Science
AAAS

Replication in Field Biology: The Case of the Frog-Eating Bat

Michael J. Ryan

Science **334**, 1229 (2011);
DOI: 10.1126/science.1214532

PERSPECTIVE

Replication in Field Biology: The Case of the Frog-Eating Bat

Michael J. Ryan

Studies conducted in the field offer unique opportunities to observe nature, but achieving true replication under natural conditions is challenging. As demonstrated by the discovery of frog eating by a charismatic bat, biology conducted in the field generally follows an interesting progression that includes discovery, demonstration, experimentation, and verification.

Darwin (*1*) proposed that elaborate courtship displays were maladaptive for survival but evolved because they enhanced mating success. He did not come to this conclusion from field observations or experimental verification of survival costs, but from examining a repeated pattern of sexual dimorphism among diverse taxa. He was right, and it was not by accident but by an informed observation of nature. Data supporting Darwin's insight accumulated in various ways. I review a series of studies to illustrate the different means by which we arrive at scientific conclusions in field biology.

Túngara frogs, *Physalaemus pustulosus*, make simple (whine only) and complex (whines with one to seven chucks) mating calls (*2*). When calling alone a male produces simple calls, and when in a chorus he makes complex calls, which are also more attractive to females (*3*). In our early studies of these frogs, we wondered why males do not always make complex calls. Harkening back to Darwin for inspiration, we assumed that these calls were costly (*4*). Support for this hypothesis arrived in the mouth of a bat. During one study, bat researcher Merlin Tuttle and I mist-netted several bats (*Trachops cirrhosus*) after they had just caught túngara frogs (*5*). Did the bats eat the frogs? Yes, that is what we saw. Was this observation replicated? Yes, we caught several bats feasting on these frogs. We quickly realized that the bats were attracted to the frogs' calls. In fact, we could increase our capture success if we broadcast frog calls near the mist nets. It became increasingly clear that the bats were targeting the frogs; in one study, we observed *Trachops* feeding on calling túngara frogs almost 100 times, at a rate of more than six per hour (*6*). These field observations were solid and supported. The bats ate the frogs. No further data, no experiments, were needed to believe it. This became a scientific fact.

Fig. 1. A frog-eating bat (*T. cirrhosus*) feeding on a túnagara frog (*P. pustulosus*). [Photo credit: A. Baugh]

So, we had discovered that frogs call and that bats eat calling frogs. This observation by itself, however, did not provide strong support for Darwin's prediction of costly sexual displays. Support for this hypothesis could come from evidence that *Trachops* were homing in on the frogs' calls. We suspected this was the case, but we had not proven it. Thus, we placed speakers in the forest that broadcast the calls, either simple or complex, of túngara frogs and counted the number of times that *Trachops* swooped over each speaker. The calls attracted the bats, especially the complex calls. This field experiment showed that not only do calling túngara frogs attract frogivorous bats, but complex calls are more attractive to both bats and female frogs: Almost two-thirds of the 249 approaches by the bats were to complex calls (*7*). This supported Darwin's claim that more attractive displays are more costly for survival.

Like many field experiments, this one was not perfect. On the positive side, we brought the experimental stimuli to the bats, and thus we were sure that the phenomenon was not a laboratory artifact. But the field experiment faced the danger of pseudoreplication (*8*). We had no idea how many bats swooped over these speakers. Therefore, we replicated the same experiment in a flight cage with individual bats and obtained similar results (*7*). These initial results have been replicated numerous times since, as we delved into the cognitive mechanisms by which the bats make decisions about acoustically based foraging preferences (*9*, *10*). Further support for this phenomenon came from a later demonstration of a prediction of the initial findings—*Trachops* has auditory specializations that allow it to be sensitive to both its own ultrasonic echolocation calls (>50,000 Hz) and the much-lower-frequency calls of frogs (<5000 Hz) (*11*).

