Appendix

A.1 Conditionally Uniform Distributions. Consider a pair of spike processes, with the property that the spike trains are conditionally uniform and independent, conditioned on (N_1, N_2), for each neuron. In the setting of discrete spike trains, this means that, conditioned on (N_1, N_2), all spike trains consistent with (N_1, N_2) are equally likely. In the setting of continuous spike trains, this means that the conditional joint probability density function of $(\boldsymbol{t}_1, \boldsymbol{t}_2)$, conditioned on (N_1, N_2), depends only on (N_1, N_2).

Conditionally uniform processes include homogeneous Poisson processes as a special case, but the class is broader than Poisson. For example, in section 3, all our conclusions were reached with (non-Poisson) examples in which N_1 and/or N_2, specifying spike counts in an interval, are constants. (See Amarasingham et al., 2012, for a more thorough development of the motivation for using conditional uniformity as a null hypothesis for spike trains, and various generalizations.)

A.2 Validity of Interval Jitter Hypothesis Testing. If the interval jitter null hypothesis (conditional uniformity) is true, the subuniformity of $p(X, R)$ as defined in equation 2.1 is explained in Amarasingham et al. (2012). This conclusion is a consequence of the fact that under the null hypothesis, $S_0, S_1, \ldots, S_K$ are exchangeable (see Amarasingham et al., 2011).[1] From this, it follows that $p(X, R)$ is subuniform (proposition A.3 in Amarasingham et al., 2011). Since $S'_0, S'_1, \ldots, S'_K$ are also exchangeable, it follows by the same logic that $p_c(X, R)$ is subuniform. Moreover, $p_c(X, R)$ is absolutely continuous. Thus, under the null hypothesis, $p_c(X, R)$ is distributed as a uniform random variable.

To address potential confusion, it is worthwhile focusing on what the assumption of exchangeability is not. Obviously it is not equivalent to the

[1] A finite collection of random variables $Y_1, Y_2, \ldots, Y_n$ is exchangeable if its joint distribution is invariant to permutations of its arguments. That is, $\Pr((Y_1, \ldots, Y_n) \in A) = \Pr((Y_{\pi(1)}, \ldots, Y_{\pi(n)}) \in A)$ for all sets A and all permutations π of the index set $(1, \ldots, n)$.

assumption that $S_0, S_1, \ldots, S_K$ are independent and identically distributed. What is more relevant, it is also not equivalent to the assumption that $S_0, S_1, \ldots, S_K$ are identically distributed. Example 1 in section 3 is a clear expression of this distinction. Consider this example on the circle instead of an interval (see remark 1). Here $S_0, S_1, \ldots, S_K$ are not exchangeable, although they are identically distributed. This alone shows that it is not sufficent to establish that $S_0, S_1, \ldots, S_K$ are identically distributed (Pipa, Grün, & van Vreeswijk, 2013). Indeed, continuing to focus on continuous time and preserving the circle construction, $S_0, S_1, \ldots, S_K$ are identically distributed for any uniform spike process and any test statistic for both spike-centered and interval jitter.[2] (Interval and spike-centered jitter are "measure preserving.") Conceptually, this is likely the source of the intuitive conflation of the two methods. At root, the distinction is one of conditional versus unconditional inference (see the discussion of this distinction in Amarasingham et al., 2011).

A.3 Sensitivity. Returning to the example of section 2, a consequence of conservatism is that the spike-centered jitter procedure is less sensitive than the interval jitter procedure in detecting nonaccidentally synchronous events, when they are present. As a demonstration, consider the following numerical experiment. Generate a third independent homogeneous Poisson spike train with rate λ_s and superpose ("inject") spikes from the third train onto the spike trains from neurons 1 and 2. This is a model of injected synchrony. (The injected spikes are perturbed slightly prior to injection so that the synchronous spikes are not perfectly instantaneous.) Then compare the sensitivity of tests specified by the critical region $\{p_c(X, R) \leq \alpha\}$ generated by the spike-centered and interval jitter procedures, respectively. Consistent with intuition, the interval jitter test is more sensitive. For one example, we repeated the experiment of Figure 2, using $\alpha = 0.05$ and $\lambda_s = 2$ Hz. The rejection event $\{p_c(X, R) \leq \alpha\}$ occurred in 2% of the trials using the spike-centered jitter procedure and 8% of the trials using the interval jitter procedure across 50,000 trials. Other parameter settings generated similar numerical conclusions (results not shown; see https://github.com/aamarasingham/bjitter for code).

