

Kenneth F. Raffa · Nathan P. Havill  
Erik V. Nordheim

## How many choices can your test animal compare effectively? Evaluating a critical assumption of behavioral preference tests

Received: 12 April 2002 / Accepted: 2 August 2002 / Published online: 5 September 2002  
© Springer-Verlag 2002

**Abstract** Behavioral choice tests comprise one of the most commonly used experimental designs in ecology. However a critical assumption of these assays, that the outcome is independent of the number of choices, has not been tested explicitly. We developed a methodology for testing this assumption, and discuss how it can be incorporated into experimental design. The model with which we performed this test consisted of an insect herbivore, the gypsy moth *Lymantria dispar* L., feeding on a clonal host plant, *Populus*. We established a dose-response feeding gradient by amending leaves of a single age class with defined concentrations of a diterpene, isopimaric acid, that exhibits feeding deterrent properties. We selected various concentrations that elicited different levels of feeding for subsequent tests in which we modified the number of choices. A sample size of 30 assay units per test generated statistically significant separations in two-way choice tests, yielded statistically significant but somewhat inconsistent results when four concentrations were offered, and failed to provide complete separation when five concentrations were offered. Other factors associated with the number of choices that affected results included specific combinations of doses, physical arrangement of choices, and total consumption per assay unit. We used our results to develop procedures for estimating the sample sizes needed to compare a specified number of choices. We based these methods on power considerations, the requirements for data transformation and inclusion of covariates. We develop a general approach for estimating the number of replicates needed to support a particular number of

choices for a test organism, and discuss factors to be considered when relating this approach to various types of behavioral choice assays.

**Keywords** Behavioral preference choice · Experimental design · Power analysis · Sample size determination · Semiochemicals

### Introduction

Evaluation of an organism's relative preferences among an array of resources is one of the most commonly used experimental designs in animal ecology and behavior. The standard approach is to provide a simultaneous choice of selections, and then measure the organism's relative responses to each. These assays are commonly referred to as "preference" tests, "choice" tests, or more colloquially as "cafeteria" tests. Such assays have been employed with all major groups of animal taxa, across a broad array of life history strategies such as herbivory, parasitism, pollination, and predation, in both laboratory and field studies and in regard to a variety of behaviors such as mating, feeding, oviposition, defense, and locomotion. Choice tests are routinely employed to quantify the effects of a broad range of environmental, hereditary, and anthropogenic influences on animal behavior (reviewed in Lockwood 1998). Motivations for conducting behavioral preference tests range from basic research questions, such as the role of induced plant responses on herbivore or parasitoid foraging, to management oriented ecological applications, such as screening crop cultivars for resistant varieties or developing semiochemicals to improve the efficacy of biological control agents.

Despite the ubiquity of multiple-choice preference tests, few studies have explicitly evaluated whether the conditions required for their validity are met. Several good reviews describe specific factors which affect the quality control of behavioral choice assays, such as plant or animal handling and conditioning, methods of measurement, correspondence between laboratory and field results, and variables that can complicate results and inter-

K.F. Raffa (✉) · N.P. Havill  
Department of Entomology, University of Wisconsin,  
Madison, WI 53706, USA  
e-mail: raffa@entomology.wisc.edu  
Tel.: +1-608-2621125, Fax: +1-608-2623322

E.V. Nordheim  
Department of Statistics, University of Wisconsin,  
Madison, WI 53706, USA

E.V. Nordheim  
Department of Forest Ecology and Management,  
University of Wisconsin, Madison, WI 53706, USA

pretations (e.g., Singer 1986; Lewis and van Emden 1986; Jones and Coleman 1988; Harris and Miller 1991; Robison and Raffa 1994, 1997; Bomford and Isman 1996; Wagner 1998; Schneidel and Bruelheide 1999). There is also a large empirical and theoretical literature on how the context in which a choice is made can influence decisions ranging from floral visitation by honey bees to product purchase by humans (Le Magnen 1999; Nowlis and Simonson 2000; Muthukrishnan and Kardes 2001; Gonzales-Vallejo 2002; Shafir et al. 2002). However, many of the underlying assumptions required for statistical hypothesis testing have not been analyzed rigorously (Chesson 1983). Some recent analyses by Marquis and Braker (1987), Raffa and Frazier (1988), Peterson and Renaud (1989), Roa (1992), Manly (1993, 1995), Horton (1995), and Lockwood (1998) have demonstrated that selection of the statistical model, units of measurement, and stopping criteria can greatly affect our interpretation of the data.

A fundamental assumption of all preference tests is that behavioral patterns are not influenced by the number of choices. That is, the discrimination among an array of treatments, and between any combination therein, is implicitly assumed to be unaffected by the number of treatments. However, to our knowledge, this basic assumption of multiple preference assays has not been tested.

Selection of the appropriate number of choices confronts researchers with some important trade-offs. It is often both logistically expedient and empirically preferable to provide the test animal with as many simultaneous choices as possible. For example, if only a subset of the options can be tested at once, the resulting incompleteness in the design can be a serious obstacle to reconstructing a comprehensive rank order. This can become even more challenging when there are constraints on the availability of test material or when special procedures are required to avoid pseudoreplication. It is often of considerable practical importance to test many choices simultaneously. For example, plant breeders need to screen many cultivars under the same conditions, and ecologists often need to test a wide array of treatment combinations to understand complex interactions, such as synergy and antagonism.

A key step towards developing practical solutions to these conflicting constraints is to determine if the sensitivity of the comparison of treatments is affected by the number of treatments available. If this effect can be quantified, then it is possible to determine the statistical power associated with each number of choices, and then calculate the corresponding number of required replicates. The objective of this research was to develop an approach for explicitly testing whether the results of a multiple choice experiment are independent of the number of choices offered, and how sample size relates to statistical power.

