Detecting social transmission in networks

William Hoppitt, Neeltje J. Boogert, Kevin N. Laland

www.elsevier.com/locate/yjtbi

Detecting social transmission in networks

William Hoppitt, Neeltje J. Boogert, Kevin N. Laland

PII: S0022-5193(10)00008-1
DOI: doi:10.1016/j.jtbi.2010.01.004
Reference: YJTBI 5826

To appear in: *Journal of Theoretical Biology*

Received date: 26 August 2009
Revised date: 4 January 2010
Accepted date: 5 January 2010

Cite this article as: William Hoppitt, Neeltje J. Boogert and Kevin N. Laland, Detecting social transmission in networks, *Journal of Theoretical Biology*, doi:10.1016/j.jtbi.2010.01.004

DETECTING SOCIAL TRANSMISSION IN NETWORKS

WILLIAM HOPPITT[a*], NEELTJE J. BOOGERT[b] AND KEVIN N. LALAND[a]

[a]School of Biology, University of St Andrews, Bute Building, Westburn Lane, St Andrews, Fife KY16 9TS, United Kingdom

[b]Department of Biology, McGill University, Montréal, Québec, Canada

*Corresponding author at: E-mail <wjeh1@st-andrew.ac.uk>

Telephone: +44 (0) 1334 467230 Fax: +44 (0) 1334 463600

Running headline: Hoppitt et al. Detecting Social Transmission

Word count: 8,546 words excluding references

Draft 2.3

Abstract

In recent years researchers have drawn attention to a need for new methods with which to identify the spread of behavioural innovations through social transmission in animal populations. Network-based analyses seek to recognize diffusions mediated by social learning by detecting a correspondence between patterns of association and the flow of information through groups. Here we introduce a new *order of acquisition diffusion analysis* (*OADA*) and develop established *time of acquisition diffusion analysis* (*TADA*) methods further. Through simulation we compare the merits of these and other approaches, demonstrating that *OADA* and *TADA* have greater power and lower Type I error rates than available alternatives, and specifying when each approach should be deployed. We illustrate the new methods by applying them to reanalyse an established dataset corresponding to the diffusion of foraging innovations in starlings, where *OADA* and *TADA* detect social transmission that hitherto had been missed. The methods are potentially widely applicable by researchers wishing to detect social learning in natural and captive populations of animals, and to facilitate this we provide code to implement *OADA* and *TADA* in the statistical package *R*.

Keywords: Social learning; Network based diffusion analysis; Culture; Traditions; Order of acquisition

Introduction

'Social learning' is broadly defined as learning that is influenced by observation of or interaction with a conspecific or its products (Heyes, 1994). Social learning can result in 'social transmission', which we define as occurring when the acquisition of information or a behavioural trait by one individual exerts a positive causal influence on the rate at which another acquires the same information or trait. Social learning appears widespread across both vertebrate and invertebrate taxa (Hoppitt and Laland, 2008; Leadbeater and Chittka, 2007), whilst experimental work has established that social transmission can result in the establishment of behavioural traditions (e.g. Galef and Allen, 1995; Whiten et al., 2005). This has lead to claims of animal cultures in natural populations of apes (McGrew, 1998; Whiten et al., 1999; van Schaik et al., 2003), cetaceans (Rendell and Whitehead, 2001; Krützen et al., 2005) and monkeys (Perry et al., 2003). However, such claims remain controversial because studies fail to adequately rule out alternative explanations for local differences in behaviour, such as local environmental differences, or genetic differences between populations (Laland and Hoppitt, 2003; Laland and Janik, 2006). There is concern that the current 'ethnographic' method, which infers social transmission only where the alternatives of genetic or environmental variation can be disregarded, will rule out genuine cases of social transmission that covary with these factors (Laland and Janik, 2006; Laland and Galef, 2009). Consequently, in recent years researchers have called for the development of quantitative methods for inferring social transmission from field and captive study data that can rule out alternative explanations for the observed effect (Laland and Janik, 2006; Laland and Galef, 2009, and chapters therein).

One type of data that has previously been used to infer social transmission in groups of animals is diffusion data, where researchers monitor the spread of a novel behavioural trait. For some time the shape of the 'diffusion curve' (the cumulative number of individuals

63 seen to perform the novel behaviour plotted against time) was used to infer social learning
64 (e.g. Lefebvre, 1995a; 1995b). The assumption was that if learning were asocial, the rate of
65 learning would be the same for all individuals, resulting in an r-shaped diffusion curve. In
66 contrast, if there were social transmission, the rate of learning would increase as the number
67 of demonstrators increased, resulting in an s-shaped curve (Reader, 2004). However, this
68 approach has been somewhat discredited, since there are a number of situations in which we
69 expect to see an s-shaped diffusion curve in the absence of social transmission (Laland and
70 Kendal, 2003; Reader, 2004), or an r–shaped curve in the presence of social transmission
71 (Franz and Nunn, 2009).

72 An alternative method is to use the order in which individuals acquire a behavioural
73 trait to infer social transmission from diffusion data, on the assumption that if social
74 transmission is operating we might expect the spread to follow the patterns of associations
75 between individuals (Boogert et al., 2008; Morrell et al., 2008). The reasoning here is that
76 individuals that are closely associated are more likely to learn from each other (Coussi-
77 Korbel and Fragaszy, 1995). A randomisation approach has already been applied to test for
78 such a pattern (Boogert et al., 2008; see also Morrell et al., 2008), but below we demonstrate
79 that this approach is vulnerable to both Type I and Type II errors.

80 Here we propose an alternative method, which we call *order of acquisition diffusion*
81 *analysis*, or *OADA*, where a model of social learning is fitted to the data by maximum
82 likelihood, and tested against a model with no social transmission[1]. Our approach is similar to
83 a method recently proposed by Franz and Nunn (2009), which they term 'network based
84 diffusion analysis' (or *NBDA*). Franz and Nunn's method exploits data on the time at which
85 individuals acquire a behavioural trait, rather than the order in which they do so. However, as

[1] In supplementary electronic material we provide code to run all analyses described in this paper in the statistical language *R 2.8.1* (*R* Development Core Team 2008). Updated versions of the code will be made available on the authors' website (http://lalandlab.st-andrews.ac.uk/)

OADA and the randomization approach of Boogert et al. (2008) are also network based diffusion analyses, for clarity we rename Franz and Nunn's approach *time of acquisition diffusion analysis* (or *TADA*), and retain *NBDA* as the more general term for network-based approaches. We see the *OADA* and *TADA* approaches as complementary, and in later sections of this paper we introduce the *OADA* model, extend Franz and Nunn's *TADA* method, and provide a full comparison of *OADA* and *TADA* models. We end by illustrating the methods by applying them to a published data set: the diffusion of novel foraging traits in groups of starlings, *Sturnus vulgaris* (Boogert et al., 2008).

Boogert et al.'s (2008) randomisation method

First, we will describe Boogert et al.'s (2008) randomisation method and illustrate its limitations. To implement this method, for each group in which a diffusion is recorded, one needs a matrix containing an appropriate measure of association between individuals (the association matrix), and the order in which individuals acquired the behavioural trait (the 'diffusion chain'). The test statistic is then simply the summed strength of associations between adjacent individuals in each diffusion chain, summed across groups. If social transmission were occurring preferentially between closely associated individuals, the test statistic is likely to be larger than if individuals were learning independently. To test this hypothesis, a null distribution is generated by randomisation (Manly, 2007): the diffusion chain is randomised for each group, and the test statistic calculated. If the diffusion of multiple behavioural traits has been observed, one can test the global null hypothesis of no social transmission by summing the test statistic across traits. Boogert et al. proposed a second test statistic where, instead of summing the associations between adjacent individuals on the diffusion chain, one sums the mean association between each individual and all individuals before it in the diffusion chain. The logic here is that an individual can learn from

111 any informed individual, not just the preceding individual on the diffusion chain. These are
112 referred to as the 'linear' and 'averaging' metrics, respectively.

113 Boogert et al.'s randomisation method is non-parametric, which has the obvious
114 advantage that researchers need to make few assumptions about the way in which social
115 transmission and asocial learning proceed in order to test the null hypothesis. A disadvantage
116 is that it does not allow inferences about the strength of social transmission to be made,
117 which might be useful for testing hypotheses about the nature of the social learning strategy
118 deployed (Laland, 2004). A more serious limitation is that it is susceptible to false positives if
119 closely associated individuals happened to have a similar rate of acquisition through asocial
120 learning. For instance, individuals of high social rank might have a higher rate of asocial
121 acquisition due to increased access to the resources required for learning. If, in addition,
122 individuals happened to associate with those of a similar social rank, this might result in a
123 false positive for the detection of social transmission (see below). An alternative approach is
124 to fit the data to a model that includes both variables representing the effects of social
125 transmission and known variables that might influence asocial learning, thereby controlling
126 statistically for the latter. Below we describe *OADA* and *TADA* methods that allow this to be
127 done.

128

129 **Order of Acquisition Diffusion Analysis (*OADA*)**

130 *Modelling social transmission*

131 Our starting model assumes that the rate at which social transmission occurs between
132 a given dyad of informed and naïve individuals is linearly proportional to the association
133 between them. This assumption is likely to be reasonable provided that *a*) the probability a
134 naïve individual observes, or is exposed to, the performance of the novel trait is proportional
135 to its association with the demonstrator, and *b*) all informed individuals are approximately

equally likely to perform the trait. The rate of acquisition of the trait through social transmission for individual *i* at time *t*, or $\lambda_{S,i}(t)$, is given by

$$\lambda_{S,i}(t) \propto (1 - z_i(t)) \sum_{j=1}^{N} \left(a_{i,j} z_j(t) \right), \quad (1)$$

where $z_i(t)$ is a binary indicator variable indicating whether *i* is naïve (0) or informed (1) at time *t*, and $a_{i,j}$ is the association between individuals *i* and *j*, in a population of size *N*.

