Anim. Behav., 1996, **52,** 423–425

A reply to Lombardi & Hurlbert

DAVID SLOAN WILSON* & LEE ALAN DUGATKIN†*
*Department of Biology, Binghamton University, State University of New York
†Department of Biology, University of Louisville, Kentucky

(Received 17 October 1995; initial acceptance 24 October 1995; final acceptance 20 February 1996; MS. number: AS-1158R)

Alas, our study of bluegill cognition (Dugatkin & Wilson 1992) has become the latest exemplar of pseudoreplication. In addition, the entire journal and editorial system has been implicated because our reply to a previous critique (Dugatkin & Wilson 1994; Lamprecht & Hofer 1994) survived the review process. In our reply to Lombardi & Hurlbert (1996), we will first re-analyse our data the way that they suggest and then comment briefly on some general issues concerning pseudoreplication. As with Lombardi & Hurlbert's (1996) critique, the reader will need to consult the original paper to follow our reply.

We re-analysed our data as recommended by Lombardi & Hurlbert (1996) (Table I). We cannot re-analyse the first comparison because we have misplaced our data on fish size. The original correlation was not significant, however, and therefore will remain so when the degrees of freedom are reduced. In general, the recommended analyses involve fewer degrees of freedom than our original analyses, resulting in higher *P* values. For two comparisons (3, 5), results that we reported as statistically significant become non-significant. Thus, solo fish may not take longer to capture their first prey than paired fish and there may be no relationship between feeding success and aggression, rather than a 'significantly positive, but very weak' relationship as we reported in our original paper. We did not ascribe much biological significance to either of these comparisons in any case.

The most important results of our study were that (1) bluegill prefer to associate with partners that they foraged best in association with over the previous 7-week period (comparisons 6, 7, 8, 10), and (2) bluegill preferred familiar over unfamiliar partners (comparison 11). These results remain statistically significant even with the reduced degrees of freedom recommended by Lombardi & Hurlbert (1996).

Thus, the statistical scrutiny that our study has received should not obscure the biological results: bluegill sunfish possess the cognitive abilities to employ strategic behaviour. Individuals can remember their feeding success with as many as five associates and use this information to prefer or avoid those associates in the future. In addition, bluegills prefer familiar to unfamiliar associates. These results are statistically significant according to our original analysis and the recommendations of Lamprecht & Hofer (1994) and Lombardi & Hurlbert (1996).

In addition to disagreeing with our own analysis, Lombardi & Hurlbert (1996) also apparently disagree with the analysis suggested by Lamprecht & Hofer (1994). Their 'proper procedure[s]' for our original analysis and for Lamprecht & Hofer's (1994) analysis differ from each other, making it difficult to know which to choose. We have elected to follow Lombardi & Hurlbert's (1996) proper procedure for our original analysis (comparison 7).

Now that we have followed Lombardi & Hurlbert's (1996) advice, we would like to comment on some basic statistical issues. We certainly do not regard ourselves as authorities on these subjects and think that we have much to learn from people such as Lombardi and Hurlbert. Pseudoreplication is a pervasive problem and we are not trying to defend all aspects of our original analysis. Nevertheless, we feel that the problem of pseudoreplication is not as black and white as implied by Lombardi & Hurlbert's (1996) critique of our study. We therefore pose the following basic questions.

Correspondence: D. S. Wilson, Department of Biology, Binghamton University, State University of New York, Binghamton, NY 13902-6000, U.S.A. (email: dwilson@bingsuns.cc.binghamton.edu). L. A. Dugatkin is at Department of Biology, University of Louisville, Louisville, KY 40292, U.S.A.

0003–3472/96/020423+03 $18.00/0

Table I. Re-analyses of Dugatkin & Wilson (1992), as suggested by Lombardi & Hurlbert (1996)

Analysis or comparison	Tank	Results (redone as suggested by Lombardi & Hurlbert 1995)
1. Correlation: size versus number eaten	1	Misplaced data
	2	Misplaced data
2. Foraging success: alone versus paired	1	Paired *t*-test, $t=-1.11$, $df=5$, $P>0.3$
	2	Paired *t*-test, $t=0.45$, $df=5$, $P>0.6$
3. Capture time: alone versus paired	1	Paired *t*-test, $t=-0.54$, $df=5$, $P>0.6$
	2	Paired *t*-test, $t=-0.28$, $df=5$, $P>0.7$
4. Number of items eaten: variation among partners (only for ‘red’)	1	$F_{4,24}=3.22$, $P>0.05$
	2	$F_{4,24}=1.29$, $P>0.05$
5. Regression: feeding success versus aggression	1	Adjusted $r^2=-0.25$, $df=4$, $P>0.9$
	2	Adjusted $r^2=-0.18$, $df=4$, $P>0.6$
6. Preference for same individual in both experiments	1	One-sample *t*-test, $t=3.13$, $df=5$, $P>0.05$
	2	One-sample *t*-test, $t=9.52$, $df=5$, $P>0.0005$
7. Preference for ‘companion in success’	1	One-sample *t*-test, $t=2.71$, $df=5$, $P<0.05$
	2	One-sample *t*-test, $t=4.58$, $df=5$, $P<0.01$
8. Number of items eaten: with chosen versus with not chosen companion	1	Paired *t*-test, $t=3.30$, $df=5$, $P<0.05$
	2	Paired *t*-test, $t=3.32$, $df=5$, $P<0.05$
10. Number of items eaten	1	Unpaired *t*-test, $t=-2.31$, $df=10$, $P<0.05$
	2	
11. Preference: familiar versus unfamiliar	1	One-sample *t*-test, $t=16.88$, $df=1$, $P<0.05$

Comparison 9 of Lombardi & Hurlbert (1996) is omitted because no ‘correct procedure’ was suggested. Comparison 12 of Lombardi & Hurlbert (1996) is also omitted (see text).

