

## ELECTRONIC SUPPLEMENTARY MATERIAL 2: Model Details

for

How New Caledonian crows solve novel foraging problems and what it means for  
cumulative culture

Corina J. Logan\*, Alexis Breen, Alex H. Taylor, Russell D. Gray, William J. E. Hoppitt

\*cl417@cam.ac.uk

### ***B1. Full model specification: first interactions***

We first fit models with continuous variables representing social effects. The corresponding full model specifies the rate of first attempted method  $l$  at locus  $k$  for individual  $i$  in group  $j$  at time  $t$  as:

$$\lambda_{ijkl}(t) = \lambda_{0,j}(t) \exp \left( O_{kl} + \varphi_{ij} + \beta_{LS} LS_{ijk}(t) + \beta_{LG} LG_{ij}(t) + \beta_{CS} CS_{ijkl}(t) \right. \\ \left. + \beta_{AS} AS_{ijkl}(t) \right) \left( 1 - z_{ijkl}(t) \right)$$

where  $\lambda_{0,j}(t)$  is an unspecified baseline function assumed to be the same for all of group  $j$  across all options;  $O_{kl}$  is a parameter allowing for differences in difficulty between the four options, with  $O_{11} = 0$  set as baseline;  $\varphi_{ij}$  is a linear predictor containing variables that vary between individuals (sex, age, and rank) and an individual random effect;  $LS_{ijk}(t)$  takes the value 1 if  $i$  interacted with locus  $k$  previously using either method prior to  $t$  (LS signifies locus specific asocial learning) and is 0 otherwise;  $LG_{ij}(t)$  takes the value 1 if  $i$  has interacted with the task at either locus using either method (LG signifies location general asocial learning) prior to time  $t$  and is 0 otherwise;  $\beta_X$  are fitted

parameters each giving the effect of a variable  $X$ ;  $z_{ijkl}(t)$  takes the value 1 if  $i$  has previously interacted with locus  $k$  using method  $l$ , or if  $i$  was a seeded demonstrator for that option, and is 0 otherwise: the  $(1 - z_{ijkl}(t))$  thus ensures that the model only models the rate of first interaction using each option. The remaining terms model social influences on learning, which we now define. The action specific variable is given by:

$$AS_{ijkl}(t) = o_{ijkl}(t),$$

where  $o_{ijkl}(t)$  is the number of observations of others interacting with the task at locus  $k$  using method  $l$ . The context specific variable is then observations summed across the relevant locus:

$$CS_{ijkl}(t) = \sum_{k=1}^2 o_{ijkl}(t)$$

We fitted models with every combination of variables given in the full model using the `coxme` (Therneau 2012) and `MuMIn` (Bartoń 2014) packages, calculating the  $AIC_c$  in each case, using the integrated likelihood and taking the number of events as the sample size. This enabled us to obtain Akaike weights for each model and thus obtain the total support for each variable (Burnham & Anderson 2002).

We then re-fitted the models with binary variables for the social effects (see main text).

$AS_{ijkl}(t)$  was replaced with  $\acute{A}S_{ijkl}(t)$ , where  $\acute{A}S_{ijkl}(t) = 1$  if  $AS_{ijkl}(t) > 0$  and is 0

otherwise, with corresponding binary variable  $\mathcal{C}S_{ijkl}(t)$  replacing  $CS_{ijkl}(t)$ . The support was found to be 39.8x greater for the set of models with binary social effects, so we report the results of the binary models in the main text

In the main text we report the support (total Akaike weight) for each variable, along with its model-averaged estimate and unconditional standard error. We also report the back-transformed (exponential) estimate and 95% unconditional confidence intervals, which are back-transformed Wald confidence intervals calculated from the unconditional standard error, and thus allow for model selection uncertainty.

## ***B2. Ruling out small-scale local enhancement***

The CS effect found in the main analysis could be explained by a small-scale local enhancement effect whereby observation of an interaction at locus  $k$  on a specific apparatus would attract observers to that location. In contrast, stimulus enhancement and observational conditioning predict that the effect would generalise between the corresponding loci on the two different task apparatuses whereas small-scale local enhancement does not. Under local enhancement, each crow's first interaction using each locus would tend to be at the same apparatus at which they had observed another crow interacting with that locus.

