Behav Ecol Sociobiol (2011) 65:37–45
DOI 10.1007/s00265-010-1039-4

REVIEW

Developing multiple hypotheses in behavioral ecology

Ned A. Dochtermann · Stephen H. Jenkins

Received: 10 October 2009 / Revised: 23 July 2010 / Accepted: 29 July 2010 / Published online: 17 September 2010

Abstract Researchers in behavioral ecology are increasingly turning to research methods that allow the simultaneous evaluation of hypotheses. This approach has great potential to increase our scientific understanding, but researchers interested in the approach should be aware of its long and somewhat contentious history. Also, prior to implementing multiple hypothesis evaluation, researchers should be aware of the importance of clearly specifying a priori hypotheses. This is one of the more difficult aspects of research based on multiple hypothesis evaluation, and we outline and provide examples of three approaches for doing so. Finally, multiple hypothesis evaluation has some limitations important to behavioral ecologists; we discuss two practical issues behavioral ecologists are likely to face.

Keywords Multiple hypotheses · Multimodel inference · Hypothesis testing · Model selection

Introduction

A research program based on the evaluation of multiple competing hypotheses has a long history (e.g., Chamberlin 1890) and can lead to a rapid increase in our understanding of organismal behavior by allowing researchers to assess multiple explanations simultaneously. Researchers are increasingly attempting to evaluate multiple hypotheses, and this increase tracks the growth in availability and understanding of statistical tools for the simultaneous evaluation of multiple hypotheses (Johnson and Omland 2004). Chief among contemporary tools are model selection procedures using Akaike's information criterion (AIC), in which its wide use in ecology and evolutionary biology is largely due to influential works by David Anderson and Kenneth Burnham (Anderson et al. 2000, 2001; Burnham and Anderson 2001; Anderson and Burnham 2002; Burnham and Anderson 2002, 2004).

Although AIC is the most frequent statistical tool currently used for the evaluation of multiple hypotheses, competing hypotheses can also be evaluated using other information criteria (IC) such as the Schwarz/Bayesian information criterion or Takeuchi's information criterion (Burnham and Anderson 2002; Johnson and Omland 2004). When considering nested hypotheses, likelihood ratio tests can also be used (Johnson and Omland 2004). Alternatively, multiple hypotheses can be evaluated based on whether specific, alternative a priori predictions are statistically supported (Table 1), an approach that may lead to more robust inferences (Lipton 2005). Factorial designs, dynamic models with quantitative predictions, and response surface methods also allow the application of multiple hypothesis evaluation to complex problems (Hilborn and Stearns 1982). Hilborn and Stearns (1982) describe these last three methods and how ecologists can design factorial experiments to assess hypotheses about a large number of potential causal factors and hypotheses in a tractable manner.

For questions in behavioral ecology where complex interactions are possible, critical experiments can be difficult to design, and the data collected are often

Communicated by L. Garamszegi

This contribution is part of the Special Issue "Model selection, multimodel inference and information-theoretic approaches in behavioural ecology" (see Garamszegi 2010).

N. A. Dochtermann (✉) · S. H. Jenkins
Department of Biology, University of Nevada,
Reno, USA
e-mail: ned.dochtermann@gmail.com

Table 1 Behavioral ecologists who have employed multiple hypothesis evaluation have most typically done so using either AIC based model ranking or evaluation of *a priori* predictions of specific hypotheses

Topic	Species	# Of hypotheses	Method	Reference
Adoption behavior	Ring-billed gulls	8	Prediction-based	Brown (1998)
Anti-predator behavior	Deer	5	Prediction-based	Bildstein (1983)
Anti-predator behavior	Gazelles	11	Prediction-based	Caro (1986)
Anti-predator behavior	Elk	38	AIC-based model ranking	Liley and Creel (2008)
Attraction/sexual selection	Flies	4	AIC-based model ranking	Stamps et al. (2005)
Attraction/sexual selection	Ghost crabs	5	AIC-based model ranking	Stamps et al. (2005)
Behavioral syndrome/ personality structure	Kangaroo rats	8	AIC-based model ranking	Dochtermann and Jenkins (2007)
Brood parasitism	Warblers	All possible	AIC-based model ranking	Stokke et al. (2008)
Cooperative breeding	Tree creepers	4	Prediction-based	Doerr and Doerr (2006)
Foraging behavior	Kangaroo rats	5	Prediction-based	Jenkins et al. (1995)
Habitat selection	Mosquitos	2	Prediction-based	Kiflawi et al. (2003)
Sexual selection/sociality	Goats	4	Prediction-based	Calhim et al. (2006)
Sociality	Swallows	3	Prediction-based	Hoogland and Sherman (1976)
Sociality/anti-predator behavior	Mountain sheep	4	Prediction-based	Bleich et al. (1997)

These approaches have been implemented to address at least ten different behavioral ecological questions, and researchers have examined as few as two mutually exclusive *a priori* hypotheses (Kiflawi et al. 2003) based on predictions to as many as 38 *a priori* hypotheses (Liley and Creel 2008) based on AIC values

correlative and observational. When dealing with correlative data and complex questions, model comparisons based on information criteria may be more tractable than experiments (Anderson et al. 2000). Using an IC-based approach, hypotheses are translated into statistical models and evaluated based on IC values which balance fit to the data and parsimony. Due to the ability of most statistical packages to calculate IC values, this approach can be implemented with relative ease. However, deeper issues remain including how researchers should go about developing and evaluating sets of alternative hypotheses (Burnham et al. 2010; Garamszegi 2010).

