

# Logic of experiments in ecology: is pseudoreplication a pseudoissue?

Lauri Oksanen

Oksanen, L. 2001. Logic of experiments in ecology: is pseudoreplication a pseudo-issue? – *Oikos* 94: 27–38.

Hurlbert divides experimental ecologists into ‘those who do not see any need for dispersion (of replicated treatments and controls), and those who do recognize its importance and take whatever measures are necessary to achieve a good dose of it’. Experimental ecologists could also be divided into those who do not see any problems with sacrificing spatial and temporal scales in order to obtain replication, and those who understand that appropriate scale must always have priority over replication. If an experiment is conducted in a spatial or temporal scale, where the predictions of competing hypotheses are convergent or ambiguous, no amount of technical impeccability can make the work instructive. Conversely, replication can always be obtained afterwards, by conducting more experiments with basically similar design in different areas and by using meta-analysis. This approach even reduces the sampling bias obtained if resources are allocated to a small number of well-replicated experiments. For a strict advocate of the hypothetico-deductive method, replication is unnecessary even as a matter of principle, unless the predicted response is so weak that random background noise is a plausible excuse for a discrepancy between predictions and results. By definition, a prediction is an ‘all-statement’, referring to all systems within a well-defined category. What applies to all must apply to any. Hence, choosing two systems and assigning them randomly to a treatment and a control is normally an adequate design for a deductive experiment. The strength of such experiments depends on the firmness of the predictions and their *a priori* probability of corroboration. Replication is but one of many ways of reducing this probability. Whether the experiment is replicated or not, inferential statistics should always be used, to enable the reader to judge how well the apparent patterns in samples reflect real patterns in statistical populations. The concept ‘pseudoreplication’ amounts to entirely unwarranted stigmatization of a reasonable way to test predictions referring to large-scale systems.

*L. Oksanen, Dept of Ecology and Environmental Science, Umeå Univ., SE-901 87 Umeå, Sweden (lauri.oksanen@eg.umu.se).*

The current attitude of ecologists to experimental work is dominated by the paper of Hurlbert (1984), emphasizing the need of genuine replication. The message of the paper can be summarized in the following three points. First, it is impossible to infer causal relationships from unreplicated experiments, because interactions between spatial and temporal variation in the system can then account for apparent treatment effects. Second, logically sound induction of causality can also be prevented by compound treatments (e.g. consistent

use of the same growth chamber for the same treatment), because even in this case, there are alternative explanations for the apparent treatment effect. Third, the use of inferential statistics without true replication is not informative, because the null hypothesis that two statistical populations are identical is trivially wrong in the living nature. The term ‘pseudoreplication’ is introduced as a stigmatizing label for experimental studies, where inferential statistics have been used in the context of unreplicated or compound treatments.

---

Accepted 14 February 2001

Copyright © OIKOS 2001

ISSN 0030-1299

Printed in Ireland – all rights reserved

Hurlbert (1984: 199–200) recognizes why lack of replication is common in many subareas of ecology. In his own words: “The most common type of ‘controlled’ experiment in ecology involves a single ‘replicate’ per treatment. This is neither surprising nor bad. Replication is often impossible or undesirable when very large scale systems (whole lakes, watersheds, rivers, etc.) are studied. When gross effects of treatments are anticipated, or when only a rough estimate of effect is required, or when the cost of replication is very great, experiments involving unreplicated treatment may also be the only or best option. What is objectionable is when tentative conclusions derived from unreplicated treatments are given an unmerited veneer of rigor by the erroneous application of inferential statistics”. In other parts of the paper, this attitude is forgotten and lack of genuine replication is treated as a consequence of ignorance. The statement on p. 193 serves as an example: “Perhaps experimental ecologists fall primarily into two groups: those who do not see any need for dispersion, and those who do recognize its importance and take whatever measures are necessary to achieve a good dose of it”. It would be at least equally justified to maintain that experimental ecologists fall primarily into two groups: those who do not see any problems with reducing spatial and temporal scales in order to obtain replication, and those who understand that experiments must be conducted in spatial and temporal scales relevant for the predictions to be tested, and replicate the experiment as well as possible within this constraint (Carpenter 1992, 1996, Korpimäki and Norrdahl 1998, Persson et al. 1999, Klemola et al. 2000a, b).

Rather than presenting value-laden ex cathedra statements, a rationally thinking scientist should try to analyze the complex issues and trade-offs, which one inevitably faces when testing hypotheses concerning the dynamics of the large-scale systems. In this analysis, one should even consider the relationship between epistemological principles – the role of empiricism in the growth of knowledge – and the practical problems of experimental design.

## Trade-offs in studies focusing on large-scale systems

### The goal and the approaches

Most scientists agree that the best guarantee for progress is an unregulated pursuit of truth, guided by human curiosity. Curiosity can be directed to many kinds of questions and systems. If we only care about our publishing rates in top journals, we should indeed all work with such questions and systems where experimentation is cheap and easy. However, the human society, which pays our bills, expects that our curiosity-guided research even produces useful results. From the

dawn of our science (Cajander 1916) to the present (Power 2001), socially aware ecologists have argued that the primary task of ecology is to provide a scientific basis for rational and sustainable use of natural resources and for the protection of biodiversity and other valued aspects of nature. According to this argument, our ultimate goal is to understand the dynamics of the large-scale ecosystems on which our lives depend. In some of these systems (e.g. rivers, see Power 1990) the central questions can be tackled in restricted spatial and temporal scales, in others, this is not possible.

The logistic and economical problems of experiments in large-scale systems can be handled in at least four different ways. First, we can use microcosms, where organisms are small and have fast dynamics, as models of large-scale systems. Second, we can focus on predictions which are testable within a limited spatial scale or which refer to initial stages of transient dynamics. Third, we can replicate the control but leave the treatment unreplicated. Fourth, we can conduct an unreplicated experiment. In the framework of the fourth alternative, the experimentalist can either (4a) refrain from computing inferential statistics, or (4b) ‘pseudoreplicate’ the experiment by computing implicit or explicit statistics, which focus on spatial differences between the treatment and the control. Hurlbert’s paper was primarily a critique of alternative (4b). To obtain a more objective perspective, it is useful to look at the limitations of the other alternatives, too.

### Alternative 1: microcosm experiments

Due to its technical feasibility, this alternative is currently quite popular (Daehler and Strong 1996, Drake et al. 1996, Lawton 1996). An illustrative example of its strengths and limitations is provided by the history of competitive exclusion. The principle was first formulated by Cajander (1909, 1916, 1926; see also Oksanen 1991), who had realized the enormous difference in the regularity of forest vegetation between Siberia and of what then was the western parts of the Russian Empire. He argued that the perceived discreteness of plant communities in Siberia was a consequence of competitive exclusion in an undisturbed environment, and of the ability of different subordinate species to survive in association with different dominants. In northern Europe, the principle of competitive exclusion got quickly accepted, possibly because the idea of relatively discrete plant associations thus created was what plant sociologists wanted to hear. The breakthrough in the English-speaking world came three decades later, in response to Gause’s (1934) microcosm experiments, where the processes, supposed to have formed the Siberian taiga under centuries of undisturbed struggle for existence, could be reproduced in a few days. However, Gause’s results did not prove that terrestrial vegetation is nor-

mally shaped by competitive exclusion between potential dominants, causing step-wise responses of the vegetation to environmental gradients. Indeed, the prevailing mood in plant ecology is that vegetational gradients are continuous, each species responding individually to environmental gradients. Sharp boundaries, created by competitive exclusion, are regarded as non-existent (Whittaker 1975) or rare (Crawley 1986).

