa and Parker point out (Nakagawa and Parker 2015), the type of replicability that researchers should aim for depends on the scope of intended inference. If we are truly interested in questions about a particular species (such as behaviour in humans, or cancer resistance in mole rats) then narrow replicability is key. But most questions in evolutionary ecology are not species-specific: rather, we are interested in testing general models. Returning to the example of dietary protein, if we were truly interested in effects in *D. melanogaster* per se then it would be sufficient to replicate studies using different *D. melanogaster* strains, different forms of protein, and different environments. But what we really want to know is how dietary protein affects animals, or at least species with relevant attributes, such as strong links between fitness and resource investment in gametes. To achieve that kind of broad insight, results must be replicable far beyond *D. melanogaster*. Such question-oriented or "conceptual" replicability is essential in all fields of science (Table 1). However, in the context of evolutionary ecology, a key feature is the potential to obtain consistent results across species, because generalizability across species is an essential foundation for testing general models. I will call this 'broad replicability'.

Previous discussions of replicability in evolutionary ecology have emphasized the importance of replicating studies using the same species and methods (Palmer 2000; Nakagawa and Parker 2015). However, because species differ, demonstrating replicable results within one species is not sufficient to infer general patterns. While some effects could be qualitatively consistent across many species, other effects are unique to particular species, or even particular populations or genotypes. When tests of general theory produce results (whether positive or negative) in one species that are not replicable in other species, it is clear that the results depend on species-specific factors or mechanisms that are not encompassed by the general theory. Palmer argued that replicating studies using different species and methods ("quasireplication") increases the risk of selective reporting (Palmer 2000). Yet, while many practical challenges certainly exist, such challenges do not diminish the importance of broad replication in evolutionary ecology because only broad replicability enables robust testing of general models.

It might be objected that, while narrow and broad replicability are both important, narrow replicability must be established before investigating broad replicability. But there is no practical or logical reason to prioritize narrow replicability over broad replicability. If a pattern or effect is predicted by well-founded theory, it warrants (indeed, requires) broad testing. Even if research on one species yields highly replicable results, studies on other species with relevant attributes are still required to draw general conclusions because biological diversity precludes generalization from research on any single species. In the context of psychological research, Yarkoni (Yarkoni 2022) pointed out that "when the manifestation of a phenomenon is highly variable across potential measurement contexts, it simply does not matter very much whether any single realization is replicable or not." In psychology, important measurement contexts include subjects, stimuli, experimenters, and other methodological variables. In evolutionary ecology, the most important measurement context is the species.

120 The appropriate phylogenetic breadth of replication depends on the research question or model
121 being tested, and generalization is invalid beyond the range of phylogenetic diversity examined.
122 While meaningful estimation of effect sizes through meta-analysis requires proportional sampling
123 of taxonomic groups (a requirement that is rarely satisfied by meta-analytic data sets
124 (Dochtermann et al. 2026)), establishing phylogenetic generalizability requires representative
125 sampling of species across the phylogeny (Figure 1). In the context of psychological research, it
126 has been suggested that authors include "constraints on generality" statements indicating how far
127 their results can be generalized (Spake et al. 2022). In evolutionary ecology, the phylogenetic
128 breadth of replicable results limits valid generalization. A finding in one species suggests that
129 similar patterns might be expected in other species with similar attributes (e.g. a finding in
130 *Drosophila* might be representative of true flies, or even all insects), but broad replication is
131 needed to verify generalizability across these groups. Note that it is neither feasible nor necessary
132 to test models in every species. If qualitatively consistent results are obtained in five species
133 belonging to five distantly related insect orders then it could be tentatively concluded that the
134 pattern generalises across insects. The key is the phylogenetic breadth of the available evidence.

135

136 **Prediction and inference**

137 Whether models are tested in one or multiple species, the questions and phenomena addressed
138 by evolutionary ecology present inferential challenges. Some studies test a single prediction: for
139 example, a single trait might be compared between groups A and B to test the prediction A > B.
140 However, such a simple pattern is unlikely to be uniquely predicted by a single biological
141 mechanism, and observing such a pattern therefore often provides weak support for the model
142 being tested. A single prediction could enable stronger inference if the prediction is quantitative
143 and precise (van Dongen et al. 2023), but such predictions are rarely possible in evolutionary
144 ecology. Basing inference on a more stringent statistical significance threshold (Benjamin et al.