Inclusion of variables influencing asocial learning

At the same time we assume that it is possible that the individual may acquire the trait through trial and error or direct interaction with the environment, uninformed by the behaviour of others. The rate of asocial learning for *i*, $\lambda_{A,i}$ can be modelled as:

$$\lambda_{A,i} \propto (1 - z_i(t)) \exp(\beta_1 x_{1,i} + \beta_2 x_{2,i} + \ldots + \beta_V x_{V,i}) \quad (2)$$

where $x_{1,i}, x_{2i,}, \ldots, x_{V,i}$ are the individual-level variables influencing asocial learning, and $\beta_1, \beta_2, \ldots, \beta_V$ are the coefficients specifying the effect of each. Exponential transformation of the linear predictor ensures that the predicted rates are always positive, which is common practise in statistical modelling of rates (Therneau and Grambsch, 2001).

The question remains of how the effects of asocial learning and social transmission are combined in the model. Here we suggest two alternative approaches: *i*) an additive model (Eqn. 3) and *ii*) a multiplicative model (Eqn. 4). If social transmission occurs as an independent process by which individuals can acquire the trait, then the total rate of acquisition, $\lambda_i(t)$, will be the sum of the rates of asocial learning and social transmission, or

$$\lambda_i(t) = \lambda_0(t) \left(1 - z_i(t)\right) \left(s \sum_{j=1}^{N} a_{i,j} z_j(t) + (1 - s) \exp\left(\sum_{k=1}^{V} \beta_k x_{k,i} \right) \right), \quad (3)$$

where $\lambda_0(t)$ is a baseline rate of acquisition common to all individuals, and *s* is a parameter determining the strength of social transmission ($0 \le s < 1$). Here $s = 0$ indicates no social

159 transmission and $s = 1$ implies all learning is social. For a natural diffusion, $s \neq 1$ since the
160 first individual must have acquired the behaviour through asocial learning. The additive
161 model is likely to be appropriate if individuals can acquire the trait as a direct consequence of
162 observation (Hoppitt and Laland, 2008), for instance, by imitation or some other form of
163 observational learning.

164 Conversely, social transmission might often operate in an 'indirect' manner (Hoppitt
165 and Laland, 2008), if the informed individual's behaviour influences the naïve individual's
166 behaviour in a manner that leads indirectly to learning. For example, Leadbeater and Chittka
167 (2007) found social transmission could speed the rate at which bumblebees (*Bombus*
168 *terrestris*) learned to discriminate differently-coloured artificial flowers, because they were
169 attracted to rewarding flowers occupied by informed conspecifics, allowing them to learn by
170 their own experience that these flowers are rewarding. Here the effect of social transmission
171 is to increase the time spent in the area in which trait acquisition can occur (local
172 enhancement, Thorpe, 1956), and so will weight the rate at which it occurs by otherwise
173 asocial means. For these, and similar cases, we suggest that a multiplicative model is more
174 appropriate, where

175 $$\lambda_i(t) = \lambda_0(t)\left(1 - z_i(t)\right)\left(s\sum_{j=1}^{N} a_{i,j} z_j(t) + (1-s)\right)\exp\left(\sum_{k=1}^{V} \beta_k x_{k,i}\right). \quad (4)$$

176 Here the (1-*s*) term ensures that the effect of social transmission is weighted relative to the
177 rate at which asocial learning occurs. The choice of model should not be seen as a nuisance.
178 In cases where the experimenter has reasonable confidence in the likely social learning
179 mechanism, the appropriate model can be selected. In other cases, both models may be used,
180 and the model that best fits the data deployed. Indeed, this exercise could potentially be seen
181 as providing information about the type of social transmission that is operating, although
182 confidence in such inferences would be enhanced by experimental validation.

Model fitting

To implement an *OADA* we only need a relative measure of the rate at which individual *i* acquires the trait at time *t* (that is, relative to other naïve individuals), or $R_i(t) = \lambda_i(t)/\lambda_0(t)$. The probability that individual *i* is the next to learn can be written as

$$p_{next,i}(t) = \frac{\lambda_i(t)}{\sum_{l=1}^{N}\lambda_l(t)} = \frac{\lambda_0(t)R_i(t)}{\lambda_0(t)\sum_{l=1}^{N}R_l(t)} = \frac{R_i(t)}{\sum_{l=1}^{N}R_l(t)}, \quad (5)$$

and the probability that it will be the n^{th} individual to acquire the trait, $p_{n,i}$, is given by

$$p_{n,i} = \frac{R_i(n)}{\sum_{l=1}^{N}R_l(n)}, \quad (6)$$

where $R_i(n)$ is *i*'s relative rate of acquisition immediately prior to the n^{th} acquisition event. We can then write

$$R_i(n) = \left(1 - z_i(n)\right)\left(s\sum_{j=1}^{N}\left(a_{i,j}z_j(n)\right) + (1-s)\exp\left(\sum_{k=1}^{V}\left(\beta_k x_{k,i}\right)\right)\right) \quad (7a)$$

and

$$R_i(n) = \left(1 - z_i(n)\right)\left(s\sum_{j=1}^{N}\left(a_{i,j}z_j(n)\right) + (1-s)\right)\exp\left(\sum_{k=1}^{V}\left(\beta_k x_{k,i}\right)\right), \quad (7b)$$

for the additive (Eqn. 3) and multiplicative models (Eqn. 4), respectively, where $z_i(n)$ is the status of individual *i* prior to the n^{th} acquisition event.

Equations 6 and 7 enable one to calculate the -log-likelihood of the observed order of acquisition data for a given set of parameters, *s* and $\beta_1, \beta_2, \ldots, \beta_n$ (e.g. see Morgan, 2009). The -log-likelihood is easily calculated for multiple groups or multiple traits by adding together the -log-likelihoods for each separate diffusion. The model is then fit by choosing the parameter values that minimise the -log-likelihood, using a suitable numerical optimisation

203 routine. In the supplementary material we provide *R* functions that fit both models (see ESM:
204 "Additional Information" part C).

205 To fit the models, we find that the optimisation algorithms used are more likely to
206 converge if we use the reparameterisation of $s' = s/(1-s)$ with $0 \le s' < \infty$. This results in an
207 additive model of

208 $$R_i(n) = \left(1 - z_i(n)\right)\left(s' \sum_{j=1}^{N} a_{i,j} z_j(n) + \exp\left(\sum_{k=1}^{V} \beta_k x_{k,i} \right) \right), \quad (8a)$$

209 and a multiplicative model of

210 $$R_i(n) = \left(1 - z_i(n)\right)\left(s' \sum_{j=1}^{N} a_{i,j} z_j(n) + 1 \right) \exp\left(\sum_{k=1}^{V} \beta_k x_{k,i} \right). \quad (8b)$$

211 To favour convergence of maximum likelihood estimation, we suggest use of Eqn. (8) for
212 model fitting, and transforming to the more intuitive parameterisation in Eqn. (7) for
213 interpretation.

214

215 *Model selection and hypothesis testing*

216 To test for social transmission, researchers can use a likelihood ratio test (LRT, see
217 Morgan, 2009 for details) to compare the fitted model with a nested null model in which *s* is
218 constrained to be zero. The significance of other parameters in the model can also be tested in
219 this way, and the model reduced in a manner analogous to a multiple regression. Confidence
220 intervals for parameters can be calculated using profile-likelihood techniques (ESM:
221 Additional Information, part D; Morgan, 2009). Researchers can also use Akaike's
222 Information Criterion (AIC) to compare alternative models with different degrees of freedom
223 (Burnham and Anderson, 2002). This has the advantage that non-nested models can be
224 compared, such as the best-fitting model containing social transmission and the best-fitting

model without social transmission, when each contains different individual level variables. Methods for dealing with tied data are given in the ESM (Additional Information, part F).

Comparison of *OADA* with *TADA*

Here we describe and extend Franz and Nunn's *NBDA* method, which we rename *TADA*, in the context of our *OADA* model, and using our notation. This facilitates a direct comparison between models reliant on order or time of acquisition.

TADA makes the same assumptions about social transmission as our model (Eqn. 1), but the models are fitted to time of acquisition data rather than to order of acquisition data, meaning the absolute rate of acquisition, $\lambda_i(t)$, is modelled, rather than the relative rate $R_i(t)$, and the baseline rate of acquisition is taken to be constant $\lambda_0(t) = \lambda_0$. Franz and Nunn suggest two approaches. The first involves fitting separate models for social transmission and asocial learning, with $\lambda_i(t) = \lambda_0$, and comparing the two models using AIC. However, this approach is only useful if the diffusion starts with informed individuals in the population, otherwise the likelihood of the model for social transmission will always be zero, since the likelihood of the first individual's acquisition is zero. Similar to *OADA*, Franz and Nunn's second approach involves fitting a two-parameter model, which allows for both social transmission and a constant rate of asocial learning.