Microcosm experiments have also been used in testing the hypothesis of exploitation ecosystems (Oksanen et al. 1981, Oksanen and Oksanen 2000, to be referred to as EEH), claiming that dynamics in the endotherm branches of terrestrial grazing webs change along major terrestrial productivity gradients, and that these changes have major impacts on terrestrial vegetation patterns (L. Oksanen 1990), on strategies of folivorous mammals (Oksanen 1992), and on their population fluctuations (Turchin et al. 2000). The study of Kaunzinger and Morin (1998) demonstrates that dynamics in a microbial food web depend on potential productivity, in accordance to the predictions of EEH. Naaem and Li (1998) show that a species-rich guild of herbivorous protists can exert uniform, strong grazing pressure on a community of unicellular algae, which corroborates the guild-population analogy proposed by Oksanen et al. (1981), and suggests that the complications created by individual characteristics of species (Abrams 1992, 1993) are primarily relevant for species-poor communities, where some plant (prey) species may be effectively defended against all herbivores (predators). These corroborating results from microcosm experiments are encouraging. However, the realism of the EEH in its primary context is still open to debate (Hunter and Price 1992, Strong 1992, Polis and Strong 1996, Oksanen et al. 1997, Polis 1999, Oksanen and Oksanen 2000).

To summarize, microcosm experiments are a natural and potentially very profitable step between mathematical modeling and empirical tests in large-scale terrestrial and aquatic systems. If, for instance, a hypothesis focuses on dynamics in autotroph-based systems, it should indeed apply to autotroph-based microcosms, too. However, corroborative evidence from microcosms only suggests that the hypothesis in question is promising and worth further work. Spatial scale and properties of plants and animals may influence even fundamental principles of community structure. After promising results from microcosms, experiments with the real thing should thus follow. If this does not happen, our views of dynamics in large-scale systems may depend primarily on changing fashions and on debating skills of scientists. Carpenter (1996), argues the technical advantages of microcosm studies even create the risk that ecology deteriorates to ‘microcosmology’, which has lost its contact to problems in large-scale systems. However, the dominance of labora-

tory and microcosm studies in leading journals peaked in 1960 and has declined since then (Ives et al. 1966). This indicates that microcosm studies have retained their support function and are not ‘eating up’ field ecology.

## Alternative 2: to focus on predictions on transient dynamics and individual behavior

This alternative became popular when experiments on asymptotic dynamics of large-scale systems became almost infeasible, due to the perceived need of replication, and many ecologists still wanted to work on macroscopic systems. Good examples are provided by the greenhouse experiment on the impact of individual voles on herb-rich boreal vegetation by Moen et al. (1993), and by the enclosure experiment of Desy and Batzli (1989) on the short-term impacts of added food and reduced predation on population dynamics of microtine rodents. The greenhouse experiment demonstrated that what seemed food was indeed edible for voles: they started to lose weight first when the standing crops of their supposed forage plants were severely reduced. The problem with such a ‘sledgehammer design’ is that it only refuted the ‘green desert’ hypothesis of White (1978). The obtained results are entirely consistent with the hypothesis of induced defense (Haukioja and Hakala 1975, Rhoades 1985, Seldal et al. 1994), because grazing in the enclosures was at once so intense that it could have conceivably overwhelmed the defensive mechanisms of plants. The enclosure experiment of Desy and Batzli, in turn, showed that in a short-time perspective, both food and predation count and the impacts of predator removal and supplemental food are additive. Unfortunately, the implications of the results are ambiguous. Even in laissez-faire predator-prey systems, where equilibrium prey density is assumed to depend on predation alone (Rosenzweig 1971), addition of prey’s food initiates transient dynamics, where prey density increases. If food is provided in a form which allows the prey to reduce its foraging time, even equilibrium prey density may increase (Abrams 1984, Oksanen et al. 2001). The positive short-term responses to predator removal, in turn, is consistent with all hypotheses of resource limitation with an implicit or explicit time delay (e.g. induced defenses or limitation by winter forage). Mortality can be additive in a short time frame but compensatory in a longer perspective.

The studies discussed above represent serious and partially successful efforts to work directly with the target system within the constraints set by replication and by the time and fiscal limits of grants and PhD projects. However, these studies also illustrate the limitations of alternative (2). There is no guarantee for the existence of critical predictions referring to short-term

dynamics or to dynamics in systems with strongly reduced spatial scales. Thus, even this technique is to be regarded as an intermediate step. It may succeed in eliminating some conjectures, but several competitors may survive the tests, and the struggle for existence between them may shift from experimental fields to debating rooms, unless we are ready to move on when the potential of alternative (2) has been exhausted.

### **Alternative 3: to compare a single treatment with replicated controls**

If the main problem is the high cost of the treatment, then a natural choice is to leave the treatment unrelicated but to replicate the controls. With  $n$  replicated controls and one treatment, we obtain  $n - 1$  degrees of freedom for testing the statistical hypothesis that the apparent treatment effect arises just by chance due to random variation within the spatial scale where the experiment is conducted. If the treatment has been randomly chosen from the statistical population in question, and if it becomes a statistical outlier after the onset of the experiment, the existence of a statistically significant treatment effect has been established. This approach has in a way been used already by Paine (1966, 1974), although his statistics are not computed accordingly. A technically correct example of this approach is the study of Reid et al. (1994) on the impact of predation on a coastal population of collared lemmings.

There are some limitations to the use of this alternative. Many statistical techniques require an estimate of variance for both the control and the treatment. Behind this technical issue lies a more fundamental problem. In principle, a treatment can influence the mean of a statistical population, its variance, or both. In the absence of replicated treatments, it is impossible to tell between these alternatives. If, for instance, we compare dynamics of target population(s) between an unrelicated predator exclosure and replicated controls, and find that the exclosure becomes an upper outlier with reference to numbers and survival rates, the statistics by themselves do not warrant the conclusion that the treatment has increased numbers and survival rates. The same response could be observed if the exclusion of predators had increased the spatial variation in densities and survival rates, and our single treatment just happened to be in the upper end of the statistical distribution of treatments. When Reid et al. (1994) conclude that predator removal has a strong, positive impact on the survival rates of collared lemmings, the conclusion is actually derived from a combination of statistical and ecological reasoning. Statistics tell that the treatment has an effect. Ecological arguments suggest that if predator removal has an effect, it is likely to be on mean survival rates, not on their spatial varia-

tion. The difference as compared to the logic of pseudoreplication (see below) is that ecological reasoning enters the scene at a later stage in the chain of arguments.