These examples illustrate how scientific facts in field biology can become established through the progression of verifiable and replicated observations, field and laboratory experiments, and the generation and testing of predictions about aspects of an animal's biology. Not all studies in field biology can progress from field observations→field experiments→lab experiments→physiological verification. But observations in the wild are real. A scientist tied to the lab, restricted only to parametric and controlled experiments, or only calculating selection gradients, would never have discovered that bats eat frogs, chimps use tools (*12*), cuckoos deposit eggs in the nests of other species of birds (*13*), and a litany of other amazing discoveries about what animals really do in the real world.

Nature is different from the laboratory, and the field researchers' experience and reputation seem to play a more important role in how we react to their findings. Results from field studies often tend to be accepted more readily without the traditional replication (*14*) required in other fields. There could

Section of Integrative Biology, University of Texas, Austin, TX 78712, USA, and Smithsonian Tropical Research Institute (STRI), Apartado 0843-03092, Balboa, Panama. E-mail: mryan@mail.utexas.edu

be several reasons for this. Many observations in nature require unusual expertise and experience, whereas laboratory experiments should be designed so they can be repeated with the same results by a naïve observer. Such an observer would lack the experience and expertise of a hardened field researcher like Jane Goodall and, if inserted into the wilds of Gombe with unhabituated chimpanzees, would have no chance of observing chimps using tools. Subsequent studies confirmed Goodall's original observations on tool use in chimps when similar abilities were observed in other chimp populations, other primates, and even New Caledonian crows (*15*). These were important, but in the end they were not needed to ensure the veracity of her original observations.

When field observations lead to field and laboratory experiments, however, rigor, controls, and replication similar to those used in the more traditional laboratory sciences are expected (*14*). Field biology recently has benefitted from an influx of technology in which audio and video recordings, remote sensing, and satellite tracking are important aids in data collection. The videos that eventually appeared of chimps using tools and bats eating frogs, for example, provided added value to the original observations. These tools enhance the reliability of field observations, allow observations at a scale not previously possible, and are now a welcome addition to the field biologist's tool kit.

Field observations are good at telling us what happens in nature. Experimentation is better at demonstrating cause and effect. Experiments in the field encompass the variables under which animals function, whereas those in the laboratory allow for some control over these variables. Each has its virtues. All findings in the field and the laboratory make predictions which, if supported, add further support to what we think we know or, if not supported, lead us to doubt our interpretations. All of this is science and, if done well, is good science.

References

1. C. Darwin, *The Descent of Man and Selection in Relation to Sex* (Murray, London, 1871).
2. M. J. Ryan, *The Túngara Frog, A Study in Sexual Selection and Communication* (Univ. Chicago Press, Chicago, 1985).
3. M. J. Ryan, *Science* **209**, 523 (1980).
4. A. S. Rand, M. J. Ryan, *Z. Tierpsychol.* **57**, 209 (1981).
5. M. D. Tuttle, M. J. Ryan, *Science* **214**, 677 (1981).
6. M. J. Ryan, M. D. Tuttle, L. K. Taft, *Behav. Ecol. Sociobiol.* **8**, 273 (1981).
7. M. J. Ryan, M. D. Tuttle, A. S. Rand, *Am. Nat.* **119**, 136 (1982).
8. S. H. Hurlbert, *Ecol. Monogr.* **54**, 187 (1984).
9. R. A. Page, M. J. Ryan, *Curr. Biol.* **16**, 1201 (2006).
10. K. L. Akre, H. E. Farris, A. M. Lea, R. A. Page, M. J. Ryan, *Science* **333**, 751 (2011).
11. V. Bruns, H. Burda, M. J. Ryan, *J. Morphol.* **199**, 103 (1989).
12. J. Goodall, *Nature* **201**, 124 (1964).
13. N. B. Davies, *Cuckoos, Cowbirds and Other Cheats* (Princeton University, Princeton, NJ, 2000).
14. A. R. Palmer, *Annu. Rev. Ecol. Syst.* **31**, 441 (2000).
15. G. R. Hunt, *Nature* **379**, 249 (1996).

