New tests for departures from random behavior in spatial memory experiments

YVES TILLÉ
Université Libre de Bruxelles, Brussels, Belgium

JONATHAN A. NEWMAN
Southern Illinois University, Carbondale, Illinois

and

SUSAN D. HEALY
University of Newcastle Upon Tyne, Newcastle Upon Tyne, England

We present statistical tests for departures from random expectation in spatial memory tasks. We consider two common protocols for spatial memory experiments. In the first one, subjects are allowed to search a fixed number of sites. In the second protocol, subjects are allowed to search until they achieve a fixed number of successes. In either of these protocols, the subjects involved may or may not revisit sites that have been previously searched or exploited. This yields four situations to consider: fixed number of sites searched or fixed number of successes, with or without revisits. We derive analytical expressions for the probability mass functions, expectations, and variances associated with each type of null hypothesis. We present three statistical tests of these hypotheses: the Kolmogorov-Smirnov test, the ordinary sign test, and the Z test. We use our results to demonstrate a priori calculation of sample sizes and statistical power and to consider a mixed model of sampling with and without replacement.

The basis for experiments investigating spatial memory capabilities of animals is that animals are able to perform spatial memory tasks better than expected by chance. In spatial memory experiments, such as radial maze experiments, a fixed number of sites is used. Some sites contain food, and others do not. Moreover, two experimental protocols can be used. In the first protocol, animals can explore a fixed number of sites. We shall refer to this protocol as the *fixed-number-of-samples/trials protocol*. The data that result from this protocol are the number of visited sites containing food. In the second protocol, animals can explore until they have found a fixed number of sites containing food. We shall refer to this second protocol as the *fixed-number-of-successes protocol*. The data that result from this protocol are the total number of sites visited.

One relevant null hypothesis is that the animals search at random until they find food-storage sites. In many cases, animals search without revisiting locations previously searched; in other cases (e.g., rats that have had their hippocampus lesioned), the animals do revisit previously searched or exploited sites. Again, two cases must be distinguished. In the first case, the hypothesis is that animals explore the sites at random and without replacement. In

the second case, we can suppose that animals explore at random with replacement. Four different situations must thus be examined. Animals can search with or without replacement in a protocol where either the number of trials or the number of successes is fixed. In order to test the hypothesis that animals search better than at random, it is necessary to know the probability mass functions of the number of successes (in the fixed-number-of-trials protocol) or of the number of trials (in the fixed-number-of-successes protocol) when animals search at random with or without replacement.

At first sight, this problem looks simple and classic. One might think that when an animal searches at random with replacement, the solution is given by the binomial distribution (in the fixed-number-of-trials protocol) and negative binomial distribution (in the fixed-number-of-successes protocol). This first impression is, however, incorrect. Indeed, when an animal explores a site, it eats the food, and, thus, the site containing food becomes empty once it is explored. From the sampling theory point of view, this problem could be presented as a sampling procedure that modifies the state of the population. Curiously, this problem seems to be unsolved. Nevertheless, results given by some Indian statisticians (Basu, 1958; Chikkagoudar, 1966; Raj & Khamis, 1958) gave a point of departure from which we were able to derive the distributions related to these problems.

If an animal searches at random without replacement, the problem is more simple. In the fixed-number-of-trials protocol, the result is given by the hypergeometric distribution, which is a classical result in sampling theory. The problem

has also been solved in the fixed-number-of-successes protocol, although we shall show that it is more unusual.

In this paper, we present the probability mass functions for all common protocols used to test spatial memory. To do this, we had to derive two new statistical distributions. The proof for these new results is given in Appendix A. The probability mass functions allow us to calculate the random expectations and variances and to use more appropriate statistics to test the null hypotheses (e.g., Kolmogorov-Smirnov test, ordinary sign test, and Z test) rather than the commonly used t test. These functions also allow us to statistically distinguish between site revisits that occur when searching with replacement (see below) and "errors" (i.e., site revisits) when searching without replacement (see below). We also use these results to develop a compound model in which the animal's behavior is assumed to be a mixture of sampling with and without replacement. Finally, we use our results to investigate the statistical power of the various test options available and comment on the choice of statistical test.

Note that Appendix B contains a listing of all symbols used in this paper. The reader may find it useful to refer to this list from time to time.

MEMORY TASKS WITH A FIXED NUMBER OF SAMPLES

In food-storing experiments (e.g., Balda & Kamil, 1989; Sherry & Vaccarino, 1989; Shettleworth & Krebs, 1982), the common protocol is to allow animals to search a fixed number of sites, and the retrievals ("successes") that the animal achieves are tallied. We shall consider two different strategies that animals could use to sample their environment: (1) at random without replacement and (2) at random with replacement. When animals search without replacement, they do not revisit sites previously visited. When sampling with replacement, the animal randomly chooses one site from among all available sites, without regard to whether that site has previously been searched. In both cases, we can calculate the distribution function, the random expectation, and the variance.

Sampling Without Replacement: The Hypergeometric Distribution

It is known (e.g., Kamil & Balda, 1985; Olson, Kamil, & Balda, 1993; Spetch & Edwards, 1986) that the number of correct choices when sampling at random has a hypergeometric distribution (e.g., Degroot, 1970). The probability mass function is given by

$$\Pr[X=r] = \frac{\binom{A}{r}\binom{B}{n-r}}{\binom{N}{n}}, r = \max(0, n-B), \dots, \min(A, n),$$

(1)

where X is the number of successes, A is the number of sites that constitute a success (e.g., sites that actually contain food), B is the number of sites that constitute failures

(e.g., sites that do not contain food), $N = A + B$ is the number of sites, and n is the number of sites that the animal is allowed to search before the trial is terminated.

The expectation of X is given by

$$\mu_X = \mathrm{E}(X) = \frac{nA}{N}.$$

(2)

The variance (e.g., Degroot, 1970) is given by

$$\sigma_X^2 = \mathrm{Var}(X) = \frac{nAB}{N^2} \frac{N-n}{N-1}.$$

(3)

To test whether animals are doing better than a simple random sampling without replacement, we can use the following test:

H_0: $\mu \le \mu_X$, the expected number of successes *is not better than* expected by chance.

H_1: $\mu > \mu_X$, the expected number of successes *is better than* expected by chance.

In this test, μ_X is given by Equation 2 and μ is the true expected number of success.

What is important about the knowledge of Equation 3 is that we can test inferences about the expected number of successes out of n sites searched using a Z test rather than a t test, as is commonly employed (e.g., Healy & Krebs, 1992; Hilton & Krebs, 1990; Olton & Samuelson, 1976; Spetch & Edwards, 1986). This is important since the Z test is more powerful (i.e., better able to reject a false null hypothesis) than the t test. This null hypothesis is thus tested with a probability level α by using a Z ratio. The following decision rule is used:

$$\text{Reject } H_0 \text{ if } Z = \frac{\overline{X} - \mu_X}{\sqrt{\sigma_X^2 / v}} > Z_{1-\alpha},$$

where $\overline{X}$ is the sample mean (i.e., the mean of the number of successes achieved for the v subjects) and $Z_{1-\alpha}$ is the $1 - \alpha$ order quantile of a standard normal variable.