### **Alternative 4: to conduct an unreplicated experiment**

In many studies focusing on large-scale systems and large organisms, high cost of treatment is not the only problem. Even the logistic problems of working with replicated controls can be formidable. The problem of obtaining genuine replicates is especially severe if the hypothesis to be tested predicts that the characteristics of ecological processes change along major environmental gradients. Currie's (1991) hypothesis on the relation between productivity and biodiversity is one example, EEH is another. As for Currie's hypothesis, only comparative 'tests' seem possible. When testing EEH, the logical unit is not a single experiment but a pair of experiments, conducted on different sides of the productivity threshold where predators are supposed to lose their controlling impact on folivorous endotherms. The crucial variable is not the treatment effect itself but treatment  $\times$  habitat interaction along a large-scale productivity gradient (Moen and Oksanen 1998). Scale cannot be reduced to make replication more feasible because in small-scale habitat mosaics, the impact of local dynamics can be overridden by 'spillover predation' (Holt 1985, T. Oksanen 1990, Oksanen et al. 1992).

As large-scale productivity gradients tend to be directional in one way or another, genuine replication of such gradients is almost impossible. There are several ways to cheat. An especially efficient one, seriously proposed by a referee of the paper of Moen and Oksanen (1998), is to conduct pairs of experiments on slopes of different mountains, thus making the set of experiments look well-dispersed, when the design is presented in two dimensions. However, altitude is just one dimension among others, and the impacts of many ecological factors correlate with altitude. Segregation along the altitudinal axis is thus at least as serious a problem as horizontal segregation between systems for which different treatment effects are predicted. Another way to mislead referees and readers is to treat the site as a random factor. This may look clean, but the statistics then focus on the question whether there are some unspecified spatial differences in the ecological processes to be studied (Underwood 1997). This question is trivial: all ecological processes show some spatial variation, and to detect it is just a matter of sample size. EEH predicts specific changes in trophic dynamics in the endotherm branches of terrestrial grazing webs at specific productivity thresholds. Hence, productivity must be treated as a fixed factor when testing EEH,

which severely reduces the feasibility of genuine replication. In the absence of replication, we have the following two choices.

*Alternative 4a: to present the results of unreplicated experiments without giving them ‘an unmerited veneer of rigor by the erroneous application of inferential statistics’*

Hurlbert's critique was not primarily directed against conducting unreplicated experiments but against the use of inferential statistics in these contexts. The most obvious alternative is provided by Bayesian statistics (Press 1989). In a nutshell, Bayesian statistics represents an attempt to quantify the surprise element, which is the essence of all good experiments. The hypothesis to be tested is assigned an a priori probability, and the probability of the predicted response is estimated under two conditions: provided that the hypothesis is true and provided that the hypothesis is false. By the rules of probability calculus, we can then estimate the a posteriori probability of the hypothesis, given different imaginable outcomes of the experiment. A good experimental project is one where different conceivable outcomes either reduce the a posteriori probability of the hypothesis (hypotheses) to the vicinity of zero (refutation) or boost it a long step towards one (hard corroboration). In subjective contexts – when planning experiments and when reviewing grant proposals and manuscripts – I have found Bayesian statistics useful. As an objective technique, however, Bayesian statistics is problematic, because of the subjective elements involved in the estimations of a priori probabilities. The proposed solution is to interview colleagues, familiar with the system in question. It is easy to see how vulnerable this procedure would be for manipulation. Hence, Bayesian statistics seems to be destined to remain useful for private purposes only.

As for inferential statistics, Hurlbert maintains that its use in the connection of unreplicated experiments is inherently erroneous, because in the absence of replication, statistics can only tell us that there are spatial differences in nature – which we know anyway. Curiously, Hurlbert's message has not penetrated the world of descriptive studies, where statistical tests of spatial and temporal differences still abound. This ‘error’ has not prevented publication in top journals. During 1985–1995, a large and stable fraction (about 20%) of ecological papers, published in the three most influential ecological journals, have been descriptive (Ives et al. 1996). When discussing the use and abuse of inferential statistics, Hurlbert forgets that inferential statistics does not only answer the question whether or not two statistical populations can be regarded as different. Statistics also allow us to evaluate, how different the two statistical populations must at least be, whether their temporal patterns can be regarded as divergent, and much else. In a statistics-free presentation of a data

set, the author presents his samples and asks the reader to blindly believe that patterns in the samples represent patterns in the sampled statistical populations. Unnecessary lack of inferential statistics is thus rudeness towards the reader.

*Alternative 4b: to ‘pseudoreplicate’ the experiment*

If an experiment is not replicated, there is no possibility to strictly statistically establish a connection between the treatment and the apparent effect. What can be done is to analyze temporal and spatial variation in the measurables of interest within the treatment and the control. With such an analysis, we can answer the question whether the time trajectories in the two areas start to diverge when the experiment is initiated. In the classical intertidal works of Menge (1972) and Paine (1966, 1974) this has been done a bit clumsily – by showing that before the onset of the experiment, the treatment and the control(s) appeared to behave as if they were a single statistical population, whereas later on, the difference between the two became significant with the same sample size. As Hurlbert correctly pointed out, the authors would have certainly found statistically significant differences between the subareas even initially, if only their samples had been large enough. It is more appropriate to accept that no two biological systems are identical and to ask what the data tell us about the magnitudes of the difference before and after the onset of the experiment.

Techniques for answering the above question have been discussed by several authors (Carpenter 1993, Osenberg et al. 1992, Stewart-Oaten et al. 1992). The majority of the proposed methods are based on sequential point estimates of differences between the two statistical populations before and after the onset of the experiment, the sampling interval being longer than the time scale of temporal autocorrelations in the data set. The statistical problems discussed in the papers arise from uncertainties of determining the appropriate time interval and from the fact that the estimated differences between the statistical populations are influenced by sampling error. Another, more conservative technique is to estimate the confidence intervals for the means of the two statistical populations before the experiment and after the experiment has started, and to compare the maximum estimate for the difference between the pre-treatment means to the minimum estimate for the difference between post-treatment means.

Whichever technique one chooses, the logic of such ‘classical pseudoreplication’ can be summarized as follows. Let  $p_0$  be the probability of obtaining the observed difference between the treatment and the control under the null hypothesis of no true treatment effect, let  $p_v$  be the probability of obtaining the difference as a consequence of sampling and measurement error and random within-site variation, and  $p_l$  be the probability that some local factor other than the treatment would

account for the observed divergence. According to basic rules of probability calculus we obtain:

$$p_0 = 1 - (1 - p_v)(1 - p_t) = p_v + p_t - p_v p_t \quad (1)$$

The value of  $p_v$  is obtained from inferential statistics as outlined above. The value of  $p_t$  must be derived from biological reasoning, involving system-specific knowledge. In situations where experienced ecologists agree that  $p_t \approx 0$ , the value of  $p_0$  emerges directly from inferential statistics.