Here, we discuss the underlying historical and philosophical bases for a research program using a framework of multiple hypothesis evaluation. We also discuss the long history of multiple hypothesis testing, including historical examples in which the approach has both aided and hindered the advancement of scientific thought. Next, we discuss one of the more challenging aspects of a framework for multiple hypothesis testing, how to develop candidate hypotheses, and we describe three approaches for developing hypotheses. Finally, we discuss some of the conceptual problems and practical limitations that may arise for behavioral ecologists when using multiple hypothesis evaluation. We focus on proper hypothesis specification (versus fishing or data dredging; Anderson and Burnham 2002), and implicit assumptions regarding hypothesis sets.

The historical roots of multiple hypothesis evaluation

In “The method of multiple working hypotheses”, T.C. Chamberlin (1890; see also Chamberlin 1897) developed the conceptual basis for subsequent methods of model selection and multimodel inference. Chamberlin asserted that the promotion of hypotheses unsupported by data independent of the creation of those hypotheses was a problematic aspect of contemporary (late nineteenth century) research. Chamberlin proposed that researchers instead develop a suite of potential explanations or hypotheses and then use data to distinguish between the different explanations. Chamberlin asserted that the evaluation of multiple hypotheses was superior to other approaches because it encouraged thoroughness and objectivity, and that the approach would result in faster development and understanding for both basic and practical research questions.

Ecologists have generally been enthusiastic about using Chamberlin’s approach as a framework for designing research, either informally or through model selection and multimodel inference (Johnson and Omland 2004; Elliott and Brook 2007). However, some geologists and historians have been critical of Chamberlin’s influence. For example, Johnson (1990) considered Chamberlin’s argument to be based on faulty logic as well as impractical (but see Railsback 1990). Moreover, Oreskes (1999) argued that Chamberlin’s influence among American geologists contributed to a delay of about 40 years in their acceptance of Alfred Wegener's theory of continental drift, while European geologists accepted the theory much earlier. Rather than following Chamberlin's method, Wegener proposed his theory and then described how it was compatible with various kinds of existing evidence (Oreskes 1999; Proctor and Capaldi 2001). Wegener's development of the theory of continental drift differed from Chamberlin's approach due to its failure to evaluate multiple hypotheses and by its failure to evaluate a hypothesis with only independent data. Chamberlin's stature in American geology contributed to American geologists overlooking the fact that there are multiple ways of doing science. Thus, despite the numerous benefits of the approaches described in this issue of *Behavioral Ecology and Sociobiology*, behavioral ecologists should remember the lesson of this episode in the history of geology.

Chamberlin's (1890) multiple hypothesis approach has often been conflated with Platt's (1964) advocacy of strong inference. While also incorporating a need for forming a priori hypotheses, Platt's approach differed from Chamberlin's by specifically advocating the development, evaluation and *falsification* of mutually exclusive hypotheses. Platt also argued that the simultaneous evaluation of *mutually exclusive* hypotheses contributes to an apparent greater rate of advancement in molecular biology and physics than in other fields.

Platt's argument caused much consternation among ecologists, illustrated by a special issue of *The American Naturalist* (1983) entitled "A round table on research in ecology and evolutionary biology". In this round table, Quinn and Dunham (1983) argued that ecological questions were too complex to yield to Platt's version of strong inference, while Simberloff (1983) suggested that ecology would be better served by rigorous application of Platt's approach to the exclusion of other ways of doing science and Salt (1983) recommended eclecticism. More recently, Krebs (2000) advocated Platt's version of strong inference as a means toward faster progress in ecology and Wolff (2000) advised graduate students in applied ecology that application of strong inference in their research would be a key to success.

There have been several criticisms of Platt's (1964) paper, the most comprehensive being that of O'Donohue and Buchanan (2001). These authors discuss eight objections to strong inference as described by Platt, ranging from inaccurate historiography to the faulty assumption that there is only one correct way to do science. Further, O'Donohue and Buchanan (2001) suggest that Platt (1964) provides no substantive evidence that strong inference has been applied more regularly in the sciences Platt argues advance faster. For behavioral ecologists, the most important point made by O'Donohue and Buchanan (2001) is their reminder that progress in science comes from multiple methods, not just using multiple alternative hypotheses and critical experiments in a falsificationist approach as described by Platt (1964). We appreciate eclecticism in science, but agree with Davis (2006) that Platt (1964) provides important inspiration for scientists to think rigorously and critically about their work, even when following Platt's specific recipe for strong inference isn't possible or appropriate.