Sampling With Replacement: The Hyperbinomial Distribution

The normal approach to sampling with replacement is to recognize that since each sample is either a success or a failure, the sample constitutes a "Bernoulli trial." The expected number of successes in n Bernoulli trials is normally given by the expectation of the binomial distribution np (where n is the number of samples and p is the proportion of samples that constitutes a success). The variance of this expression is npq, where q is $1 - p$. However, in the case of spatial memory experiments, these expressions are not useful because once a site containing food is emptied, it is not refilled and hence cannot be exploited again (as is the assumption behind the use of the binomial distribution). In Appendix A (see Theorem 1), we derive a new distribution, which we call the *hyperbinomial* distribution. The probability mass function of the number of successes in n total samples when sampling with replacement is given by (see Appendix A, Theorem 1)

$$\Pr[\xi = r] =$$

$$\binom{A}{r}\frac{1}{N^n}\sum_{i=r}^{\min(n,B+r)} i!\binom{B}{i-r}S_n^{(i)}, \quad r = 0,\dots,\min(n,A), \quad (4)$$

where ξ denotes the number of successes, and A, B, N, and n are defined as they were for Equation 1. $S_n^{(i)}$ is the Stirling number of the second kind (Abramowitz & Stegun, 1957, p. 824) and is defined simply as

$$S_n^{(r)} = \frac{1}{r!}\sum_{k=0}^{r}(-1)^{r-k}\binom{r}{k}k^n.$$

The expectation of ξ is given by

$$\mu_\xi = E(\xi) = A\left(1 - \frac{(N-1)^n}{N^n}\right). \quad (5)$$

The variance is given by

$$\sigma_\xi^2 = \text{Var}(\xi) =$$

$$A\left(\frac{(N-1)^n + (A-1)(N-2)^n}{N^n} - A\frac{(N-1)^{2n}}{N^{2n}}\right). \quad (6)$$

Again, since the variance of the distribution is known, we can use a Z test in place of a t test.

To test whether animals are doing better than a simple random sampling with replacement, we can also use a mean test:

H_0: $\mu \le \mu_\xi$, the expected number of successes *is not better than* expected by chance.

H_0: $\mu > \mu_\xi$, the expected number of successes *is better than* expected by chance.

In this test, μ_ξ is given by Equation 5 and μ is the true expected number of success.

This null hypothesis is thus tested with a probability level α by using a Z ratio. The following decision rule is used:

$$\text{Reject } H_0 \text{ if } Z = \frac{\overline{X} - \mu_\xi}{\sqrt{\sigma_\xi^2/v}} > Z_{1-\alpha},$$

where $\overline{X}$ is the sample mean (i.e., the mean of the number of successes obtained for the v animals).

MEMORY TASKS WITH A FIXED NUMBER OF SUCCESSES

In spatial memory experiments that utilize a radial arm maze, one common protocol (e.g., Balda & Kamil, 1988; Hilton & Krebs, 1990; Ilersich, Mazmanian, & Roberts, 1988; Olson, Kamil, & Balda, 1993; Roberts & van Veldhuizen, 1985; Spetch & Honig, 1988) allows the animals to continue sampling until they have achieved a certain number of successes (commonly, but not necessarily, this number is all possible successes). In this case, Equations 1–6 are not useful in calculating the random expectation. What we need to know is how many sites we should expect the animals to search before they find this num-

ber of successes. Again, we must consider two cases: sampling without replacement and sampling with replacement.

Sampling Without Replacement: The Negative Hypergeometric Distribution

When animals sample without replacement, it can be shown that the number of sites sampled until the rth success has a negative hypergeometric distribution (see Johnson, Kotz, & Kemp, 1992, p. 241):

$$\Pr[Y = r+k] = \frac{\binom{N-r-k}{B-k}\binom{r+k-1}{k}}{\binom{N}{A}}, \quad k = 0,1,\dots,B,$$

$$(7)$$

where Y is the number of sites searched to find exactly r successes, and A, B and N are as defined above. The expectation of Y is given by

$$\mu_Y = E(Y) = r\frac{N+1}{A+1}. \quad (8)$$

The variance is given by

$$\sigma_Y^2 = \text{Var}(Y) = \frac{r(A+1-r)(N+1)B}{(A+1)^2(A+2)}. \quad (9)$$

In this case, a Z test can still be used. To test whether animals do better than a simple random sampling without replacement, we test the following hypotheses:

H_0: $\mu \ge \mu_Y$, the expected number of sites searched *is not less than* expected by chance.

H_1: $\mu < \mu_Y$, the expected number of sites searched *is less than* expected by chance.

Note that μ denotes here the expected number of trials to obtain r successes. If animals search more efficiently than using a simple random sampling, the number of sites visited will be smaller than μ_Y.

This null hypothesis is tested with a probability level α by using a Z ratio. The following decision rule is used:

$$\text{Reject } H_0 \text{ if } Z = \frac{\overline{Y} - \mu_Y}{\sqrt{\sigma_Y^2/v}} < Z_\alpha,$$

where $\overline{Y}$ is the sample mean (i.e., the mean of the number of trials for the v animals).

Sampling With Replacement: The Negative Hyperbinomial Distribution

Now, suppose that the animals sample with replacement. Normally, the number of failures until the rth success is given by the negative binomial distribution. However, we cannot use the negative binomial result since food-storage sites are not refilled after each success. To deal with this situation, we present a new distribution, which we call the *negative hyperbinomial* distribution. In

Appendix A (Theorem 2), we derive the probability mass function for this distribution. We show that the probability mass function of the number of sites searched until the rth success, when sampling with replacement, is given by

$$\Pr[\eta = r + k] =$$
$$\binom{A}{r}\frac{r}{N^{r+k}}\sum_{i=0}^{\min(k,B)}(i+r-1)!\binom{B}{i}S_{r+k-1}^{(i+r-1)}, \quad k = 0,1,\dots,$$
(10)

where η is the number of sites searched, r is the number of successes at which the sampling is terminated, and A, B, and N are defined as above. The expectation of η is given by

$$\mu_\eta = E(\eta) = N\sum_{j=0}^{r-1}\frac{1}{A-j}. \quad (11)$$

The variance is given by

$$\sigma_\eta^2 = \mathrm{Var}(\eta) = N\sum_{j=0}^{r-1}\frac{B+j}{(A-j)^2}. \quad (12)$$

Again, a Z test can still be used. To test whether animals do better than a simple random sampling with replacement, we test the following hypotheses:

H_0: $\mu \geq \mu_\eta$, the expected number of sites searched *is not less than* expected by chance.

H_1: $\mu < \mu_\eta$, the expected number of sites searched *is less than* expected by chance.

Again, μ denotes the expected number of trials to obtain r successes. This null hypothesis is tested with a probability level α by using a Z ratio. The following decision rule is used:

$$\text{Reject } H_0 \text{ if } Z = \frac{\bar{Y} - \mu_\eta}{\sqrt{\sigma_\eta^2/v}} < Z_\alpha,$$

where $\bar{Y}$ is the sample mean (i.e., the expected number of trials for the v animals).

Example 1

In this first example (Edwards, 1994), there are 8 sites. One site contains food and the seven others do not: $N = 8$, $A = 1$, $B = 7$. Twenty animals ($v = 20$) are allowed to search until they find the site containing food ($r = 1$). Results are given for the first day and after 6 days of training in Table 1.

The sample mean number of trials $\bar{Y}$ needed to find the site equals 4.45 for the first day and 1.25 after 6 days of training.

To test whether animals do better than a simple random sampling with replacement, using Equations 11 and 12, we get $\mu_\eta = 8$ and $\sigma_\eta^2 = 56$. For the first day, $Z = -2.12153$; after 6 days, $Z = -4.0339$. These two Z ratios are smaller than the 0.05 order quantile of a standard normal variable ($Z_{0.05} = -1.64$), and, thus, H_0 is rejected in both cases. Thus, we can say that since the first-day animals are doing better than a simple random sampling with replacement.

To test whether animals do better than a simple random sampling without replacement, using Equations 8 and 9, we get $\mu_Y = 9/2$ and $\sigma_Y^2 = 21/4$. For the first day, $Z = -0.09759$; after 6 days, $Z = -6.34335$. On the first day, the Z ratio is larger than the 0.05 order quantile of a standard normal variable ($Z_{0.05} = -1.64$); however, after 6 days, the Z ratio is smaller than $Z_{0.05} = -1.64$. So, on the first day, H_0 is not rejected; however, after 6 days, H_0 is rejected. Thus, we can conclude that, after 6 days, animals perform better than sampling without replacement.