### An evaluation of the alternatives

Out of the alternatives discussed above, only one – 4a – is genuinely suboptimal. With or without replication, inferential statistics helps the reader to distinguish patterns from scatter and provides an objective estimate for  $p_v$ . Hence, the decision to refrain from using inferential statistics in the context of unreplicated experiments is just as rational as the decision of a hiker not to use a trail at all, when he knows that the trail ends a bit before the target of the hike. All other alternatives have their strengths and their realms of application. Alternative (1) provides a natural first step. Alternative (2) is a potentially useful next stage, provided that the predictions of some contesting hypotheses diverge in small spatial and/or temporal scales. Ultimately, however, one has to proceed to large-scale experiments. In this context, genuine replication is the ideal design for an individual scientist. However, agencies allocating grant money should think twice before giving full funding for a proposal which is extremely costly due to the combined effects of large spatial and temporal scales and replication. As pointed out by Carpenter (1992), our collective rate of progress could be enhanced if the resources were allocated to a larger number of unreplicated experiments. This would provide more material for meta-analyses, and the greater number of separate experiments would reduce the sampling bias in the selection of the experimental systems.

As the prefix ‘pseudo’ implicitly refers to false pretenses, it might be profitable to restrict the use of the term pseudoreplication to situations where an experimentalist tries to give a false picture of the experimental design. On the other hand, one could apply a Finnish proverb and maintain that a label does not defame an approach if the researchers using the approach do not defame the label. Call it pseudoreplication or whatever you wish, but an unreplicated test of strong and critical predictions is likely to be more instructive than a well-replicated test of weak and trivial ones, and computing inferential statistics is just courtesy towards the reader.

## The epistemological dimension

### On Hurlbert's epistemological position

Revealing for Hurlbert (1984) philosophy is that even compound treatments (case B-4 in Hurlbert's Fig. 1) are regarded as pseudoreplication. He is indeed right when claiming that no clean statistical inferences of causality can be induced from a growth chamber experiment where the same growth chamber is always used for producing a given set of environmental conditions. However, the reasoning could be pursued further. The essence of compound treatments is that statistically significant differences between the treatment and the control can emerge, due to unintended side effects of the treatment. This is unfortunately the state of affairs in all experiments. When we conduct an experiment, we inevitably do all kinds of things in addition to the manipulation of the supposed causal factor. Sometimes, the unintended side effects are obvious – like in an enclosure experiment conducted by myself and H. Henttonen (unpubl.), where all bilberry twigs in our vole exclosures died, as hoarfrost, forming in the steel

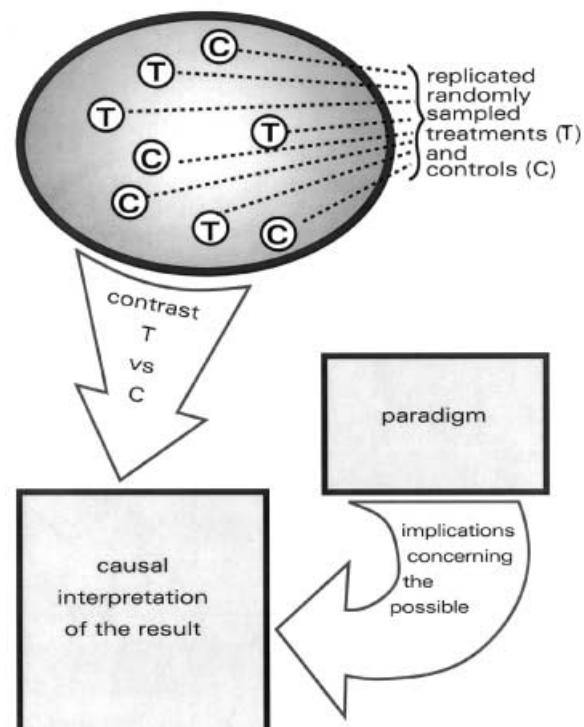


Fig. 1. The logic of an inductive experiment. Solid statistical inference requires both random sampling from the entire statistical population to be studied and replication. Causal interpretation is possible by interpreting the contrast between treatments and controls in the light of the prevailing paradigm.

net, even excluded snow – sometimes the unintended side effects are more difficult to detect. A classical example of the impossibility to eliminate even those side effects, which the experimentalists are well aware of, is the ‘fence effect’ inherent in predator exclusion experiments. When excluding predators, we even prevent movements of prey, which may have population dynamical consequences (Krebs et al. 1969). We can provide sink areas, accessible to animals with high motivation to disperse (Desy and Batzli 1989), but the one-way gates thus produced do not adequately simulate the possibilities of both emigration and immigration, characteristic for open systems (Ostfeld 1994). The use of semi-permeable exclosures (Erlinge 1987, Krebs et al. 1992, Reid et al. 1994), in turn, severely weakens the design, by allowing predation outside the enclosure to indirectly influence dynamics inside, by maintaining a dispersal-provoking density gradient.

Let us face it. If the concept pseudoreplication is used in the broader sense, including compound treatments, then all experiments are pseudoreplicated, though we do not always have enough information to understand how. In applied research, this does not matter, because the ultimate goal of experiments is to establish a statistical connection between the treatment and the outcome. The questions asked by the experimentalist are: ‘How much does a given treatment (e.g. fertilization of cropfields and forests, medical treatment of people) help in achieving a societal goal (larger crops, healthier people). This question can be rationally approached by means of a replicated experiment, because the primary interest is in the statistical connection itself, while causal interpretation is of secondary importance and can be derived from the prevailing paradigm (Fig. 1). However, the fundamental discovery of Popper (1933, 1934) is that causality cannot ever be reliably inferred from any kind of data. Hence, as a method for basic sciences, inductionism has been dead for decades, and its resurrection in ecology in 1984 is truly amazing.

### The Bayesian perspective on experimental work

The overwhelmingly dominating epistemology of basic sciences is the hypothetico-deductive approach. It can be divided into two main variants: the Bayesian variant, which emphasizes corroboration of seemingly unlikely predictions, and the Popperian variant, which emphasizes falsification. It is natural to start from Bayesianism (Russell 1943), both for historical reasons and because the classical papers of community ecologists have been written in the Bayesian spirit (see below). Concerning the proper structure of the scientific process, advocates of both variants agree. By definition, empirical sciences need empirical inputs – either in the form of conflicts between predicted and observed data or in the form of previously undetected or unexplained

empirical patterns. Form these empirical inputs, combined with a pre-existing body of corroborated theories, the scientist induces new conjectures, which are formalized to allow a rigorous analysis of their logical structures and deduction of their empirical consequences. These consequences, referred to as predictions, must concern all elements belonging to a well-defined category. The normal scope of experiments is to arrange a situation where the predicted and observed responses of the elements can be compared. As the predictions concern all elements, the behavior of any single element provides an adequate test.