10.1126/science.1214532

PERSPECTIVE

Improving Validation Practices in "Omics" Research

John P. A. Ioannidis[1] and Muin J. Khoury[2]*

"Omics" research poses acute challenges regarding how to enhance validation practices and eventually the utility of this rich information. Several strategies may be useful, including routine replication, public data and protocol availability, funding incentives, reproducibility rewards or penalties, and targeted repeatability checks.

The exponential growth of the "omics" fields (genomics, transcriptomics, proteomics, metabolomics, and others) fuels expectations for a new era of personalized medicine. However, clinically meaningful discoveries are hidden within millions of analyses (*1*). Given this immense biological complexity, separating true signals from red herrings is challenging, and validation of proposed discoveries is essential.

Some fields already employ stringent replication criteria. For example, in genomics, genome-wide association studies demand high statistical significance (P values $< 5 \times 10^{-8}$) and perform large-scale replication efforts within international consortia (*2*). Conversely, other fields continue to perform "mile-long, inch-thick" research (*3*), in which many factors are tested once ("discovered") but are rarely further validated. Studies in gene expression profiling and transcriptomics sometimes try to validate the results using different assays within single populations as well as statistical techniques such as cross-validation, which do not require the evaluation of additional, independent samples. However, such methods do not guarantee good performance across different populations. Moreover, very often cross-validation overestimates classifier performance, probably because biases are introduced in the process (*4*, *5*). Independent external validation usually yields more conservative results, but may also be inflated because of optimism, selective reporting, and other biases (*5*, *6*). Independent external validation by completely different teams remains rare.

Even strong replication of omics results does not automatically imply the potential for successful adoption in clinical or public health practice. Demonstrating clinical validity requires evaluation of the predictive value in real-practice populations, whereas clinical utility requires evaluation of the balance of benefits and harms associated with the adoption of these technologies for different intended uses (*7*). Ideally, randomized clinical trials are needed to assess whether omics information improves patient outcomes. Long-term, large-scale trials, such as those under way for Oncotype^{DX} (a diagnostic test that analyzes a panel of 21 genes within a breast tumor to assess the likelihood of disease recurrence and/or patient benefit from chemotherapy) and MammaPrint (a breast cancer signature of 70 genes) also require careful consideration of design issues (*8*, *9*), because information on available classifiers constantly changes and new classifiers are proposed. There is at least one recent unfortunate example, where gene signatures were moved into clinical trial experimentation with insufficient previous validation. Three trials of gene signatures to predict outcomes of chemotherapy in treating non–small-cell lung cancer and breast cancer were suspended in 2011 after the realization that their supporting published evidence was nonreproducible (*10*).

Many scientists now demand reproducible omics research (*11*). This requires access to the full data, protocols, and analysis codes for published studies so that other scientists can repeat analyses and verify results. Fortunately, several public data repositories exist, such as the Gene Expression Omnibus, ArrayExpress, and the Stanford Microarray Database. There have also been many calls for diverse comprehensive study registries, such as for tumor biomarkers, a field riddled with uncertainty because of suboptimal study design and data quality, and a poor replication record (*12*, *13*). Many leading journals are now working to adopt policies to make public deposition of data and protocols a prerequisite for publication (*14*). However, the practice of making this information accessible is applied inconsistently; furthermore, it is challenging to verify that complete data and protocols are indeed

[1]Stanford Prevention Research Center, Department of Medicine and Department of Health Research and Policy, Stanford University School of Medicine and Department of Statistics, Stanford University School of Humanities and Sciences, Stanford, CA 94305, USA. [2]Office of Public Health Genomics, Centers for Disease Control and Prevention, Atlanta, GA 30333, USA; and Division of Cancer Control and Population Sciences, National Cancer Institute, National Institutes of Health, Bethesda, MD 20852, USA.

*To whom correspondence should be addressed. E-mail: muk1@cdc.gov