Erratum

Erratum in paper JR Kendal, RL Kendal and KN Laland (2007) *Quantifying and Modelling Social Learning Processes in Monkey Populations* [Cuantificación y Modelado de Procesos de Aprendizaje Social en Poblaciones de Monos] originally published in Vol. 7 (2), pp.123-138.

The erratum is in page 133 Subsection 'Model Selection', paragraph 1, lines 7 and 8

It read:	*Should read:*
(…) 'best model' (case (i) AICc = 38.235,Akaike weighting, $w_{m,a}$> 0= 0.749; case (ii) AICc= 155.175, $w_{m,a}$> 0> 0.999).	(…) 'best model' (case (i) AICc = 38.235,Akaike weighting, $w_{m,a > 0}$= 0.749; case (ii) AICc= 155.175, $w_{m,a > 0}$> 0.999).

We apologize for this inconvenient.

Quantifying and Modelling Social Learning Processes in Monkey Populations

Jeremy R Kendal, Rachel L Kendal, Kevin N Laland

University of Cambridge, UK

ABSTRACT

Putative traditions in animal populations remain unsubstantiated in the absence of methods to isolate the mechanisms of social transmission in social groups. Here we address this problem by quantifying the effect of two social learning processes, namely stimulus enhancement and observational learning, on the adoption of a novel extractive foraging behavior in groups of *callitrichid* monkeys. We simulate the effect of these processes in a model for the spread of a novel behaviour and also select which model-parameters provide the best fit to the monkey data. Our analysis reveals evidence for asocial processes and stimulus enhancement but not observational learning. The latency to solve the tasks co-varied with the strength of the asocial, but not social, processes. Further, only asocial parameters were required for the model to fit the data. Both model and monkey diffusion data exhibited deceleratory diffusion curves. We discuss the relationship between both the asocial and social processes and the diffusion dynamics.
Keywords: animal behavior, culture, learning, social learning, traditions.

RESUMEN

La existencia de supuestas tradiciones en las poblaciones animales sigue sin resolverse en ausencia de métodos que permitan aislar los mecanismos de transmisión social en los grupos sociales. Aquí tratamos este problema mediante la cuantificación del efecto que jugaron dos procesos de aprendizaje social, el realce del estímulo y el aprendizaje observacional, en la adopción de un comportamiento novedoso de recolección en grupos de monos de la familia *Callitrichidae*. Simulamos el efecto de estos procesos en un modelo de cara a explicar la extensión de un comportamiento novedoso y para seleccionar que parámetros del modelo proporcionaban el mejor ajuste a los datos del mono. Nuestro análisis revela evidencias para los procesos asociales y el realce estimular, pero no para aprendizaje observacional. Las latencias empleadas en la solución de las tareas covarió con la fuerza de los procesos asociales, pero no con la de los procesos sociales. Más aún, el modelo solo requirió los parámetros de los factores no sociales para ajustar los datos. El modelo y los datos de difusión mostraron ambos curvas desaceleradas de difusión. Finalmente, discutimos la relación entre los procesos asociales y sociales y la dinámica de la difusión cultural.
Palabras clave: comportamiento animal, cultura, aprendizaje, aprendizaje social, tradiciones.

*Correspondence concerning this article should be addressed to the first author: School of Biology, University of St Andrews, St Andrews, Fife, KY16 9TS, U.K. Email: jeremykendal@gmail.com

Traditionally, social and asocial learning processes have been distinguished in the laboratory by exposing naïve individuals ('observers') to the behavior of trained individuals ('demonstrators') in observer-demonstrator pairs. Following removal of the demonstrators, the behavior of the observers is recorded and compared to the behavior of animals in control conditions, such as non-observers (see, for example, Galef, 1988; Heyes *et al.*, 1994; Zentall, 2004). While this method is of great utility in investigating the psychological processes that underlie social learning, it sheds little light on whether or how social learning might propagate behavior patterns in natural or captive animal populations and its use is usually impractical in such contexts. Moreover, experiments have established that the demonstration of social learning in paired demonstrator-observer designs does not guarantee the propagation of the target behavior through a population (Laland & Plotkin, 1993), thus compromising the applicability of lab-based findings to natural or captive animal populations.

In natural animal populations, patterns of intra- and inter-population variation in behavioral repertoires have been studied to identify behavioral traditions in capuchin monkeys (Perry *et al.*, 2003), cetaceans (Rendell & Whitehead, 2001; Krützen *et al.*, 2005) and apes (McGrew, 1998; Whiten *et al.*, 1999; van Schaik *et al.*, 2003). Galef (2004), however, has expressed concern that the validation of a tradition often remains unresolved without determining how the purported traditional behavior develops in an individual; specifically, whether it is learned socially. What is required then, is a technique to measure the influence of social learning processes in naturalistic social settings. Here, we show how the asocial processes of intrinsic movement and asocial learning, and the social processes of stimulus enhancement and observational learning, can be quantified using data collected from captive groups of callitrichid monkeys.

The effect of these asocial and social processes on the spread of a novel behavior through a population can be examined using a simple mathematical model. The validity of the model can be evaluated by comparing the generated patterns of spread of the novel behavior with those observed in the groups of animals.