There are inherent strengths and weaknesses to both *TADA* and *OADA* methods. The fundamental difference is the type of data that is modelled, time or order. We demonstrate below that time of acquisition data typically possesses more power to detect a social transmission effect, which is the major advantage of *TADA*. However, *TADA* requires assumptions about the specific distribution of latencies: Franz and Nunn assume an exponential distribution, where the rate of acquisition at a given time is dependent only on the status of other individuals in the group. In contrast, *OADA* makes the less onerous

assumption that the ratio of acquisition rates between two individuals is dependent only on the variables included in the model. The flexibility of this 'proportional hazards' assumption has lead to the preference of the Cox proportional hazards model as the most widely used method for analysing time to event data (Therneau and Grambsch, 2000). The similarity of *OADA* to the Cox model is described in the ESM (Additional Information, part A). Below we show that the vulnerability of *TADA* and *OADA* to Type I error varies, and that each is more reliable than the other in some contexts.

In its initial form, Franz and Nunn's *TADA* is also susceptible to the same problems of confounding variables as Boogert et al.'s randomisation method. Accordingly, here we extend *TADA* to include individual level variables influencing rate of acquisition. By the above reasoning, the additive model can be written as

$$\lambda_i(t) = \lambda_0\left(1 - z_i(t)\right)\left(s\sum_{j=1}^{N} a_{i,j} z_j(t) + (1-s)\exp\left(\sum_{k=1}^{V}\beta_k x_{k,i}\right)\right), \quad (9a)$$

and the multiplicative model as

$$\lambda_i(t) = \lambda_0\left(1 - z_i(t)\right)\left(s\sum_{j=1}^{N} a_{i,j} z_j(t) + 1 - s\right)\exp\left(\sum_{k=1}^{V}\beta_k x_{k,i}\right), \quad (9b)$$

where λ_0 determines the overall rate of asocial acquisition, and s parameterises the social transmission effect relative to the rate of asocial acquisition. As for *OADA*, we find the reparameterisation $s' = s/(1-s)$ works better for maximum likelihood estimation. We have also found this reparameterisation preferable to independent parameters for the rate of social and asocial transmission, since in the latter case the estimators for each are highly negatively correlated (Morgan, 2009). Setting $\lambda_0 = 1/L_0$ can facilitate convergence of the optimisation routines. The model can either be fitted by treating time as a continuous variable or by splitting time into a number of discrete steps, depending on the way in which the data was collected (details are given in the ESM: Additional Information, part E). Functions to

implement this extended version of TADA for the multiplicative and the additive models, using both discrete and continuous methods of fitting, are given in the ESM: Additional Information part E.

Simulation details

We compared how the *OADA*, *TADA* and randomization models performed under different circumstances. All simulations considered the diffusion of a single learned behavioural trait through a single hypothetical group of animals of size *N*. Where the rate of acquisition of the trait was affected by an individual-level variable, this was generated by drawing a value for each individual from a normal distribution (x~N(0,1)). We simulated an association matrix for the population by first generating a matrix of associations that was normally distributed with a specified correlation, *c*, with the magnitude of the differences in the individual-level variable. To make the matrix more realistic, we made the matrix symmetrical by setting $a_{i,j}{}' = a_{j,i}{}' = (a_{i,j} + a_{j,i})/2$. We then transformed the associations to vary between 0 and 1 by ranking the values and dividing each by the maximum rank. To explore the effect of different levels of connectedness within the group, we set associations less than a threshold value, *T*, to zero, and explored how the magnitude of *T* affected the utility of the models.

Order and time of acquisition data were simulated according to either the additive model (Eqn. 8a) or the multiplicative model (Eqn. 8b) for specified values of λ_0, s and β. At each point in the diffusion chain, a value was drawn from an exponential distribution with an appropriate rate parameter for each naïve individual (determined by Eqn. 9a or 9b). The individual with the lowest value was taken to be the next individual to solve the task, with the intervals between solving events determined by the value itself. The data was then analysed

using the additive and multiplicative *OADA*, the additive and multiplicative *TADA*, randomisation tests using Boogert et al.'s linear metric (1000 randomisations) and averaging metric (100 randomisations only, due to larger computation time). The simulations were usually run 10,000 times for each combination of simulation parameter values. This was reduced to 1,000 times when there were individual-level variables due to the increased computation time required to fit NBDA models.

Where there were no individual-level effects, we considered a variety of group sizes (N=10, 20, 50), and social transmission effect sizes (s=0, 0.2, 0.4, 0.6, 0.8, 0.99), and recorded the power of each technique to detect social transmission. Since there were no individual-level variables, the multiplicative and additive models are equivalent in this case.

We explored individual-level effects in simulations in which group size was fixed at 20, β=10, and there were a range of social transmission effect sizes (s=0, 0.4, 0.8) and levels of correlation between the association matrix and differences in the individual-level variable (c=0,0.4,0.8). We recorded the statistical power to detect social transmission at the 5% significance level and the *OADA* and *TADA* models preferred by AIC_c.

In another series of simulations, we allowed the baseline rate of acquisition, $\lambda_0(t)$, to vary within a diffusion, either *i*) at random or *ii*) systematically. For *i*), to determine the initial baseline acquisition rate a number was drawn from a normal distribution with mean=log (0.0002) and a standard deviation of 0,2,4,6 or 8 and then exponentially transformed. This process was repeated to generate a new baseline acquisition rate after each acquisition event. For *ii*), the baseline hazard rate either increased or decreased with successive acquisition events, with $\lambda_0(t) = 0.0002\exp(\varepsilon p(t))$, where $p(t)$ is the proportion of demonstrators in the population at time t, and ε determines the strength of the effect. We considered ε=-4, -3, -2, -1, 0, 1, 2, 3, 4, s= 0 or 0.8, and β = 0.

322 To explore the effect of network connectedness, we altered the threshold value, *T*,
323 under which simulated associations were set to zero (*T*=0, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9),
324 decreasing the number of non-zero associations in the network as *T* increases. Here we
325 assumed s= 0.4 or 0.8, N= 20 and $\beta = 0$. Unless otherwise indicated, $T=0.8$, λ_0=0.0002.
326
327 **Application of the models to Boogert et al. (2008)**
328 We go on to illustrate the methods by applying *OADA* and TADA to a published
329 dataset. Boogert et al. (2008) presented three captive groups of five starlings (*Sturnus*
330 *vulgaris*) with six different artificial foraging tasks. Each task was presented separately for
331 several sessions. The time (measured cumulatively over sessions) at which each individual
332 first contacted each task and first solved each task was recorded. Associations between
333 individuals were calculated as the proportion of discrete point samples a given dyad was
334 within pecking distance. In addition, a number of individual-level variables were recorded: *a*)
335 a measure of asocial learning ability, *b*) two measures of neophobia: *i*) the latency to feed in a
336 novel environment, and *ii*) average latency to feed next to three novel objects, *c*) two
337 measures of social rank: *i*) competitive rank: time spent dominating a limited resource and *ii*)
338 agonistic rank calculated as David's scores based on agonistic interactions (deVries et al.,
339 2006). The aims of the study were to investigate which individual-level variables predicted
340 the diffusion dynamics, and whether the order of acquisition of task solution followed
341 patterns of association. Boogert et al. pursued the former aim by fitting linear mixed models
342 (LMMs) or generalised linear mixed models (GLMM) to data on the number of times an
343 individual was first to solve a task within its group, and the latency to solve the task
344 (excluding the first solver), each with the individual-level variables as predictors. The
345 question of whether order of acquisition followed patterns of association was tested using the

346 randomisation approach described above. Boogert et al reported their analysis showed no
347 evidence for social transmission.

348 Here we implement an alternative approach that uses *OADA* and our extended version
349 of *TADA*, comparing the results of each with the original findings. The methods were applied
350 to the data from all diffusions, across all groups and tasks, in a global analysis. To test for
351 social transmission, we first identified the combination of individual level variables best able
352 to account for the data, in the absence of social transmission. We fitted models with all
353 possible combinations of individual-level variables and recorded AIC_c in each case. We
354 selected the two best models and used these as null models to test for social transmission,
355 assuming additive and multiplicative functions. We then fit a model with separate social
356 transmission parameters for each group. We used a LRT to test each of these against zero,
357 and dropped those that were not significant at the 5% level. We quantified the significance of
358 the terms left in the model by dropping each from the model and using a LRT. To assess
359 whether the social transmission parameter differed between specific groups, we fitted a null
360 model with the parameter constrained to be equal for each group, and used a LRT to compare
361 this to a model where they were unconstrained. We also obtained approximate confidence
362 intervals for each parameter using profile likelihood techniques (see ESM: Additional
363 Information, part D). The same approach was used to fit *TADA*. Individual level variables
364 representing an effect of 'group' and 'task' were considered alongside those considered in
365 *OADA*.

366

367 **Results**

368 *Comparison in the absence of individual-level effects*

369 In the absence of individual-level effects, and for a given group and effect size, *TADA*
370 typically had more statistical power to detect social transmission than did *OADA*, while both

371 of these methods were more powerful than the averaging and linear randomization methods
372 (Fig. 1 a-b). In the case of the randomization methods, the averaging metric usually provided
373 more power than the linear metric, especially for larger group sizes, where social
374 transmission is less likely to occur between adjacent individuals in the diffusion chain. In
375 most cases, power increased with group size, except for the randomisation method with the
376 linear metric. As expected, statistical power also increased with the strength of social
377 transmission.
378
379 *Effect of individual-level variables*
380 When there was no correlation between the individual-level variable and association,
381 the type I error rates were appropriate (~5%) for all methods (Fig. 1d). However, the power
382 to detect an effect using *OADA* or *TADA* was greatly increased by inclusion of the variable in
383 the model (see Fig. 2).
384 As the correlation between the individual-level variable and association increased,
385 type I error rates were greatly inflated for all methods that did not include an individual-level
386 variable (see Fig. 1d). However, inclusion of the individual-level variable in both *OADA* and
387 *TADA* methods restored type I error to an appropriate rate, for both multiplicative and
388 additive models. When social transmission and asocial learning were additive, power to
389 detect social transmission was little affected so long as the additive model was fitted to the
390 data (see Fig. 2 a-b). In contrast, when social transmission and asocial learning combined
391 multiplicatively, power was markedly reduced, though again, there was more statistical
392 power when the appropriate multiplicative model was used, rather than the additive model
393 (see Fig. 2 c-d). AIC_c was generally a successful criterion in selecting the appropriate model
394 (additive versus multiplicative, see ESM: Additional Information, part B).