2018) does not solve this problem. Likewise, combining tests of a simple prediction from multiple studies through meta-analysis does not strengthen inference when the prediction itself is not unique to a particular model. Inferring a complex mechanism on the basis of a simple emergent pattern is always risky, and perhaps especially so in evolutionary ecology.

Some studies derive and test multiple predictions from the model of interest. For example, distinct treatment effects might be predicted for each of several traits. As long as each prediction is independent and follows clearly from the model under scrutiny, testing for a complex pattern comprising multiple predictions can enable stronger inference because few (if any) alternative mechanisms could plausibly generate the predicted pattern. In the terminology of philosophy of science, such complex predictions are "riskier" and enable greater "test severity" (Mayo and Spanos 2006; Mayo 2018; van Dongen et al. 2023). Prediction complexity could thus be used to gauge the strength of scientific inference in evolutionary ecology (Figure 2). As a simple rule, the proportion of independent patterns clearly predicted by the model being tested, and found to be consistent with results, could provide a measure of support for the model. The power to test distinct predictions could also be taken into account. For example, morphological traits can often be measured with higher repeatability than behavioural or life history traits can, and results for different types of traits could therefore be considered to contribute unequally to overall inference.

Whether predictions are simple or complex, testing any prediction is informative only to the extent that the prediction is unique to a single model (Mayo and Spanos 2006; Mayo 2018). Thus, ideally, two or more biologically plausible alternative models could be tested within the same study based on contrasting predictions (van Dongen et al. 2023). When results fail to clearly discriminate the alternative models, further research is required to resolve the ambiguity (Figure 2).

Note that testing complex predictions is not the same as multiple testing, where several tests of the same prediction are carried out within a study. For example, if support for a model is inferred from a predicted pattern in any one of several response variables, the false-positive rate will be elevated above the desired level (typically, $\alpha = 0.05$). To maintain the desired false-positive rate, correction for multiple testing is required. However, biological responses are seldom truly equivalent, and are often subject to interactions or trade-offs. For example, a positive effect on fitness might involve increased fecundity but reduced longevity. Thus, rather than treating each measure as equivalent, it may be better to derive separate predictions for each response variable and base inference on the full set of results. Of course, it is essential that such predictions are clearly based on theory and made prior to obtaining results, rather than devised after the fact to support a favoured interpretation. Predictions based on theory are key to testing models and making sense of empirical data (Shavit and Ellison 2017; Betts et al. 2021). Pre-registration of predictions could help ensure honest inference.

Given the challenges of biological diversity and complexity, replicability should, in most cases, be gauged on the basis of qualitative agreement across studies rather than effect sizes or levels of statistical support. For studies testing simple predictions, similar effect direction can be regarded as evidence of replicability, even if the pattern lacks statistical support. The results of the replicated studies could be combined through meta-analytic approaches (Jennions et al. 2013). For complex predictions, a result could be regarded as replicable if the overall conclusion is consistent between studies, even if the observed patterns are not identical. Broad replicability, in particular, will often have to be based on qualitative consistency of conclusions between studies because studies on distantly related species are unlikely to yield identical results.

## Conclusions

Biological diversity presents special inferential challenges for evolutionary ecology, and requires a distinct concept of replicability. While both narrow (within-species) and broad (across-species) replicability are important, broad replicability is key to testing general ecological and evolutionary models. Well-grounded theory should be tested broadly even if initial tests in one species fail to provide support. Inference can be strengthened by testing complex predictions that discriminate between alternative biological models. In most cases, replicability should be gauged on the basis of qualitative agreement rather than effect sizes or level of statistical support. Several strategies suggested in the literature (Palmer 2000; Mayo and Spanos 2006; Nakagawa and Parker 2015; Boose and Lerner 2017; Ellison 2017; Fidler et al. 2017; Vanderbilt and Blankman 2017; Mayo 2018; Feng et al. 2019; Powers and Hampton 2019; Fraser et al. 2020; Filazzola and Cahill Jr 2021; Jenkins et al. 2023; van Dongen et al. 2023; Gómez-Llano and De Lisle 2026) could enhance progress in evolutionary ecology (Table 2). Finally, as the history of science clearly shows, no single study can provide the final word on any scientific question. Any study, no matter how well-designed and executed, should be seen as just one additional piece of evidence that could contribute to progress in its scientific field.