## Materials and methods

### Overall approach

We conducted a series of feeding preference trials using an insect herbivore. We chose a test model that reduced other sources of

variation to the fullest extent possible: a single age class of a laboratory-cultured folivore, feeding on leaves of a single age class, on a clonal plant, treated with defined concentrations of a feeding deterrent. In preliminary experiments, we applied varying concentrations to leaves, in order to establish a range of statistically different consumption rates. Once this rank order of feeding preference was established, we conducted subsequent trials in which we provided various choices from among these concentrations. We varied the number of simultaneous choices, determined the effect of the number of choices on our capability of resolving differences between pairs of treatments, and performed power calculations to determine the number of replicates the insect needs to generate the correct (based on the dose-response experiment) sequence under each condition. Secondly, we also varied the arrangement of treatments, to determine if position affected either overall results or the resulting power computations.

### Selection of study system

We set as requirements for our study system that insect feeding on control foliage is relatively uniform, the feeding deterrent acts in a dose-dependent fashion, the test compound is nontoxic within the relevant range, and the test compound is relatively nonvolatile and therefore does not obscure treatment differences by airborne effects.

The gypsy moth, *Lymantria dispar* L. (Lepidoptera: Lymantriidae) is a polyphagous folivore. Laboratory colonies maintained by the United States Department of Agriculture (USDA) provide high quality control and uniformity. Gypsy moth larvae are well suited for feeding preference assays and have already been used to detect variation due to host species, host genetics, environmental parameters, and wound induction (Mauffette et al. 1983; Martinat and Barbosa 1987; Havill and Raffa 2000). We used clonal material for our plant substrate. Hybrid poplar NC5271 (*P. nigra* 'char-kowiensis' × *P. nigra* 'Caudina') is highly preferred by gypsy moth larvae, and there is very low among-plant variation within uniform leaf age categories (Robison and Raffa 1997; Havill and Raffa 2000). The diterpene isopimaric acid occurs in the foliage of 2nd-year shoots of *Larix* and other trees in the family Pinaceae. *Larix* foliage from current-year shoots is a preferred feeding substrate of the gypsy moth, but foliage from 2nd-year shoots is less preferred (Kruse and Raffa 1997). Isopimaric acid reduces gypsy moth larval feeding in a dose-dependent fashion, does not affect growth or survival at moderate concentrations, does not occur in the foliage of our test tree, and is nonvolatile under laboratory conditions (melting point =160°C) (Powell and Raffa 1999).

### Plant and insect culture

Trees were established from clonal material growing in a common garden at the University of Wisconsin West Madison Agricultural Research Station. Softwood cuttings (20 cm long) were dipped in Hormex No. 8 rooting powder (0.8% Indole-3-butyric acid; Brooker Chemical, North Hollywood, Calif.) and planted in saturated Fafard No. 2 potting soil (Fafard, Agawam, Mass.) in 15.12-l plastic pots. Trees from these cuttings were grown in a University of Wisconsin-Madison greenhouse at 24°C and with a 16:8 h (L:D) light regime. The trees were fertilized with 15 g/plant Osmocote 19-6-12 slow release fertilizer (Sierra Chemical, Milpitas, Calif.) plus micronutrients, and flood irrigated when needed. Because larval performance can vary significantly with leaf age on *Populus* (Bingaman and Hart 1993; Robison and Raffa 1997), all experiments were standardized for leaf position. The most apical fully unfolded leaf was designated as leaf 1 and numbering continued sequentially down the stem. Trees for all experiments were approximately 60 days old.

We used *L. dispar* larvae from egg masses of strain NJSS obtained from the USDA APHIS Methods Development Center Otis AFB, Mass. Before hatching, pre-chilled egg masses were surface sterilized in a sodium hypochlorite solution (2,060 ml dH<sub>2</sub>O, 21 ml polyoxyethylene sorbitan monooleate, and 40 ml bleach) for 5 min,

rinsed 3 times with dH<sub>2</sub>O, and allowed to dry. Egg masses were then placed in circular 14.0-cm × 3.9-cm clear plastic containers (TriState Plastics, Dixon, Ky.) and reared in an environmental growth chamber at 16:8 h (L:D) photoperiod, 240°C and 50–70% RH. Upon hatching, the larvae were provided with one or two 2-cm<sup>3</sup> cubes of artificial diet (ICN Biomedicals, Aurora, Ohio), which were replaced every 2–3 days. Preliminary choice tests showed that larvae fed more consistently if they were not switched directly from artificial diet to assay leaf material. Therefore, the artificial diet was replaced with poplar leaves from positions 4–7, 4 days prior to feeding assays, and larvae were allowed to feed freely.

### Experimental conditions

We used a standard leaf disk procedure for all assays. The arena consisted of a 9-cm-diameter plastic petri dish with a thin coat of paraffin wax on the bottom. Each dish contained a filter paper moistened with distilled water to prevent leaf desiccation and shrinkage. Leaf disks (16 mm diameter) were cut intervenally with a cork borer. One leaf from position no. 4–6 was used from each tree. The average weight of a leaf disk was 52.4 mg. Disks were arranged evenly around the periphery of each dish, and were anchored with minuten pins. We were able to cut 12 disks out of each leaf. All 12 disks from a leaf were used across one full replication of an experiment (e.g. a two-way choice test with six different combinations of doses). However, the five-way test with six combinations (see below) required 30 disks for a complete replication, so we cut 10 disks out of each leaf and randomly assigned groups of 5 disks to each dish.

Crystallized isopimaric acid was dissolved in HPLC grade methanol at different concentrations to provide a range of doses to apply to leaf disks. The surface of each leaf disk was evenly spread with 80 µl of solution and allowed to air dry until the methanol had evaporated and a thin layer of isopimaric acid crystal was left on the leaf disk. Control disks received 80 µl of methanol alone.