In principle, experiments have no logical priority over spontaneous experimental situations, and neither replication nor control are necessary parts of a critical experiment. Indeed, the most influential ‘experiment’ of the past century – the bending of Mercury’s light by the gravitational field of the sun – was a spontaneous experimental situation, which was neither controlled nor replicated. To have a control whenever feasible is nevertheless rational, because the control allows the reliability of the measuring techniques to be tested and the aberrant nature of the response to be clearly visualized. The control can also serve as a ‘dummy treatment’, where the experimentalist tries to reproduce as many side-effects of the treatment as possible, in order to demonstrate that these side effects do not account for the response of the system. (If both the treatment and the control show the behavior predicted for the treatment only, the experiment is a technical failure, as spontaneous changes or side-effects of the treatment then obviously account for the response.) To control the experiment is logically necessary if past observations suggest that the plausible range of spontaneous temporal variation in the category of systems exceeds the predicted magnitude of the treatment effected, because lack of control would then create a substantial risk for spurious corroboration or spurious falsification. As practically all ecological systems are characterized by occasional, very dramatic changes, control can indeed be regarded as an essential part of ecological experiments. Replication is logically necessary, if the observed background variation is both large enough and sufficiently heterogeneous in space to potentially override the predicted contrast between the treatment and the control. The logical structure of an experimental test of deduced predictions is illustrated in Fig. 2.

The classical intertidal experiments have been clearly conducted in Bayesian spirit. Their focus is on abstract ideas on ecological processes – competition between basal organisms (Connell 1961a, b) or between top predators (Menge 1972), or the impact of predation and herbivory on competitive interactions between basal organisms (Connell 1961a, Paine 1966, 1974, 1980, Lubchenko 1980). These ideas are first speculatively connected to observed spatial patterns in community

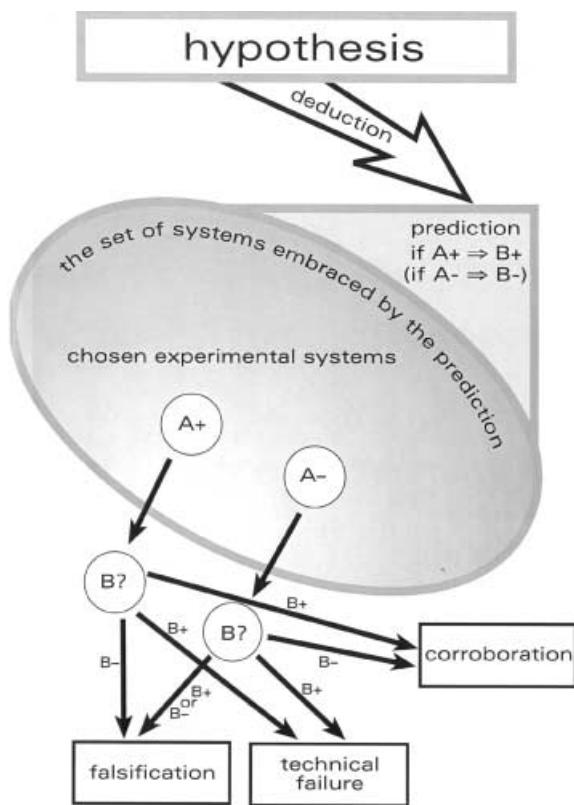


Fig. 2. The logic of an experiment conducted in the spirit of the hypothetico-deductive epistemology. Predictions concerning all systems belonging to a well-defined category are deduced from the hypothesis. A typical prediction states that under the treatment conditions 'A', any system belonging to the category must shift to state 'B'. Such a shift must be very unlikely to happen because of reasons unrelated to the hypothesis to be tested (e.g. due to environmental stochasticity). If this shift is observed in the experimental system, subjected to conditions 'A', but not in a reference system, subjected to a 'dummy treatment' (similar procedures except for the critical treatment condition 'A'), the hypothesis is corroborated. If the shift to state 'B' is not observed under conditions 'A', the hypothesis is falsified regardless of what happens in the reference system. If a shift to 'B' is observed in both systems, the experiment is a technical failure (undesired side effects of the treatment are likely to account for the apparent treatment effect).

structure and to generally accepted ecological laws (Turchin 2001). From the conjectures thus developed, the authors deduce predictions, applying to broad and clearly defined categories of systems (e.g. rocky intertidals of the Pacific Northwest or New England). To test these conjectures, the supposed, strong interactors are removed and the predicted, drastic changes in the rest of the community are observed. As the predicted changes are of a totally different order of magnitude than the variation normally observed in these systems, and the treatment and the control(s) are selected from sites similarly exposed to plausible confounding factors (storms), no need for replication is perceived, and noth-

ing is done to conceal the lack of replication. Except for Lubchenko's study, where a genuinely replicated competitor removal experiment is nested within a comparative, unreplicated herbivory study, the word 'replication' is not even mentioned. Inferential statistics are presented in the form of error bars, standard errors or standard deviations.

Rather than serving as examples of erroneous use of inferential statistics, these 'pseudoreplicated' experiments are designed and reported entirely in agreement with the basic principles of hypothetico-deductive science. In a Bayesian perspective, it is easy to understand why such 'unrigorous' studies have had such enormous impact on our thinking on community ecology and why they are visibly discussed in standard ecology texts (e.g. Begon et al. 1996), without any caveats referring to 'shortcomings' in the experimental design. First, the predictions to be tested are genuine. They have existed before the experiment, which prevents selective focusing on corroborating evidence. Second, the predictions are strong. Assuming that the conjecture is correct, there had been little room for excuses, if results had not corresponded to predictions. Third, the predictions are critical, i.e. have a very low a priori probability of corroboration. In a Bayesian perspective, these three attributes of an instructive empirical work can be combined to a single sentence. A good empirical study (whether experimental or observational) is a study which creates a large difference between the a priori and the a posteriori probability of the hypothesis to be tested. Provided that the a priori probability of the hypothesis is not zero, we do not need to worry about the debatable aspects of Bayesian statistics, because the a priori probability of the hypothesis is only a scaling factor.

The Bayesian approach has been under heavy criticism. The central argument is that the probability of a hypothesis to be a perfect truth is always zero (Popper 1963, Lakatos 1968, 1972), and zero, multiplied by any number, is still zero. This argument is derived from the assumption that every hypothesis is drawn from an infinite population of logically consistent and empirically plausible hypotheses. Thus, hitting the exactly right one is just as impossible as having a physical constant with an integer value. The reasoning is based on experience from physical sciences, where the objects of theories, such as elementary particles and electron orbits, are very far from the actual observations of the empiricist. In ecology, this distance is much smaller: we can actually see predators killing prey. Indeed, the mortality caused by predation might be compensatory, and observed changes in predation rates might be actually caused by some other factors, but it is not obvious that there would be infinitely many alternative explanations for each observed scenario. If infinity is replaced by any real number – no matter how large – the above argument falls apart and Bayesian convergence towards the truth becomes a real possibility.

## The falsificationist perspective

Philosophers of science generally agree that the Bayesian approach is a step forwards from inductionism, but further steps must be taken to provide a logically defensible framework for basic sciences. The choices involved have been vividly discussed in the 1960s and early 1970s, and the classics published then still retain their actuality. Popper (1963) insists that falsification is the only logically defensible way to go. Corroboration is for him a by-product of an empirical test, while only falsification really teaches us something new. In Popper's (1957) view, speculations create new ideas like mutations create new genes – and falsification eliminates those which are not fit (see also Platt 1964). Popper thus sees the scientific process as evolution lifted to a new level. In the past, when 'ideas' (ways to structure the environment) were direct expressions of genes, unfit ideas were eliminated by eliminating their carriers. During the evolution of humans, the emergence and spreading of new ideas became extremely fast, due to our capacity of speculation and communication, and the elimination of unfit ideas became a social process. In Popper's opinion, falsificationism represents an attempt to further streamline this process, freeing the struggle for existence between ideas from the constraints of social interactions between humans.