Platt's paper has only been cited five times and Chamberlin's not at all in *Behavioral Ecology and Sociobiology*, *Animal Behaviour*, and *Behavioral Ecology* (ISI Web of Science, 8/2009); yet the general approach of Platt and Chamberlin has been utilized by behavioral ecologists investigating a wide variety of topics (Table 1). An early application of multiple hypothesis evaluation in behavioral ecology comes from Hoogland and Sherman's (1976) study of social behavior in bank swallows (*Riparia riparia*). Based on a priori predictions, they evaluated three general alternative hypotheses proposed by Alexander (1974) for the adaptive value of social living: that sociality increases foraging efficiency, that sociality decreases predation risk, and that swallows necessarily live together due to the localization of a required resource.

Hoogland and Sherman (1976) tested these hypotheses in competition as described by Platt (1964) and concluded that bank swallow sociality decreases predation risk. Hoogland (1981) reprised the approach in research on prairie dogs (*Cynomys leucurus* and *C. ludovicianus*) and suggested that predation risk was an important cause of the evolution of sociality for these species. Sherman (1977) also applied this approach to studies of alarm calling in Belding's ground squirrels (*Spermophilus beldingi*), finding that predation risk and kin selection were likely causal factors. These applications of multiple hypothesis evaluation greatly advanced our understanding of the evolution of sociality, and the three papers have been cited more than 800 times (ISI Web of Science, 8/2009).

One of the most difficult challenges in using strong inference in evolutionary ecology is the complexity of causation of behavioral, ecological, and evolutionary phenomena (Jenkins 2004). For example, hypothesized causes may not be mutually exclusive. A factor may be necessary but not sufficient or sufficient, but not necessary to cause an outcome; these possibilities emphasize the importance of specifying the role of factors in the hypotheses being considered. Multiple factors may jointly contribute to an outcome, in which case, our task is to estimate the relative importance of factors, not eliminate all but one from further consideration. Multimodel inference and model averaging based on information theory as described by Burnham and Anderson (2002) are particularly welcome additions to the arsenal of statistical tools for model selection for dealing with this situation.

However, discussions of model averaging have focused on single parameters (e.g., Burnham and Anderson 2002), and model averaging's application to interactions requires additional research.

A form of complex causation especially important in behavioral biology is the existence of proximate and ultimate causes of behaviors and other organismal traits. Mayr (1961) contrasted mechanistic (proximate) and evolutionary (ultimate) causes of organismal traits, and Tinbergen (1963) extended this description to four levels of causation: environmental triggers and neuroendocrine mechanisms, developmental mechanisms, current utility, and evolutionary history, with the first two being proximate causes and the last two being ultimate causes.

In general, hypotheses about causation at different levels may often be complementary (e.g., Holekamp and Sherman 1989). If complementary hypotheses are treated as alternatives in model selection or even multimodel inference, then results can be misleading or simply wrong. For example, the assertion that a simple drive for provisioning by adults leads to cooperative breeding (Jamieson 1989) ignores the evolutionary pressures that also affect the expression of cooperative breeding (Emlen et al. 1991). Multiple hypotheses should only be evaluated in competition with actual competitors. Therefore, behavioral biologists who plan to use a multiple hypothesis framework and the methods discussed in this issue of *Behavioral Ecology and Sociobiology* need to consider the possibility that their hypotheses expressed as models may not be true competitors.

Developing suites of hypotheses in evolutionary ecology

Despite the advantages of using multiple working hypotheses to structure research, developing hypothesis sets is not easy (Chamberlin 1890; Anderson and Burnham 2002; Eberhardt 2003; Steidl 2006). The number of potential statistical models accumulates geometrically with the number of potential variables even without considering interactions (2^k statistical models are possible with k variables, excluding interactions). In addition to this large number of possibilities, the proper consideration of interacting factors and the proper recognition of levels of causation complicate the ability to create meaningful sets of candidate hypotheses. Potential hypotheses should be consistent with the natural history of study organisms, and we discuss three methods by which researchers can develop such hypotheses and the statistical models implied by these hypotheses. The approaches we advocate are exploratory analyses and model simplification, the use of previous research, and considering the predictions of available theory. An important distinction between these three methods is that because analyses using hypotheses derived from previous research or theory can be conducted as confirmatory analyses, they both avail the ability to draw more general inferences than those possible with exploratory analyses.

Developing hypotheses based on exploratory analyses

One useful source of information for formulating hypotheses sets are data already at hand. Exploratory approaches can lead to the identification of key parameters (or parameter combinations) which can then be tested with independent data. Statistical packages that test all possible models can also be used to generate suites of top models as long as an ecological justification for their subsequent consideration is clear. It is important to note, however, that exploratory analyses, including stepwise modeling or conducting model selection based on all the statistical models possible, should be used with caution as general inferences and parameter estimates cannot be made reliably from exploratory methods (Grace 2006). Inferences from such approaches apply to the data used in model selection/simplification and cannot be extended further without confirmatory analyses (Zhang 1992; Chatfield 1995; Anderson et al. 2001). Hypotheses developed from exploratory analyses can be reexamined with confirmatory analyses, allowing generalizable inferences. Confirmatory analyses can be conducted with either independently acquired data sets or using data withheld from the initial analysis (Hurvich and Tsai 1990; Quinn and Keough 2002). However, this approach may be limited due to logistical constraints on the ability to collect sufficient data (Quinn and Keough 2002).