In this case, 20 animals are used, but, normally, we require at least 30 animals ($v \geq 30$) to use a Z test. However, it is very common in spatial memory experiments to have sample sizes of less than 30. Where the experimenter is unwilling to make the normality assumption required of such a parametric test, there are alternative nonparametric tests available (see Examples 3 and 4).

Remarks on Example 1. Note that the data are paired because they come from the same animals at two distinct times. For this reason, if we want to test whether the animals improve their performance, we must take into account the lack of independence between the successive observations. In this case, we might use the paired t test or the Wilcoxon signed rank test (see, e.g., Gibbons, 1993, pp. 4–5), both of which can be used for paired data. In Example 1, the Z test is applied at a precise time to verify whether animals are doing better than expected by chance. In this case, the lack of independence does not greatly affect the validity of the test. (However, it would be a problem if we used a Z test to test the results every day until we found a significant difference.)

In this example, and the examples that follow, we are concerned with testing the one-tailed hypothesis that the animal performs better than expected by chance. A reviewer pointed out that two-tailed tests may also be of interest. Indeed, as that reviewer noted, animals following a "win-shift-lose-stay" strategy in any of the protocols presented here, could reliably lead to results that are

Table 1
The Number of Sites Searched Until the Animals Discovered the Site Containing Food

Trials	Animal Number																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Day 1	6	5	5	5	4	3	5	4	1	7	4	5	1	2	7	6	3	4	5	7
Day 6	1	1	1	1	1	1	2	2	2	1	1	1	1	2	2	1	1	1	1	1

Note—There were 8 sites in total. Data from Edwards (1994).

worse than expected by chance. Extending the Z test (or any of the tests that follow) to consider the two-tailed hypothesis can be easily accomplished and is not affected by the calculation of the random expectation.

CALCULATING SAMPLE SIZES

Equations 1, 4, 7, and 10 give the probability mass functions for the four hypotheses of interest. Knowledge of the probability mass function allows us to calculate exact probabilities for any event or subset of events that we desire. One property that knowledge of the probability mass function allows us to calculate is the theoretical variance of the distribution. Knowledge of the theoretical variance allows us to calculate, a priori, the sample size needed to detect a difference of some particular size.

Example 2

Suppose, as in Example 1 (i.e., animals search among eight possible sites for one hidden item, they search until they find the site containing food), that we want to test whether animals perform better than would be expected if they were using a simple random sampling design without replacement until an item is found ($N = 8, A = 1, B = 7, r = 1$).

Assume that we wish, a priori, to be able to detect a significant difference if our observed sample mean is, for example, $d = 1$ less than expected if animals are searching using a simple random sampling design without replacement. In that case, we should need

$$-1.64 > \frac{d}{\sqrt{\sigma_Y^2}/\sqrt{v}} = \frac{-1}{\sqrt{21/4}\ \sqrt{v}}$$

$$v > \left(\frac{1.64\sqrt{21/4}}{-1}\right)^2$$

$$v > 14.1204.$$

That is, if we wished to be able to detect a departure from random performance with a magnitude of 1 site sampled, then we would need a sample size of at least 15 animals to detect this difference at the $\alpha = 0.05$ level on a one-tailed test. Similar calculations can be performed for any of the situations demonstrated below.

KOLMOGOROV-SMIRNOV TEST

There is some question as to whether parametric tests, such as the Z and t tests, are appropriate for testing the

hypothesis in the case of actual spatial memory experiments. The logic of using a parametric test is that we are dealing with distributions of sample means, which the central limit theorem tells us must be approximately normally distributed. The distribution of sample means converges to a normal distribution as the sample size becomes large. It is generally accepted that the sample size must be greater than 30 to ensure convergence. For experimenters who are leery of making the normality assumption without such a large sample size (it is unusual to have such large sample sizes for these types of experiments), it may be more appropriate to use a nonparametric test, such as the Kolmogorov-Smirnov test.

The Kolmogorov-Smirnov test is desirable in that it does not require normality—it is distribution-free. This test is applicable because the exact distribution under the null hypothesis is calculable (see Conover, 1980; Gibbons, 1985).

Example 3

To illustrate the use of this test, consider the following experiment. Ten animals are allowed to find two food items hidden among 16 sites ($A = 2, B = 14, N = 16$) (Edwards, 1994). The animals are allowed to sample until they have found both items ($r = A = 2$), at which point the experiment is terminated. Table 2 gives the number of trials required for two successes for the first day and after 6 days of training.

In this case, the small number of animals used does not ensure convergence, so we may prefer not to use a Z test. We want to test whether animals search better than if they were using a simple random sampling design without replacement. So, the hypotheses are:

H_0: $F(x) \leq F_0(x)$, the cumulative distribution function *is not greater than* the distribution function when sampling without replacement.

H_1: $F(x) > F_0(x)$, the cumulative distribution function *is greater than* the distribution function when sampling without replacement.

$F(x)$ denotes the cumulative distribution function of the distribution from which the data come, and $F_0(x)$ denotes the cumulative distribution function under the null hypothesis—in this case,

$$F_0(x) = \sum_{j=r}^{x} \Pr[Y = j],$$

where Y is defined as in Equation 7. Table 3 gives, for the first and sixth day, the observed cumulative distribution

Table 2
The Number of Sites Searched Until the
Animals Discovered the Two Sites Containing Food

	Animal Number									
Trials	1	2	3	4	5	6	7	8	9	10
Day 1	10	14	15	12	12	9	10	8	9	6
Day 6	4	3	2	4	2	3	3	2	4	4

Note—There were 16 sites in total. Data from Edwards (1994).

function $F_v(x)$, the cumulative distribution function under the null hypothesis $F_0(x)$, and the differences between $F_v(x)$ and $F_0(x)$. The Kolmogorov-Smirnov statistic is given by $T^- = \max_x[F_v(x)-F_0(x)]$.

On the first day, $T^- = 0.25$, the tabulated value for $v = 10$ and $\alpha = 0.05$ is 0.369, and thus the decision is: "fail to reject H_0." After 6 days of training, $T^- = 0.95$, the tabulated value does not change, and the decision is: "reject H_0."

Thus, with no prior experience with the experimental set-up, the animals do no better than expected by chance. However, after 6 days of experience with the set-up, the animals do considerably better than random.

Exact Method for Critical Values

The Kolmogorov-Smirnov test is normally used with continuous data, not discrete data of the type produced by spatial memory experiments. However, it can be shown that for discrete distributions, the Kolmogorov-Smirnov test becomes very conservative (Pettit & Stephens, 1977). Thus, if the test suggests rejection of the null hypothesis using the tabulated value, then we can have a great deal of confidence in this result. However, when the difference between our observed and expected distributions is not large, we will not be able to afford the use of such a conservative test. To overcome this problem, Conover (1972) presents an algorithm for calculating the exact critical value for the Kolmogorov-Smirnov test. This method can be up to three times more powerful than using tabulated values for the Kolmogorov-Smirnov test, especially when the sample size is small. The only problem with this method is that it can require significant computation time, even for large computers. This algorithm is clearly explained in Conover (1980, pp. 350-353).

ORDINARY SIGN TEST (MEDIAN TEST)

Another way to test the hypothesis that animals are searching at random (with or without replacement) is to use the ordinary sign test to test hypotheses about the median performance (see Gibbons, 1985, pp. 100–104). This test is distribution-free; moreover, it is the most powerful test for testing the value of a quantile (e.g., the median; see Lehmann, 1986, pp. 106-107).