These simulations demonstrate that the inclusion of individual-level variables in the analysis of diffusion data is highly desirable, both with respect to controlling type I error rates, and maximising statistical power. This is an advantage that both *OADA* and our extension of *TADA* have over the randomisation techniques. Again, we see that *TADA* has more power than *OADA* in each case. Our analysis also lends confidence that the procedure we recommend will select a model (multiplicative or additive) appropriate to the data.

Varying baseline rate of acquisition

We manipulated the baseline rate of acquisition, both by increasing the variance of the underlying distribution (Fig. 3a) and by allowing it to increase or decrease as the diffusion proceeded (Fig. 3b). In all cases the power and type 1 error rates remained approximately constant for the *OADA* method (see Fig. 3), as we would anticipate, since the baseline hazard function does not change the relative rate of acquisition. In contrast, *TADA* was very sensitive to changes in the baseline acquisition rate. When the baseline acquisition rate varied at random, statistical power dropped as the variance of the underlying distribution of rates increased (see Fig. 3a), whereas the Type I error rate increased. When the baseline acquisition rate decreased systematically throughout the diffusion, it obscured a social transmission effect from the *TADA* method (see Fig. 3b). Conversely, when the baseline acquisition rate increased, this resulted in an increase in Type I error for *TADA* (but see below).

These simulations illustrate the relative strengths and weaknesses of *OADA* and *TADA*. If there are fluctuating variables influencing the rate of acquisition that affect all individuals equally, then *OADA* is preferable to *TADA*. Likewise, if there is a factor that causes a systematic decrease in the baseline acquisition rate, *OADA* may be more likely to detect social transmission. This might occur if, for example, an increasing number of

informed individuals depletes the resources necessary for trait acquisition, or increases the number of opportunities for scrounging, which might inhibit acquisition (Giraldeau and Lefebvre, 1987). The increase in Type I error for an increasing baseline acquisition rate could be seen as a problem with *TADA* if there is reason to believe that a variable is influencing trait acquisition in this way. However, a systematic increase in baseline acquisition rate could be a direct result of the increased number of informed individuals, which would mean it is a case of social transmission by our definition. This shows that *OADA* is only sensitive to social transmission if it results in a difference in the relative rate of acquisition by individuals, whereas *TADA* is also sensitive to absolute changes in the rate of acquisition (Fig. 3b).

Number of connections in the network

Network connectedness (the number of non-zero associations) had a different effect on *OADA* and *TADA* (see Fig. 4). For *TADA*, power either remained approximately constant (s=0.8) or declined (s=0.4) as connectedness went down (increasing *T*). In contrast, for *OADA*, the power increased in both cases, appearing to converge with the power for TADA when the proportion of zero associations was large. This is because *OADA* will detect social transmission when it results in large differences between the rates at which individuals acquire the trait, and works best when opportunities for social learning differ greatly between individuals at a given time. In contrast, *TADA* is also sensitive to the acceleration in the rate of acquisition which occurs as a result of an increased number of informed individuals. This effect will be more pronounced when there are many connections between individuals, offering many opportunities for social transmission.

Application of the models to Boogert et al. (2008)

Where the magnitude of the social transmission parameter was constant across groups, the best predictive *OADA* model included object neophobia and asocial learning, and no social transmission (henceforth Model 1: $AIC_c = 138.39$), but a model with latency to feed in a novel environment as sole predictor was almost as good (henceforth Model 2: $AIC_c = 138.40$). Social transmission was not statistically significant when added to either model as an additive effect (Model 1: LR=0, p=1; Model 2: LR=0.03, p=0.870) or a multiplicative effect (Model 1: LR=0.321, p=0.571; Model 2: LR=0.468, p=0.494). However, when the social transmission parameter was allowed to vary between groups, we found a significant effect on group 1 in all models ($p<0.05$, see Table 1), but no evidence for an effect on groups 2 or 3 ($p>0.5$, see Table 1). For the additive model the social transmission effect on group 1 was also found to be significantly stronger than a putative effect on group 3 (Model 1: LR=5.64 p=0.018; Model 2: LR=15.95, $p<0.001$) but not than that on group 2 (Model 1: LR=0.65 p=0.420, Model 2: LR=1.02, p=0.312). The same result was found for the multiplicative model: group 1 versus group 3: Model 1: LR=5.15 p=0.023; Model 2: LR=4.30, p=0.038; group 1 versus group 2: Model 1: LR=0.91 p=0.340; Model 2: LR=0.25, p=0.614. The best model, as judged by AIC_c included object neophobia and asocial learning performance as individual-level variables, with an additive social transmission effect for group 1 only ($AIC_c = 135.08$), although a multiplicative model worked almost as well ($AIC_c = 135.14$). The AIC_c when all individual-level variables were dropped from the final model was 136.15, which is preferred to an additive model including latency to feed in a novel environment. None of the individual-level variables were significant at the 5% level when dropped from any of the final models. See Table 1 for full details of the best fitting *OADA* models.

The best predictive *TADA* model excluding social transmission included latency to feed in a novel environment as a sole predictor ($AIC_c = 1175.56$), so this was used as the null

model to test for social transmission. When social transmission was added to the null model it was highly significant for both the additive (LR=15.54, df=1, p<0.001) and multiplicative model (LR=16.75, df=1, p<0.001). There was no evidence of a difference in the effect of social transmission between groups for either the additive (LR=0.27, df=2, p=0.872) or multiplicative model (LR=1.30, df=2, p=0.523). The best model, as judged by AIC_c , included latency to feed in a novel environment as an individual-level variable, with a common multiplicative social transmission effect for all groups (AIC_c = 1161.02), although an additive model worked almost as well (AIC_c = 1162.24). In contrast to the *OADA* model, there is a clear indication that latency to feed in a novel environment has a negative relationship with individuals' rates of acquisition (Additive model: LR=5.04, df=1, p=0.025; Multiplicative model: LR=6.80, df=1, p=0.009). See Table 2 for full details of the best fitting *TADA* models.

Consistent with Boogert et al.'s original conclusions, when all groups were analysed together, there was no evidence of social transmission using the randomisation methods used by Boogert et al. (linear metric= 206.5; p=0.170; averaging metric= 204.2, p=0.149)[2]. However, when groups were analysed separately (this was not done by Boogert et al.), both randomization metrics provided evidence for social transmission in group 1 (linear metric=72, p=0.013; averaging metric=69.1, p=0.012; new metric: G1=25.3, p=0.012), but no evidence for groups 2 and 3 (p>0.15 in all cases).

Whereas the randomisation tests used by Boogert et al. failed to find evidence of social transmission, based on the order of acquisition data, our *OADA* method found evidence for social transmission in Group 1. When the randomisation methods were reapplied to the data from each group separately, the same results were found. However, unlike *OADA*, the

[2] Note that the P values given here are one-tailed, whereas Boogert et al. calculated two-tailed P values by doubling the one-tailed P value, though there was a mistake in the calculation of the P value corresponding to the averaging statistic causing it to be reported as half its estimated value.

493 randomisation methods do not enable us to construct confidence intervals on the effect of
494 social transmission in each group. The 95% confidence intervals from *OADA* reveal that the
495 data provide no resolution to distinguish social transmission from asocial learning in group 2,
496 whereas in groups 1 and 3, the data are consistent with a lower and upper limit on social
497 transmission respectively (see Table 1).

498 In contrast to *OADA*, *TADA* provided evidence of social transmission in all groups,
499 with no evidence of differences between them. This is probably the result of increased power
500 resulting from inclusion of time of acquisition data, which is reflected in the narrower 95%
501 confidence intervals for social transmission (see Table 2). The findings of *TADA* and *OADA*
502 are less contradictory than they might appear at first. The *TADA* confidence intervals for
503 social transmission are within the *OADA* confidence intervals for the effect for Groups 1 and
504 2 and overlap with the *OADA* confidence intervals for Group 3. The only real discrepancy is
505 the finding from *OADA*, that social transmission was significantly stronger in Group 1 than it
506 was in Group 3.

507 The results concerning individual-level variables are qualitatively similar for both
508 *OADA* and *TADA*. *TADA* suggested that an individual's latency to feed in a novel
509 environment was the best predictor of time of acquisition. In *OADA* this variable was also
510 found to be a good predictor of the order of acquisition, though a model including object
511 neophobia and asocial learning ability was approximately as good. However, when the model
512 included social transmission for Group 1, none of these variables were significant at the 5%
513 level. In Boogert et al.'s original analysis, significant differences in latency to solve were
514 found between tasks (not significant in *TADA*), but no other variable was found to be
515 significant (however, latency to feed in a novel environment was found to be correlated with
516 the latency to contact the task). The critical differences between the *TADA* presented here and
517 Boogert et al.'s analysis are: *a*) social transmission was accounted for; *b*) first-solvers were

not excluded from the analysis; *c*) individuals not solving the task were modelled as non-solvers rather than assigned a 'ceiling' value, which can distort an analysis of latencies (Crawley, 2002); and *d*) we compared all possible subsets of variables, rather than using backward selection, which can be misleading when predictors are correlated (Weisberg, 1980).