Symonds and Johnson (2008) provide an example of how exploratory analyses can be used to develop suites of hypotheses which could later be independently tested. Symonds and Johnson (2008) examined all possible combinations of the factors affecting species richness and species evenness in Australian birds. While no strong support was found for any single hypothesis, strongly influential parameters were identified. Moreover, several explanatory hypotheses were identified, each sharing similar statistical support (Symonds and Johnson 2008). These hypotheses for the causal patterns leading to evenness and richness can now be subjected to further analyses and be tested with independent data.

Similarly, Whittingham et al. (2006) examined nine habitat characteristics that potentially explain the distribution of yellowhammers (*Emberiza citronella*), a passerine in which populations in Europe are in decline. Whittingham et al. (2006) fit all possible models (with the exception that they did not include interactions) to data for yellowhammer distributions, resulting in over 500 possible models. This exploratory analysis demonstrated the importance of several habitat factors which may now be subjected to further scrutiny. These examples also highlight the need to distinguish between the exploratory nature of considering all the statistical models possible and the greater generality of inferences availed by confirmatory analyses.

In general, when researchers have sufficient sample sizes to conduct all-model or stepwise approaches, they should strongly consider the option of holding some data aside. This will allow researchers to subsequently evaluate multiple hypotheses, based on their exploratory analyses, in a confirmatory manner (Quinn and Keough 2002). However, to allow the greatest generalizability of inferences, models supported by exploratory analyses should be evaluated with independently gathered data (Guthery et al. 2005).

Developing hypotheses based on previous research

Published research with ecologically similar organisms can also provide researchers with multiple competing hypotheses for consideration. These alternative hypotheses may each have similar support in the literature, in which case, the simultaneous evaluation of multiple hypotheses will allow clearer understanding of ecological and evolutionary patterns than a case by case evaluation of hypotheses. This approach is consistent with Fisher's (1925: Chapter 5) recommendation that when no single hypothesis of causation is strongly supported, conclusions be subjected to repeated independent testing. This approach was also used in the evaluation of multiple hypotheses for social behavior by Hoogland and Sherman (1976) and would contribute to the need for behavioral ecologists to more frequently replicate results (Kelly 2006).

Jenkins et al. (1995) provide an example of the utility of multiple hypothesis evaluation in their examination of how Merriam's kangaroo rats (*Dipodomys merriami*) spatially distribute food stores (seed caches). How *D. merriami* distribute seed caches is ecologically important because kangaroo rats use food stores to survive periods of low or unpredictable food availability, and different spatial strategies for food storing may play a role in species coexistence (Jenkins et al. 1995; Price and Mittler 2006). Jenkins et al. (1995) identified a suite of five causal hypotheses for how and why *D. merriami* store food based on studies of food storing in other species (Table 2).

Each hypothesis was based on previous research and related to the trade-offs *D. merriami* may experience. For example, if, as in magpies (*Pica pica*; Clarkson et al. 1986), *D. merriami* experience considerable competition for food at primary sources, then it would be beneficial to maximize harvest rates initially and store food quickly near sources of production. Alternatively, if caches are lost to individuals who use cues to detect the presence of competitors (Shaw 1934), then it would be beneficial to store food away from an individual's burrow. The five proposed hypotheses (Table 2) were evaluated not using model selection based on information theory but rather by specifying mutually exclusive predictions for each competing hypothesis. Experimental results were then compared to these a priori predictions in order to determine which hypothesis was supported. Jenkins et al. (1995) concluded that *D. merriami* initially maximize food acquisition and then move food stores and remember their locations.

Dochtermann and Jenkins (2007) also used previous research and the natural history of *D. merriami* to develop hypotheses about behavioral syndrome (personality) structure. Based on the synthetic review of Sih et al. (2004), eight hypotheses about how four behavioral traits (food hoarding, intra-individual variation, boldness, and aggression) might covary were identified. The different hypotheses of behavioral covariance were translated into statistical hypotheses using structural equation modeling and evaluated using AIC-based model selection (see Figure 1 in Dochtermann and Jenkins 2007). These hypotheses of syndrome structure ranged from behaviors all being expressed independently to all behaviors covarying together. Based on this analysis of a priori hypotheses of how behaviors covary, Dochtermann and Jenkins (2007) found support for the inference that *D. merriami* exhibit personality structure.

These two examples addressed very different research questions and employed different statistical approaches (analysis of variance versus structural equation modeling coupled with AIC-based model comparison). However, both Jenkins et al. (1995) and Dochtermann and Jenkins (2007) examined questions for which several hypotheses had been previously proposed. Distinguishing between these hypotheses could only be achieved by simultaneously evaluating multiple hypotheses.