Second-instar gypsy moth larvae were starved for 24 h, and one larva was placed in the center of each dish. Dishes were covered and sealed with Parafilm (American National Can, Greenwich, Conn.), and the larvae were allowed to feed for 48 h at 16:8 h (L:D) and 240°C. After 48 h leaf disks were electronically scanned and consumption was measured using the software MacFolia (Regent Instruments 1996).

### Preliminary assays and determination of dose levels

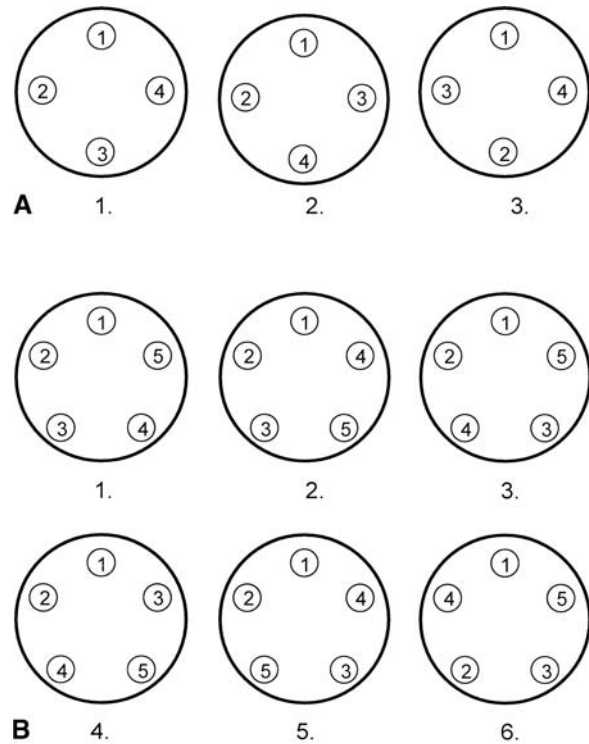
We conducted a preliminary series of two-way choice assays to determine appropriate doses for subsequent tests. Disks were placed approximately 1 cm from the edge of the dish, 180° from each other. We paired each of five doses (0.25%, 1.0%, 2.0%, 4%) versus a control, and a control versus a control. The latter pair served as a check against potentially toxic rather than purely behavioral interactions.

We also conducted a series of no-choice assays, which consisted of single leaf disks placed in the center of each dish. Thirty larvae were tested at each dose. Doses were selected from the above experiment based on: (1) significantly different feeding on treated versus control leaves, and (2) no reduction in feeding on the control tissue, relative to the units that contained only control tissue.

### Feeding preference assays with variable numbers of choices

Based on the results of the preliminary experiments, the treatments selected for subsequent choice assays were 0.0, 0.25, 1.0, and 2.0% (see Results). We performed three different feeding assays. These were:

1. A two-way choice test which included assay units containing all six pairs of the above four doses: Thirty dishes of each combination were prepared.
2. Four-way choice tests, using the above four treatments, in which the disks were arranged 90° from each other and ap-



**Fig. 1a, b** Schematic diagram showing the spatial arrangement of choice assays. Dose numbers on leaf disks refer to the following: **a** four-way choice test: 1=0.0, 2=0.25; 3=1.0; 4=2.0; **b** five-way choice test: 1=0.0, 2=0.125, 3=0.25, 4=1.0, 5=2.0

proximately 1 cm from the dish edge: We used 30 dishes for each of the three distinct spatial configurations shown in Fig. 1a. Configurations were considered distinct if they do not have the same order of treatments; hence patterns that differ only by direction (clockwise versus counterclockwise) were not considered distinct.

3. Five-way choice tests using the above four treatments plus 0.125%: This latter dose was selected based on results from the preliminary assays, which suggested that this treatment might lead to significant separation from the neighboring concentrations. This presumption was substantiated by additional two-way choice tests ( $n=30$ ): 0.0 versus 0.125;  $P=0.0315$ ; 0.125 versus 0.25;  $P=0.0141$ . In five-way tests, there are 12 distinct configurations, using the same criteria as described above. We arbitrarily selected 6 of these for our experiment (Fig. 1b). Disks were spaced evenly in each dish and 30 dishes were used per configuration.

### Statistical analyses

The relative merits of both raw and relative (i.e., proportion of total feeding per leaf disk) consumption data have been advocated for behavioral choice tests, and both are recognized as having positive and negative attributes (Lockwood 1998). We conducted both types of analyses, which yielded the same basic conclusions. We report analyses based on raw consumption data because total consumption did not vary greatly among dishes within each set of 30 trials, and because we believe it more readily conveys the patterns and quantities of consumption under different doses and assay conditions. Feeding consumption data for each dose in each experiment were examined for univariate normality (PROC UNIVARIATE; SAS Institute 1990). Roa (1992) suggests this method as an important partial test of the assumption of multivariate normality of the sampling population for multivariate analyses. For all tests,



consumption data were log-transformed to achieve normality and homogeneity of variance.

The preliminary two-way choice tests were analyzed qualitatively to determine the appropriate range of doses for the subsequent tests. In the preliminary no-choice test, the effects of isopimaric acid concentrations on larval feeding were analyzed by one-way analysis of variance (PROC GLM; SAS Institute 1990).

Initial overall analyses of the multiway choice tests with two, four, and five choices (assays 1, 2, and 3) were performed using multivariate analytical methods. For the two-way tests with all possible combinations of 4 doses, the multivariate approach reduces to paired *t*-tests for each combination. Differences in feeding among doses for the four-way and five-way choice tests were analyzed with multivariate analyses of variance (MANOVA) using PROC GLM, with the log feeding values for the doses within a dish treated as vectors (Manly 1995). Mean comparisons between pairs of doses were performed using multivariate contrasts. For example, for the four-way choice test, multivariate contrasts were performed with the (joint) null hypothesis:  $m_4 - m_3 = 0$ ;  $m_3 - m_2 = 0$ ;  $m_2 - m_1 = 0$ , where  $m_m$  denotes the mean for dose  $k$ . This was done separately for each spatial combination of doses. In addition, separate analyses were performed with the null hypothesis:  $m_4 - m_1 = 0$ ;  $m_4 - m_2 = 0$ ;  $m_4 - m_3 = 0$ , thus allowing determination of mean comparisons for all pairs of the four doses. Similar tests were performed for the five-way choice test. In order to compare the total feeding per dish for the different configurations in the four-way and five-way tests, we used one-way analysis of variance.

