Appendix: Algorithm for Finding Pause Onsets and Offsets

Here, we outline a Bayesian algorithm for the detection of pause onsets and offsets within a trial. Being Bayesian, the algorithm combines prior knowledge with new data to gradually adjust an estimate of some quantity of interest, here the instantaneous probability of the cell's firing, p_s. Prior knowledge is acquired from statistics measured *across* all trials within the pre-CS, CS-US and post-US periods. The essence of the algorithm is to use our prior distributions to build two different models of p_s at each time and then combine their estimates into a model of the location of a pause offset or onset.

Having taken into account data from all trials, the priors bias a statistical model to favor the overall trend of spiking within a given interval. This is particularly important for the estimate of the firing rate during the pauses because there is often no or very little data (no spikes) during a pause. Improving the estimates of the firing rates improves the detection and localization of the onsets and offsets of the pauses.

There is, however, no hard prior on the direction of the change. If the spike rate during the CS-US sub-interval is substantially higher than that during the pre-CS sub-interval, as happens on some trials, the change-finding algorithm will indicate a change when the increase in the firing rate appears to occur, even though what is expected is a decrease. This reflects a fundamental fact about Bayesian calculations: data clearly favoring one conclusion (e.g., an increase in the firing rate) will override priors that bias toward the opposite conclusion (a decrease in the firing rate), provided the prior is not a hard prior. A hard prior assigns 0 probability to possibilities that are (taken to be) known a priori to be 0. We do not know a priori that there must be a reduction in the firing rate during the CS-US interval on any given trial, because we sometimes observe the opposite.

The algorithm first discretizes trial time into 1-ms-wide bins referenced to 0 at CS onset. The peri-stimulus intervals across all trials are then divided into three sub-intervals each with duration T equal to the CS-US interval: a pre-onset interval terminating at CS onset ($t = -T, \ldots, 0$), a CS-US interval ($t = 0, \ldots, T$), and a post-US interval ($t = T, \ldots, 2T$), which begins when a US would have occurred in a training trial. The algorithm will seek changes in spiking rate within two larger intervals containing the putative change-points: the "onset" interval ($t = -T, \ldots, 0.7T$) and "offset" interval ($t = 0.7T, \ldots, 2T$). For concision, we will only describe change-point detection in the onset interval, and it is assumed hereafter that all time indices come from this interval. The procedure is identical for the offset interval.

The algorithm converts vectors of recorded spike times in each trial to binary vectors $x(t) \in \{0, 1\}$ such that $x(t) = 1$ when a spike occurs in the bin centered at t and $x(t) = 0$ otherwise. We will model this spike train as a sequence of independent Bernoulli random variables with parameter p_s assumed to change

at most once, so that, away from a change-point, spikes approximate a Poisson process of constant rate. We assume the appearance of a change-point is itself a Bernoulli random variable modeled by a prior distribution with parameter p_c. Should a change occur, the location L of a change-point is therefore also a random variable. We identify the event in which there is no change in p_s with the event $L = 0.7T$ (the change-point happens after the interval in consideration).

Two Beta prior distributions are then constructed for p_s by using the spike counts from the pre-CS and CS-US intervals across all trials under analysis (in the case of offset detection, priors come from spike counts in the CS-US and post-US intervals). These Beta priors have parameters $(\alpha_{\text{pre-CS}}, \beta_{\text{pre-CS}})$, and $(\alpha_{\text{CS-US}}, \beta_{\text{CS-US}})$, respectively. Recall that these Beta distributions have probability densities proportional to

$$p_s^{\alpha_I}(1-p_s)^{\beta_I}, \tag{1}$$

where $I = \text{pre-CS}, \text{CS-US}$. These distributions represent our prior knowledge about p_s as derived from gross, cross-trial statistics about spike number. For example, the CS-US prior puts more weight on lower p_s, since it was built from data during an interval when–overall–spike counts were lower. A posterior distribution on p_s refines this prior knowledge with new data available at time t by multiplying the prior distribution with a likelihood function.

It is a convenient property of a Bernoulli random variable with a Beta prior that the posterior on its parameter is completely determined by cumulative sample counts (since the Beta prior is a "conjugate" prior in that it shares a mathematical form with the posterior) which we collect into the quantities,

$$\begin{aligned} S_B(t) &= \sum_{t' \le t} x(t') & F_B(t) &= t - S_B(t) \\ S_A(t) &= \sum_{t' > t} x(t') & F_A(t) &= 1.7T - t - S_A(t), \end{aligned}$$

so that $S_B(t)$ and $S_A(t)$ count the number of spikes before and after t, respectively, and $F_B(t)$ and $F_A(t)$ count the non-spikes. Then, two posterior models on p_s at each t are constructed using the pre-CS and CS-US priors:

$$\begin{aligned} \mathbb{P}_B(p_s|\{x(t')\}_{t' \le t}) &\propto p_s^{\alpha_{\text{pre-CS}}+S_B(t)}(1-p_s)^{\beta_{\text{pre-CS}}+F_B(t)} \\ &= \mathbb{P}_B(\{x(t')\}_{t' \le t}, p_s),, \end{aligned} \tag{2}$$

and

$$\begin{aligned} \mathbb{P}_A(p_s|\{x(t')\}_{t' > t}) &\propto p_s^{\alpha_{\text{CS-US}}+S_A(t)}(1-p_s)^{\beta_{\text{CS-US}}+F_A(t)} \\ &= \mathbb{P}_A(\{x(t')\}_{t' > t}, p_s). \end{aligned} \tag{3}$$

Note that Eqs. 2 and 3 differ from Eq. 1 only in their exponents, owing to the Beta prior's conjugacy. It is helpful to consider these equations as the estimate on p_s of a Bayesian observer. Eq. 2 represents the estimate on p_s at t for

an observer acquiring data from before t using prior knowledge from the cross-trial pre-CS period. Eq. 3, on the other hand, represents the estimate of an alternative observer acquiring data from after t using prior knowledge from the cross-trial CS-US interval.