Example 4

Consider the following experiment on 6 hummingbirds using a radial arm maze protocol (Healy & Hurly, 1995). Each bird was presented with eight flowers containing sucrose solution. When a bird had visited four of the flowers, it was interrupted by the experimenter and chased away. On return, the bird was then observed until it had visited the four flowers that still contained sucrose solution. Table 4 contains the number of successes achieved in the first four flowers visited. In this case, we have $A = 4$, $B = 4$, $N = 8$, and $n = 4$; 6 animals were used to conduct the experiment ($v = 6$).

We want to test whether animals search better than if they were using a simple random sampling design without replacement. Table 5 gives the cumulative observed function $F_v(x)$ and the cumulative function $F_0(x)$ under the hypothesis that animals search at random without replacement:

$$F_0(x) = \sum_{j=0}^{x} \Pr[X = j],$$

where X is defined as in Equation 1.

Examining Table 5, we see that the median under the hypothesis that animals search at random without replacement is $M_0 = 2$. If M denotes the median of the distribution from which the data come, the null hypothesis to test is $H_0: M \leq 2$ versus $H_1: M > 2$.

From Table 4, we observe that 3 animals get exactly two successes, 3 get more than two successes, and 0 less than two successes. To do the ordinary sign test, we consider only the three cases where the number of successes is different from two. Under H_0, the probability of getting three results larger than two in three experiments is given by

$$p = \binom{3}{3}.5^3 = .125.$$

Since $p > \alpha = .05$, H_0 is not rejected. Thus, we fail to reject the null hypothesis that animals search at random without replacement.

DISTINGUISHING "ERRORS" FROM SAMPLING WITH REPLACEMENT

We sometimes have a case where animals do revisit sites that they have previously investigated or depleted, but we do not believe that these animals are sampling with replacement. Rather, we think that the animals are sampling without replacement, but that they make occasional "errors." These revisits are often simply treated as errors and ignored in subsequent analyses. That is, researchers test the null hypothesis that the animals were sampling with-

Table 3
Kolmogorov-Smirnov Test for the Data in Table 2

Trials	Day 1 $F_v(x)$	Day 6 $F_v(x)$	$F_0(x)$	Day 1 $F_v(x)-F_0(x)$	Day 6 $F_v(x)-F_0(x)$
1	0.0	0.0	0.00	0.00	0.00
2	0.0	0.3	0.01	−0.01	0.29
3	0.0	0.6	0.03	−0.03	0.58
4	0.0	1.0	0.05	−0.05	**0.95**
5	0.0	1.0	0.08	−0.08	0.92
6	0.1	1.0	0.13	−0.03	0.88
7	0.1	1.0	0.18	−0.08	0.83
8	0.2	1.0	0.23	−0.03	0.77
9	0.4	1.0	0.30	0.10	0.70
10	0.6	1.0	0.38	0.23	0.63
11	0.6	1.0	0.46	0.14	0.54
12	0.8	1.0	0.55	**0.25**	0.45
13	0.8	1.0	0.65	0.15	0.35
14	0.9	1.0	0.76	0.14	0.24
15	1.0	1.0	0.88	0.13	0.13
16	1.0	1.0	1.00	0.00	0.00

Note—Shows the cumulative distribution of the data and the distribution expected under the null hypothesis.

Table 4
The Number of Successes in the First Four Sites Searched

	Bird Number					
	1	2	3	4	5	6
Trials	2	4	2	3	2	3

Note—Data from Healy and Hurly (1995).

out replacement, even when there are some sites that are revisited. In this section, we consider this problem further.

Perhaps the simplest way to approach this problem is to test whether the number of sites that are revisited is different than we expect by chance under the null hypothesis of sampling *with* replacement. In the case of a fixed number of trials, the distribution is obtained immediately from Equation A1 (see Appendix A). Equation A1 is the probability mass function for the number of unique storage sites searched when making n random samples with replacement from a population of size N. If we let Ψ denote the number of sites visited more than once, in a random sample with replacement of size n, from a population of size N, then the probability mass function for the number of sites visited more than once is given by

$$\Pr[\Psi = r] = \Pr[\Phi = n - r], \qquad (13)$$

where Φ is defined as in Equation A1. Thus, we can use Equation A1 to conduct a Kolmogorov-Smirnov test as above, or we can use Equations A2 and A4 (see Appendix A) to conduct a Z test (whichever is justified) to test the null hypothesis that the animals are revisiting sites at a frequency expected by chance when sampling at random with replacement. If we reject this hypothesis and find that the animals are revisiting sites, but less frequently than we expect by chance, then we may choose to ignore the revisits. We would then test the null hypothesis that, aside from the errors, the animals were performing better than we would expect if they were sampling at random without replacement. To do this, we proceed as above.

We note that the probability that r sites were visited k times is given by Feller (1968, p. 112). This problem is called the *classical occupancy* problem. This bivariate probability mass function could allow us a more precise approach to testing the errors hypothesis. We shall not develop this problem further in the present paper.

In the case of a fixed number of successes, if we denote ζ as the number of trials needed to visit exactly r distinct sites, we can derive the probability mass function of ζ from Equation 10 by taking $A = N$ and $B = 0$. The probability mass function of ζ is given by

$$\Pr[\zeta = r + k] = \frac{N!}{(N-r)! \, N^{r+k}} S_{r+k-1}^{(r-1)}, \qquad k = 0, 1, \dots .$$

This problem is known as *inverse sampling with replacement* and was studied by Basu (1958). The expectation of ζ is given by

$$E(\zeta) = N \sum_{j=0}^{r-1} \frac{1}{N-j}.$$

The variance is given by

$$\mathrm{Var}(\zeta) = N \sum_{j=0}^{r-1} \frac{j}{(N-j)^2}.$$

If τ denotes the number of sites visited more than once that are needed to visit exactly r distinct sites, we have

$$\Pr[\tau = k] = \Pr[\zeta = r + k], k = 0, 1, \dots$$

As in any of the above examples, these results can be used to construct a Z test, a Kolmogorov-Smirnov test, or an ordinary sign test.

COMPOUND MODEL

Another means of dealing with the "problem of errors" would be to construct a compound model. The logic of such a model proceeds as follows. The two simple models (sampling with and without replacement) may appear to be an oversimplification of animal behavior. Indeed, in most of the cases, we may expect that animals search without replacement but occasionally make mistakes. This kind of behavior suggests the use of a compound model. The simplest compound model is formed by assuming that, in the case of a fixed number of trials, the probability mass function of the number of successes is a linear combination between the probability mass functions of the with- and without-replacement schemes. For instance, for a fixed number of trials, we can define variable W such as

$$\Pr[W = r] = \lambda \Pr[X = r] + (1 - \lambda) \Pr[\xi = r],$$
$$r = 1, \dots, n, \text{ with } 0 < \lambda < 1, \qquad (14)$$

where $\Pr[X = r]$ and $\Pr[\xi = r]$ are defined as in Equations 1 and 4. The expectation and variance come directly:

$$E[W] = \lambda E[X] + (1 - \lambda) E[\xi]$$

and

$$\mathrm{Var}[W] = \lambda \mathrm{Var}[X] + (1 - \lambda) \mathrm{Var}[\xi] + \lambda (E[X] - E[W])^2$$
$$+ (1 - \lambda)(E[\xi] - E[W])^2.$$

In the same way, we can define a variable U for which the probability mass function is linear combination between the with- and without-replacement distributions for the fixed-number-of-successes scheme by using Equations 8 and 11 in place of Equations 1 and 4 above.