A.4 Exact Tests and Subuniform *p*-Values. The motivation for subuniform (as opposed to uniform) *p*-values in hypothesis testing can be understood through several routes.

A.4.1 An Elementary Example. As a general nontechnical example of nonuniformity, keep in mind the most familiar elementary examples in

[2] A more general observation is that under these conditions, (S_0, S_j) are pairwise exchangeable for any j in $\{1, 2, \ldots, K\}$, for both interval and spike-centered jitter.

which p-values are strictly subuniform as a consequence of discreteness. For example, consider $H_0 : X \sim \text{Binomial}(N, .5)$. A textbook p-value is $\Pr(Y \geq X|X)$, where Y and X are independent and identically distributed (i.i.d.), which computes the area to the right of the curve under the probability mass function. This p-value is subuniform but not uniform.

A.4.2 Nonrandomized Testing. More technically, in the standard general approach for defining p-values for nonrandomized (deterministic) tests (Lehmann & Romano, 2005), one begins with a nested family of hypothesis tests, associated with a continuum of significance values: for each significance level α, $0 \leq \alpha \leq 1$, one associates a critical region T_α, a subset of the sample space that satisfies $\Pr(X \in T_\alpha) \leq \alpha$. This is the standard definition of an exact hypothesis test. It is called nonrandomized if T_α is deterministic. A nested family of hypothesis tests is one in which the corresponding critical regions are nested:

$$T_\alpha \subset T_{\alpha'} \quad \text{if } \alpha < \alpha'. \tag{A.1}$$

In such a setting, the p-value is formally defined as

$$\hat{p}_{\{T_\alpha\}} = \hat{p}_{\{T_\alpha\}}(X) = \inf\{\alpha : X \in T_\alpha\}, \tag{A.2}$$

where we have used the subscript $\{T_\alpha\}$ to emphasize the dependence of the p-value on the choice of nested hypothesis testing family. In this case, $\hat{p}_{\{T_\alpha\}}$ is guaranteed to be subuniform (lemma 3.3.1 in Lehmann & Romano, 2005). Thus, a nested family of hypothesis tests determines the p-value distribution (and it is subuniform).

What is more, any subuniform random variable can be used to construct a nested family of hypothesis tests. Given a subuniform random variable $q(X)$ (see equation 3.1), define $T'_\alpha = \{x : q(x) \leq \alpha\}$. One can verify directly that the resulting hypothesis tests are exact and nested and, further, in this case

$$\hat{p}_{\{T'_\alpha\}}(X) = \inf\{\alpha : X \in T'_\alpha\} = q(X). \tag{A.3}$$

Thus, subuniform random variables and nested families of hypothesis tests are in exact correspondence (in nonrandomized testing). It follows, for example, that if a given random variable is not subuniform for some distribution, it cannot be a p-value for any hypothesis testing system that includes that distribution in the null hypothesis.

A.4.3 Randomized Testing. In our case, the function of interest is $p(X, R)$, where R (encoding the Monte Carlo surrogates) is a random variable that depends on X, so the construction of T'_α must be generalized.

The analogous generalization (Lehmann & Romano, 2005) $\{X : p(X, R) \leq \alpha \text{ with probability one}\}$ might be degenerate in our setting. A workaround, implicit here, is to incorporate the randomization mechanism into the sample space and then apply the nonrandomized testing framework. In our setting, consider the sample outcome to be (X, R). As established earlier, $p(X, R)$ is subuniform. Then define the critical region T'_α:

$$T'_\alpha = \{(x, r) | p(x, r) \leq \alpha\}. \tag{A.4}$$

The consequence is the same: the resulting tests are exact ($\Pr((X, R) \in T'_\alpha) \leq \alpha$) and the p-value is $p(X, R)$ (and subuniform), even though $p(X, R)$ will inherit the randomness of R (see the appendix to Amarasingham et al., 2012; Habiger & Pena, 2011, also develop this interpretation in greater generality). Thus, the p-value $p(X, R)$ is stochastic, but the associated tests are still exact.