To evaluate how sensitivity in comparing distinct doses relates to the number of choices, we considered each pair of treatments (doses) separately within each multi-way test. Differences between consumption values for each pair of treatments were computed for each dish, and the means and standard deviations of these differences were calculated. A paired *t*-test was conducted to test the hypothesis of no difference between treatment means. This test was performed for each pair of doses, for each spatial orientation of disks, for each of the two-, four-, and five-choice experiments. The *t*-score,  $T = \bar{y} / (s / \sqrt{n})$ , was calculated for each test, where  $\bar{y}$  and  $s / \sqrt{n}$  are the mean and standard error, respectively, of the differences (with  $n$ , the number of dishes, equal to 30). To obtain an overall measure of performance for the different numbers of choices, we averaged the *T*-scores over the different configurations, 1, 3, and 6 combinations respectively for two-, four-, and five-choice tests.

To estimate relative sample size requirements for the different numbers of choices, we used the expression

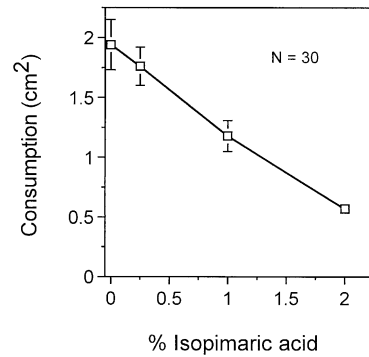
$$n = 10 \left( \frac{\sigma}{\mu_d} \right)^2 \quad (1)$$

(Snedecor and Cochran 1993, p. 103) where  $\sigma$  is the population standard deviation of the differences (between log-transformed consumption values) corresponding to a given pair of doses and  $\mu_d$  is the population magnitude of the true difference of interest. Ten is a multiplier calculated from *Z*-scores that depend on the probabilities of type I and type II error of interest. Ten corresponds very closely to a probability of type I error ( $\alpha$ ) of 0.05 and a power of 0.90. The values used for  $\sigma / \mu_d$  in equation 1 were those that correspond to the average *T*-scores in Table 4; thus we have applied the equation in a somewhat novel way by using the empirical ratio of the estimate of  $\sigma$  and  $\mu_d$ . This allows us to capture the "signal-to-noise" ratio suggested by our data. (We ignored the fact that the values used for  $\sigma$  in equation 1 were based on sample data since this would have minimal impact on the results).

## Results

### Preliminary assays and determination of dose levels

The results of the preliminary two-way choice tests of each dose versus a control suggested that the most ap-



**Fig. 2** Effect of isopimaric acid on feeding by *Lymantria dispar* larvae on *Populus* foliage under no-choice conditions.  $n=30$  for each dose.  $df=3, 116$ ;  $F=18.35$ ;  $P<0.0001$

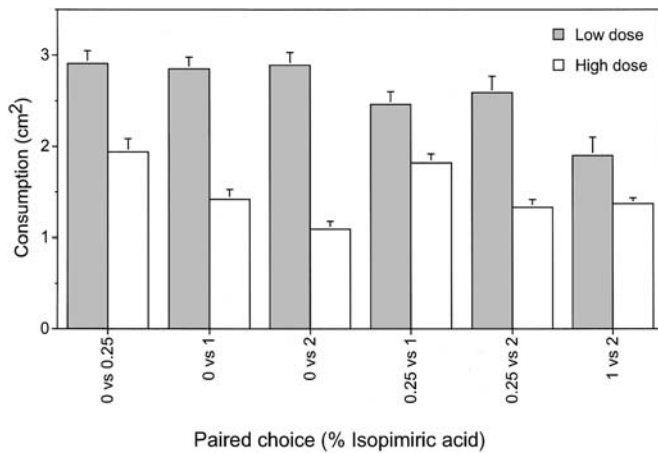
propriate doses of isopimaric acid for the comparative assays ranged from 0.25 to 2.0%. Each of these doses reduced feeding by *L. dispar* on treated disks relative to the controls without reducing feeding on control disks. For example, disks treated with 2.0% isopimaric acid were consumed only 29% as much as controls. At the highest concentration, 4.0%, there was reduced feeding on treated leaves relative to untreated leaves in the same assay units, but these controls had only 79% the feeding on untreated disks in dishes that contained no treated tissue. This suggested possible sublethal toxic effects at this dose, so the 4.0% treatment was omitted from subsequent tests.

Under no-choice conditions, isopimaric acid reduced feeding by *L. dispar* larvae in a dose-dependent fashion (Fig. 2). Consumption was significantly lower at 1.0% and 2.0% than in the controls ( $P<0.05$ ). However, the conditions imposed by a no-choice test were less sensitive than two-way tests, in that 0.25% in the no-choice test did not result in a significant inhibition to feeding, but it did relative to the control in the above paired choice test.

### Feeding preference assays with variable numbers of choices

In the two-way choice tests using the treatments selected from the preliminary dose-finding assays, all three concentrations of isopimaric acid reduced feeding by *L. dispar* larvae relative to controls (Fig. 3). Moreover, each pairwise comparison indicated a significant difference in feeding, with lower consumption always occurring at the higher dose.