Developing hypotheses based on theory

Behavioral ecological research should be informed by ecological and evolutionary theory, and ideally, researchers will draw on relevant theory when developing suites of hypotheses. As a branch of evolutionary ecology, this attention to theory is essential for ensuring that appropriate questions are asked and that proper inferences are drawn. Theory will also often suggest multiple hypotheses. For example, animals have finite time which can be allocated to different purposes. Competing requirements (energy acquisition, finding mates, and surviving) lead to constraints on how animals can spend their time. The potential evolutionary pressures faced by individuals generate predictions as to how time should be budgeted. Similarly, optimal foraging theory proposes that organisms should behave differently based on environmental conditions and resulting trade-offs (e.g.,

Table 2 The five hypothesized spatial patterns of seed storage by Merriam's kangaroo rats (*Dipodomys merriami*) proposed by Jenkins et al. (1995)

	Hypotheses of seed caching behavior	Prediction	Reference[a]
(1)	Larder defense: Kangaroo rats store seeds in a single location to minimize energetic (transportation) costs and predation risk	Seeds are all/mostly stored in a single location	Daly et al. (1992); Reichman et al. (1986)
(2)	Scatterhoarding near burrow: Kangaroo rats can defend areas and scatter hoard to minimize larderhoarding costs but keep energetic costs low and minimize predation risk	Seeds are stored in numerous locations close to burrows	Shaw (1934); Hawbecker (1940); Congdon (1974); Blaustein and Risser (1976)
(3)	Scatterhoarding away from burrow: If competitors use burrows as cues to find seeds, Kangaroo rats will store seeds away from their burrow. This increases potential costs of recovering caches.	Seeds are stored in numerous locations away from burrows	Vanderwall (1994)
(4)	Scatterhoarding near food sources: Kangaroo rats rapidly take seeds from sources to maximize harvest	Seeds are stored close to food sources	Clarkson et al. (1986); Jenkins and Peters (1992)
(5)	Memory of caching locations: Kangaroo rats remember where they store food	Caches will be widely spaced to minimize the effectiveness of random search by cache thieves	Tinbergen et al. (1967); Jacobs and Liman (1991); Jacobs (1992)

These hypotheses generated specific predictions based on the experimental design (see Jenkins et al. (1995) for additional detail), allowing Jenkins et al. to distinguish between the different hypotheses based on results of conventional statistical approaches (e.g. ANOVA)

[a] See Jenkins et al. (1995) for complete bibliographic information

Stephens and Krebs 1986), and life history theory proposes that organisms face numerous trade-offs in how energy is allocated between survival and reproduction (e.g., Roff 2002). These trade-offs result in alternative and often mutually exclusive hypotheses.

As an example of the utility of theory in developing a priori hypotheses, Johnson (2002) generated 17 hypotheses about how selective pressures influence life history traits. Johnson (2002) examined the role of trade-offs in the evolution of life history strategies of the live-bearing fish *Brachyrhaphis rhabdophora* and evaluated four potential selective pressures and five life history traits for 27 populations of *B. rhabdophora*. Johnson's (2002) hypotheses were based on available theory regarding the evolution of life history strategies and represented combinations of direct and indirect effects. Johnson (2002) translated his biological hypotheses into structural equation models that were then evaluated based on AIC values. Johnson (2002) determined that the direct effects of extrinsic mortality, density dependence, resource availability, and habitat stability shape the evolution of life history strategies in *B. rhabdophora*.

Limitations of multiple hypothesis evaluation

Like any scientific approach, the use of multiple hypothesis evaluation has potential limitations that require consideration. Our discussion of the practical problems with multiple hypothesis evaluation centers on its implementation when using information criteria.

Anderson and Burnham (2002) outlined several problems relevant to multiple hypothesis evaluation using information criteria (specifically AIC values). Here, we discuss the two problems we consider most relevant to behavioral ecologists: the inclusion of too many models and determining whether ecologically relevant models were included. While these issues were initially discussed within the context of AIC use, they are general to any approach based on information criteria, and we encourage readers to return to the earlier discussion of these and other concerns in Anderson and Burnham (2002).

The problem of too many models

As discussed earlier, when working within a framework of multiple hypotheses, it may be possible to generate a large number of statistical models without consideration of the biological hypotheses these statistical models represent. For example, with only four potential causal factors, over 250 possible combinations of variables and their interactions are possible.

Some current statistical packages allow the generation of all the statistical models possible with the potential causal factors (e.g., Spatial Analysis in Macroecology v3.1; the 'leaps' package in R and using Proc Reg in SAS). However, this sort of approach in which models are considered without a priori justification should be avoided because including a large number of variables beyond those for which there is a priori justification is essentially an exploratory analysis. As discussed earlier, with exploratory analyses extending interpretations beyond the data set analyzed can lead to spurious inferences (Hurvich and Tsai 1990; Zhang 1992; Chatfield 1995; see Forstmeier and Schielzeth 2010, this issue, for discussion of additional statistical concerns). Whenever possible, confirmatory analyses, which could include the evaluation of multiple a priori hypotheses, are preferred over exploratory analyses (Grace 2006). Thus, researchers must refrain from including every conceivable model in analyses unless these analyses are properly framed as exploratory and the scope of inferences narrowed.