The simulations presented above suggest that we should prefer *TADA* to *OADA* because of its greater power provided we are happy to assume that the baseline rate of acquisition is constant. We can think of no reason to reject this assumption in Boogert et al.'s diffusion experiment: the diffusions were conducted under laboratory conditions, reducing the possibility for external influences on the birds' rate of acquisition. In addition, there were multiple versions of each task available and each was replenished as soon as it was solved, ensuring that informed individuals could not block naïve individuals from accessing the task. In any case, the blocking of naïve individuals would result in a decrease in the power of *TADA* to detect an effect (see above), whilst the failure of *OADA* to find a social transmission effect for groups 2 and 3 is likely to be a result of the reduced power of the analysis relative to *TADA*.

In summary, these new more powerful methods lead us to the conclusion that there is strong evidence for social transmission in all three groups of starlings, a finding starkly contrasting with that of Boogert et al. (2008).

Discussion

The above simulations bring home the desirability of including individual-level variables in an analysis to detect social transmission from diffusion data. The analyses establish that the inclusion of individual-level variables both increases statistical power and reduces type I error rates. In addition, the sensitivity of the diffusion analyses to network

543 structure prompts us to recommend that researchers use methods that can generate confidence
544 intervals for the strength of social transmission, rather than relying on a rejection/acceptance
545 procedure. For these reasons, our *OADA* and refined *TADA* methods are preferable to
546 established randomisation approaches.

547 The simulations clearly show that *TADA* yields more statistical power than
548 *OADA*. Consequently, in choosing which approach to utilise, we suggest that researchers use
549 *TADA* unless there is good reason to suppose that the baseline rate of acquisition has changed
550 over time. This might be the case, if, for example, the availability of a resource necessary to
551 acquire the trait has varied over time, in which case the weaker assumption of proportional
552 hazards is more appropriate, and the *OADA* method should be deployed (also see below). In
553 principle, one could modify *TADA* to incorporate a non-constant baseline rate of acquisition.
554 However, the success of this method would depend on the researcher choosing an appropriate
555 baseline function. *OADA* has the advantage that it is insensitive to the shape of the baseline
556 function.

557 The power of either method will depend critically on the association measure used in
558 the analysis. Both models are built on the assumption that the rate of transmission between
559 individuals is proportional to the association between them, and, if our interest is in testing
560 for the presence or absence of social transmission, an association measure should be chosen
561 for which this is likely to be true. We suggest that researchers utilise the association measure
562 that is most relevant to the experimental context. For example, in the analysis of the diffusion
563 of foraging task solutions (Boogert et al., 2008) presented above, a measure of association
564 that reflects how often individuals feed together might have been preferable to the general
565 proximity measure that was used. Note that the estimated effect of social transmission
566 depends on the scaling of the association measures used, so if the effect of social
567 transmission is compared between populations or species, either the same association

measure needs to be used, or a case needs to be made that each association measure quantifies opportunities for social learning on a common scale. Franz and Nunn (2009) suggest an alterative approach: that different measures of association, reflecting different social and individual variables, can be used to fit separate NBDA's in order to identify which factors are important in determining diffusion dynamics. A third possible future application of NBDA is to use order or time of acquisition data to infer network structure. This could be of use in cases where it is known or assumed that behaviour is transmitted socially, but social transmission-relevant association data is difficult to acquire. For example, in humpback whales (*Megaptera novaeangliae*) novel vocalisations are easily recorded, but association data is likely to be difficult to obtain in high latitudes (Noad et al., 2000).

There is clearly scope for a far more extensive investigation of how network structure influences both the overall rate of social transmission (Franz and Nunn, 2009) and the power of *OADA* and *TADA* to detect it. The simulations presented here must be viewed as a relatively crude first step. Nonetheless they are sufficient to show that network structures that promote social transmission (e.g. where all individuals are connected) are not necessarily the same as those that make it more likely to be detected, especially by *OADA*. Consequently, if researchers are to use these methods to make comparisons of the levels of social transmission between groups or species, which might have different network structures, we recommend that they obtain power estimates or (preferably) confidence intervals for the social transmission effect, rather than relying solely on presence/absence arguments based on hypothesis tests.

As discussed above, if all individuals have equal opportunity to learn from each other, *OADA* will have no power to detect social learning. In *TADA*, this situation can be modelled by setting all associations to 1, in which case *TADA* is effectively reduced to a diffusion curve analysis, since it is only sensitive to the acceleratory effect that an increasing number

of informed individuals has on the rate of acquisition. However, in principle, our extended version of *TADA* may constitute an improved method for diffusion curve analysis (*DCA*), since it can statistically control for individual-level variables, which might otherwise obscure the underlying pattern.

The sensitivity of *TADA* to acceleration in the rate of acquisition could also be seen as a weakness. It has been noted that *DCA* is vulnerable to false positives if the latency to acquire a trait by asocial learning has a unimodal distribution (Reader, 2004), and *TADA* is also vulnerable under these circumstances. A unimodal distribution of latencies can arise if the process of trait acquisition has multiple steps, each of which is completed at a similar constant rate (Kendal, 2003). For example, to solve a foraging task an individual might first have to approach the task, and then interact with it in an appropriate way. If each of these component processes occurs at a similar constant rate, the overall latency to solve the task asocially would have an approximately gamma distribution with shape parameter k=2, which would in turn result in an apparent acceleration in the rate of acquisition.

Though the models presented here, and the original *TADA* presented by Franz and Nunn (2009), assume a linear relationship between association and rate of social transmission, the methods could be adapted to accommodate other models of social transmission. For instance, the models could be refined to detect social transmission from the spatial spread of a behavioural trait through time (e.g. Fisher and Hinde, 1949). Here one merely needs to propose a relationship between the rate of transmission and the distance between individuals. If this is linear, or the distances can be transformed to linearise the relationship, researchers can use the above methods to fit the model.

The possibility that NBDA might allow us to infer something about the mechanism of social transmission is an issue worth pursuing. Given the fact that currently the ability to detect specific social learning mechanisms is restricted to the experimental laboratory, a

method that could infer learning mechanisms from diffusion data could be extremely valuable. Above we suggested that if social transmission operates indirectly through social influences such as local enhancement, the multiplicative model is likely to provide a better fit to the data. In contrast, we suggest that if social transmission operates directly as an independent learning process, such as imitation, the additive model might provide a better fit. However, such findings should only be taken as suggestive of mechanism at this stage, since there are a number of issues that might complicate this apparent dichotomy. For instance, if asocial and social learning ability covary between individuals, the multiplicative model might fit the data well even if the mechanism is additive in nature. Future extensions of NBDA could investigate these issues by including the effect of individual-level variables on the rate of social transmission (s). There is also the possibility that a number of social transmission processes, both direct and indirect, might operate in parallel. Either of these processes might result in a lack of resolution between multiplicative and additive models, as observed in our reanalysis of Boogert *et al.*'s data.

There are further improvements that can be made to the models in their current form. As it stands, if the models are fit to multiple diffusions involving the same individuals, they assume that the rate of acquisition by the same individual on different tasks is independent, conditional on the variables included in the model. In principle, this assumption could be dropped by incorporating a random effect for individuals. However, this is currently only implemented for the multiplicative *OADA* method, using our multiCoxFit function (see ESM: Additional Information, part A), which fits a Cox Proportional Hazards model with 'frailty' or 'cluster' terms (Therneau and Grambsch, 2000). A more general model would allow the user to specify a correlation structure between the rates of acquisition, for example, a spatial correlation structure (cf. Pinheiro and Bates, 2000). In the spatial analysis described above,

642 this might allow us to control for the fact that two proximate individuals acquire the trait at a
643 similar time because they have similar access to the resources necessary for trait acquisition.
644 NBDA appears to be a relatively novel approach to the statistical analysis of network
645 data. Statistical methods have been developed to investigate properties of flow through
646 networks, such as telecommunication interactions and traffic flow on roads and the internet.
647 (Kolaczyk, 2009). In contrast to NBDA, such models are more concerned with estimation of
648 the strength of connections in the network, rather than testing for the presence of flow against
649 an alternative hypothesis. In addition, such models assume that flow involves the continued
650 transfer of material between nodes, rather than the switching of nodes from one state to
651 another that is a feature of NBDA. In this respect, NBDA bears more resemblance with
652 epidemiological models of the spread of a disease (e.g. Keeling, 1999) or the spread of
653 rumours and fashions (Newman et al., 2006). However, such models *assume* that disease or
654 information spreads through connections in the network, and usually aim to investigate
655 theoretically the effect of network structure on the dynamics of spread (e.g. Meyers et al.
656 2006). In contrast, NBDA aims to test whether trait acquisition does spread through a given
657 network, given real data on network connections and the pattern of trait acquisition. We are
658 not aware of any equivalent epidemiological models that allow statistical inference about the
659 transmission process based on an observed network (see Kolaczyk, 2009, p279).
660 Nonetheless, existing network models (e.g. Newman et al., 2006) could be used to
661 investigate the effect of network structure on the spread of a behavioural trait as a result of
662 social transmission. However, modifications might be necessary. For example,
663 epidemiological models usually assume individuals move from 'susceptible' to 'infected' and
664 then 'recovered' categories, sometimes then moving back to the 'susceptible' category
665 (Watts, 1999). Whilst 'naïve' and 'informed' categories correspond closely to 'susceptible'
666 and 'infected' categories, there is no obvious role for a 'recovered' category in the diffusion

667 of many behavioural traits. A move back to the 'susceptible' category is only applicable if
668 individuals forget the behavioural trait. In addition, we have assumed that a behavioural trait
669 can arise spontaneously in an individual through asocial learning, a feature which is absent
670 from epidemiological models. Watts' (2002) model of information cascades in networks
671 suggests another way in which NBDA could be formulated. In his model, individuals adopt a
672 trait when they are connected to a threshold number of individuals displaying that trait.
673 NBDA could be modified to investigate the factors that make an individual more likely to be
674 an early adopter (low threshold), or one of the early (medium threshold) or late majority
675 (high threshold).