When four treatments were provided simultaneously, all spatial configurations generated significant overall treatment effects. When the treatments were arranged in configurations 1 and 3 (Fig. 1a), complete separation of doses was observed at  $P<0.05$  (Table 1). However, larvae did not discriminate between the two highest concentrations when treatments were arranged in configuration 2. An interesting secondary result was that total consumption per dish varied



**Fig. 3** Effect of isopimaric acid on relative feeding by *L. dispar* larvae on *Populus* foliage under two-way choice conditions. Assays include two separate doses as controls, and paired assays between these three doses.  $n=30$  for each assay. 0.0 versus 0.25:  $df=1,29$ ;  $F=50.83$ ,  $P=0.0001$ ; 0.0 versus 1.0:  $df=1,29$ ;  $F=67.17$ ,  $P=0.0001$ ; 0.0 versus 2.0:  $df=1,29$ ;  $F=136.40$ ,  $P=0.0001$ ; 0.25 versus 1.0:  $df=1,29$ ;  $F=21.40$ ,  $P=0.0001$ ; 0.25 versus 2.0:  $df=1,29$ ;  $F=34.87$ ,  $P=0.0001$ ; 1.0 versus 2.0:  $df=1,29$ ;  $F=16.65$ ,  $P=0.0003$

significantly ( $P=0.038$ ) among the three configurations. Specifically, total feeding for configuration 3 was higher than that for 1 or 2. However, the overall patterns among the four doses were similar in all cases. That is, feeding was reduced by increasing doses of isopimaric acid.

**Table 1** Consumption of leaf disks by *Lymantria dispar* larvae in a four-way choice test. Configuration numbers refer to every possible combination of four doses of isopimaric acid (see Fig. 1a). Means  $\pm$  SE followed by capital letters down a column, and lower

Configuration	$n$	Dose				$df$ ; $F$ ; $P$	Total feeding/dish*
		0.00	0.25	1.0	2.0		
1	30	1.91 $\pm$ 0.18a	1.18 $\pm$ 0.12b	0.76 $\pm$ 0.07c	0.63 $\pm$ 0.03d	3,27; 21.30; 0.0001	4.48 $\pm$ 0.27 A
2	30	1.92 $\pm$ 0.16a	1.24 $\pm$ 0.13b	0.74 $\pm$ 0.08c	0.68 $\pm$ 0.05c	3,27; 31.27; 0.0001	4.58 $\pm$ 0.28 A
3	30	2.07 $\pm$ 0.19a	1.41 $\pm$ 0.17b	1.09 $\pm$ 0.10c	0.82 $\pm$ 0.05d	3,27; 26.30; 0.0001	5.39 $\pm$ 0.30 B
All combinations	90	1.97 $\pm$ 0.10a	1.28 $\pm$ 0.08b	0.86 $\pm$ 0.05c	0.71 $\pm$ 0.03d	3,87; 79.90; 0.0001	

\* $df=2,58$ ;  $F=3.47$ ;  $P=0.038$

**Table 2** Consumption of leaf disks by *L. dispar* in a five-way choice test. Configuration numbers refer to different spatial combinations of five doses of isopimaric acid (see Fig. 1b). Means  $\pm$  SE followed by capital letters down a column, and lower case

Configurations	$n$	Dose (%)					$df$ ; $F$ ; $P$	Total feeding/dish*
		0.0	0.125	0.25	1.0	2.0		
1	30	1.35 $\pm$ 0.17a	0.97 $\pm$ 0.14b	0.95 $\pm$ 0.13b	0.55 $\pm$ 0.03c	0.53 $\pm$ 0.02c	4,26; 8.50; 0.0002	4.37 $\pm$ 0.39 A
2	30	1.04 $\pm$ 0.16a	0.82 $\pm$ 0.10a	0.81 $\pm$ 0.08ab	0.64 $\pm$ 0.05bc	0.58 $\pm$ 0.04c	4,26; 4.76; 0.0052	3.89 $\pm$ 0.29 A
3	30	1.48 $\pm$ 0.19a	1.18 $\pm$ 0.13b	0.89 $\pm$ 0.12c	0.68 $\pm$ 0.07cd	0.63 $\pm$ 0.04d	4,26; 8.40; 0.0002	4.86 $\pm$ 0.43 A
4	30	1.13 $\pm$ 0.17a	0.90 $\pm$ 0.13ab	0.81 $\pm$ 0.09b	0.64 $\pm$ 0.07bc	0.61 $\pm$ 0.04c	4,26; 3.24; 0.0277	4.09 $\pm$ 0.37 A
5	30	1.35 $\pm$ 0.19a	1.08 $\pm$ 0.14b	0.82 $\pm$ 0.14bc	0.68 $\pm$ 0.09c	0.59 $\pm$ 0.06c	4,26; 6.96; 0.0006	4.52 $\pm$ 0.54 A
6	30	1.01 $\pm$ 0.14a	0.89 $\pm$ 0.10a	0.79 $\pm$ 0.09a	0.60 $\pm$ 0.04b	0.58 $\pm$ 0.03b	4,26; 5.79; 0.0018	3.87 $\pm$ 0.31 A
All	180	1.63 $\pm$ 0.07a	1.37 $\pm$ 0.05b	1.25 $\pm$ 0.04c	1.03 $\pm$ 0.03d	0.99 $\pm$ 0.02e	4,176; 33.56; 0.0001	

\* $df=5,85$ ;  $F=1.08$ ;  $P=0.3779$

When confronted with five simultaneous choices, there were significant overall treatment effects for all configurations (Table 2). However, gypsy moth larvae were never able to discriminate among all of the choices. In most configurations, three means overlapped. Part of this was due to the addition of the 0.125 level, which was often difficult to discriminate from the 0 and/or 0.25 levels in other than two-choice tests. However, even when only levels from the four-way choice test are considered, the five-way results demonstrated a reduced ability to discriminate between pairs of choices. For example, in the five-way tests, there was no statistically significant discrimination between 0.25 and 1.0 for four of the six configurations. As with the four-way choice test, the proper rank order was always maintained. There were no significant differences in total consumption per dish among the six configurations, due to greater variabilities within configurations in the five-choice than four-choice trials.