The main criticism against Popper's ideal is that strict falsificationism would be just as useful for the evolution of ideas as a full-scale nuclear war would be for the evolution of biota: practically everything would be eliminated (Kuhn 1970). As an alternative, Kuhn suggests that scientists should normally concentrate on fitting pieces of empirical evidence to the prevailing framework of theories (paradigms). This proposition opens the door for dogmatism, which is rampant in ecology anyway. The Kuhnian perspective is thus hardly attractive for us. Feyerabend (1970) argues that it is impossible to find rational norms for the scientific process. For him, a convincing piece of scientific work is simply whatever work which manages to convince peers. Feyerabend emphasizes that there is always an element of fine art in truly innovative science. This seems agreeable, but philosophers of science should be able to propose something more practical in order to be worth the resources invested in their salaries. A more balanced but still crisp and clear version of falsificationism is provided by Lakatos (1972). Like Popper, Lakatos emphasizes falsifications, but adds to the picture the hierarchical structure of conjectures. He argues that it is perfectly legitimate to start by revising the most specific and peripheral parts of a falsified conjecture. As for the central parts, forming the hard core of a research program, the essential question is whether there is another, competing program, which has higher success rate in generating corroborated conjectures. In biological terms, Lakatos replaces Popper's 'natural

selection by nuclear war', where theories stand or fall, when struck by the blast of empirical tests, by normal struggle for existence, leading to gradual competitive exclusion of some research programs – or to competitive coexistence, where each program finds its own niche.

The difference between the sophisticated falsificationism of Lakatos and the Bayesian perspective matters little for the practical issues of experimental design. The phrase 'low a priori probability of corroboration' means that colleagues, who know the system but are either unaware of the new conjecture or regard it as unrealistic, expect that the observed results will differ from the predicted ones. These expectations of informed peers derive from their view of nature, guided by accepted theories. Hence, corroboration of seemingly implausible predictions of one conjecture almost automatically amounts to falsification of another. The intertidal studies provide again excellent case points. When seen in a Bayesian perspective, the experiments corroborate genuine, strong and critical predictions of the Cajander–Gause theory of community ecology, according to which communities are true associations, where biotic interactions exclude many species, potentially capable of dealing with the physical environment, and have profound impact on abundance relationships in the community and on the relationship between community structure and environmental gradients. On the other hand, the results can be also seen as falsifications of the continuum conjecture, according to which intertidal communities are but haphazard sets of coexisting organisms, whose patterns of distribution and abundance along environmental gradients reflect relationships between the physical environment and the individual optima of each species (Gleason 1926, Whitaker 1975). In other words, strong corroboration of one conjecture and at least implicit falsification of another are just two sides of the same coin.

## Conclusions

### General viewpoints

The design of experiments is a tactical question, which cannot be rationally discussed without first tackling the strategical issue of whether the study is conducted in a deductive or inductive spirit. Both approaches have their roles in science. Inductive experiments can provide new, unexpected insights. Using evolutionary analogy, such experiments can be regarded as breeding grounds for promising mutants. Experimental tests of deduced predictions, in turn, intensify the struggle of existence between conjectures and research programs, increasing the likelihood that only the fittest will survive. The logic of induction requires replication, as pointed out by Hurlbert (1984). If the experiment is based on deductive

logic, the rules of the game are entirely different, as explained above, and replication is not an essential part of experimental design. The catch is that discussions of those results, which are not directly related to predictions, are not legitimate in this logical framework. The scope of a deductive experiment is not to provide a basis for induction from data but to allow the experimentalist to check ‘yes’ or ‘no’ boxes in a pre-existing test protocol. The rest of the discussion must focus on the implications of the test protocol upon the plausibility of the hypothesis (hypotheses) to be tested – as was the case in the intertidal studies discussed above. Below, I will summarize my personal reflections concerning the practical implications of the above viewpoints for the planning of experiments, for funding decisions, for the review process of submitted manuscripts, and for training of ecologists.

### To the experimentalist

When planning your experiment, be clear and consistent about the position of your work in the logical framework of the scientific process, and choose the experimental techniques accordingly. In experiments based on the deductive logic, preferentially focus on clear-cut cases, where the predicted response of the system vastly exceeds the range of background variation, and replication is unnecessary. Nevertheless, replicate the experiment, if feasible – it never hurts. Whether the experiment has been replicated or not, try to accompany all presentations of results with implicit or explicit inferential statistics. Statistics-free point estimates present patterns in samples. The reader is interested in patterns in the statistical populations, and you should allow him to judge what your data tell about these populations. If you conduct an experiment in the deductive spirit, and discover something unexpected and exciting in your results, it is your duty to clearly point out where your study becomes inductive, and to treat the results accordingly. And last but not least, please read Hurlbert’s (1984) paper. In spite of its shortcomings, the paper contains much useful information on how experiments should ideally be designed in a constraint-free situation.

### To the boards of grant allocation agencies

The problem to allocate limited resources in a way which maximizes our collective rate of progress is a tough one, and has not been made easier by the perceived need to replicate even those experiments which must be carried out in large spatial and temporal scales to retain their critical characteristics. However, in the context of deductive work, where the predicted changes are dramatic enough to override spontaneous diver-

gence within the spatial scale of the experimental design, there is no real need for replication. The money could be better spent on a larger number of unreplicated experiments, dispersed throughout the geographical area where a given set of predictions applies. Sampling bias would be reduced, while replication could be obtained afterwards by means of meta-analysis.

### To referees and editors

The term pseudoreplication has been so much abused that its value in a review is questionable. Referees should preferentially refrain from using it. Instead, they should specify perceived statistical problems, and editors should require such a practice. Replicated experiments with compound treatments should never be referred to as pseudoreplicated, because all treatments are inherently compound. As long as pseudoreplication is regarded as a bad word, the use of the concept in the context of compound treatments provides a blanket excuse for suppressing the publication of any experimental paper which the referee happens to dislike. It is reasonable to require that the author is explicit about his/her epistemological position and about the design of the experiment, but to require that inferential statistics should not be used in the context of unreplicated experiments is plain nonsense. If an author presents data which fulfill the basic requirements for computing interval estimates, referees and editors should insist that this computation must be done and the results must be clearly presented, either as error bars in graphs or in numerical form.

