

*Ecology*, 86(4), 2005, pp. 1068–1071  
© 2005 by the Ecological Society of America

## EXPERIMENTAL VENUE AND ESTIMATION OF INTERACTION STRENGTH: REPLY

David K. Skelly<sup>1</sup>

Since the 1960s, experimentation has progressed from a rarity to the inferential underpinning of most ecological studies. As experimental techniques proliferate it is critical to ask whether and how the decisions we make as experimenters affect outcomes.

In their Comment, Chalcraft et al. (2005) present a multifaceted argument against an experiment (Skelly 2002) in which I evaluated the effect of venue on estimates of interaction strength. Much of their critique boils down to a disagreement regarding the hypothesis evaluated in the original paper (Skelly 2002). The remainder of the comment raises a number of questions concerning the conduct of the experiment. I appreciate the opportunity to discuss issues so critical to strategies of experimental design and hope that this exchange will contribute to improvements in ecological experimentation.

### *Experimental evaluation of venue*

In the context of ecological experiments, “confounding” is the situation in which a clear conclusion regarding a study hypothesis is prevented by the existence of multiple plausible mechanisms (Underwood 1997). Chalcraft et al. (2005) contend that my experiment is confounded because the experimental venues I used are likely to vary in a variety of attributes. If I had used my experiment to evaluate the effect of water source (well vs. pond) on the interaction between species, their Table 1 would be extremely useful in highlighting some alternative potential interpretations of my findings. Because I was testing a different hypothesis, this allegation by Chalcraft et al. is misplaced.

An experimental venue (e.g., laboratory container, mesocosm, field enclosure) represents a suite of choices made by an experimenter. These decisions can simultaneously affect many aspects of the environment experienced by study subjects. I made it clear throughout

Manuscript received 5 October 2004; accepted 2 November 2004. Corresponding Editor: A. Sih.

<sup>1</sup> School of Forestry & Environmental Studies, and Department of Ecology & Evolutionary Biology, Yale University, New Haven, Connecticut 06511 USA.  
E-mail: david.skelly@yale.edu

my paper that I was evaluating the influence of venue as it has been practiced. Prior to carrying out the experiment I searched the literature for characteristic methods used by aquatic ecologists. From a sample of representative studies in each venue I used the dominant practice (e.g., well water within cattle tank mesocosms) and where there was variation I chose a midpoint value (e.g., the amount of leaf litter placed in mesocosms and field enclosures). When I found that experimental outcome differed between venues, the result had meaning because the venues I used have meaning.

It is likely that a variety of factors contributed to the difference I observed. I was careful not to attribute my findings to one mechanism over another because I believe that the design does not warrant such speculation. However, because the venues I used are designed to represent practices characteristic of hundreds of prior experiments, my results have relevance regardless of the specific mechanism underlying the difference. Chalcraft et al. assert that I should have designed my experiment to minimize differences between mesocosms and field enclosures. It would have been illogical to design an experiment to try to eliminate the impact of venue on outcome before we knew, from an experiment, that such an effect exists.

Now that we have clear evidence that venues widely used by ecologists can have profound influences on the outcome of experiments, I believe it is important to determine why. In their Table 1, Chalcraft et al. provide a nice summary of some potential factors to investigate in future studies.

### *Mesocosms are used to make inferences regarding natural systems*

Chalcraft et al. (2005) argue that comparing mesocosm- and field-enclosure-based experiments is misguided because experiments in different venues are designed to ask different questions. Specifically, they suggest that experiments in mesocosms are not intended to provide inferences regarding particular natural populations and communities. The published literature suggests otherwise.

The target species in my experiment (*Pseudacris* (= *Hyla*) *crucifer*) has been the subject of several previous mesocosm experiments. Following one of them, Morin (1983:134) drew the following conclusions:

“*Hyla crucifer* and *Hyla gratiosa* may both rely on predators to reduce competition among tadpoles and to ensure their larval success. In the Sandhills, spring breeding *H. crucifer* and summer-breeding *H. gratiosa* are among the most abundant anurans metamorphosing from Grassy Pond, which contains high

densities of salamanders. . . . These observations are consistent with patterns in tank communities.”

