Kershenbaum, A., T.M. Freeberg, and D.E. Gammon. 2015. Estimating vocal repertoire size is like collecting coupons: A theoretical framework with heterogeneity in signal abundance. *Journal of Theoretical Biology* 373:1-11.

Vocal repertoire size is an important behavioural measure in songbirds and mammals with complex vocal communication systems, and has traditionally been used as an indicator of individual fitness, cognitive ability, and social structure. Estimates of asymptotic repertoire size have typically been made using curve fitting techniques. However, the exponential model usually applied in these techniques has never been provided with a theoretical justification based on probability theory, and the model has led to inaccurate estimates. We derived the precise expression for the expected number of distinct signal types observed for a fixed sampling effort: a variation of what is known in the statistical literature as the “Coupon Collector's problem”. We used empirical data from three species (northern mockingbird, Carolina chickadee, and rock hyrax) to assess the performance of the Coupon Collector model compared to commonly used techniques, such as exponential fitting and repertoire enumeration, and also tested the different models against simulated artificial datasets with the statistical properties of the empirical data. We found that when signal probabilities are dissimilar, the Coupon Collector model provides far more accurate estimates of repertoire size than traditional techniques. Enumeration and exponential curve fitting greatly underestimated repertoire size, despite appearing to have reached saturation. Application of the Coupon Collector model can generate more accurate estimates of repertoire size than the commonly used exponential model of repertoire discovery, and could go a long way towards re-establishing repertoire size as a useful indicator in animal communication research.