**Table 1: Definitions of reproducibility and replicability**

**Reproducibility**

- *Computational reproducibility*: potential to reproduce a set of quantitative results using the same data and code
- *Analytical reproducibility*: potential to reproduce a set of quantitative results using the same data but different code
- *Conclusive reproducibility*: potential to obtain the same qualitative results from the same data using a different valid analysis (Gómez-Llano and De Lisle 2026)

**Replicability (general)**

- *Exact or direct replicability*: potential to obtain consistent results across studies using identical methods
- *Conceptual replicability*: potential to obtain consistent results across studies using different but appropriate methods

**Replicability (evolutionary ecology)**

- *Narrow replicability*: potential to obtain consistent results within a species
- *Broad replicability*: potential to obtain consistent results across species

**Table 2: Strategies to enhance progress in evolutionary ecology**

- Test well-founded models in multiple species even if initial investigations in one species provide no support
- Test complex predictions that discriminate between alternative biological models
- To gauge narrow replicability, investigate diverse populations and genotypes within a species using a range of reasonable methodologies and ecologically relevant environments
- To gauge broad replicability, investigate diverse model and non-model species with relevant attributes using a range of reasonable methodologies and ecologically relevant environments
- Gauge replicability based on qualitative agreement rather than effect size or level of statistical support
- Use appropriate study designs, describe methods fully, and collect and analyse data carefully
- Provide data and code in open repositories
- Encourage, fund and publish both narrow and broad replications

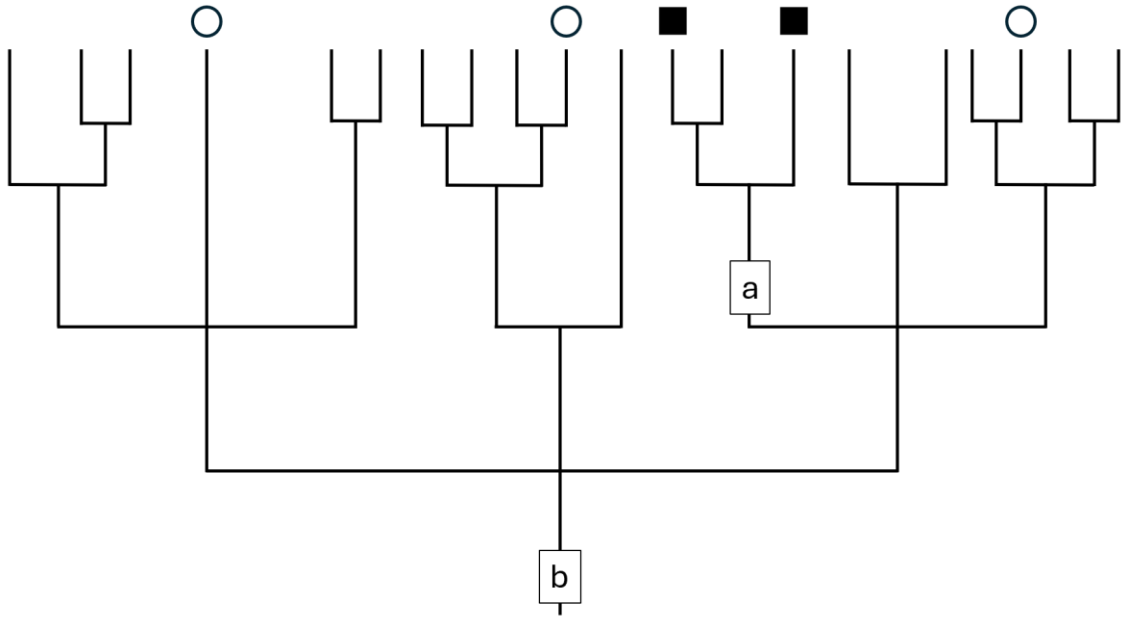


**Figure 1.** Generalizability based on broad replication follows the logic of phylogenetic inference. Replicable results represented by the filled square can be tentatively generalized to monophyletic group a, while replicable results represented by the open circle can be tentatively generalized to monophyletic group b. Generalization only extends to species with relevant attributes. For example, if the filled square represents results that support a model of the evolution of visual sexual signals, the model can be regarded as supported for species within monophyletic group a that reproduce sexually, have eyes, and court in daylight. However, even for species with relevant attributes, it is possible that the results would not replicate in some unstudied species or lineages. Research on additional, hitherto unstudied species can therefore help verify broad replicability and refine phylogenetic generalization. Note that it is nearly always invalid to generalize results to vaguely defined or paraphyletic categories of organisms such as "invertebrates".