We assessed whether each crow's choice of apparatus for its first interaction with each locus was influenced by the apparatuses at which it had seen other crows interacting, if any. We used a generalised linear mixed model (GLMM) with a binary response variable

(1 = right apparatus; 0 = left apparatus) and logit link function, with individual as a random effect. The two binary predictor variables were a) whether the subject had previously seen another crow interacting with the right apparatus and b) whether the subject had previously seen another crow interacting with the left apparatus. If local enhancement was responsible for the CS effect, we expect a) to be positive and b) to be negative.

We found that this was not the case (see Figure 3 in the main text): in the absence of observation there was an underlying tendency to prefer the left box (probability of choosing left = 0.90; 95% U.C.I. = 0.56 - 0.99), with some evidence that the odds of using the right box were increased by observations at both the right box (support = 66%; estimate = 4.96x; 95% U.C.I. = 0.90 - 27.5) and the left box (support = 50%; estimate = 5.51x; 95% U.C.I. = 0.47 - 65.1). The pattern could be explained by proximity: the left box was located nearer to the perches, which is where crows approached the table from. The key point is that we would expect a clear bias in favor of the apparatus observed if local enhancement was responsible for the CS effect observed, and this pattern is not shown.

### ***B3. Full model specification: ‘discovered’ to ‘solved’ transition***

If imitation or emulation were operating, crows would, through observation, be learning something about *how* to interact with the task to obtain food, so we might expect observation to impact on how quickly the crows obtain food using a specific option. The model described in section B1 models *i*'s rate of transition from a naïve state for method *l*

at locus  $k$  to having ‘discovered’ that option: i.e. a state in which  $i$  is attempting to extract food from the task at locus  $k$  using method  $l$ . We can further model the rate at which crows transition from the ‘discovered’ state to a ‘solved’ state, in which they have successfully extracted food from the task at locus  $k$  using method  $l$ , analogous to the model developed by Atton et al. (2012). We modeled  $\lambda_{S,ijkl}(t)$  the rate of solving at locus  $k$  using method  $l$  by individual  $i$  in group  $j$  as:

$$\lambda_{S,ijkl}(t) = \lambda_{0S,j}(t) \exp \left( O_{kl} + \varphi_{ij} + \beta_{LS} LS_{ijk}(t) + \beta_{LG} LG_{ij}(t) + \beta_{CS} CS_{ijkl}(t) \right. \\ \left. + \beta_{AS} AS_{ijkl}(t) \right) z_{ijkl}(t) \left( 1 - y_{ijkl}(t) \right)$$

where most terms are defined the same as for the model of first interaction, above, except  $\lambda_{0S,j}(t)$  is an unspecified baseline rate function for solving for group  $j$ , and  $y_{ijkl}(t) = 1$  when  $i$  has solved the task at locus  $k$  using method  $l$  prior to time  $t$ , or if  $i$  is a seeded demonstrator for that option, and is 0 otherwise. Therefore, the  $z_{ijkl}(t) \left( 1 - y_{ijkl}(t) \right) = 1$  if  $i$  has interacted with the task at locus  $k$  using method  $l$  prior to time  $t$ , but not solved the task at locus  $k$  using method  $l$  prior to time  $t$ , and is 0 otherwise. This term ensures that we are modeling the rate of transition from the ‘discovered’ to the ‘solved’ state. We used the same procedure as described in section B1 to fit and compare models, again considering binary versions of each social effect.

#### ***B4. Relationship to existing statistical models for analyzing diffusion data***

Here we provide information about how the modeling approach described in sections B1 and B3 relates to those models previously used in diffusion studies of social learning in non-human animals, primarily to aid other social learning researchers wishing to choose a modeling approach for their own data. The task was designed to be similar to that used by Hoppitt et al. (2012) in their diffusion experiment on meerkats: i.e., the task is designed such that the pattern of generalization of social effects among options can be used to infer the social learning mechanisms in operation. Hoppitt et al. (2012) developed a Stochastic Mechanism Fitting Model (SMFM), which does not just model the social influences on the rate of first interaction with each option, but also the successive interactions with each option. The SMFM enabled Hoppitt and colleagues to distinguish transient social effects on behaviour from mechanisms that enable learning directly from observing others.

