Let's first consider the case where one makes 3 attempts and gets a payoff of 0 each time. What can we say about the outcome of our next 3 attempts? And in the case where one makes 15 attempts and gets a payoff of 0 each time, what can we then say about the outcome of our next 15 attempts? Our estimated expected payoff is 0 in both cases, but in which case are we most confident about that expectation?

In the most general form, we can use Hoeffding's inequality for a distribution-free example of constructing confidence bands around our estimate. In this case, because 0 is a natural lower limit for the games we consider, we are only considering an upper-confidence bound.

$$p\left(\mathbb{E}[x] \geq \epsilon\right) = e^{-2n\epsilon^2} = \delta$$

Solving for ϵ in terms of δ gives us our upper confidence bound for any level of confidence, δ we choose:

$$\epsilon = \sqrt{\frac{2}{n} \log\left(\frac{1}{\delta}\right)}$$

The expected payoff, y, of n trials for any of the possible true means:

$$\mathbb{E}[y] = n\mathbb{E}[x] = n\epsilon$$

Which is an increasing function in n:

$$n\sqrt{\frac{2}{n} \log\left(\frac{1}{\delta}\right)} = \sqrt{2n \log\left(\frac{1}{\delta}\right)}$$

Therefore, while you learn *more* about your expected payoff from a single attempt if you play the bandit more times, you learn *less* about the your expected payoff in the next n rounds after playing the bandit n times, as n increases.