Morin (1983) further concluded that *P. crucifer* is a weak competitor in nature and persists in natural ponds as a fugitive species (sensu Hutchinson 1951). This finding was reinforced in a subsequent mesocosm experiment (Morin and Johnson 1988) examining competition between *P. crucifer* and *Rana sylvatica*, the neighbor species in my experiment.

In their chapter entitled “From cattle tanks to Carolina Bays: the utility of model systems for understanding natural communities”, Resetarits and Fauth (1998: 135) note that a long list of prior mesocosm experiments on amphibians were “. . . carefully grounded in field data” and “. . . designed so that their specific results would be applicable to the natural ponds they mimicked.” The authors go on to argue that mesocosm-based experiments should be used more often to make a “. . . direct contribution to understanding specific natural systems. . . ”.

In the most recent comprehensive review of anuran community ecology (Alford 1999), experiments conducted in whole natural ponds, mesh field enclosures, and mesocosms are aggregated under a category entitled “Field Studies.” Findings from these different venues are used, collectively, to conclude that “The combination of high population densities and relatively low niche differentiation suggests that interspecific competition may be relatively common in tadpole assemblages” (Alford 1999:270).

These examples make it clear that experimenters and those assessing the literature are using mesocosm experiments and field experiments to make both general and specific inferences about the natural world. Far from being misguided, there is a great need for more experiments enabling comparisons of the effectiveness of different venues for the common purposes to which they are applied.

#### *Field enclosures*

Chalcraft et al. (2005) suggest that the presence of diffusible chemical cues or resources within natural ponds renders the results of my experiment inconclusive since cues or resources could have moved into or between enclosures assigned to different treatments. Designing experiments to preserve independence and clear inference is a general challenge for experimental ecologists and one certainly not restricted to mesh field enclosures (Scheiner 2001). In any experimental setting it is possible to think of means by which movement of material or information among experimental units may compromise their independence or influence the experimental outcome. For example, experiments designed to evaluate oviposition site choice in mesocosm

arrays (e.g., Binckley and Resetarits 2002, 2003) are subject to comparable questions because breeding individuals may be affected by the presence or behavior of conspecifics in nearby mesocosms assigned to different treatments.

In the field-enclosure portion of my experiment, I repeated the same manipulation in four different natural ponds. That the overall results were consistent across this sample of environments suggests that any influence on outcome stemming from flux of nutrients or chemical cues transcended the multiple contexts into which the enclosures were placed.

Chalcraft et al. assert that the absence of density effects in field enclosures may have resulted from the influence of predator cues on all enclosures, meaning that I was not estimating the “true treatment effect.” Actually, there were density effects in field enclosures; growth of wood frogs was reduced by higher wood frog stocking density. Whatever the influence of predator cues, wood frogs influenced their own performance in a situation where they had no influence on the target species. The presence of intraspecific density effects on wood frogs also counters the argument that flux of nutrients or other resources among enclosures compromised treatment differences among enclosures.

Finally, Chalcraft et al. raise the issue of pseudo-replication of the analysis, claiming that enclosures within a single pond were not statistically independent. In response, I reanalyzed my primary result, first selecting at random one enclosure from each of the four ponds where the field portion of the study was conducted (and representing one enclosure per density treatment). The reanalysis confirmed the original study conclusions. Namely, final mass of spring peepers was strongly influenced by venue (ANOVA:  $MS = 108767$ ,  $F_{1,12} = 70.1$ ,  $P < 0.001$ ) and density of wood frog neighbors (ANOVA:  $MS = 30035$ ,  $F_{3,12} = 19.4$ ,  $P < 0.001$ ). Spring peepers were heavier in mesocosms and were also heavier at lower wood frog densities. An interaction between venue and wood frog density indicated that the effect of density was stronger in mesocosms than in field enclosures (ANOVA:  $MS = 7565$ ,  $F_{3,12} = 4.87$ ,  $P = 0.019$ ).

#### *Experimental realism*

Chalcraft et al. (2005) voice doubts about the evaluation of experimental realism used in my study. I found that a relationship between wood frog density and body mass based on responses from experimental mesocosms provided a poor foundation for predicting body-mass responses of natural wood frog populations at field densities. Their criticisms are valid if I underestimated the density or body mass of wild-caught wood frogs in a way that led me to falsely conclude

that wood frog growth patterns in mesocosms are unrealistic.