This model may look unrealistic, but our discussion of this model can be extended to any compound model with

Table 5
Sign Test for the Data in Table 4

Successes	$F_v(x)$	$F_0(x)$
0	0.000	0.0143
1	0.000	0.2429
2	0.500	0.7571
3	0.833	0.9857
4	1.000	1.0000

Note—Cumulative distribution of the data and the distribution expected under the null hypothesis.

at least one unknown parameter. Indeed, if a compound model is used to test a hypothesis, parameter λ will be a "nuisance" parameter. To test the compound hypothesis, we must use a two-step procedure. In the first step, we conduct an experiment to estimate λ. In the second step, we conduct another experiment to test the model. If we suppose that animals have a behavior that is a mixture of sampling with and without replacement, the expected number of successes will be intermediate between the expectations of both replacement protocols. So, if animals are searching according to the model given by Equation 14, in the best case (i.e., animals make no mistakes, $\lambda = 1$), the expectation will be just sampling without replacement (Equation 1). Thus, it makes sense to first test the simple (noncompound) model. We can test directly that animals are doing better than a simple random sampling without replacement (the best case). The estimation of parameter λ is not necessary to test this null hypothesis. If and only if that test fails, then a compound model is called for. If a more complex and probably more realistic compound model is used, this model will also have at least one nuisance parameter to estimate, and the estimation of this parameter will interfere with the test in such a way that it may be more useful to test if animals are searching at random without replacement.

WHICH EXPERIMENTAL PROTOCOL TO USE?

The examination of experimental protocols and of the varied ways to test a hypothesis leads to a fundamental question: How do we best organize a spatial memory experiment? This question leads directly to many practical questions. Is it preferable to use a fixed-trials or a fixed-samples protocol? It may be tempting to use a fixed-number-of-successes protocol with $r = A$, because this test may yield more "information." We shall show that, from a statistical viewpoint, this first impression is false. Another important problem is the allocation between the sites containing food and the sites that do not contain food (ratio A/B). For the fixed-number protocol, which value should we choose for n? For the fixed number of successes, which value should we choose for r?

These are questions of experimental design, which can be posed as a power problem. The power of a statistical test is the probability of rejecting H_0 given that the alternative hypothesis, H_1, is true (for a fixed level of α). We shall see that a general optimal design (most powerful) does not exist but that, for each alternative hypothesis, one can determinate a particular optimal test. To illustrate this problem, we will consider two examples. For simplicity, in these two examples, we will illustrate this procedure using a Z test. Similar calculations can be made for other statistical tests.

Example 5

Suppose that we want to test the hypothesis H_0 that animals search at random with replacement against

H_1 that animals search in accordance with the compound model given by Equation 14 with a fixed value for λ.

In the fixed-number-of-trials protocol, hypothesis H_0 will be rejected if

$$\overline{X} \geq L_1 = \mu_\xi + Z_{1-\alpha} \frac{\sigma_\xi}{\sqrt{v}},$$

where $Z_{1-\alpha}$ is the quantile of the $1 - \alpha$ order of a standard normal distribution function. The power of H_1 is given by

$$\text{power}_{\text{FNT}}(\alpha \mid H_1) = 1 - F\left(\frac{L_1 - \text{E}(W)}{\sqrt{\text{Var}(W)/v}}\right).$$

$F(\bullet)$ is the probability distribution function for a standard normal random variable.

In the fixed-number-of-successes protocol, hypothesis H_0 will be rejected if

$$\overline{Y} \leq L_2 = \mu_\eta - Z_{1-\alpha} \frac{\sigma_\eta}{\sqrt{v}}.$$

The power for H_1 is given by

$$\text{power}_{\text{FNS}}(\alpha \mid H_1) = F\left(\frac{L_2 - \text{E}(U)}{\sqrt{\text{Var}(U)/v}}\right).$$

If we calculate the power for the fixed-number-of-successes protocol using Equation 14 (but replacing Equations 1 and 4 with Equations 8 and 11) with $\lambda = 0.6$, $N = 8$, and $v = 30$, we find that the most powerful experiment consists of taking $A = N = r = 8$ with a power of 0.950412. For the fixed-number-of-trials protocol, the best experiment consists of taking $A = N = n = 8$ with a power of 0.714743. The fixed-number-of-successes protocol is, in this case, more powerful. So, perhaps somewhat counterintuitively, the best experiment consists of beginning with food in all the sites and letting the animals search until all the sites have been visited.

Example 6

Suppose, in the fixed-number-of-samples protocol, that an animal is able to retrieve directly a limited number, c, of sites containing food and next searches at random without replacement. If X_c denotes the number of successes after sampling n sites, we have

$$\Pr[X_c = r] = \begin{cases} I_{\{r=n\}}, r = 1, \ldots, \min(A,n), \text{ if } n \leq c \\ \\ \dfrac{\dbinom{A-c}{r-c}\dbinom{B}{n-r}}{\dbinom{N-c}{n-c}}, r = \max(c, n-B), \ldots, \\ \hspace{4cm} \min(A,n), \text{ if } n > c, \end{cases}$$

where I_D takes the value 1 if D is true and 0 if not. Thus,

$$E(X_c) = \begin{cases} n, & \text{if } n \leq c, \\ c + \dfrac{(n-c)(A-c)}{(N-c)}, & \text{if } n > c, \end{cases}$$

and

$$Var(X_c) = \begin{cases} 0, & \text{if } n \leq c, \\ \dfrac{(n-c)(A-c)B}{(N-c)^2} \dfrac{N-n}{N-c-1}, & \text{if } n > c. \end{cases}$$

Suppose now that we want to test the null hypothesis H_0 that animals search at random without replacement against the alternative hypothesis H_1 that animals' behavior can be described by the probability mass function of variable X_λ, where

$$\Pr[X_\lambda = r] = \lambda\Pr[X_c = r] + (1-\lambda)\Pr[X = r],$$

where X is defined as in Equation 1. Thus,

$$E(X_\lambda) = \lambda E(X_c) + (1-\lambda)E(X).$$

The most powerful Z test for these two hypotheses can be determined. Indeed, with the Z test, we shall reject H_0 if

$$\overline{X} > L_3 = E(X) + Z_{1-\alpha}\sqrt{\dfrac{Var(X)}{v}}.$$

The power is the probability that we reject H_0 given that H_1 is true. Thus,

$$\text{power}_{FNT}(\alpha \mid H_1) = 1 - F\left(\dfrac{L_3 - E(X_\lambda)}{\sqrt{Var(X_\lambda)/v}}\right).$$

For the alternative hypothesis H_1, it is thus possible to determine the best choice of A and B, and then to obtain the most powerful test (see below).

Suppose, in the fixed-number-of-successes protocol, that an animal is able to retrieve directly a limited number, c, of sites containing food and next searches at random without replacement. If Y_c denotes the number of trials needed to find r sites containing food, we have

$$\Pr[Y_c = r + k] = \begin{cases} I_{\{k=0\}}, & \text{if } r \leq c, \\ \dfrac{\dbinom{N-r-k}{B-k}\dbinom{r+k-1-c}{n-r}}{\dbinom{N-c}{A-c}}, & \text{if } r > c, \end{cases}$$

with

$$E[Y_c] = \begin{cases} r, & \text{if } r \leq c, \\ (r-c)\dfrac{N-c+1}{A-c+1}, & \text{if } r > c, \end{cases}$$

and

$$Var[Y_c] = \begin{cases} 0, & \text{if } r \leq c, \\ \dfrac{(r-c)(A-c+1-r)(N-c+1)B}{(A-c+1)^2(A-c+2)}, & \text{if } r > c. \end{cases}$$

We can also define the variable Y_λ in the following way:

$$\Pr[Y_\lambda = r + k] = \lambda\Pr[Y_c = r + k] \\ + (1 - \lambda)\Pr[Y = r + k],$$

where Y is defined as in Equation 7. The expectation and variance of Y_λ are easily derived. The most powerful Z test for these two hypotheses can be determined. Indeed, with the Z test, we shall reject H_0 if

$$\overline{Y} < L_4 = E(Y) - Z_{1-\alpha}\sqrt{\dfrac{Var(Y)}{v}}.$$

Thus,

$$\text{power}_{FNT}(\alpha \mid H_1) = F\left(\dfrac{L_4 - E(Y_\lambda)}{\sqrt{Var(Y_\lambda)/v}}\right).$$

If, for example, $N = 100$, $v = 30$, and $\lambda = 0.2$, we find that the best fixed-number-of-trials protocol experiment is $A = n = 4$ with a power of 0.988358. The best fixed-number-of-successes experiment protocol is $A = r = 4$ with a power of 0.96032. Unlike Example 5, here, the fixed number of trials protocol is preferable. But even if a fixed-number-of-successes protocol was to be used, the best choice of r is not $r = A = N$, as in Example 5.