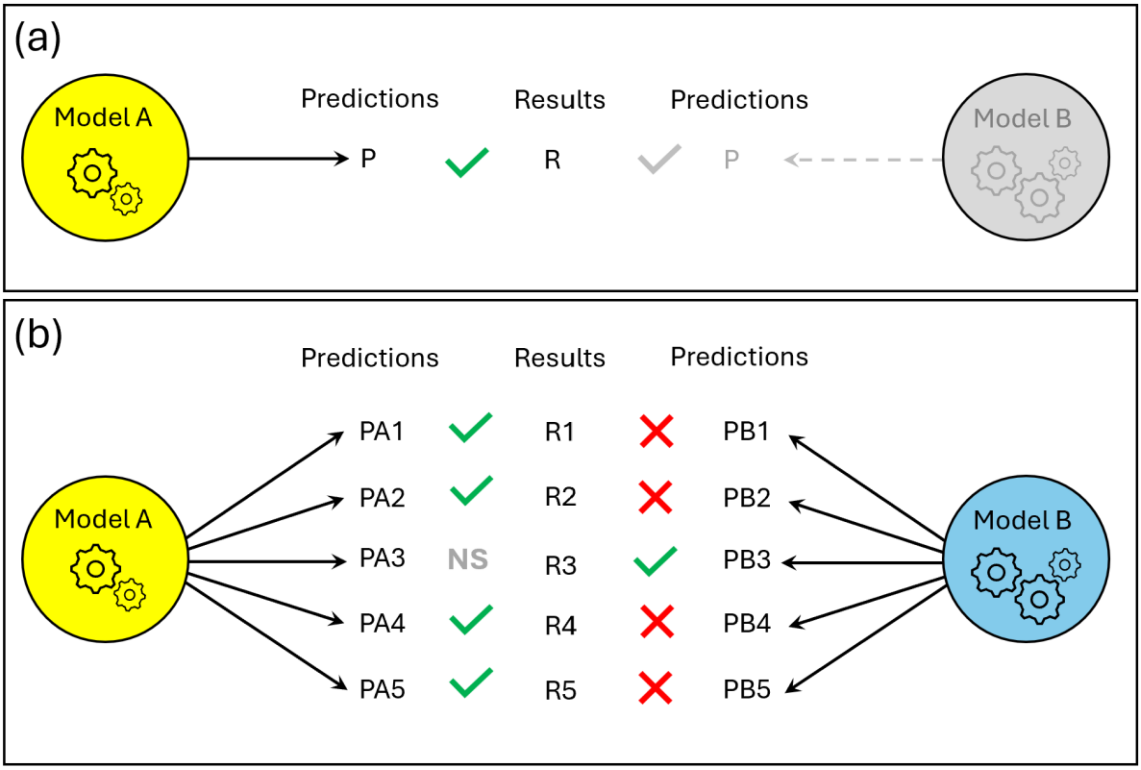


**Figure 2. (a)** A simple pattern, such as a positive treatment effect on trait Z1, is unlikely to be uniquely predicted by a single underlying mechanism. If the aim of the study is to test Model A by testing for predicted pattern P, support for this prediction (green checkmark) will often provide weak support for Model A because other underlying mechanisms, such as Model B, which is not explicitly tested in the study, could yield the same prediction. **(b)** A more complex pattern, such as distinct effects on traits Z1 … Z5, is more likely to be uniquely predicted by a single underlying mechanism. Observing the complex pattern predicted by Model A will therefore provide stronger support for Model A as long as the predicted effects are independent and clearly predicted by the model. Nonetheless, testing predictions from a single biological model against a statistical null hypothesis limits inference because the null hypothesis is often biologically uninformative. Inference can be strengthened by deriving and testing separate sets of predictions from multiple plausible biological models (e.g. Model A versus Model B). If the predictions derived from these models are independent and contrasting then the results could support one model while rejecting the other model. Complications can arise if complex sets of predictions are neither fully consistent nor fully inconsistent with the models being tested. In this example, four predictions (PA1, PA2, PA4, PA5) from Model A are supported by results (green checkmarks) but results are unclear for prediction PA3 because the predicted pattern is not supported statistically (NS). For Model B, four predictions (PB1, PB2, PB4, PB5) are contradicted by results (red crosses) while prediction PB3 is supported. In this case, it could be concluded that Model A has stronger support because most of Model A's predictions are supported and none are clearly contradicted by results, while most of Model B's predictions are contradicted by results. However, further research is needed to understand the unclear/conflicting results for predictions PA3/PB3. Such inferential complexity, typical of studies in evolutionary ecology, makes it challenging to test models and to assess both narrow and broad replicability.

