

Appendix S1

R-code for estimating the rate of mis-identification in a two-species, three-observer setting.

The data is summarized into 8 numbers $n_{i,j,k}$ = the number of oxpeckers identified as species i by observer 1, species j by observer 2, and species k by observer 3, with i, j , and k taking value 1 or 2. These values are manually entered into the R environment by typing `n111 = ...`, and so forth.

In practice, the same observer performed the identification challenge twice (with a two week interval). We applied the same observer mis-identification probability, allowing us to spare some parameters relative to if we had three different observers.

The negative log-likelihood function was then:

```
nll <- function(theta) {
  p1l = plogis(theta[1]) ## proportion of species 1
  p1 = plogis(theta[2])  ## probability of correct identification species 1 by observer 1
  p2 = plogis(theta[3])  ## probability of correct identification species 2 by observer 1
  q1 = plogis(theta[4])  ## probability of correct identification species 1 by observer 2
  q2 = plogis(theta[5])  ## probability of correct identification species 2 by observer 2
  P111 = p1l*p1*p1*q1 + (1-p1l)*(1-p2)*(1-p2)*(1-q2)
  ## proba obs 1 says species 1 twice, obs 2 also says species 1
  P112 = p1l*p1*p1*(1-q1) + (1-p1l)*(1-p2)*(1-p2)*q2
  P121 = p1l*p1*(1-p1)*q1 + (1-p1l)*(1-p2)*p2*(1-q2)
  P122 = p1l*p1*(1-p1)*(1-q1) + (1-p1l)*(1-p2)*p2*q2
  P221 = p1l*(1-p1)*(1-p1)*q1 + (1-p1l)*p2*p2*(1-q2)
  P222 = p1l*(1-p1)*(1-p1)*(1-q1) + (1-p1l)*p2*p2*q2
  P211 = p1l*(1-p1)*p1*q1 + (1-p1l)*p2*(1-p2)*(1-q2)
  P212 = p1l*(1-p1)*p1*(1-q1) + (1-p1l)*p2*(1-p2)*q2

  lik = n111*log(P111) + n121*log(P121) + n211*log(P211) + n221*log(P221) +
        n112*log(P112) + n122*log(P122) + n212*log(P212) + n222*log(P222)
  return(-(lik))
}

s=nlm(nll,rnorm(5,2,1),hessian=TRUE) ## minimize starting from a random initial value
```

We minimized the negative log-likelihood using the `nlm` function in R with default options, and back-transformed the estimates from the logit to the natural scale.

In practice, after a first round of identification, the estimated rates of mis-identification were not satisfactory (>5%). Thus GP conducted a second round of identification, taking into account the outcome of the previous round and focusing in particular on cases with disagreements between the three previous passages.