Table 3 shows the paired analysis for each combination of doses in four-way and five-way choice tests. The number of choices affected the sensitivity of comparisons between pairs of treatments. For example, only half of the comparisons between 0% and 0.25% were significant at  $P<0.05$  when there were five choices, whereas all were significant when there were four choices. Likewise, these ratios were 0.33 and 0.66 at 1% versus 0.25%, 0.66 and 1.0 at 2.0% versus 0.25%, and 0.0 and 0.66 at 2.0% versus 1.0%. Some of this variation appears to be associated with the configuration of choices (Table 3). That is,

case letters across rows are significantly different ( $P<0.05$ ). Consumption data were log-transformed prior to statistical analysis. Original untransformed values are presented

letters across rows are significantly different ( $P<0.05$ ). Consumption data were log-transformed prior to analysis. Original untransformed values are presented

**Table 3** Paired *T*-values for choices between different feeding substrates within arenas containing variable numbers and configurations of choices (*n*=30). Within each paired choice of doses, the

total number of choices within the arena (5 or 4) are listed as sub-headings. Configurations correspond to those in Fig. 1

	% Isopimaric acid									
	0			0.125		0.25			1.0	
	5	4	2	5	5	4	2	5	4	2
0.125	3.16									
	0.93									
	1.82									
	1.43									
	1.45									
	0.32									
0.25	2.09	4.02	2.73	0.16						
	0.83	3.95		0.11						
	3.05	2.72		2.71						
	1.94			0.43						
	3.40			2.55						
	1.24			1.60						
1.0	5.48	7.43	6.30	3.84	3.42	4.07	5.44			
	2.43	9.28		1.71	1.94	4.75				
	5.19	4.68		4.71	1.44	1.65				
	3.45			1.75	1.77					
	4.43			3.74	1.13					
	2.73			3.65	2.27					
2.0	5.23	10.19	5.82	3.70	3.85	6.21	5.66	0.46	2.30	2.91
	3.07	12.77		2.91	3.06	5.29		1.91	0.50	
	5.46	7.97		5.15	2.02	3.79		0.42	2.67	
	3.44			2.04	2.09			0.16		
	4.71			3.06	1.84			1.43		
	3.15			4.00	2.24			0.68		

*P*<0.1: 1.699; *P*<0.05: 2.045; *P*<0.01: 2.756

within a given number of choices, paired comparisons gave different results depending on how the choices were arranged. In the five-way test for example, configuration 1 yielded significant differences among all pairs of choices across all treatment combinations except 1% versus 2%, whereas configuration 4 only yielded significant differences between 0% versus 0.25%, 1.0%, or 2.0%, and 2.0% versus 0.125% or 0.25%.

The averaged results are shown in Table 4 (excluding 0.125% because it was not common to all tests), which again lists *T*-values according to the configuration of choices. In this overall summary, there was more sensitivity (higher *T*-scores) for the two-way and four-way than five-way choice tests for each pairwise comparison. There was no clear pattern comparing the two-way and four-way results. However, the four-choice assays appeared more sensitive (i.e., had higher *t*-values) when differences involving the 0% dose were considered, whereas the two-choice results were more sensitive for the other comparisons. In the five-way choice tests, treatment effects changed from *P*<0.01 to nonsignificant. This suggests that the sensitivity of choice tests depends not only on the number and configuration of choices, but also on the specific comparisons.

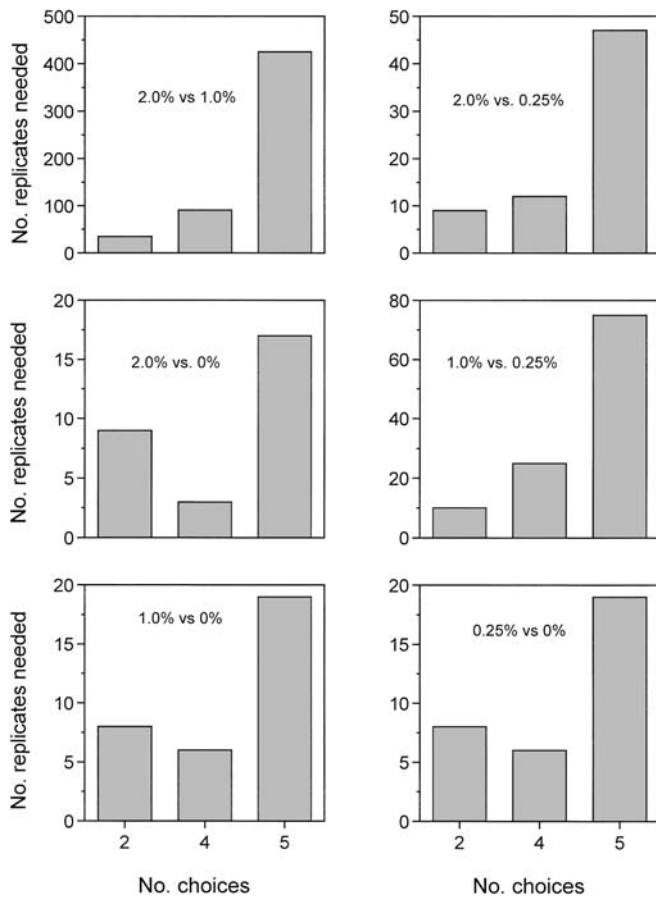
The estimated numbers of samples required for various pairs of treatment comparisons amongst different numbers of choices are shown in Fig. 4. These values are based on power calculations computed from our data using equa-

**Table 4** Paired *T*-values for choices between different feeding substrates by *L. dispar*, within arenas containing variable numbers of choices. Data are the averages of various configurations of choice