(1) When can repeated behaviour patterns by the same individual be treated as statistically independent events? Lombardi & Hurlbert (1996) seem to imply that the answer to this question is ‘never’. We think that this is too extreme and that the answer will depend on the nature of the behaviour pattern. For example, if an individual flips a coin many times, these may be regarded as independent events because the outcome of one flip does not influence the outcome of the next flip. As we will describe below, some of our comparisons involve behaviour patterns that are similar to coin flipping, which might justify treating each trial as an independent event.

(2) Any experiment involves an actual number of independent events that can be either underestimated or overestimated by the statistical analysis. It sounds rigorous and conservative to underestimate the number of independent events, but that merely trades the likelihood of a type I error (of inappropriately rejecting the null hypothesis) for the likelihood of a type II error (of inappropriately accepting the null hypothesis). If we want to be conservative, we should set the acceptable value of *P* to 0.01 or 0.001. We should not retain a *P* value of 0.05 and make a series of assumptions that severely underestimates the probable number of independent events in our study.

Both of these issues are nicely illustrated by Lombardi & Hurlbert’s re-analysis of our strongest result, in which bluegill showed a preference for familiar over unfamiliar associates (comparison 11). Twelve fish were separated into two tanks ($N=6$/tank) for a period of several months. Preference was measured in an apparatus that consisted of three adjacent 10-litre tanks. The focal fish was placed in the centre tank, a member of its own group was randomly placed in one side tank and a member of the other group was placed in the other side tank. The centre fish was said to prefer the fish that it spent the most time in proximity to. Each fish was tested three times on separate days and on different pairs of familiar and unfamiliar fish ($N=36$ trials). The familiar fish was chosen in 35 out of the 36 trials. Before continuing, we invite the reader to think about how he or she would analyse these data. We assumed that each trial represented an independent event, similar to an individual flipping a coin. If an individual randomly chooses the familiar associate during its first trial, there is no way that this random effect can carry over to the next trial, inflating the sample size. Therefore, if an individual chooses the familiar associate three times in a row, these should count as three independent events. More generally, if an individual bluegill cannot choose or has no preference for familiar versus unfamiliar associates (the null hypothesis), its choice can in no way be biased towards familiar associates over multiple trials.

It is important to emphasize that our test only demonstrates that some fish choose familiar over unfamiliar associates and does not address the issue of differences between individual fish. Suppose that some fish prefer familiar associates but others choose at random. If we tested a single individual many times, our result would depend on which type of individual we happened to pick. If we tested many fish once, our result would depend on our sample size and the proportion of the two types in the population. If we tested several fish several times, the result would depend on our sample size of individuals, the sample size of trials per individual, and the proportion of the two types in the population. In all of these designs, however, a statistically significant bias for the combined trials indicates that at least some individuals in the population choose familiar associates at above-chance levels. This minimal biological statement can be justified statistically, even when variation among individuals is 'nonzero'.

Suppose that we wanted to be more conservative and assume that repeated trials of the same individual are not independent events. Then we would be left with a sample size of 12 fish, all of whom chose the familiar associate more than half the time. The probability of this happening by chance (assuming a null hypothesis of no preference) is $P=0.5^{12}=0.0002$, which is still highly significant. Lombardi & Hurlbert (1996) make an even more conservative assumption, that all 18 trials within a single tank are not independent events and must be lumped into a single data point. This leaves us with a sample size of two tanks, one with a score of $18/18=1.00$ and another with a score of $17/18=0.944$, which they test against a value of 0.5 in a one sample t-test. This test yields a value of $P=0.04$ which, 'while still significant in the conventional sense, is very different from the $P<0.001$ that Dugatkin & Wilson obtained by pooling data for the replicate tanks' (Lombardi & Hurlbert 1996, page 420).

It is arguable that we overestimated the degrees of freedom by treating each trial as an independent event. It is also arguable that Lombardi & Hurlbert (1996) underestimated the degrees of freedom by lumping all 18 trials within a single tank into a single data point, making a very strong result appear marginally significant statistically. We will let the reader decide which test is most reasonable. Our main point is that statistical questions such as this are no more black and white than the biological issues that they are intended to address.

REFERENCES

Dugatkin, L. A. & Wilson, D. S. 1992. The prerequisites of strategic behaviour in the bluegill sunfish. *Anim. Behav.*, **44,** 223–230.

Dugatkin, L. A. & Wilson, D. S. 1994. Choice experiments and cognition: a reply to Lamprecht & Hofer. *Anim. Behav.*, **47,** 1459–1461.

Lamprecht, J. & Hofer, H. 1994. Cooperation among sunfish: do they have the cognitive abilities? *Anim. Behav.*, **47,** 1457–1458.

Lombardi, C. M. & Hurlbert, S. H. 1996. Sunfish cognition and pseudoreplication. *Anim. Behav.*, **52,** 419–422.