### To organizers of ecology curricula

There is a reason for second thoughts, when Hurlbert’s paper has been so widely accepted, in spite of its implicit dependence on a totally outdated epistemology. A probable reason lies in the way how methodological training is organized. Courses in philosophy of science and courses in statistics and experimental design are normally taught by entirely different teachers, with little knowledge of and even less interest in each other’s fields. Hurlbert’s paper is a striking example of this segregation. In a major methodological paper with 22 text pages, 34 explicitly statistical works are cited, and the list of references includes even one philosophy-related novel. However, there is not a single reference to epistemological books or papers. Conversely, scientists with an interest in philosophical issues have a tendency to regard statistics as uninteresting hack work, to be done quickly when needed and to be forgotten thereafter. My list of references can indeed be regarded as an indication of the latter bias. Yet, rational solutions to

the tactical problems of experimental design derive from strategical views on the relation between empiricism and growth of knowledge. Conversely, philosophy of science sways in its own spheres without being useful, unless it is explicitly connected to the day-to-day problems faced by active scientists. Hopefully, future ecologists will have access to courses produced jointly by philosophers of science and statisticians, to provide a good point of departure for methodological decisions – and maybe for feedbacks from ecology to philosophy of science.

**Acknowledgements** – The impulse to write this paper was provided by comments of an anonymous referee and the correspondence thus initiated, which made it clear to me that even considerate ecologists, with good grasp of both statistics and philosophy of science, can regard the two areas as largely unrelated to each other. In the discussion on the Bayesian approach I have used some crisp formulations from Steve Fretwell's unpublished manuscript 'The effective scientist'. Lennart Persson updated me on aquatic literature and pointed out the perspectives opened by meta-analysis. Mary Power, Tarja Oksanen, and Jon Moen helped me to make my formulations more communicative. The study has been supported by grants from NFR (Swedish Council for Natural Sciences) and SJFR (Swedish Council for Agriculture and Forestry).

## References

- Abrams, P. 1984. Foraging time optimization in food webs. – *Am. Nat.* 124: 106–124.
- Abrams, P. 1992. Predators that benefit prey and prey that harm predators: unusual effects of interacting foraging adaptations. – *Am. Nat.* 140: 573–600.
- Abrams, P. 1993. Effects of increased productivity on abundances of trophic levels. – *Am. Nat.* 141: 351–371.
- Begon, M., Harper, J. L. and Townsend, C. R. 1996. *Ecology: individuals, populations and communities*. 3rd ed. – Blackwell.
- Cajander, A. K. 1909. Über die Waldtypen. – *Acta For. Fenn.* 1: 1–175.
- Cajander, A. K. 1916. Metsähoidon perusteet. I Kasvibiologian ja kasvimaantieteen pääpiirteet. (Foundations of Forestry. I The main features of plant geography and plant biology, in Finnish.) – Werner Söderström, Porvoo, Finland.
- Cajander, A. K. 1926. The theory of forest types. – *Acta For. Fenn.* 29: 1–108.
- Carpenter, S. R. 1992. Replication and treatment strength in whole-lake experiments. – *Ecology* 70: 453–463.
- Carpenter, S. R. 1993. Statistical analysis of the ecosystem experiments. – In: Carpenter, S. R. and Kitchell, J. F. (eds), *The trophic cascade in lakes*. Cambridge Univ. Press, pp. 26–42.
- Carpenter, S. R. 1996. Microcosm experiments have limited relevance for community and ecosystem ecology. – *Ecology* 77: 677–680.
- Connell, J. H. 1961a. The effects of competition, predation by *Thais lapillus*, and other factors on natural populations of the barnacle *Balanus balanoides*. – *Ecol. Monogr.* 31: 61–104.
- Connell, J. H. 1961b. The influence of interspecific competition and other factors on the distribution of the barnacle *Chtalamus stellatus*. – *Ecology* 42: 710–723.
- Crawley, M. J. 1986. The structure of plant communities. – In: Crawley, M. J. (ed.), *Plant ecology*. Blackwell, pp. 1–50.
- Currie, D. J. 1991. Energy and large scale patterns of animal and plant species richness. – *Am. Nat.* 137: 27–49.
- Daehler, C. C. and Strong, D. J. 1996. Can you bottle nature? The role of microcosms in ecological research. – *Ecology* 77: 663–664.
- Desy, E. A. and Batzli, G. O. 1989. Effects of food availability and predation on prairie vole demography: a field experiment. – *Ecology* 70: 411–421.
- Drake, J. A., Huxel, G. R. and Hewitt, C. L. 1996. Microcosms as models for generating and testing community theory. – *Ecology* 77: 670–677.
- Erlinge, S. 1987. Predation causing non-cyclicity in a microtine population. – *Oikos* 50: 347–352.
- Feyerabend, P. K. 1970. *Against method*. – Minnesota Univ. Press.
- Gause, G. F. 1934. The struggle for existence. – Williams & Wilkins.
- Gleason, H. A. 1926. The individualistic concept of the plant association. – *Bull. Torrey Bot. Club* 53: 7–26.
- Haukioja, E. and Hakala, T. 1975. Herbivore cycles and periodic outbreaks: formulation of a general hypothesis. – *Rep. Kevo Subarctic Res. Stn.* 12: 1–9.
- Holt, R. D. 1985. Population dynamics in two patch environments: some anomalous consequences of an optimal habitat distribution. – *Theor. Popul. Biol.* 28: 181–208.
- Hunter, M. D. and Price, P. W. 1992. Playing chutes and ladders: bottom-up and top-down forces in natural communities. – *Ecology* 73: 724–732.
- Hurlbert, S. H. 1984. Pseudoreplication and the design of ecological field experiments. – *Ecol. Monogr.* 54: 187–211.
- Ives, A. R., Foufopoulos, J., Klopper, E. D. et al. 1966. Bottle or big-scale studies: how do we do ecology? – *Ecology* 77: 681–685.
- Kaunzinger, C. M. K. and Morin, P. J. 1998. Productivity controls food chain properties in microbial communities. – *Nature* 395: 495–497.
- Klemola, T., Norrdahl, K. and Korpimäki, E. 2000a. Do delayed effects of overgrazing explain population cycles in voles? – *Oikos* 90: 509–516.
- Klemola, T., Koivula, M., Korpimäki, E. and Norrdahl, K. 2000b. Experimental tests of predation and food hypotheses for population cycles of voles. – *Proc. R. Soc. Lond. B* 267: 351–356.
- Korpimäki, E. and Norrdahl, K. 1998. Experimental reduction of predators reverses the crash phase of small mammal cycles. – *Ecology* 79: 2448–2455.
- Krebs, C. J., Keller, B. L. and Tamarin, R. H. 1969. *Microtus* population biology: demographic changes in fluctuating populations of *Microtus ochrogaster* and *M. pennsylvanicus* in southern Indiana. – *Ecology* 50: 587–607.
- Krebs, C. J., Boonstra, R., Boutin, S. et al. 1992. What drives the snowshoe hare cycle in Canada's Yukon? – In: McCullough, D. M. and Barrett, R. (eds), *Wildlife 2001: Populations*. Elsevier, pp. 886–896.
- Kuhn, T. S. 1970. *The structure of scientific revolutions*, 2nd ed. – Univ. of Chicago Press.
- Lakatos, I. 1968. Changes in the problem of inductive logic. – In: Lakatos, I. (ed.), *The problem of inductive logic. Proceedings of the International Colloquium in the Philosophy of Science, London, 1965*, Vol. 2. North Holland.
- Lakatos, I. 1972. Falsification and the methodology of scientific research programmes. – In: Lakatos, I. and Musgrave, A. (eds), *Criticism and the growth of knowledge*. Cambridge Univ. Press, pp. 91–196.
- Lawton, J. H. 1996. The ecotron facility at Silwood Park: the value of "big bottle" experiments. – *Ecology* 77: 665–669.
- Lubchenko, J. 1980. Algal zonation in the New England rocky intertidal community: an experimental analysis. – *Ecology* 61: 333–344.
- Menge, B. A. 1972. Competition for food between intertidal starfish species and its effect on body size and feeding. – *Ecology* 53: 635–644.