A.4.4 Monte Carlo Approximation. For completeness, it is worth pointing out that in the discrete setting, a third interpretation of equation 2.1 is that $p(X, R)$ is a Monte Carlo approximation of the p-value $p^*(X)$, associated with the deterministic permutation test, a nonrandomized test (Harrison et al., 2015; Ernst, 2004):

$$p^*(X) = \frac{\sum_{\{X' : C_\Delta(X') = C_\Delta(X)\}} \mathbb{1}\{f(\boldsymbol{t}'_1, \boldsymbol{t}'_2) \geq S_0\}}{\#\{X' : C_\Delta(X') = C_\Delta(X)\}}, \tag{A.5}$$

which borrows from the fact that $p(X, R) \to p^*(X)$ as $K \to \infty$ (# represents cardinality in equation A.5). But this interpretation does not make clear, as it is above, that $p(X, R)$ is subuniform for all K, not simply in that aymptotic limit.

In summary, p-values for exact tests are subuniform, and a random variable that is subuniform (for H_0) can be used to construct exact tests for H_0, such that that subuniform random variable is the p-value. In that sense, subuniformity of a random variable (in H_0) is a necessary and sufficient condition for exact hypothesis tests of H_0. Note, finally, that if a random variable is subuniform and (absolutely) continuous, then it must be uniform. This motivates the trick for converting subuniform p-values to uniform p-values in sections A.2 and 5.

A.5 p-Value Transformation. The standard elementary approach for constructing p-values for a random variable X when the null hypothesis or conditional null hypothesis is a unique (and computable) distribution is as follows. Construct a random variable Y that has the same distribution as X and is also independent. Then let $p(X) = \Pr(Y \geq X | X)$ (this is an application of the definition of a p-value, equation A.2, in this setting). It is

readily apparent that $p(X)$ is subuniform (see equation 3.1; see also lemma 3.3.1 in Lehmann & Romano, 2005). The corresponding tests have power for alternative distributions that tend to be "greater" than that of X. However, $p(X)$ will not be strictly uniform, if X is discrete.

Here is an intuitive construction for generating strictly uniform p-values, which involves randomizing the data. For simplicity, consider the case that X is integer-valued. Construct a new random variable $X_c = X + \epsilon_X$, where ϵ_X is independent of X and uniformly distributed on $[-1/2, 1/2]$. Analogous to the above, suppose X_c and Y_c are independent and identically distributed. Compute $p'(X) = \Pr(Y_c \geq X_c | X_c)$. Then $p'(X)$ will be subuniform and absolutely continuous. Thus, $p'(X)$ will be uniform.

To compute $p'(X)$, note that

$$\begin{aligned} p'(X) &= \Pr(Y_c \geq X_c | X_c) \\ &= [(X + 1/2) - X_c] \Pr(Y = X | X) + \Pr(Y > X | X). \end{aligned} \tag{A.6}$$

Since $[(X + 1/2) - X_c]$ is uniformly distributed on [0,1], this is equivalent to drawing a sample U independently from a uniform distribution on [0,1] and computing

$$p'(X) = \Pr(Y_c \geq X_c | X_c) = U \cdot \Pr(Y = X | X) + \Pr(Y > X | X). \tag{A.7}$$

Since, by construction, $Y > X \iff Y_c > X_c$, it follows that tests constructed from $p(X)$ and $p'(X)$ will behave identically except for those outcomes exactly at the critical threshold. There is no loss of generality if X is not an integer, or if $1/2$ above is replaced with a different constant b. The main ingredient is only that b and the bounds on ϵ_X are chosen so as to preserve $Y > X \iff Y_c > X_c$.