To evaluate whether I underestimated larval density, I examined observations of wood frog egg density from 44 different ponds at the 3800-ha Yale-Myers Forest (Connecticut, USA) collected between 1999 and 2004. I determined that, excluding all zero observations and assuming 400 eggs per mass (Freidenburg 2003, Halverson 2004), the average density of wood frog eggs was 50 eggs/m<sup>2</sup> (based on a total of 185 pond-years of data). This value represents the typical density of wood frogs at the time of laying; larval densities must be equal to or lower than this estimate. Substituting the egg-count-based estimate for that based on pipe sample larval density estimates (average of 52 larvae/m<sup>2</sup> measured in four ponds), the point in my original Fig. 2 would move very slightly to the left (not right) and would remain within the confidence bounds of the field-enclosure relationship. I also determined that in less than 8% of pond-years would wood frog egg density match or exceed that predicted by the mesocosm relationship (~145 eggs/m<sup>2</sup>). The “optimistic” estimate of 550 individuals/m<sup>2</sup> generated by Chalcraft et al. (2005: Fig. 1) has never been observed, even at the egg stage, at the Yale-Myers Forest during many years of sampling a wide diversity of wetlands.

In my study (Skelly 2002), I estimated body mass of wild-caught wood frogs in the set of four ponds where the experimental enclosures were placed. If these populations have uncharacteristically slow growth and low body mass compared with a larger sample of populations, I may have falsely concluded that wood frogs in mesocosms are unrealistically large. I evaluated potential biases in my wood frog body-size estimates using data collected between 1997 and 2000 in a set of 16 ponds including the 4 experiment ponds (Halverson et al. 2003). The response for each pond was the average snout-vent length across multiple sample periods each year. I asked whether body sizes in the experiment ponds differed from those in the remaining 12 ponds not sampled (hereafter “reference ponds”). Average sizes were virtually identical between experiment ponds and reference ponds (two sample *t* test: *df* = 14, *P* = 0.96).

Additional analyses presented here provide no suggestion of bias in the estimation of attributes of wild amphibian populations and confirm the original interpretation of the data. Growth patterns in mesocosms provided a poor basis for predicting attributes of wild populations compared with predictions based on data from field enclosures. These results reinforce the original interpretation: mesocosms in my experiment were unrealistic relative to field enclosures.

### Conclusions

Experiments have tremendous power to inform that is, unfortunately, matched by their power to mislead. The choice of venue is a decision every experimenter must make and one that is likely to influence the accuracy of inferences about the natural world. Lawton (2000) has contended that experiments should never be conducted in mesocosms when it is possible to carry them out in the field. While it is clear that mesocosms offer opportunities for experimentation that are impossible or unethical in the field, mesocosms may be more broadly useful than Lawton suggests and many field experiments may be less realistic than he implies. At this point we do not know—and we will never know—if, as Chalcraft et al. (2005) argue, we experiment in single venues to the exclusion of manipulations allowing cross-venue evaluations. Chalcraft et al. suggest these multi-venue experiments will be less efficient; I suggest an alternative standard of experimental effectiveness is more likely to lead to deeper understanding of ecological systems.

There has been a great deal written on experimental venue (Diamond 1986, Hairston 1989, Peters 1991, Morin 1998). Much of it has been based on logical argument and little of it has been founded on the evaluation of data. In my experiment (Skelly 2002) and in a prior review (Skelly and Kiesecker 2001) I have adopted a different approach. There is a wealth of information hidden in prior experiments and a huge array of opportunities for new experiments that will help us make each of the venues we use better able to meet the challenges before us. If we do deconstruct our experimental venues, we are likely to be in for some surprises; not all of them will be good. That we should do it anyway is overwhelmingly supported by the results of an earlier catharsis in ecology. During recent decades ecologists have become remarkably astute in the use of statistical approaches in the design and analysis of our experiments. Few would argue that the rise of statistical sophistication has been detrimental to our science. I have no doubt that ecology will become an even better science when we put the same degree of critical attention into the design of our experimental settings.