The problem of too many models is potentially lessened when researchers are attempting to evaluate hypotheses based on explicit predictions (as done in Jenkins et al. 1995; see Table 1 for additional examples). In this approach, hypotheses can only be considered if there is a theoretical or natural history justification. Further, hypotheses can only be considered using this approach if they generate concrete and discrete predictions.

Despite our encouragement to minimize the number of hypotheses being considered, numerous biological hypotheses and resulting statistical models may be necessary because biological systems are complex with interacting mechanisms and causation operating at multiple levels. In such instances, exploratory analyses can be used to initially reduce the number of models being considered, followed by independent testing of a subset of most supported hypotheses.

Determining whether relevant models were included

One of the key limitations of using multiple hypothesis evaluation is the implicit assumption that biologically relevant effects are included in hypothesis sets, and none of the three approaches discussed for generating hypotheses guarantees this. Proper knowledge of the natural history of the organisms being studied is one way to improve the likelihood that appropriate models and appropriate causal factors are examined. Nonetheless, model selection approaches including those based on information criteria are restricted to the a priori defined model set; a model or models will be selected as “best” within the set of those considered regardless of how well it actually describes a system. Thus, there is the possibility of the selected “best” model(s) explaining only a small amount of the variation present, potentially leading to inappropriate inferences.

One way to address this issue is to include some model of null or random expectations in the model set. If the null model is supported to a similar degree as other candidate models, this suggests that none of the candidate models is appropriate and that unidentified causal factors are of key biological importance. This approach is employed by population ecologists who include statistical models that lack any of the potential causal factors of interest when estimating population sizes, survival, or other population characteristics.

Another way of dealing with this issue is to use a measure of explained variation. Depending on the form of the statistical model corresponding to a hypothesis, researchers can use the coefficient of determination (R^2) to assess whether an overall model explains a biologically important proportion of variation. Maddala (1983), Magee (1990), Cox and Snell (1989) and Nagelkerke (1991) extended the calculation of R^2 to models fit using maximum likelihood, which can be used when AIC values are calculated:

$$R^2 = 1 - \left(\frac{L\left(\widehat{\theta}\right)_{null}}{L\left(\widehat{\theta}\right)_{x}} \right)^{\frac{2}{n}}$$

where $L\left(\widehat{\theta}\right)_{null}$ is the log likelihood of the null model and $L\left(\widehat{\theta}\right)_{x}$ is the log likelihood of the statistical model corresponding to a particular biological hypothesis. This measure can be interpreted in the same manner as traditionally calculated coefficients of determination. Cohen’s (1992) discussion of the strength of effect sizes can then be used as a guide in combination with the research question being considered for interpreting whether candidate models explain a biologically substantive proportion of the available variation (but see Abelson 1985).

Concluding remarks

Testing multiple hypotheses based on Chamberlin’s (1890) framework, using either Platt’s (1964) strong inference approach or model comparison based on an information criterion, has considerable potential to advance our understanding of the behavioral ecology of animals and of evolutionary ecology in general. However, researchers should be cognizant of the history of these approaches, with the examples of potential difficulties, and should also be aware of the intense debates the approaches have engendered. For multiple hypothesis testing to increase our understanding of behavioral ecology, researchers must also allocate considerable energy to the formulation of a priori hypotheses. Because of the difficulty of specifying a priori hypotheses, we have outlined three methods for doing so. Finally, we have discussed two key problems researchers should be aware of prior to initiating data collection (or at least analyses). While statistical advances and the increasing ease with which researchers can analyze multiple hypotheses have encouraged increased use of the approach, we caution researchers to think carefully about their research questions, their analytical choices and especially their choices of candidate hypotheses.

Acknowledgments We thank Laszlo Garamszegi for his encouragement in writing this manuscript, participants of the “ABCD” statistical symposium held at the 2008 meeting of the International Society for Behavioral Ecology for vigorous and insightful discussion, and three anonymous reviewers for critical and insightful comments.

References

Abelson RP (1985) A variance explanation paradox: when a little is a lot. Psychol Bull 97:129–133

Alexander RD (1974) The evolution of social behavior. Annu Rev Evol Syst 5:325–383

Anderson DR, Burnham KR (2002) Avoiding pitfalls when using information-theoretic methods. J Wildl Manage 66:912–918

Anderson DR, Burnham KP, Thompson WL (2000) Null hypothesis testing: problems, prevalence, and an alternative. J Wildl Manage 64:912–923

Anderson DR, Burnham KP, Gould WR, Cherry S (2001) Concerns about finding effects that are actually spurious. Wildl Soc Bull 29:311–316

Bildstein KL (1983) Why white-tailed deer flag their tails. Am Nat 121:709–715

Blaustein AR, Risser AC (1976) Interspecific interactions between 3 sympatric species of Kangaroo rats (*Dipodomys*). Anim Behav 24:381–385

Bleich VC, Bowyer RT, Wehausen JD (1997) Sexual segregation in mountain sheep: resources or predation? Wildl Monogr 134:1–50

Brown KM (1998) Proximate and ultimate causes of adoption in ring-billed gulls. Anim Behav 56:1529–1543