	% Isopimaric acid			No. of choices
	0	.25	1.0	
0.25	2.73 **			2
	3.56 ***			4
	2.09 *			5
1.0	6.30 ***	5.44 ***		2
	7.13 ***	3.49 ***		4
	3.95 ***	2.00 *		5
2.0	5.82 ***	5.66 ***	2.91 ***	2
	10.31 ***	5.10 ***	1.82 *	4
	4.18 ***	2.52 **	0.84	5

\**P*<0.1: 1.699; \*\**P*<0.05: 2.045; \*\*\**P*<0.01: 2.756

tion 1. We consider the specific values to be less important than relative values for purposes of this evaluation. The general trend is for the required sample size to rise with the number of choices, particularly when both choices include some isopimaric acid. For example, the sample size needed to separate 2.0% from 1.0% isopimaric acid at *P*<0.05 ranges from *n*=35 with 2 choices to *n*=425 with five choices. In our system, there is a substantial loss in sensitivity when using a five-choice test compared with



**Fig. 4** Sample sizes needed to detect significant differences among pairs of various treatment combinations when two, four, or five, choices ranging from 0% to 2.0% isopimaric acid are presented simultaneously

two-choice or four-choice tests, with more subtle differences existing between the two-choice and four-choice tests. It appears that the two-choice test is more sensitive for comparisons involving alternatives between two doses of isopimaric acid, but four-choice tests were more sensitive for comparisons involving the control.

## Discussion

These results demonstrate that the ability of behavioral choice tests to discern biological differences is influenced by the number of concurrent choices. This effect was non-linear, which seems likely for other systems. However, the specific relationships between sensitivity and the number of choices are likely to vary among test organisms, substrates, physiological mechanisms, types of stimuli, and other factors. Although the optimal number of choices has received little attention in ecological experimentation, it is currently a topic of debate in the education literature, as it pertains to standardized tests taken by humans (Delgado and Prieto 1998 ; Abad et al. 2001).

The attention that researchers need to place on the number of choices depends on whether they need a true

rank order, or merely wish to identify extremes. For example, in the five-way assay  $n=30$  was always sufficient to discriminate the least from most preferred substrate and provide general categories of preference. Hence, this might be adequate for the initial rounds of a cultivar screening program. Likewise, if one only wanted to identify plants possessing feeding deterrent properties, a five-way choice might be adequate. However to obtain a true rank order of behavioral preference for purposes of regression against some other variable of ecological interest (e.g., plant growth, plant tolerance, adult oviposition, larval growth, chemical structure), a substantially larger  $n$  would be needed. A quantitative justification of both the numbers and specific combinations of choices is particularly important when the absence of statistical difference is used to construct a subsequent argument (see Parkhurst 2001). For example, the question of whether early instar *L. dispar* larvae can differentiate between substrates containing 1.0% versus 2.0% isopimaric acid will be answered differently depending on whether two or five choices are provided and will not be answered with certainty when there are four concurrent choices, with 30 replicates.

Given these limitations on the independence of results on the number of choices, some possible adjustments include reducing the number of treatments tested concurrently or increasing the number of replicates. Increasing the number of replicates poses some additional problems in field studies, however, because these experiments also assume that the organism being attracted or repelled is uniformly distributed across the sampling universe. This assumption can become less tenable as the sampling universe is expanded to accommodate additional replicates.

In addition to the number of choices, the spatial configuration of various treatments can exert substantial effects. This highlights the lack of independence among choices and reinforces the need to randomize treatment positions. Further observational studies are needed to determine underlying mechanisms for this effect, to identify primary sources of variation affecting total consumption (e.g., premoult periods within assay time spans, genetic variation), and to separate meaningful effects of orientation from random noise. Such configuration effects should be considered when making comparisons across studies, particularly laboratory and field experiments. In laboratory assays, a variable number of choices is typically accommodated within the same experimental unit, which alters the space between choices. The differences we observed in behavioral outcomes associated with different configurations suggest that such changes in between-treatment distance could have important effects. In contrast, field assays typically hold the space between treatments constant, and alter the sampling universe. Although this avoids configuration effects, it rests on the assumption of equivalent distributions within blocks, an assumption whose validity may likewise vary with the number of treatments.

Although multiple choice tests provide a powerful tool for evaluating animal behavior, and indeed are more



sensitive than no-choice tests, our results reinforce the need for biological understanding of the test system. System-specific features of the organisms, behaviors, stimuli, and their interactions need to be understood to assure the validity of choice assays under varying conditions. Our methodological approach allows an assessment of the impact of the number of choices that can be used in a wide range of situations.

If it is impractical to perform a detailed prior study of the effect of the number of choices, we provide the following as a possible general approach: (1) conduct a standard cafeteria experiment using a biologically reasonable number of alternatives; (2) if any pair of alternatives is statistically indistinguishable, initiate a pairwise comparison between them. If this latter trial is also indistinguishable, then there is no difference across methodologies. If there are statistical differences, then (3) compute *t*-scores as described under "Statistical Analyses"; (4) determine estimates of requisite sample sizes using Eq. 1; and (6) repeat the initial assays at the emergent recommended sample size.

**Acknowledgements** This work was supported by McIntire-Stennis and the University of Wisconsin-Madison College of Agricultural and Life Sciences. Nichole Broderick conducted valuable preliminary assays. Daniel Robison, Sylvio Codella, Kimberly Wallin, Jamie Powell, and Brian Aukema, Department of Entomology, and the Behavioral Interdisciplinary Discussion Group, UW-Madison, provided helpful suggestions on our approach to this problem. Thomas Tabone, CALS Computing Center, UW-Madison, assisted with SAS programs. The USDA APHIS Methods Development Laboratory generously provided gypsy moth egg masses. Critical reviews by Marc Mangel and Peter Nonacs provided valuable suggestions that improved this manuscript.