In this study, we were interested primarily in the option specificity of the social learning mechanisms, and not the transient versus direct learning distinction. Consequently, we decided to adopt a simpler model which models only the transition from a naïve state to a ‘discovered’ state and the transition from the ‘discovered’ state to the ‘solved’ state for each option. This is similar in structure to the multistate model developed by Atton et al (2012), which they applied to a diffusion experiment on sticklebacks. However, Atton and colleague’s model is an extension of Network Based Diffusion Analysis (NBDA: Franz & Nunn 2009), so the predictor variables corresponding to the social effects are the social network connections to other individuals that have already discovered or solved each of the task options. The assumption, as with most applications of NBDA, was that

the network connection from individual  $j$  to individual  $i$ ,  $a_{ij}$ , acts as a proxy for the rate at which  $i$  has opportunities for social learning from  $j$ .

In this study, we were able to obtain a record of who is likely to have observed each interaction with the task, and so we can model the effects of observation directly (as in Hoppitt et al.'s SMFM) without having to rely on social network connections as a proxy. Our model is therefore similar to that developed by Hobaiter et al. (2014), who also directly modeled the effect of observations on the rate at which chimpanzees first performed two novel foraging innovations, though Hobaiter and colleagues do not require the multistate model used here. Hobaiter et al. describe their model in the context of a dynamic NBDA, noting that NBDA can be expanded to operate on a dynamic network that changes over time: with the  $a_{ij}$  replaced with  $a_{ij}(t)$ . In their case the  $a_{ij}(t)$  terms of the dynamic network are simply the number of times that  $i$  has observed  $j$  prior to time  $t$ . Our model can also be seen as part of the dynamic NBDA framework: above, in the interests of accessibility, we describe our model in the context of a Cox model framework since this is a more widely used and less specialised statistical model. In contrast, Hobaiter and colleagues consider a number of function forms for their dynamic NBDA, only some of which fit within the Cox model framework. Hobaiter et al. discuss the advantages of using a dynamic NBDA over a standard NBDA when the data are available, and also compare the performance of the model to an NBDA using a static network.

### ***B5. Electronic Supplementary Material 1 References***

- Atton, N., Hoppitt, W., Webster, M. M., Galef, B. G. & Laland, K. N. 2012 Information flow through threespine stickleback networks without social transmission. *Proc. R. Soc. B* **279**, 4272-4278. (DOI 10.1098/rspb.2012.1462)
- Bartoń K. 2014. MuMIn: Multi-model inference. R package version 1.10.0. <http://CRAN.R-project.org/package=MuMIn>.
- Burnham, K. P. & Anderson, D. R. 2002 *Model selection and multimodel inference: a practical information-theoretic approach*. New York: Springer.
- Franz, M. & Nunn, C. L. 2009 Network-based diffusion analysis: a new method for detecting social learning. *Proc. R. Soc. B* **276**, 1829-1836. (DOI 10.1098/rspb.2008.1824)
- Heyes, C. M. 1994 Social learning in animals: categories and mechanisms. *Biol. Rev.* **69**, 207-231. (DOI 10.1111/j.1469-185X.1994.tb01506.x)
- Hobaiter, C., Poisot, T., Zuberbuhler, K., Hoppitt, W. & Gruber, T. (2014) Social network analysis shows direct evidence for social transmission of tool use in wild chimpanzees. *PLOS Biology* **12**, e1001960. (DOI 10.1371/journal.pbio.1001960)
- Hoppitt, W., Samson, J., Laland, K. N. & Thornton, A. 2012 Identification of learning mechanisms in a wild meerkat population. *PLOS ONE* **7**, e42044. (DOI 10.1371/journal.pone.0042044)
- Therneau, T. 2012 coxme: Mixed Effects Cox Models.. R package version 2.2-3. <http://CRAN.R-project.org/package=coxme>.