### Acknowledgments

Susan Bolden, Kealoha Freidenburg, Nicole Freidenfelds, Anders Halverson, Manja Holland, Linda Puth, Mark Urban, and many others helped collect field data. Thanks to Keith Berven, Kealoha Freidenburg, Anders Halverson, Eric Lee, Linda Puth, Os Schmitz, Andy Sih, and Mark Urban for commenting on an earlier draft. My research has been supported by the National Science Foundation, the National Institutes of Health, and the John Simon Guggenheim Foundation.

### Literature Cited

Alford, R. A. 1999. Ecology: resource use, competition, and predation. Pages 240–278 in R. W. McDiarmid and R. Al-

- tig, editors. Tadpoles: the biology of anuran larvae. University of Chicago Press, Chicago, Illinois, USA.
- Binckley, C. A., and W. J. Resetarits. 2002. Reproductive decisions under threat of predation: squirrel treefrog (*Hyla squirella*) responses to banded sunfish (*Enneacanthus obesus*). *Oecologia* **130**:157–161.
- Binckley, C. A., and W. J. Resetarits. 2003. Functional equivalence of non-lethal effects: generalized fish avoidance determines distribution of gray treefrog, *Hyla chrysocelis*, larvae. *Oikos* **102**:623–629.
- Chalcraft, D. R., C. A. Binckley, and W. J. Resetarits. 2005. Experimental venue and interaction strength: comment. *Ecology* **86**:1061–1067.
- Diamond, J. 1986. Overview: laboratory experiments, field experiments and natural experiments. Pages 3–22 in J. Diamond and T. J. Case, editors. Community ecology. Harper and Rowe, New York, New York, USA.
- Freidenburg, L. K. 2003. Spatial ecology of the wood frog, *Rana sylvatica*. Dissertation. University of Connecticut, Storrs, Connecticut, USA.
- Hairston, N. G. 1989. Ecological experiments: purpose, design, and execution. Cambridge University Press, Cambridge, UK.
- Halverson, M. A. 2004. Relatedness and the ecology and evolution of the wood frog. Dissertation. Yale University, New Haven, Connecticut, USA.
- Halverson, M. A., D. K. Skelly, J. M. Kiesecker, and L. K. Freidenburg. 2003. Forest mediated light regime linked to amphibian distribution and performance. *Oecologia* **134**: 360–364.
- Hutchinson, G. E. 1951. Copepodology for the ornithologist. *Ecology* **32**:571–577.
- Lawton, J. H. 2000. Community ecology in a changing world. Ecology Institute, Oldendorf/Luhe, Germany.
- Morin, P. J. 1983. Predation, competition, and the composition of larval anuran guilds. *Ecological Monographs* **53**: 119–138.
- Morin, P. J. 1998. Realism, precision, and generality in experimental ecology. Pages 50–70 in W. R. Resetarits, Jr. and J. Bernardo, editors. Ecological experiments: issues and perspectives. Oxford University Press, New York, New York, USA.
- Morin, P. J., and E. A. Johnson. 1988. Experimental studies of asymmetric competition among anurans. *Oikos* **53**:98–407.
- Peters, R. 1991. A critique for ecology. Cambridge University Press, Cambridge, UK.
- Resetarits, W. R., Jr., and J. E. Fauth. 1998. From cattle tanks to Carolina bays: the utility of model systems for understanding natural communities. Pages 133–151 in W. J. Resetarits, Jr. and J. Bernardo, editors. Ecological experiments: issues and perspectives. Oxford University Press, New York, New York, USA.
- Scheiner, S. M. 2001. Theories, hypotheses, and statistics. Pages 3–13 in S. M. Scheiner and J. Gurevitch, editors. Design and analysis of ecological experiments. Oxford University Press, Oxford, UK.
- Skelly, D. K. 2002. Experimental venue and estimation of interaction strength. *Ecology* **83**:2097–2101.
- Skelly, D. K., and J. M. Kiesecker. 2001. Venue and outcome in ecological experiments: manipulations of larval anurans. *Oikos* **94**:198–208.
- Underwood, A. J. 1997. Experiments in ecology. Cambridge University Press, Cambridge, UK.