These two examples are very simple, but they show that there is no optimal experiment to test the two simple hypotheses we presented. The optimal protocol can only be defined as a function of the alternative hypothesis. To construct a powerful experiment, it is thus important to guess how animal "memory" will be expressed. A calculation of power can also be useful to determine the best protocol. A general discussion of tests of power can be found in Cox and Hinkley (1982). Conover (1972) discusses a calculation of power for the Kolmogorov-Smirnov test.

CHOICE OF STATISTICAL TEST

In the past, significance tests of the null hypothesis in spatial memory experiments were often conducted using a t test. The t test is a parametric test that uses the sample variance as the estimate of the theoretical variance. We have shown that if parametric statistics are to be used, it is possible to use the Z test since the theoretical variance is now known and does not need to be estimated from the data. This provides a more powerful test of the null hypothesis and also allows us to calculate, a priori, the sample size necessary to detect a departure from random behavior, of any given size. A mean test is not the only possible solution. If the aim is to test whether animals search better than at random, we can formulate different null hypotheses. It is possible to conduct a test on the mean (Z test), the distribution function (Kolmogorov-

Smirnov test), or the median (ordinary sign test). These tests may lead to different conclusions because they do not test exactly the same null hypotheses. The nonparametric tests are particularly useful in the case of the small sample sizes that are often used in spatial memory experiments. The Kolmogorov-Smirnov test has two advantages. First, it is valid for small sample sizes regardless of how the data are distributed. Second, because it is a goodness-of-fit test, it is more sensitive than a test of departure from the sample mean. That is, it is possible that our sample mean will not deviate from the theoretical mean, but the animals may still be behaving differently from random. A goodness-of-fit test can detect this, whereas a test of the mean cannot. If Conover's (1972) method is to be used, then, unfortunately, the Kolmogorov-Smirnov test can become difficult, especially if Equation 4 or Equation 10 is used. As an alternative to the Kolmogorov-Smirnov test, a more easily implemented and very powerful test is the ordinary sign test (Lehmann, 1986). Indeed, once the median of the distribution is known, this test can be done by hand using a simple binomial table.

DISCUSSION

In this paper, we provide the mathematical tools necessary to test for departures from random behavior using any of the described common experimental protocols. In the past, only tests for a fixed number of samples were conducted, and then only in cases of sampling without replacement (using the well-known hypergeometric distribution). Experiments that were conducted using a fixed-number-of-successes protocol had to be analyzed as if they were conducted using a fixed-number protocol, so that the data were in a form that could be readily assessed with statistics. The results presented in this paper alleviate this problem. Furthermore, we present results that allow us to distinguish between sampling with replacement and errors that occur when animals are otherwise sampling without replacement. Whatever the conclusion (i.e., sampling with or without replacement), we have presented tools to test the appropriate null hypothesis. We have also presented results that allow the testing of compound models, as well as a priori calculation of statistical power that is essential for conducting well-designed experiments.

Finally, we have described several hypothesis tests in this paper, but we must be very careful to understand exactly what each is testing. Suppose that we detect a departure from the random expectation, have we really demonstrated spatial memory? The answer is no. What we have demonstrated is a departure from the random expectation, not the mechanism for accomplishing that departure. For example, an animal's performance might be different from the random expectation if it had a site bias such that it only stored and retrieved food from a certain area in the experimental arena. Then, a calculation of random expectation based on all available storage sites is not appropriate, since the animal only uses some

smaller proportion of the total available. For this reason, experiments such as that conducted by Kamil and Balda (1985) are particularly well designed. In that experiment, Kamil and Balda allowed the animals (Clark's nutcrackers) to store only in sites selected by the experimenters (at random), but they had to retrieve from among all sites. In this way, any site bias will not enhance the animal's performance, and the calculations of random expectation presented in the present paper will be useful in demonstrating spatial memory.

REFERENCES

ABRAMOWITZ, M., & STEGUN, I. A. (1957). *Handbook of mathematical functions*. New York: Dover.

BALDA, R. P., & KAMIL, A. C. (1988). The spatial memory of Clark's nutcrackers (*Nucifraga columbiana*) in an analogue of the radial arm maze. *Animal Learning & Behavior, 16*, 116-122.

BALDA, R. P., & KAMIL, A. C. (1989). A comparative study of cache recovery by three corvid species. *Animal Behavior, 38*, 486-495.

BASU, D. (1958). On sampling with and without replacement. *Sankhyā, 20*, 287-294.

CHIKKAGOUDAR, M. S. (1966). A note on inverse sampling with equal probabilities. *Sankhyā, A28*, 93-96.

CONOVER, W. J. (1972). A Kolmogorov goodness-of-fit test for discontinuous distributions. *Journal of the American Statistical Association, 67*, 591-596.

CONOVER, W. J. (1980). *Practical nonparametric statistics*. New York: Wiley.

COX, D. R., & HINKLEY, D. V. (1982). *Theoretical statistics*. Chapman and Hall.

DEGROOT, M. H. (1970). *Optimal statistical decisions*. New York: McGraw-Hill.

EDWARDS, G. R. (1994). *The creation and maintenance of spatial heterogeneity in plant communities: The role of plant-animal interactions*. D. Phil. Thesis, University of Oxford.

FELLER, W. (1968). *An introduction to probability and its applications*. New York: Wiley.

GIBBONS, J. D. (1985). *Nonparametric statistical inference*. New York: Wiley.

GIBBONS, J. D. (1993). *Nonparametric statistics: Quantitative applications in the social sciences*. Newbury Park: Sage University Papers.

HEALY, S. D., & HURLY, T. A. (1995). Spatial memory in rufous hummingbirds (*Selasphorus rufus*): A field test. *Animal Learning & Behavior, 23*, 63-68.

HEALY, S. D., & KREBS, J. R. (1992). Comparing spatial memory in two species of tit: Recalling a single positive location. *Animal Learning & Behavior, 20*, 121-126.

HILTON, S. C., & KREBS, J. R. (1990). Spatial memory of four species of *Parus*: Performance in an open-field analogue of a radial maze. *Quarterly Journal of Experimental Psychology, 42B*, 345-368.

ILERSICH, T. J., MAZMANIAN, D. S., & ROBERTS, W. A. (1988). Foraging for covered and uncovered food on a radial maze. *Animal Learning & Behavior, 16*, 388-394.

JOHNSON, N. L., KOTZ, S., & KEMP, A. W. (1992). *Univariate discrete distributions*. New York: Wiley.

KAMIL, A. C., & BALDA, R. P. (1985). Cache recovery and spatial memory in Clark's Nutcrackers. *Journal of Experimental Psychology: Animal Behavior Processes, 11*, 95-111.

KONIJN, H. S. (1973). *Statistical theory of sample survey design and analysis*. Amsterdam: North-Holland.

LEHMANN, E. L. (1986). *Testing statistical hypotheses*. New York: Wiley.

OLSON, D. J., KAMIL, A. C., & BALDA, R. P. (1993). Effects of response strategy and retention interval on performance of Clark's Nutcrackers in a radial maze analogue. *Journal of Experimental Psychology: Animal Behavior Processes, 19*, 138-148.