676 Currently, methods for analysing diffusion data tend to assume that individuals fall
677 into one of two binary categories, 'naïve' or 'informed', and that both social transmission and
678 asocial learning result in a transition from the naïve to the informed state. Linked to this is the
679 assumption that all informed individuals demonstrate the trait at the same rate once they are
680 informed. *OADA* and *TADA* are no exceptions to these assumptions. In many cases the
681 reduction to 'naïve' and 'informed' categories is a useful simplification that enables us to
682 model social transmission in a relatively straight-forward manner. However, it is worth
683 noting that there may be some cases where this simplification is not appropriate, and that
684 both *OADA* and *TADA* might fail to adequately model the underlying process. An
685 individual's rate of performance of a trait is, in reality, a complex function of its own history
686 of trait performance, observation and reward. Accordingly, we envisage that the process of
687 acquisition may sometimes be better captured by a learning rule, such as Rescorla-Wagner
688 (e.g. Kendal et al., 2009). However, this would make modelling a diffusion a much more
689 challenging task, especially when data is limited.

690 Nonetheless, we envisage that the novel *OADA* method presented here, as well as
691 Franz and Nunn's (2009) *Network Based Diffusion Analysis* and our *TADA* extensions of it,

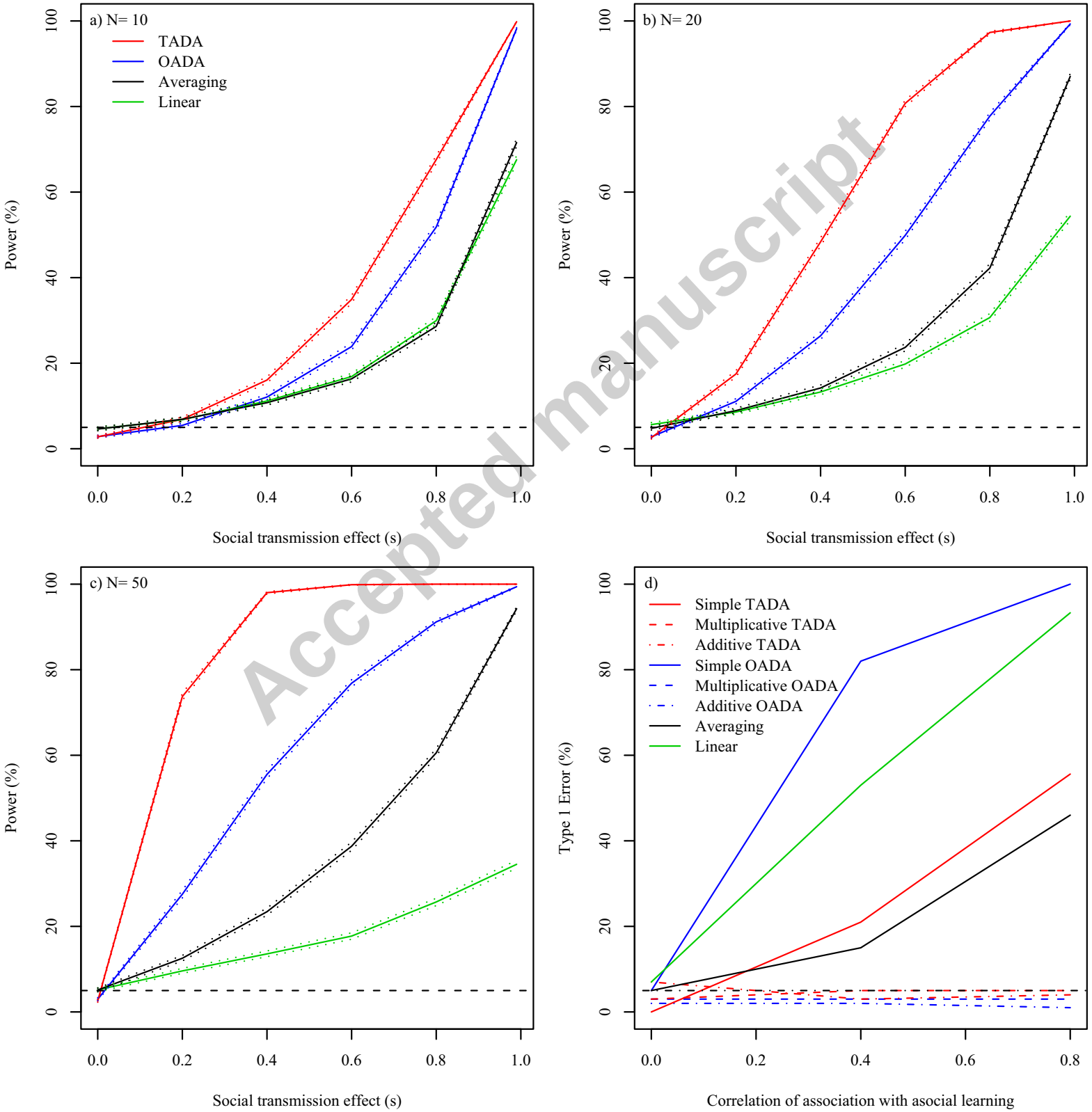
692 will provide a useful toolkit for those wishing to detect and quantify social transmission in
693 networks of animals, in captivity and the field. We hope that these methods will rejuvenate
694 interest in collecting and analysing diffusion data, and add statistical rigour to the study of
695 social transmission and culture be it in nonhuman animals or in humans.
696
697 **Acknowledgements**
698 WH was supported by a BBSRC grant (BB/D015812/1), NB by a McGill Milton Leong
699 Fellowship and KNL by grants from the BBSRC (BB/C005430/1 and BB/D015812/1) and an
700 ERC Advanced Grant (EVOCULTURE, ref 232823). We would like to thank the members of
701 the Laland Lab Journal Club for useful comments on an earlier draft, and to Laurel Fogarty,
702 Tess Hanrahan and Joel Higgin for 'test-driving' the R code.
703
704 **References**
705 Boogert, N.J., Reader, S.M., Hoppitt, W., Laland, K.N., 2008. The origin and spread of
706 innovations in starlings. Anim. Behav. 75, 1509-1518.
707 doi:10.1016/j.anbehav.2007.09.033.
708 Burnham K.P. & Anderson D.R., 2002. Model selection and multimodel inference: A
709 practical information-theoretic approach. 2nd edition. Springer, New York.
710 Coussi-Korbel, S. and Fragaszy, D.M., 1995. On the relation between social dynamics and
711 social learning. Anim. Behav. 50, 1441-1453.
712 Crawley, M.J., 2002. Statistical Computing: an Introduction to Data Analysis using S-Plus.
713 Wiley, Chichester, U.K.
714 Fisher, J. and Hinde, R.A., 1949. The opening of milk bottles by birds. Br. Birds 42, 347–
715 357.

