**Estimation of statistical power for modulation of neuronal firing rates in the basal forebrain by reward expectation**

*Panna Hegedüs and Balázs Hangya*

‘Lendület’ Laboratory of Systems Neuroscience, Institute of Experimental Medicine, Hungarian Academy of Sciences

In an ideal experiment, during the planning phase the experimenters establish the sample size to be collected that is optimally suited to answer the research questions addressed. This is important to avoid statistical pitfalls, like collecting data until the effect reaches statistical significance, which policy is prone to detect random fluctuations in the empirical statistics of the data as false positives.

This is best done by estimating statistical power, that is, by determining the sample size with which a desired effect is statistically detected with a sufficiently high probability (Button et al., 2013; Thomas and Juanes, 1996). Setting a threshold for statistical power is arbitrary, but a power of 0.8 is usually accepted in biological experiments. The problem is that in neurophysiology a reasonable power estimation is most of the time prevented by a lack of clear expectations about the ‘effect sizes’ or the ‘effects’ themselves.

Time from reinforcement (s)

Lick rate (Hz)

Time from predictive cue (s)



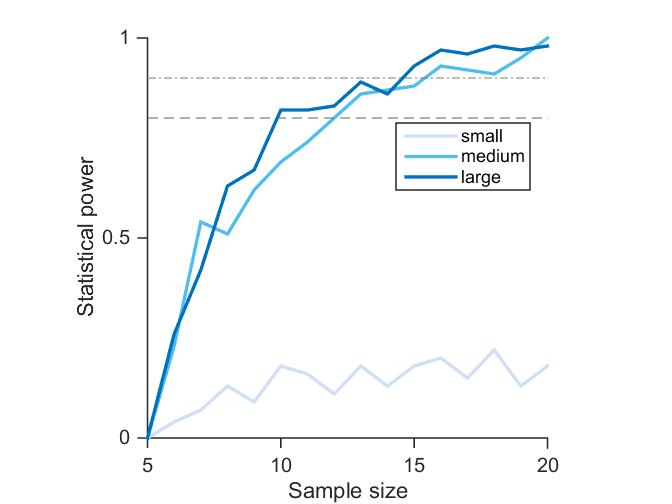
**Figure 1. Head-fixed auditory Pavlovian task.** A brief flash of light indicates the start of each trial. After a variable delay period, one of two different auditory cues is presented. The cues are either indicating likely reward or likely punishment. The difference in anticipatory licking indicates that the animal has learned the contingencies of the task.

Nevertheless, we attempted to perform a power analysis for recording basal forebrain neurons in Pavlovian conditioning based on the results of our previous paper on cholinergic activity in an auditory detection task (Hangya et al., 2015).

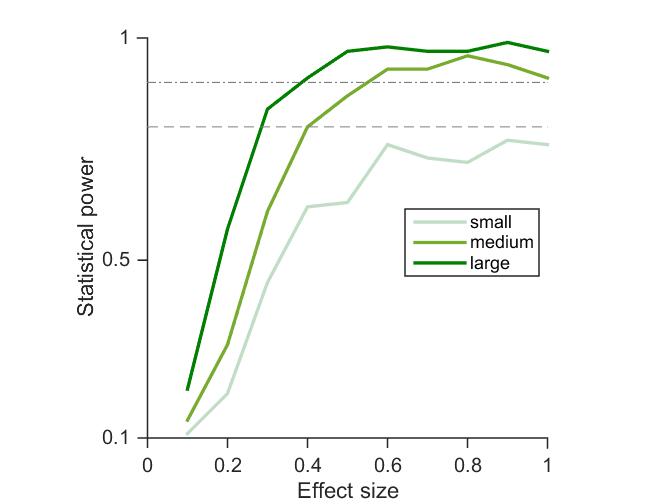
Briefly, head-fixed mice are trained on a head-fixed auditory Pavlovian task, in which two cue tones predict either likely reward or likely punishment. The animals perform differential anticipatory licking for the cues, demonstrating that they have learned the contingencies of the task (Fig.1). Identified cholinergic or PV-expressing GABAergic neurons are recorded from the HDB nucleus (horizontal limb of the diagonal band of Broca) of the basal forebrain.

Based on our previous paper we expect that neuronal activity in the HDB might be modulated by reward expectations. We estimated the expected effect size based on the extent of modulation of HDB cholinergic activity by reinforcement surprise after water reward in an operant learning task (Fig. 6F in (Hangya et al., 2015)). Normalized firing rate averaged for the two faint and two loud tones showed a 50% increase (0.61 to 0.89). Therefore, we ran our simulations at 10%, 50% and 100% effect sizes as ‘small’, ‘medium’ and ‘large’ effect. As reward often elicited a single spike per trial in HDB cholinergic neurons with short latency (27 ms), small jitter (11 ms) and variable reliability depending on the signal-to-noise ratio of the reward predicting cue (average, 0.45), we modeled the expected effect using these parameters on top of a 5 Hz baseline Poisson firing (Fig.1G in (Hangya et al., 2015)). We estimated that 60% of HDB neurons may show a modulation and assumed no modulation in the remaining 40%; nevertheless this estimation probably represents a lower bound, since modulation of firing rate by surprise or reward expectation may not reach statistical significance in all recorded neurons. To be on the conservative side, we modeled a relatively short session with only 100 rewarded trials. Statistical significance was investigated by performing an ROC analysis in a 50 ms window and tested with Wilcoxon sign-rank test at alpha = 0.05. A hundred experiments were simulated to provide an estimate of statistical power, that is, the proportion of simulated experiments with statistically significant effect.

We found that at medium effects, a sample size of 12 was sufficient to reach 80% statistical power and a sample size of 16 reliably achieved 90% power. A small 10% effect present in 60% of the neurons remained largely undetected up to a sample size of 20, which was the maximal tested (Fig.2). Next, we tested statistical power as a function of effect size for small (n=10), medium (n=15) and large (n=20) samples (Fig.3.). The medium sized sample reached 0.8 power at an effect size of 40% in 6 out of 10 neurons.



**Figure 2. Statistical power as a function of sample size.** Small effect: 10% firing rate increase in 60% of the neurons. Medium effect: 50% firing rate increase in 60% of the neurons, corresponding to the expected effect size. Large effect: 100% firing rate increase in 60% of the neurons. Lines, 80 and 90% statistical power.



**Figure 3. Statistical power as a function of effect size.** Statistical power as a function of expected effect size when the number of neurons with a detectable effect was fixed at 60%. The function is plotted for sample sizes of 10 (‘small’), 15 (‘medium’) and 20 (‘large’).

In summary, a sample size of 12 neurons was sufficient to reach 0.8 statistical power even when the effect was only present in 60% of all neurons and trial number for individual cells was kept low.

**References**

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., et al. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nat. Rev. Neurosci.* 14, 365–376. doi:10.1038/nrn3475.

Hangya, B., Ranade, S. P., Lorenc, M., and Kepecs, A. (2015). Central Cholinergic Neurons Are Rapidly Recruited by Reinforcement Feedback. *Cell* 162, 1155–1168. doi:10.1016/j.cell.2015.07.057.

Thomas, L., and Juanes, F. (1996). The importance of statistical power analysis: An example from Animal Behaviour. *Anim. Behav.* 52, 856–859. doi:10.1006/anbe.1996.0232.