We have expressed these posteriors before and after each t as proportional to the joint density on spikes and p_s, but we can obtain the marginal likelihoods on spikes alone by integrating away p_s:

$$\begin{aligned}\mathbb{P}_B(\{x(t')\}_{t'\leq t}) &= \int_0^1 \mathbb{P}_B(\{x(t')\}_{t'\leq t}, p_s)\ dp_s \\ &= \text{Beta}(\alpha_{\text{pre-CS}} + S_B(t), \beta_{\text{pre-CS}} + F_B(t)),\end{aligned} \tag{4}$$

$$\begin{aligned}\mathbb{P}_A(\{x(t')\}_{t'>t}) &= \int_0^1 \mathbb{P}_A(\{x(t')\}_{t'>t}, p_s)\ dp_s \\ &= \text{Beta}(\alpha_{\text{CS-US}} + S_A(t), \beta_{\text{CS-US}} + F_A(t)),\end{aligned} \tag{5}$$

where $\text{Beta}(\alpha, \beta) = \int_0^1 x^\alpha (1-x)^\beta\ dx$ is the beta function. Eqs. 4 and 5 represent the likelihood of the data before (after) t expressed as an average over estimates of the true parameter p_s according to observers with prior knowledge acquired from before (after) the true change.

We are now prepared to state our likelihood model for the whole interval conditioned on change-point location, which take to be the product of Eqs. 4 and 5:

$$\mathbb{P}_C(x|L=t) = \mathbb{P}_B(\{x(t')\}_{t'\leq t})\mathbb{P}_A(\{x(t')\}_{t'>t}). \tag{6}$$

We emphasize that $\mathbb{P}_c$ is a function of t. Here, t controls where we place a change-point which divides the data into two sections respectively modeled by the "before" and "after" observers. Intuitively, the maximum of this likelihood as a function of t should occur at the veridical change point, $t = 0$.

The algorithm returns several quantities based on $\mathbb{P}_C$. First, it returns the maximum likelihood estimate and expected value of t under the model specified by Eq. 6:

$$t^* = \arg\max_t \mathbb{P}_C(x|L=t) \tag{7}$$

$$\mathbb{E}_{\mathbb{P}_C}[t] = \sum_t t\ \mathbb{P}_C(x|L=t). \tag{8}$$

The second quantity is the expected location of the change-point, inasmuch as it is a weighted average of the bin indices in which large weight is given to those t which explain the data according to Eq. 6. Second, the algorithm returns the posterior odds on the existence of a change-point somewhere in the sequence. Accordingly, a likelihood ratio of a model assuming a change-point somewhere

in the sequence and a model assuming no change-point is calculated:

$$\begin{aligned} R_{\text{likelihood}} &= \frac{\sum_t \mathbb{P}_{\mathbb{C}}(x, L=t)}{\mathbb{P}_C(x|L=0.7T)} \\ &= \frac{\sum_t \mathbb{P}_{\mathbb{C}}(x|L=t)\mathbb{P}(L=t)}{\mathbb{P}_C(x|L=0.7T)} \\ &= \frac{\frac{1}{T}\sum_t \mathbb{P}_{\mathbb{C}}(x|L=t)}{\mathbb{P}_C(x|L=0.7T)}. \end{aligned} \tag{9}$$

In the second line, we took advantage of the fact that a joint distribution is a conditional times a prior. In the third line, we assume that L has a uniform prior. The denominator of the ratio is the likelihood of the the data assuming a change-point after $0.7T$ (that is, not in the sequence). The prior odds on the presence of exactly one change-point in the sequence are then computed, using the fact the number of change-points is binomially distributed with parameter p_c:

$$\begin{aligned} R_{\text{prior}} &= \frac{\binom{1.7T}{1} p_c^1 (1-p_c)^{n-1}}{\binom{1.7T}{0} p_c^0 (1-p_c)^n} \\ &= \frac{1.7T p_c}{(1-p_c)} \end{aligned} \tag{10}$$

The posterior odds on the presence of a change-point are thus

$$R_{\text{posterior}} = R_{\text{likelihood}} \times R_{\text{prior}}. \tag{11}$$

The algorithm also delivers the raw estimates of the firing rates in the pre-, during-, and post-CS-US sub-intervals. These raw estimates are simply the number of spikes within each sub-interval divided by the duration of the sub-interval. It also delivers the duration of the longest inter-spike interval between its estimates of pause onset and pause offset. We use this as a measure of the depth of the pause; the longer this interval, the deeper the pause. Lastly, it delivers the latency between its estimate of pause onset (or CS onset, whichever is greater) and the onset of the longest inter-spike interval. We take the inverse of this latency as a measure of the abruptness of pause onset; the shorter this latency, the more abruptly the firing rate drops to its lowest level. When the latency is 0, the inverse is infinite, which is to say that the onset is a step drop to the deepest level.

We refer to putative onsets and offsets because we may reject some of them in the light of the other statistics returned by the algorithm. If, for example, the raw rate of firing during the CS-US sub-interval is greater than raw firing rates during the flanking pre- and post- intervals, then we may decide not to include those trials when computing the distributions of pause parameters. We may also decide not to include a trial if the weight of the evidence for a pause onset and/or a pause offset favors the hypothesis that there was no change at the putative onset and/or at the putative offset. This post-algorithmic data

exclusion capability enables us to study the effects on the distributions of pause statistics of either including or excluding trials where a rise rather than a fall in firing rate was observed during the time when a pause was expected and trials where the evidence for a pause is weak or non-existent.