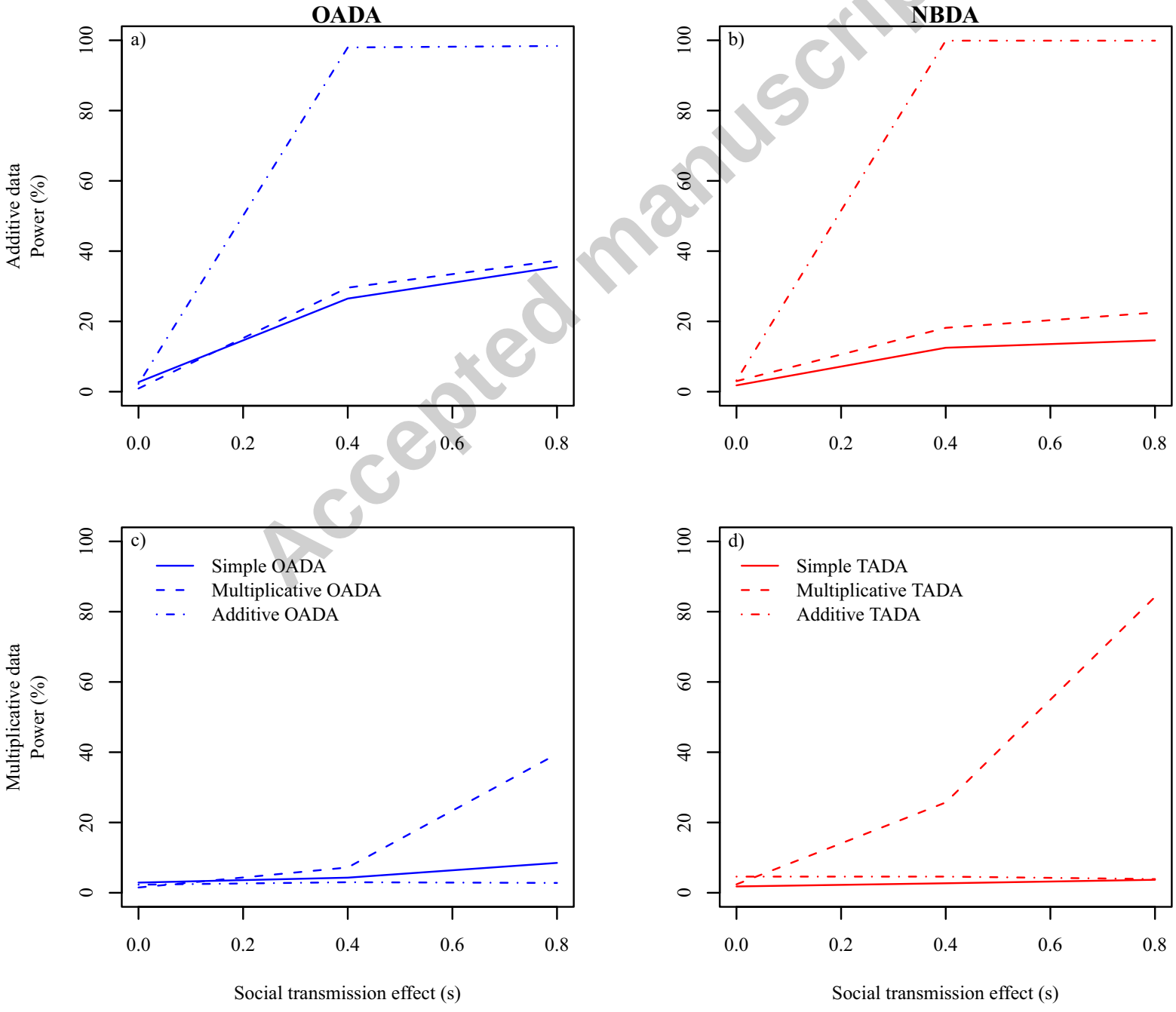
716 Franz, M., Nunn, C.L., 2009. Network-based diffusion analysis: a new method for detecting
717 social learning. Proc. R. Soc. B-Biol. Sci. 276, 1829-1836.,
718 doi:10.1098/rspb.2008.1824.
719 Galef, B.G., Jr., Allen, C., 1995. A new model system for studying behavioral traditions in
720 animals. Anim. Behav. 50, 705-717.
721 Giraldeau, L-A., Lefebvre, L., 1987. Scrounging prevents cultural transmission of food-
722 finding behavior in pigeons. Anim. Behav. 35, 387-394.
723 Heyes, C.M., 1994. Social-learning in animals: categories and mechanisms. Biol. Rev. 69,
724 207-231.
725 Hoppitt, W., Laland, K.N., 2008. Social processes influencing learning in animals: a review
726 of the evidence. Adv. Study Behav. 38, 105-165., doi: 10.1016/S0065-
727 3454(08)00003-X.
728 Keeling, M.J., 1999. The effects of local spatial structure on epidemiological invasions. Proc.
729 R. Soc. B-Biol. Sci. 266, 859-867,
730 Kendal, J.R., 2003 Ph.D. Thesis. University of Cambridge
731 Kendal, R.L., Kendal, J.R., Hoppitt, W., Laland, K.N., 2009. Identifying social learning in
732 animal populations: a new 'option-bias' method. PLoS ONE 4, e6541.,
733 doi:10.1371/journal.pone.0006541.
734 Kolaczyk, E.D. 2009. Statistical Analysis of Network Data: Methods and Models. Springer,
735 NY.
736 Krutzen, M., Mann, J., Heithaus, M.R., Connor, R.C., Bejder, L., Sherwin, W.B., 2005.
737 Cultural transmission of tool use in bottlenose dolphins. Proc. Natl. Acad. Sci. USA
738 102, 8939-8943.
739 Laland, K.N., 2004. Social learning strategies. Learn. Behav. 32, 4-14.

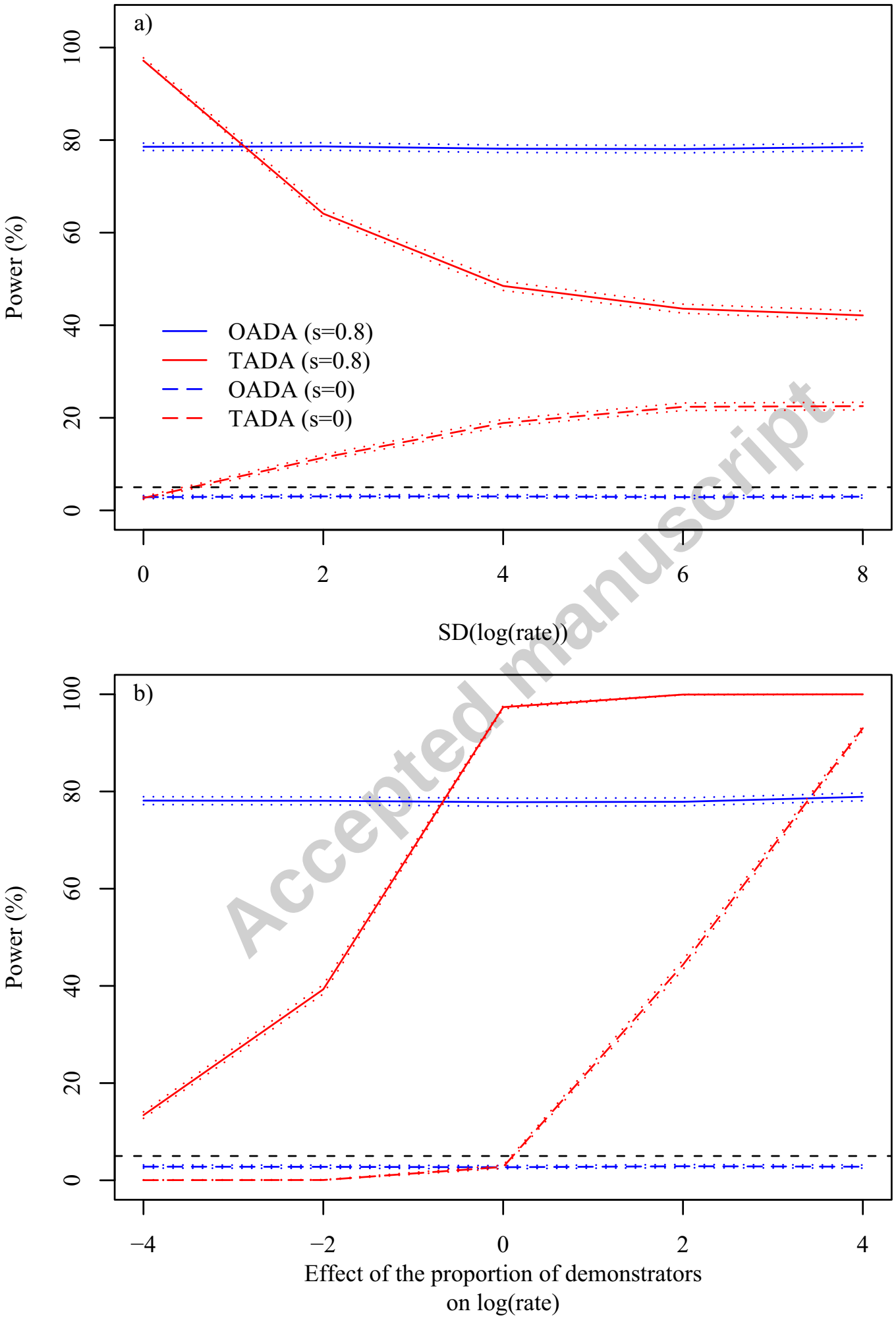
740 Laland, K.N., Galef, B.G., Jr., 2009. The Question of Animal Culture. Harvard University
741 Press, Cambridge, MA.
742 Laland, K.N., Hoppitt, W., 2003. Do animals have culture? Evol. Anthro. 12, 150-159.,
743 doi:10.1002/evan.10111.
744 Laland, K.N., Janik, V.M., 2006. The animal cultures debate. Trends Ecol. Evol. 21, 542-
745 547., doi:10.1016/j.tree.2006.06.005.
746 Laland, K.N., Kendal, J.R., 2003. What the models say about social learning. In: Fragaszy, D.
747 and Perry, S. (Eds), The Biology of Traditions: Models and Evidence. Cambridge
748 University Press, Cambridge, UK, pp. 33-55.
749 Leadbeater, E., Chittka, L., 2007. The dynamics of social learning in an insect model, the
750 bumblebee (*Bombus terrestris*). Behav. Ecol. Sociobiol. 61, 1789-1796.,
751 doi:10.1007/s00265-007-0412-4.
752 Lefebvre, L., 1995a. The opening of milk bottles by birds- evidence for accelerating learning
753 rates, but against the wave-of-advance model of cultural transmission. Behav.
754 Process. 34, 43-53.
755 Lefebvre, L., 1995b. Culturally-transmitted feeding-behaviour in primates. Primates 36, 227-
756 239.
757 Manly, B.F.J., 2007. Randomization, Bootstrap and Monte Carlo Methods in Biology. 3rd
758 Edition. Chapman and Hall/CRC, London.
759 McGrew, W.C., 1998. Culture in nonhuman primates? Annu. Rev. Anthropol. 27, 301-328.
760 Meyers, L.A., M.E.J. Newman, B. Pourbohloul (2006) Predicting epidemics on directed
761 contact networks. Journal of Theoretical Biology 240: 400-418.
762 Morgan, B.J.T., 2009. Applied Stochastic Modelling. 2nd Edition. CRC, London.