OLTON, D. S., & SAMUELSON, R. J. (1976). Remembrance of places

passed: Spatial memory in rats. *Journal of Experimental Psychology: Animal Behavior Processes, 2,* 97-116.

PETTIT, A. N., & STEPHENS, M. A. (1977). A Kolmogorov-Smirnov goodness of fit statistic with discrete and grouped data. *Technometrics, 19,* 205-210.

RAJ, D., & KHAMIS, S. D. (1958). Some remarks on sampling with replacement. *Annals of Mathematical Statistics, 29,* 550-557.

ROBERTS, W. A., & VAN VELDHUIZEN, N. (1985). Spatial memory in pigeons on the radial maze. *Journal of Experimental Psychology: Animal Behavior Processes, 19,* 138-148.

SHERRY, D. F., & VACCARINO, A. L. (1989). Hippocampal aspiration dis-rupts cache recovery in black-capped chickadees. *Behavioral Neuroscience, 103,* 308-318.

SHETTLEWORTH, S. J., & KREBS, J. R. (1982). How marsh tits find their hoards: The role of site preference and spatial memory. *Journal of Experimental Psychology: Animal Behavior Processes, 11,* 241-260.

SPETCH, M. L., & EDWARDS, C. A. (1986). Spatial memory in pigeons (*Columba livia*) in an open-field feeding environment. *Journal of Comparative Psychology, 16,* 388-394.

SPETCH, M. L., & HONIG, W. K. (1988). Characteristics of pigeons' spatial working memory in an open-field task. *Animal Learning & Behavior, 16,* 123-131.

APPENDIX A
Sampling With Replacement

In Appendix A, we derive Equations 4, 5, 6, 10, 11, and 12. Let us start by stating the problem generally. Suppose that we are sampling from boxes. In some boxes, there is a reward; in others, there is not. Suppose that when a reward is sampled, it is removed from the box. Suppose further that we judge a sample to be a "success" if we receive a reward and a failure otherwise. We seek an expression for the distribution function of the number of successes resulting from sampling a fixed number of boxes. We also seek the expectation and the variance of this random variable.

Before deriving these expressions, we need to know the distribution of the number of *distinct* boxes explored. This is different from the number of boxes explored, in that we are not interested in boxes that have already been sampled at least once. More formally, we ask, What is the distribution of distinct boxes sampled in a simple random sample with replacement in a finite population? This is a classic problem that has been treated within the context of sampling theory by Basu (1958), Raj and Khamis (1958), Chikkagoudar (1966), and Konijn (1973, chap. IV). This problem was also solved within the context of the "classical occupancy problem" of "urn models" (see Johnson, Kotz, & Kemp, 1992, p. 418). We give here some results needed for further developments.

If Φ denotes the number of distinct boxes sampled by drawing n samples at random with replacement from a population of size N, then

$$\Pr[\Phi = r] = \frac{N!}{(N-r)! N^n} S_n^{(r)}, \quad r = 1, \ldots, \min(n, N), \tag{A1}$$

where $S_n^{(r)}$ is the Stirling number of the second kind.

$$E[\Phi] = N\left(1 - \frac{(N-1)^n}{N^n}\right). \tag{A2}$$

$$E[\Phi(N-\Phi)] = N(N-1)\left(\frac{(N-1)^n}{N^n} - \frac{(N-2)^n}{N^n}\right). \tag{A3}$$

$$\mathrm{Var}[\Phi] = \frac{(N-1)^n}{N^{n-1}} + (N-1)\frac{(N-2)^n}{N^{n-1}} - \frac{(N-1)^{2n}}{N^{2n-2}}. \tag{A4}$$

THEOREM 1

Suppose that we are sampling from boxes with replacement. In some boxes, there is a reward; in others, there is not. Suppose that when a reward is sampled, it is removed from the box. Suppose further that we judge a sample to be a success if we receive a reward and a failure otherwise. Let there be A boxes that contain rewards and B boxes that contain no rewards ($N = A + B$).

If ξ denotes the number of successes in n samples then,

(i)
$$\Pr[\xi = r] = \binom{A}{r}\frac{1}{N^n}\sum_{i=r}^{\min(n, B+r)} i!\binom{B}{i-r}S_n^{(i)}, \quad r = 0, \ldots, \min(n, A). \tag{A5}$$

(ii)
$$E[\xi] = A\left(1 - \frac{(N-1)^n}{N^n}\right). \tag{A6}$$

(iii)
$$\mathrm{Var}[\xi] = A\left(\frac{(N-1)^n + (A-1)(N-2)^n}{N^n} - A\frac{(N-1)^{2n}}{N^{2n}}\right). \tag{A7}$$

PROOF. (i) Let Φ be the number of distinct boxes sampled. The probability that we obtain r successes in n samples can be obtained from the hypergeometric distribution by conditioning on Φ:

$$\Pr[\xi = r \mid \Phi] = \frac{\binom{A}{r}\binom{B}{\Phi - r}}{\binom{N}{\Phi}}, \quad r = \max(0, \Phi - B), \dots, \min(\Phi, A),$$

and thus,

$$E[\xi \mid \Phi] = \Phi \frac{A}{N},$$

and

$$\mathrm{Var}[\xi \mid \Phi] = \Phi \frac{AB}{N^2} \frac{N - \Phi}{N - 1}.$$

We obtain the unconditional probability by

$$\Pr[\xi = r] = \sum_{i=r}^{\min(n, B+r)} \Pr[\xi = r \mid \Phi = i] \Pr[\Phi = i], \quad r = 0, \dots, \min(n, A).$$

However, from Equation A1 we know $\Pr[\Phi = i]$, so we get

$$\Pr[\xi = r] = \sum_{i=r}^{\min(n, B+r)} \frac{\binom{A}{r}\binom{B}{i-r}}{\binom{N}{i}} \frac{N!}{(N-i)! N^n} S_n^{(i)}$$

$$= \binom{A}{r} \frac{1}{N^n} \sum_{i=r}^{\min(n, B+r)} i! \binom{B}{i-r} S_n^{(i)}, \quad r = 0, \dots, \min(n, A).$$

(ii) We obtain the unconditional expectation from Equation A2:

$$E[\xi] = E\big[E[\xi \mid \Phi]\big]$$

$$= E\left[\Phi \frac{A}{N}\right]$$

$$= A\left(1 - \frac{(N-1)^n}{N^n}\right).$$

(iii) Finally, the unconditional variance can be obtained using Equations A3 and A4:

$$\mathrm{Var}[\xi] = E\big[\mathrm{Var}[\xi \mid \Phi]\big] + \mathrm{Var}\big[E[\xi \mid \Phi]\big]$$

$$= \frac{AB}{N^2} \frac{1}{N-1} E[\Phi(N - \Phi)] + \left(\frac{A}{N}\right)^2 \mathrm{Var}[\Phi]$$

$$= \frac{AB}{N} \frac{1}{N-1} N(N-1)\left(\frac{(N-1)^n - (N-2)^n}{N^n}\right)$$

$$+ \left(\frac{A}{N}\right)^2 \left(\frac{(N-1)^n + (N-1)(N-2)^n}{N^{n-1}} - \frac{(N-1)^{2n}}{N^{2n-2}}\right)$$

$$= A\left(\frac{(N-1)^n + (A-1)(N-2)^n}{N^n} - \frac{A(N-1)^{2n}}{N^{2n}}\right).$$

THEOREM 2

Suppose that we are sampling from boxes with replacement. In some boxes, there is a reward; in others, there is not. Suppose that when a reward is sampled, it is removed from the box. Suppose further that we judge a sample to be a success if we receive a reward and a failure otherwise. Let there be A boxes that contain rewards and B boxes that contain no rewards ($N = A + B$). If η denotes the number of samples needed to obtain exactly r success, then