## References

- Abad FJ, Olea J, Ponsoda V (2001) Analysis of the optimum number of alternatives from the Item Response Theory. *Psicothema* 13:152–158
- Bingaman BR, Hart ER (1993) Clonal and leaf age variation in *Populus* phenolic glycosides: implications for host selection by *Chrysomela scripta* (Coleoptera: Chrysomelidae). *Environ Entomol* 22:397–403
- Bomford MK, Isman MB (1996) Desensitization of fifth instar *Spodoptera litura* to azadirachtin and neem. *Entomol Exp Appl* 81:307–313
- Chesson J (1983) The estimation and analysis of preference and its relationship to foraging models. *Ecology* 64:1297–1304
- Delgado AR, Prieto G (1998) Further evidence favoring three-option items in multiple-choice tests *Eur J Psychol Assess* 14:197–201
- Gonzalez-Vallejo, C (2002) Making trade-offs: A probabilistic and context-sensitive model of choice behavior. *Psychol Rev* 109:137–155
- Harris MO, Miller JR (1991) Quantitative analysis of ovipositional behavior – effects of a host-plant chemical on the onion fly (Diptera, Anthomyiidae). *J Insect Behav* 4:773–792
- Havill NP, Raffa KF (2000) Effects of eliciting treatment and genotypic variation on induced resistance in *Populus*: impacts on gypsy moth (Lepidoptera: Lymantriidae) development and feeding behavior. *Oecologia* 120:295–303
- Horton DR (1995) Statistical considerations in the design and analysis of paired-choice assays. *Environ Entomol* 24:179–192
- Jones CG, Coleman JS (1988) Leaf disk size and insect feeding preference: implications for assays and studies on induction of plant defense. *Entomol Exp Appl* 47:167–172
- Kruse JJ, Raffa KF (1997) Effects of selected midwestern larval host plants on performance by two strains of the gypsy moth parasitoid *Cotesia melanoscela* (Hymenoptera: Braconidae). *Environ Entomol* 26:1155–1166
- Le Magnen J (1999) Effect of a multiplicity of foot stimuli on the amount eaten by the rat. *Appetite* 33:36–39
- Lewis AC, van Emden HF (1986) Assays for insect feeding. In: Miller JR, Miller TR (eds) *Insect – plant interactions*. Springer, Berlin Heidelberg New York, pp 65–120
- Lockwood JR (1998) On the statistical analysis of multiple-choice feeding experiments. *Oecologia* 116:475–481
- Manly BJF (1993) Comments on design and analysis of multiple-choice feeding- preference experiments. *Oecologia* 93:149–152
- Manly, BFJ (1995) Measuring selectivity from multiple choice feeding-preference experiments. *Biometrics* 51:709–715
- Marquis RJ, Braker HE (1987) Influence of method of presentation on results of plant- host preference tests with two species of grasshopper. *Entomol Exp Appl* 44:59–63
- Martinat PJ, Barbosa P (1987) Relationship between host-plant acceptability and suitability in Newly eclosed first instar gypsy moths, *Lymantria dispar* (L.) (Lepidoptera: Lymantriidae). *Ann Entomol Soc Am* 80:141–147
- Mauffette Y, Lechowicz MJ, Jobin L (1983) Host preferences of the gypsy moth, *Lymantria dispar* (L.), in southern Quebec. *Can J For Res* 13:53–60
- Muthukrishnan AV, Kardes FR (2001) Persistent preferences for product attributes: the effects of the initial choice context and uninformative experience. *J Consumer Res* 28:89–104
- Nowlis SM, Simonson I (2000) Sales promotions and the choice context as competing influences on consumer decision making. *J Consumer Psychol* 9:1-16
- Parkhurst DF (2001) Statistical significance tests: equivalence and reverse tests should reduce misinterpretation. *BioScience* 51:1051–1057.
- Peterson CH, Renaud PE (1989) Analysis of feeding preference experiments. *Oecologia* 80:82–86
- Powell JS, Raffa KF (1999) Sources of variation in the concentration and composition of foliar monoterpenes in tamarack (*Larix laricina*) seedlings: the roles of nutrient availability, time of season, and plant architecture. *J Chem Ecol* 25:1771–1797
- Raffa KF, Frazier JL (1988) A generalized model for quantifying behavioral desensitization to antifeedants. *Entomol Exp Appl* 46:93-100
- Regent Instruments (1996) MacFOLIA Reference, version 3.4. Regent Instruments, Quebec, Canada
- Roa R (1992) Design and analysis of multiple-choice feeding-preference experiments. *Oecologia* 89:509–515
- Robison DJ, Raffa KF (1994) Characterization of hybrid poplar clones for resistance to the forest tent caterpillar. *For Sci* 40:686–714
- Robison DJ, Raffa KF (1997) Effects of constitutive and inducible traits of hybrid poplars on forest tent caterpillar feeding and population ecology. *For Sci* 43:252–267
- SAS Institute (1990) Language and procedures, version 6. SAS Institute, Cary, N.C.
- Schneidel U, Bruelheide H (1999) Selective slug grazing on montane meadow plants. *J Ecol* 87:828–838.
- Shafir S, Waite TA, Smith, BH (2002) Context-dependent violations of rational choice in honeybees (*Apis mellifera*) and gray jays (*Perisoreus canadensis*). *Behav Ecol Sociobiol* 51:180–187
- Singer MC (1986) The definition and measurement of oviposition preference in plant-feeding insects. In: Miller JR, Miller TR (eds) *Insect-plant interactions*. Springer, Berlin Heidelberg New York, pp 65–94
- Snedecor GW, Cochran WG (1993) *Statistical methods*. Iowa State University Press, Ames
- Wagner WE (1998) Measuring female mating preferences *Anim Behav* 55:1029–1042