- Moen, J., Gardfjell, H., Oksanen, L. et al. 1993. Grazing by food-limited microtine rodents on a productive experimental plant community: does the "green desert" exist? – *Oikos* 68: 401–413.
- Naaem, S. and Li, S. 1998. Consumer species richness and autotroph biomass. – *Ecology* 79: 2603–2615.
- Oksanen, L. 1990. Predation, herbivory, and plant strategies along gradients of primary productivity. – In: Grace, J. and Tilman, D. (eds), *Perspectives on plant consumption*. Academic Press, pp. 445–474.
- Oksanen, L. 1991. A century of community ecology: how much progress? – *Trends Ecol. Evol.* 6: 294–296.
- Oksanen, L. 1992. Evolution of exploitation ecosystems. I. Predation, foraging ecology and population dynamics in herbivores. – *Evol. Ecol.* 6: 15–23.
- Oksanen, L. and Oksanen, T. 2000. The logic and realism of the hypothesis of exploitation ecosystems. – *Am. Nat.* 155: 703–723.
- Oksanen, L., Fretwell, S. D., Arruda, J. and Niemelä, P. 1981. Exploitation ecosystems in gradients of primary productivity. – *Am. Nat.* 118: 240–261.
- Oksanen, L., Aunapuu, M., Oksanen, T. et al. 1997. Outlines of food webs in a low arctic tundra landscape in relation to three theories of trophic dynamics. – In: Gange, A. C. and Brown, V. K. (eds), *Multitrophic Interactions in terrestrial Systems*. Blackwell, pp. 351–373.
- Oksanen, T. 1990. Exploitation ecosystems in heterogeneous habitat complexes. – *Evol. Ecol.* 4: 220–234.
- Oksanen, T., Oksanen, L. and Gyllenberg, M. 1992. Exploitation ecosystems in heterogeneous habitat complexes II: impact of small-scale heterogeneity on predator-prey dynamics. – *Evol. Ecol.* 6: 383–398.
- Oksanen, T., Oksanen, L., Jedrzejewski, W. et al. 2001. Predation and the dynamics of the bank vole, *Clethrionomys glareolus*. – In: Bujalska, G. and Hansson, L. (eds), *Bank vole biology: recent advances in the population biology of a model species*. Pol. J. Ecol. 48 (suppl.): 197–217.
- Osenberg, C. W., Holbrook, S. J. and Schmitt, R. J. 1992. Implications for the design of environmental assessment studies. – In: Griffmann, P. M. and Yoder, S. E. (eds), *Perspectives on the marine environment*. USC Sea Grant Program, Los Angeles, pp. 75–89.
- Ostfeld, R. S. 1994. The fence effect reconsidered. – *Oikos* 70: 340–348.
- Paine, R. T. 1966. Food web complexity and species diversity. – *Am. Nat.* 100: 65–75.
- Paine, R. T. 1974. Intertidal community structure: experimental studies on the relationship between a dominant competitor and its principal predator. – *Oecologia* 15: 93–120.
- Paine, R. T. 1980. Food webs, linkage, interaction strength and community infrastructure. – *J. Anim. Ecol.* 49: 667–685.
- Persson, L., Byström, P., Wahlström, E. et al. 1999. Interactions among size-structured populations in a whole-lake experiment: size- and scale-dependent processes. – *Oikos* 87: 139–156.
- Platt, J. R. 1964. Strong inference. – *Science* 146: 347–352.
- Polis, G. A. 1999. Why are parts of the world green? Multiple factors control productivity and the distribution of green biomass. – *Oikos* 86: 3–15.
- Polis, G. A. and Strong, D. R. 1996. Food web complexity and community dynamics. – *Am. Nat.* 147: 813–846.
- Popper, K. 1933. Ein Kriterium des empirischen Charakters theoretischer System. – *Erkenntnis* 3: 426–427.
- Popper, K. 1934. Logik der Forschung: zur Erkenntnistheorie der modernen Naturwissenschaft. – *Schriften zur wissenschaftlichen Weltanschauung* 9. – Wien.
- Popper, K. 1957. The aim of science. – *Ratio* 1: 24–35.
- Popper, K. 1963. Conjectures and refutations. – Harper & Row.
- Power, M. E. 1990. Effects of fish in river food webs. – *Science* 250: 811–814.
- Power, M. E. 2001. Field biology, food web models, and management: challenges of context and scale. – *Oikos* 94: 118–129.
- Press, S. J. 1989. *Bayesian statistics*. – Wiley & Sons.
- Reid, D. G., Krebs, C. J. and Kenney, A. 1994. Limitation of collared lemming population growth at low densities by predation mortality. – *Oikos* 387–394.
- Rhoades, D. W. 1985. Offensive-defensive interactions between herbivores and plants: their relevance in herbivore population dynamics and ecological theory. – *Am. Nat.* 125: 205–238.
- Rosenzweig, M. L. 1971. Paradox of enrichment: the destabilization of exploitation ecosystems in ecological time. – *Science* 385–387.
- Russell, B. 1943. Reply to criticisms. – In: Schilp, P. A. (ed.), *The philosophy of Bertrand Russell. The Library of Living Philosophers*, Evanston, IL, pp. 681–741.
- Seldal, T., Andersen, K.-J. and Höglstedt, G. 1994. Grazing-induced proteinase inhibitors: a possible cause for lemming population cycles. – *Oikos* 70: 3–11.
- Stewart-Oaten, A., Bence, J. R. and Osenberg, C. W. 1992. Assessing effects of unreplicated perturbations: no simple solutions. – *Ecology* 73: 1396–1404.
- Strong, D. R. 1992. Are trophic cascades all wet? Differentiation and donor-control in speciose ecosystems. – *Ecology* 73: 747–754.
- Turchin, P. 2001. Does population ecology have general laws? – *Oikos* 94: 17–26.
- Turchin, P., Oksanen, L., Ekerholm, P. et al. 2000. Are lemmings prey or predators? – *Nature* 405: 562–564.
- Underwood, A. J. 1997. Experiments in ecology. Their logical design and interpretation using analysis of variance. – Cambridge Univ. Press.
- White, T. R. C. 1978. The importance of a relative shortage of food in animal ecology. – *Oecologia* 3: 71–86.
- Whittaker, R. H. 1975. *Communities and ecosystems*, 2nd ed. – Macmillan.