## References


Benjamin, D. J., J. O. Berger, M. Johannesson, B. A. Nosek, E. J. Wagenmakers, R. Berk, K. A. Bollen, B. Brembs, L. Brown, C. Camerer, D. Cesarini, C. D. Chambers, M. Clyde, T. D. Cook, P. De Boeck, Z. Dienes, A. Dreber, K. Easwaran, C. Efferson, E. Fehr, F. Fidler, A. P. Field, M. Forster, E. I. George, R. Gonzalez, S. Goodman, E. Green, D. P. Green, A. G. Greenwald, J. D. Hadfield, L. V. Hedges, L. Held, T. Hua Ho, H. Hoijtink, D. J. Hruschka, K. Imai, G. Imbens, J. P. A. Ioannidis, M. Jeon, J. H. Jones, M. Kirchler, D. Laibson, J. List, R. Little, A. Lupia, E. Machery, S. E. Maxwell, M. McCarthy, D. A. Moore, S. L. Morgan, M. Munafó, S. Nakagawa, B. Nyhan, T. H. Parker, L. Pericchi, M. Perugini, J. Rouder, J. Rousseau, V. Savalei, F. D. Schönbrodt, T. Sellke, B. Sinclair, D. Tingley, T. Van Zandt, S. Vazire, D. J. Watts, C. Winship, R. L. Wolpert, Y. Xie, C. Young, J. Zinman, and V. E. Johnson. 2018. Redefine statistical significance. Nature Human Behaviour 2:6-10.

Betts, M. G., A. S. Hadley, D. W. Frey, S. J. K. Frey, D. Gannon, S. H. Harris, H. Kim, U. G. Kormann, K. Leimberger, K. Moriarty, J. M. Northrup, B. Phalan, J. S. Rousseau, T. D. Stokely, J. J. Valente, C. Wolf, and D. Zárrate-Charry. 2021. When are hypotheses useful in ecology and evolution? Ecology and Evolution 11:5762-5776.

Boose, E. R. and B. S. Lerner. 2017. Replication of Data Analyses: PROVENANCE IN R *in* A. Shavit, and A. M. Ellison, eds. Stepping in the same river twice : replication in biological research. Yale University Press, New Haven.

Brecht, K. F., E. W. Legg, C. Nawroth, H. Fraser, and L. Ostojić. 2021. The Status and Value of Replications in Animal Behavior Science. Animal Behavior and Cognition 8:97-106.

Dochtermann, N. A., M. A. Sekhar, and S. Nakagawa. 2026. Ungeneralizable generalizations? A meta-meta-analysis of the influence of taxonomic bias on the study of behaviour. Animal Behaviour 236:123542.

Ellison, A. M. 2017. Best Practices for Creating Replicable Research *in* A. Shavit, and A. M. Ellison, eds. Stepping in the same river twice : replication in biological research. Yale University Press, New Haven.

Feng, X., D. S. Park, C. Walker, A. T. Peterson, C. Merow, and M. Papeş. 2019. A checklist for maximizing reproducibility of ecological niche models. Nature Ecology & Evolution 3:1382-1395.

Fidler, F., Y. E. Chee, B. C. Wintle, M. A. Burgman, M. A. McCarthy, and A. Gordon. 2017. Metaresearch for Evaluating Reproducibility in Ecology and Evolution. BioScience 67:282-289.

Filazzola, A. and J. F. Cahill Jr. 2021. Replication in field ecology: Identifying challenges and proposing solutions. Methods in Ecology and Evolution 12:1780-1792.