Burnham KP, Anderson DR (2001) Kullback–Leibler information as a basis for strong inference in ecological studies. Wildlife Res 28:111–119

Burnham KP, Anderson DR (2002) Model selection and multimodel inference: a practical information-theoretic approach, 2nd edn. Springer, New York

Burnham KP, Anderson DR (2004) Multimodel inference: understanding AIC and BIC in model selection. Sociol Method Res 33:261–304

Burnham K, Anderson D, Huyvaert K (2010) AICc model selection in the ecological and behavioral sciences: some background, observations, and comparisons. Behav Ecol Sociobiol. doi:10.1007/s00265-010-1029-6

Calhim S, Shi J, Dunbar RIM (2006) Sexual segregation among feral goats: testing between alternative hypotheses. Anim Behav 72:31–41

Caro TM (1986) The functions of stotting in Thomson's gazelles: some tests of the predictions. Anim Behav 34:663–684

Chamberlin TC (1890) The method of multiple working hypotheses. Science 15:92–96

Chamberlin TC (1897) The method of multiple working hypotheses. J Geol Sci 5:837–848

Chatfield C (1995) Model uncertainty, data mining and statistical inference. J R Stat Soc Ser A-Stat Soc 158:419–466

Clarkson K, Eden SF, Sutherland WJ, Houston AI (1986) Density dependence and magpie food hoarding. J Anim Ecol 55:111–121

Cohen J (1992) A power primer. Psychol Bull 112:155–159

Congdon J (1974) Effect of habitat quality on distributions of three sympatric species of desert rodents. J Mammal 55:659–662

Cox DR, Snell EJ (1989) The analysis of binary data, 2nd edn. Chapman and Hall, London

Daly M, Jacobs LF, Wilson MI, Behrends PR (1992) Scatter hoarding by Kangaroo rats (*Dipodomys merriami*) and pilferage from their caches. Behav Ecol 3:102–111

Davis RH (2006) Strong inference-rationale or inspiration? Perspect Biol Med 49:238–249

Dochtermann NA, Jenkins SH (2007) Behavioural syndromes in Merriam's kangaroo rats (*Dipodomys merriami*): a test of competing hypotheses. Proc R Soc B 274:2343–2349

Doerr ED, Doerr VAJ (2006) Comparative demography of tree-creepers: evaluating hypotheses for the evolution and maintenance of cooperative breeding. Anim Behav 72:147–159

Eberhardt LL (2003) What should we do about hypothesis testing? J Wildl Manage 67:241–247

Elliott LP, Brook BW (2007) Revisiting chamberlin: multiple working hypotheses for the 21st century. Bioscience 57:608–614

Emlen ST, Reeve HK, Sherman PW, Wrege PH, Ratnieks FLW, Shellmanreeve J (1991) Adaptive versus nonadaptive explanations of behaviour: the case of alloparental helping. Am Nat 138:259–270

Fisher RA (1925) Statistical methods for research workers. Oliver and Boyd, London

Forstmeier W, Schielzeth H (2010) Cryptic multiple hypotheses testing in linear models: overestimated effect sizes and the winner's curse. Behav Ecol Sociobiol. doi:10.1007/s00265-010-1038-5

Garamszegi LZ (2010) Information-theoretic approaches to statistical analysis in behavioural ecology: an introduction. Behav Ecol Sociobiol. doi:10.1007/s00265-010-1028-7

Grace JB (2006) Structural equation modeling and natural systems. Cambridge University Press, Cambridge

Guthery FS, Brennan LA, Peterson MJ, Lusk JJ (2005) Information theory in wildlife science: critique and viewpoint. J Wildl Manage 69:457–465

Hawbecker AC (1940) The burrowing and feeding habits of *Dipoclomys venustus*. J Mammal 21:388–396

Hilborn R, Stearns SC (1982) On inference in ecology and evolutionary biology: the problem of multiple causes. Acta Biotheor 31:145–164

Holekamp KE, Sherman PW (1989) Why male ground squirrels disperse. Amer Sci 77:232–239

Hoogland JL (1981) The evolution of coloniality in white-tailed and black-tailed prairie dogs (Sciuridae: *Cynomys leucurus* and *Cynomys ludovicianus*). Ecology 62:252–272

Hoogland JL, Sherman PW (1976) Advantages and disadvantages of bank swallow (*Riparia riparia*) coloniality. Ecol Monogr 46:33–58

Hurvich CM, Tsai CL (1990) The impact of model selection on inference in linear regression. Am Stat 44:214–217

Jacobs LF (1992) Memory for cache locations in Merriam kangaroo rats. Anim Behav 43:585–593

Jacobs LF, Liman ER (1991) Gray squirrels remember the locations of buried nuts. Anim Behav 41:103–110

Jamieson IG (1989) Behavioral heterochrony and the evolution of birds helping at the nest: an unselected consequence of communal breeding. Am Nat 133:394–406

Jenkins SH (2004) How science works: evaluating evidence in biology and medicine. Oxford University Press, New York