763 Morrell, L.J., Croft, D.P., Dyer, J.R.G., Chapman, B.B., Kelley, J.L., Laland, K.N., Krause, J.
764 2008. Association patterns and foraging behaviour in natural and artificial guppy
765 shoals. Anim. Behav. 76, 855-864., doi:10.1016/j.anbehav.2008.02.015.
766 Newman, M., Barabasi, A-L., Watts, D.J., 2006. The Structure and Dynamics of Networks.
767 Princeton University Press, Oxford, UK.
768 Noad, M.J., Cato, D.H., Bryden, M.M., Jenner, M.N., Jenner, K.C.S., 2000. Cultural
769 revolution in whale songs. Nature. 408 (6812), 537.
770 Perry, S., Manson, J.H., 2003. Traditions in monkeys. Evol. Anthro. 12, 71-81.,
771 doi:10.1002/evan.10105.
772 Pinheiro, J.C., Bates, D.M., 2000. Mixed-Effects Models in S and S-Plus. Springer, New
773 York.
774 Reader, S.M., 2004. Distinguishing social and asocial learning using diffusion dynamics.
775 Learn. Behav. 32, 90-104.
776 Rendell, L., Whitehead, H., 2001. Culture in whales and dolphins. Behav. Brain Sci. 24 (2)
777 309-324.
778 Therneau, T.M., Grambsch, P.M., 2001. Modeling Survival Data: Extending the Cox Model.
779 Springer, New York.
780 Thorpe, W.H., 1956. Learning and Instinct in Animals. Methuen, London.
781 van Schaik, C.P., Ancrenaz, M., Borgen, G., Galdikas, B., Knott, C.D., Singleton, I., Suzuki,
782 A., Utami, S.S., Merrill, M., 2003. Orangutan cultures and the evolution of material culture.
783 Science 299(5603), 102-105.
784 Watts, D.J., 2002. A simple model of global cascades on random networks. PNAS, 99, 5766-
785 5771. doi: 10.1073/pnas.082090499.
786 Weisberg, S., 1980. Applied Linear Regression. Wiley, Chichester, U.K.

787 Whiten, A., Goodall, J., McGrew, W.C., Nishida, T., Reynolds, V., Sugiyama, Y., Tutin,
788 C.E.G., Wrangham, R.W., Boesch, C., 1999. Cultures in chimpanzees. Nature
789 399(6737), 682-685.
790 Whiten, A., Horner, V., de Waal, F.B.M., 2005. Conformity to cultural norms of tool use in
791 chimpanzees. Nature 437(7059), 737-740., doi:10.1038/nature04047.







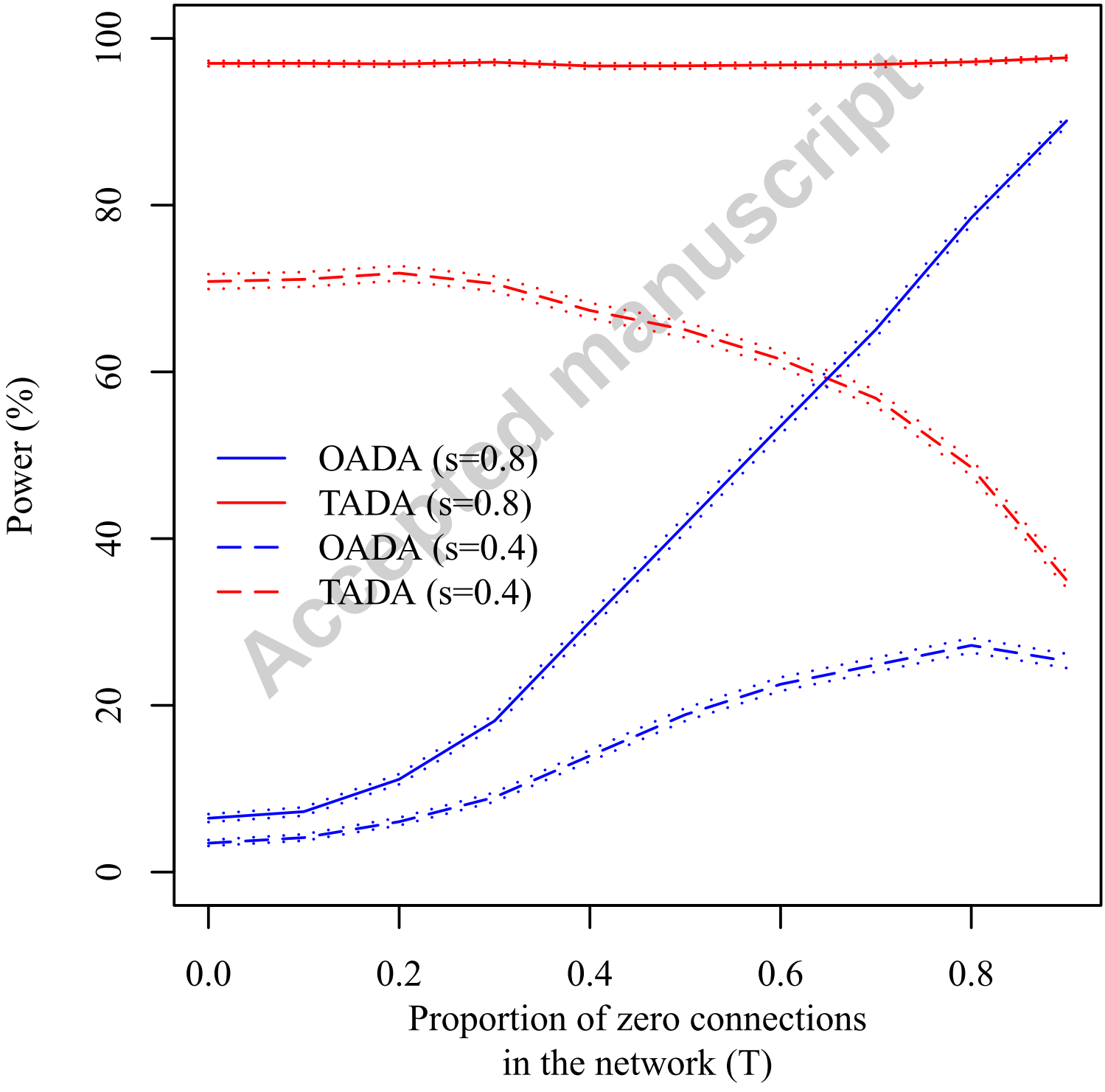


Table 1

Model used	Null model	Null model AIC_c	Estimated social transmission effect (s) 95% CI LRT (H_0: $s'=0$)[a]			Estimated effects for individual-level variables LRT (H_0: $\beta=0$)[a]			AIC_c for a model including social transmission for groups:		
			Group 1	Group 2	Group 3	ON[c]	AL[c]	LFNE[c]	1	1 & 2	All
Additive OADA	Object neophobia +Asocial learning	138.4	≈1 [b] [0.08,1) $\chi^2_1 = 5.53$, p = 0.018	0.06 [0,1) $\chi^2_1 = 0.12$, p = 0.734	0 [0,0.13] $\chi^2_1 = 0$, p = 1	-0348 $\chi^2_1 = 2.87$ p = 0.090	-0.350 $\chi^2_1 = 3.18$ p = 0.075		135.08	137.26	139.63
	Latency to feed in a novel environment	138.4	≈1 [b] [0.06,1) $\chi^2_1 = 4.15$, p = 0.042	0.12 [0,1) $\chi^2_1 = 0.51$, p = 0.477	0 [0,0.22] $\chi^2_1 = 0$, p = 1			-0.112 $\chi^2_1 = 1.89$ p = 0.169	136.39	138.10	140.39
Multiplicative OADA	Object neophobia +Asocial learning	138.4	≈1 [b] [0.05,1) $\chi^2_1 = 5.47$, p = 0.019	0.02 [0,1) $\chi^2_1 = 0.02$, p =0.898	0 [0,0.20] $\chi^2_1 = 0$, p = 1	-0.247 $\chi^2_1 = 2.61$ p = 0.106	-0.322 $\chi^2_1 = 3.18$ p = 0.075		135.14	137.42	139.79
	Latency to feed in a novel environment	138.4	≈1 [b] [0.03,1) $\chi^2_1 = 4.68$, p = 0.031	0.18 [0,1) $\chi^2_1 = 0.39$, p = 0.531	0 [0,0.24] $\chi^2_1 = 0$, p = 1			-0.117 $\chi^2_1 = 2.42$ p = 0.120	135.86	137.69	140.37

[a] LRT's are for significant parameters dropped from the final model, and for non-significant parameters when added to the final model. In all cases the final model included a social transmission effect for group 1 only. df=1 in all cases.

[b] The likelihood continued to increase towards the upper bound 1, though s=1 has a likelihood of zero, since the first acquisition would be impossible (e.g. see Appendix D Fig D.1).

[c] ON= Object neophobia; AL= Asocial learning ability; LFNE= Latency to feed in a novel environment.

Table 2

Model used	Null model	Null model AIC_c	Estimated social transmission effect (s) 95% CI LRT (H_0: s'=0)a	Estimated effects for individual-level variable (LFNE[b]) 95% CI LRT (H_0: β=0)[a]	AIC_c including social transmission
Additive TADA	Latency to feed in a novel environment	1175.56	0.16 (0.08, 0.43) $\chi^2_1 = 15.54$, $p < 0.001$	-0.23 (-0.39, -0.04) $\chi^2_1 = 5.04$, p =0.025	1162.24
Multiplicative TADA	Latency to feed in a novel environment	1175.56	0.17 (0.08, 0.33) $\chi^2_1 = 16.75$, $p < 0.001$	-0.16 (-0.36, -0.01) $\chi^2_1 = 6.80$, p =0.009	1161.02

[a] LRT's are for significant parameters dropped from the final model.
[b] LFNE= Latency to feed in a novel environment.