Fraser, H., A. Barnett, T. H. Parker, and F. Fidler. 2020. The role of replication studies in ecology. Ecology and Evolution 10:5197-5207.

Godoy, P. L., D. M. Casali, J. Meachen, A. López-Arbarello, and M. D. D'Emic. 2026. The reproducibility crisis in phylogenetic analyses. Journal of Vertebrate Paleontology:e2638040.

Gómez-Llano, M. and S. P. De Lisle. 2026. Demonstrations of reproducibility are not reproducibility. Nature Ecology & Evolution 10:615-616.

Jenkins, G. B., A. P. Beckerman, C. Bellard, A. Benítez-López, A. M. Ellison, C. G. Foote, A. L. Hufton, M. A. Lashley, C. J. Lortie, Z. Ma, A. J. Moore, S. R. Narum, J. Nilsson, B. O'Boyle, D. B. Provete, O. Razgour, L. Rieseberg, C. Riginos, L. Santini, B. Sibbett, and P. R. Peres-Neto. 2023. Reproducibility in ecology and evolution: Minimum standards for data and code. Ecol Evol 13:e9961.

Jennions, M. D., C. J. Lortie, and J. Koricheva. 2013. Using Meta-analysis to Test Ecological and Evolutionary Theory. Pp. 0. Handbook of Meta-analysis in Ecology and Evolution. Princeton University Press.

Kelly, C. D. 2019. Rate and success of study replication in ecology and evolution. PeerJ 7:e7654.

Lind, P. A. 2019. Repeatability and Predictability in Experimental Evolution. Pp. 57-83 *in* P. Pontarotti, ed. Evolution, Origin of Life, Concepts and Methods. Springer International Publishing, Cham.

Mayo, D. 2018. Statistical Inference as Severe Testing. Cambridge University Press.

Mayo, D. G. and A. Spanos. 2006. Severe Testing as a Basic Concept in a Neyman-Pearson Philosophy of Induction. The British Journal for the Philosophy of Science 57:323-357.

Mundinger, C., N. K. E. Schulz, P. Singh, S. Janz, M. Schurig, J. Seidemann, J. Kurtz, C. Müller, H. Schielzeth, V. T. von Kortzfleisch, and S. H. Richter. 2025. Testing the reproducibility of ecological studies on insect behavior in a multi-laboratory setting identifies opportunities for improving experimental rigor. PLOS Biology 23:e3003019.

Nakagawa, S. and T. H. Parker. 2015. Replicating research in ecology and evolution: feasibility, incentives, and the cost-benefit conundrum. BMC biology 13:88.

Palmer, A. R. 2000. Quasi-Replication and the Contract of Error: Lessons from Sex Ratios, Heritabilities and Fluctuating Asymmetry. Annu. Rev. Ecol. Syst. 31:441-480.

Powers, S. M. and S. E. Hampton. 2019. Open science, reproducibility, and transparency in ecology. Ecological Applications 29:e01822.

Shavit, A. and A. M. Ellison. 2017. Toward a Taxonomy of Scientific Replication *in* A. Shavit, and A. M. Ellison, eds. Stepping in the same river twice : replication in biological research. Yale University Press, New Haven.

Spake, R., R. E. O'Dea, S. Nakagawa, C. P. Doncaster, M. Ryo, C. T. Callaghan, and J. M. Bullock. 2022. Improving quantitative synthesis to achieve generality in ecology. Nature Ecology & Evolution 6:1818-1828.

van Dongen, N., J. Sprenger, and E. J. Wagenmakers. 2023. A Bayesian perspective on severity: risky predictions and specific hypotheses. Psychonomic bulletin & review 30:516-533.

Vanderbilt, K. and D. Blankman. 2017. Reliable Metadata and the Creation of Trustworthy, Reproducible, and Re-usable Data Sets *in* A. Shavit, and A. M. Ellison, eds. Stepping in the same river twice : replication in biological research. Yale University Press, New Haven.

Yang, Y., E. van Zwet, N. Ignatiadis, and S. Nakagawa. 2024. A large-scale in silico replication of ecological and evolutionary studies. Nature Ecology & Evolution 8:2179-2183.

Yarkoni, T. 2022. The generalizability crisis. Behavioral and Brain Sciences 45:e1.