Jenkins SH, Rothstein A, Green WCH (1995) Food hoarding by Merriam's kangaroo rats—a test of alternative hypotheses. Ecology 76:2470–2481

Jenkins SH, Peters RA (1992) Spatial patterns of food storage by Merriam kangaroo rats. Behav Ecol 3:60–65

Johnson JG (1990) Method of multiple working hypotheses: a chimera. Geology 18:44–45

Johnson JB (2002) Divergent life histories among populations of the fish *Brachyrhaphis rhabdophora*: detecting putative agents of selection by candidate model analysis. Oikos 96:82–91

Johnson JB, Omland KS (2004) Model selection in ecology and evolution. Trends Ecol Evol 19:101–108

Kelly CD (2006) Replicating empirical research in behavioral ecology: how and why it should be done but rarely ever is. Q Rev Biol 81:221–236

Kiflawi M, Blaustein L, Mangel M (2003) Predation-dependent oviposition habitat selection by the mosquito *Culiseta longiareolata*: a test of competing hypotheses. Ecol Lett 6:35–40

Krebs CJ (2000) Hypothesis testing in ecology. In: Boitani L, Fuller TK (eds) Research techniques in animal ecology. Columbia University Press, New York, pp 1–14

Liley S, Creel S (2008) What best explains vigilance in elk: characteristics of prey, predators, or the environment? Behav Ecol 19:245–254

Lipton P (2005) Testing hypotheses: prediction and prejudice. Science 307:219–221

Maddala GS (1983) Limited-dependent and qualitative variables in econometrics. Cambridge University Press, Cambridge

Magee L (1990) R^2 measures based on Wald and likelihood ratio joint significance test. Am Stat 44:250–253

Mayr E (1961) Cause and effect in biology. Science 134:1501–1506

Nagelkerke NJD (1991) A note on a general definition of the coefficient of determination. Biometrika 78:691–692

O’Donohue W, Buchanan JA (2001) The weaknesses of strong inference. Behav Philos 29:1–20

Oreskes N (1999) The rejection of continental drift. Oxford University Press, New York

Platt JR (1964) Strong inference. Science 146:347–353

Price MV, Mittler JE (2006) Cachers, scavengers, and thieves: a novel mechanism for desert rodent coexistence. Am Nat 168:194–206

Proctor RW, Capaldi EJ (2001) Improving the science education of psychology students: better teaching of methodology. Teach Psychol 28:173–181

Quinn JF, Dunham AE (1983) On hypothesis testing in ecology and evolution. Am Nat 122:602–617

Quinn GP, Keough MJ (2002) Experimental design and data analysis for biologists. Cambridge University Press, Cambridge

Railsback LB (1990) “Method of multiple working hypotheses: a chimera”—comment. Geology 18:917–918

Reichman OJ, Fattaey A, Fattaey K (1986) Management of sterile and moldy seeds by a desert rodent. Anim Behav 34:221–225

Roff DA (2002) Life history evolution. Sinauer Associates, Inc., Sunderland

Salt GW (1983) Roles: their limits and responsibilities in ecological and evolutionary research. Am Nat 122:697–705

Shaw WT (1934) The ability of the giant kangaroo rat as a harvester and storer of seeds. J Mammal 15:275–286

Sherman PW (1977) Nepotism and evolution of alarm calls. Science 197:1246–1253

Sih A, Bell AM, Johnson JC, Ziemba RE (2004) Behavioral syndromes: an integrative overview. Q Rev Biol 79:241–277

Simberloff D (1983) Competition theory, hypothesis testing, and other community ecological buzzwords. Am Nat 122:626–635

Stamps J, McElreath R, Eason P (2005) Alternative models of conspecific attraction in flies and crabs. Behav Ecol 16:974–980.

Steidl RJ (2006) Model selection, hypothesis testing, and risks of condemning analytical tools. J Wildl Manage 70:1497–1498

Stephens DW, Krebs JR (1986) Foraging theory. Princeton University Press, Princeton

Stokke BG, Hafstad I, Rudolfsen G, Moksnes A, Moller AP, Roskaft E, Soler M (2008) Predictors of resistance to brood parasitism within and among reed warbler populations. Behav Ecol 19:612–620

Symonds MRE, Johnson CN (2008) Species richness and evenness in Australian birds. Am Nat 171:480–490

Tinbergen N (1963) On aims and methods of ethology. Zeit Tierpsych 20:410–433

Tinbergen N, Impekoven M, Franck D (1967) An experiment on spacing-out as a defence against predation. Behaviour 28:307–321

Vanderwall SB (1994) Seed fate pathways of antelope bitterbrush: dispersal by seed-caching yellow pine chipmunks. Ecology 75:1911–1926

Whittingham MJ, Stephens PA, Bradbury RB, Freckleton RP (2006) Why do we still use stepwise modelling in ecology and behaviour? J Anim Ecol 75:1182–1189

Wolff JO (2000) In my opinion—reassessing research approaches in the wildlife sciences. Wildl Soc Bull 28:744–750

Zhang P (1992) Inference after variable selection in linear regression models. Biometrika 79:741–746