(i) $$\Pr[\eta = r + k] = \binom{A}{r} \frac{r}{N^{r+k}} \sum_{i=0}^{\min(k, B)} (i + r - 1)! \binom{B}{i} S_{r+k-1}^{(i+r-1)}, \quad k = 0, 1, \dots \qquad (A8)$$

(ii)
$$\sum_{k=0}^{\infty}\binom{A}{r}\frac{r}{N^{r+k}}\sum_{i=0}^{\min(k,B)}(i+r-1)!\binom{B}{i}S_{r+k-1}^{(i+r-1)}=1.\tag{A9}$$

(iii)
$$E[\eta]=\sum_{j=0}^{r-1}\frac{N}{A-j}.\tag{A10}$$

(iv)
$$\text{Var}[\eta]=N\sum_{j=0}^{r-1}\frac{B+j}{(A-j)^2}.\tag{A11}$$

PROOF. (i) The probability of obtaining the rth success on the $n=r+k$ sample equals the probability that exactly k failures precede the rth success. This event occurs iff among the first $r+k-1$ samples there are exactly k failures and the $(r+k)$th sample results in success. Since the probability of obtaining k failures on the first $r+k-1$ samples is given by Equation A5, we can write

$$\Pr[\eta=r+k]=\binom{A}{r-1}\frac{1}{N^{r+k-1}}\sum_{i=r-1}^{\min(k,B)+r-1}i!\binom{B}{i-r+1}S_{r+k-1}^{(i)}\frac{A-(r-1)}{N},\quad k=0,1,\ldots,$$

$$=\binom{A}{r}\frac{r}{N^{r+k}}\sum_{i=0}^{\min(k,B)}(i+r-1)!\binom{B}{i}S_{r+k-1}^{(i+r-1)},\quad k=0,1,\ldots.$$

(ii) Next, we calculate

$$\sum_{k=0}^{\infty}\Pr[\eta=r+k]=\sum_{k=0}^{\infty}\binom{A}{r}\frac{r}{N^{r+k}}\sum_{i=0}^{\min(k,B)}(i+r-1)!\binom{B}{i}S_{r+k-1}^{(i+r-1)}$$

$$=\binom{A}{r}\frac{r}{N}\sum_{i=0}^{B}(i+r-1)!\binom{B}{i}\sum_{k=i}^{\infty}\frac{S_{r+k-1}^{(i+r-1)}}{N^{r+k-1}}$$

$$=\binom{A}{r}\frac{r}{N}\sum_{i=0}^{B}(i+r-1)!\binom{B}{i}\sum_{k=i+r-1}^{\infty}\frac{S_{k}^{(i+r-1)}}{N^{k}}.$$

Since

$$\sum_{k=a}^{\infty}\frac{1}{N^{k}}S_{k}^{(a)}=\frac{(N-a-1)!}{(N-1)!},\quad N>a,$$

we get

$$\sum_{k=0}^{\infty}\Pr[\eta=r+k]=\binom{A}{r}\frac{r}{N}\sum_{i=0}^{B}(i+r-1)!\binom{B}{i}\frac{(N-i-r)!}{(N-1)!}$$

$$=\sum_{i=0}^{B}\frac{\binom{A}{r}\binom{B}{i}}{\binom{N}{r+i}}\left(\frac{r}{r+i}\right)$$

$$=\sum_{i=0}^{B}\Pr[Y=r+i]=1,$$

where Y is defined as in Equation 7.

(iii) Equation A9 can be written

$$\sum_{k=0}^{\infty}\frac{r}{N^{r+k}}\sum_{i=0}^{\min(k,B)}(i+r-1)!\binom{B}{i}S_{r+k-1}^{(i+r-1)}=\frac{r!}{\prod_{j=0}^{r-1}(N-B-j)}.\tag{A12}$$

By differentiating Equation A12 with respect to N, we get

$$\sum_{k=0}^{\infty}-(r+k)\frac{r}{N^{r+k+1}}\sum_{i=0}^{\min(k,B)}(i+r-1)!\binom{B}{i}S_{r+k-1}^{(i+r-1)}=\frac{r!}{\prod_{j=0}^{r-1}(N-B-j)}\sum_{j=0}^{r-1}\frac{-1}{N-B-j}.$$

Thus,

$$E[\eta] = \binom{A}{r} \sum_{k=0}^{\infty} \frac{(r+k)r}{N^{r+k}} \sum_{i=0}^{\min(k,B)} (i+r-1)! \binom{B}{i} S_{r+k-1}^{(i+r-1)} = \sum_{j=0}^{r-1} \frac{N}{A-j}.$$

(iv) By taking the second derivative of Equation A12 with respect to N, we get

$$\sum_{k=0}^{\infty} \frac{(r+k)(r+k+1)r}{N^{r+k+2}} \sum_{i=0}^{\min(k,B)} (i+r-1)! \binom{B}{i} S_{r+k-1}^{(i+r-1)} = \frac{r!}{\prod_{j=0}^{r-1}(N-B-j)} \left\{ \left(\sum_{j=0}^{r-1} \frac{1}{N-B-j} \right)^2 + \left(\sum_{j=0}^{r-1} \frac{1}{(N-B-j)^2} \right) \right\}.$$

Thus,

$$E[\eta(\eta+1)] = \binom{A}{r} \sum_{k=0}^{\infty} \frac{(r+k)(r+k+1)r}{N^{r+k}} \sum_{i=0}^{\min(k,B)} (i+r-1)! \binom{B}{i} S_{r+k-1}^{(i+r-1)} = \left(\sum_{j=0}^{r-1} \frac{N}{A-j} \right)^2 + \sum_{j=0}^{r-1} \frac{N^2}{(A-j)^2}.$$

So, the variance is given by

$$Var[\eta] = E[\eta(\eta+1)] - E[\eta] - E[\eta]^2 = N \sum_{j=0}^{r-1} \frac{B+j}{(A-j)^2}.$$

APPENDIX B
Symbols Used Throughout the Paper

Symbol	Meaning
A	Number of storage sites that contain food
B	Number of storage sites that do not contain food
N	Total number of storage sites
n	Number of storage sites sampled
r	Number of storage sites sampled that contain food
v	Sample size (i.e., number of animals)
X, μ_X, σ_X^2	The random variable, expectation, and variance of the number of successes in n samples given sampling without replacement
$\xi, \mu_\xi, \sigma_\xi^2$	The random variable, expectation, and variance of the number of successes in n samples given sampling with replacement
Y, μ_Y, σ_Y^2	The random variable, expectation, and variance of the number of sites sampled until the rth success given sampling without replacement
$\eta, \mu_\eta, \sigma_\eta^2$	The random variable, expectation, and variance of the number of sites sampled until the rth success given sampling with replacement
$\overline{X}$	Sample mean (over the v animals) of the number of successes in the fixed-number-of trials protocol
$\overline{Y}$	Sample mean (over the v animals) of the number of trials in the fixed-number-of-successes protocol
$S_n^{(i)}$	Stirling number of the second kind
$F_0(x)$	Cumulative distribution function of the null hypothesis for the Kolmogorov-Smirnov test
T^-	Kolmogorov-Smirnov test statistic
M_0	Median of the distribution under the null hypothesis
Φ	Random variables representing the number of distinct sites visited in n samples given sampling with replacement
Ψ	Random variables representing the number of sites visited more than once in n samples given sampling with replacement
ζ	Random variable representing the number of trials needed to visit exactly r distinct sites given sampling with replacement
τ	Random variable representing the number of sites visited more than once that are needed to visit exactly r distinct sites given sampling with replacement
λ	Weighting factor for the compound model
W	Random variable representing the number of successes in n samples given that sampling is a mixture between sampling with replacement and sampling without replacement
$Z_\alpha, Z_{1-\alpha}$	Quantiles of a standard normal variable