While there has been little development of mathematical models that might explain the diffusion dynamics of novel behavioral traits, there has been even less of an attempt to fit such models to empirical data sets in non-human animals. One exception is that of Lefebvre and Palameta (1988), who present a stochastic model of the diffusion dynamics of extractive foraging behavior in populations of pigeons. They showed that information can diffuse through a population, open to migration, as a result of frequency-dependent social learning. Their model accurately predicted that diffusion was more limited in a closed than an open population as the majority of individuals scrounged from the extractive foragers.

Here, we describe how data from animal populations can be used to derive estimates for asocial and social learning parameters that can then be applied to a model for the spread of a novel behavior. We illustrate the method using data taken during the diffusion of novel extracting foraging behavior in groups of callitrichid monkeys. We assess the validity of the model by comparing both the generated latency to perform the novel behavior and the shape of the diffusion curves with that of the callitrichid data.

METHODS

Subjects and Apparatus

We studied 20 single-species groups of zoo-housed callitrichids (5 L. chrysomelas, 5 L. rosalia and 5 L. chrysopygus, 3 C. argentata and 2 C. geoffroyi), totalling 81 individuals. The subjects, ranging from 6.5 months to 18.5 years, were in group sizes (2-9 individuals) and compositions (mated pairs or families) within the bounds of those seen in the wild.

Each group was exposed to a number of extractive foraging tasks over separate trials in a randomised order and with at least an hour between trials. The two tasks analysed here, labelled the 'round-box' and 'flip-top' boxes, were both opaque white plastic boxes of different shapes, with spatially separated options, or doors, that the animals could open to extract raisins, contained within (also see Kendal *et al.*, 2005a). When the boxes were closed, subjects had limited visual and olfactory access to the raisins. The options could not be used simultaneously and were distinguished by colors (blue and either green, red or yellow) visible to all (di- and trichromatic) individuals. The task options were equivalent in every other respect. There was no effect of colour preference across all groups for either task and there was no scrounging. The tasks were designed to be solved using foraging actions, natural to all genera, such as employed when turning over bark, exploring crevices and rummaging in leaf litter.

Subjects were marked with animal marking inks for individual identification. Prior to task presentation each group of subjects was provided with raisins to ensure that they were familiar with the food reward.

Behavioral data and statistical analysis

Each trial began with the presentation of a novel task to a group and laste© Inte 30 minutes or until all of the food reward (multiple raisins) had been extracted, whichever occurred sooner. For each of 29 such trials (15 trials and 14 trials for each of the two respective tasks, where 9 of the 20 groups provided data for both tasks), the identity of subjects that were 'at proximity', that is within approximately a 50cm radius of the task, was measured every 10 seconds. We also recorded the latency for each individual to solve the task, that is, to have extracted a raisin. For each trial, we calculated the median latency to solve the task across individuals, rather than the mean, to avoid systematic underestimation induced by ceiling values.

We conjecture that the spread of the extractive foraging behavior may be influenced by movement to the task and by learning[1] at the task. We refer to asocial and social influences on movement of naïve individuals to the task (i.e. 'naïve' are those yet to solve the task) as 'intrinsic movement' and 'stimulus enhancement', respectively, and asocial and social influences on learning at the task as 'asocial learning' and 'observational learning', respectively. Furthermore, the rate of spread may also be influenced by movement of naïve individuals away from the task, affecting their opportunity to learn the extractive foraging behaviour, and also demonstrators' movement to and from the

task, influencing the rate of demonstration. We calculate values for each of these seven parameters to quantify their effects over an average 10s time period[2] per trial (see Table 1). For instance, stimulus enhancement is defined as the probability that the demonstration of a successful task manipulation between time t=0 and t=10 seconds results in the movement of a naïve individual from 'at a distance' at t=10 to 'at proximity' to the task at t=20.

The social parameters were defined to measure the influence of a demonstration during the preceding time period and thus were not calculated for time steps where no such demonstration occurred, and visa versa for asocial parameters. Parameters measuring the rate of movement away from the task and demonstrators' (i.e. individuals that had solved the task) movement towards the task were assumed to be independent of the frequency of demonstrations. Our derivations of the social processes, stimulus enhancement and observational learning, are consistent with commonly used definitions (e.g. Galef 1988; Whiten and Ham, 1992; Zentall 1996).

The parameter values for the two social processes were compared with those of the default, asocial process that may influence movement to the task and learning at the task. Thus, we examined the difference in parameter values between stimulus (or local) enhancement (s) and intrinsic movement (m) and also between observational learning (b) and asocial learning (a). These comparisons were made using paired t-tests to account for their non-independence (pairing the average parameter values per trial for each comparison).

For each trial, values for each of the seven parameters, averaged across trials, were applied to a simple model for the spread of novel behavior at a food patch (see Box 1). We tested whether simulated latencies and the shape of simulated diffusion curves were consistent with the monkey data. Here, the simulated and observed data are non-independent as parameter values used in the simulations are derived from observed data (see Table 1).

We performed a model selection procedure to compare the predictive power between model variants, (see Box 1). We allowed parameter values to vary freely (between 0 and 1 ±0.05) so that, in this exercise, the simulated latencies were independent of the monkey data. We used Akaike's Information Criterion for small sample sizes (AICc) to compare the fit to the monkey data (based on residual sum of squares) of different models. The Akaike weight, w_i, was calculated to estimate the probability that model i was the correct model out of the set of model variants (Johnson & Omland, 2004). We also assessed whether the 'best model' can be used as a tool to predict parameter values by comparing parameter values that minimised the error between the simulated best model and observed latencies with parameter values calculated from the observed data (Table 1).

The best-fit function of the observed diffusion curves were analysed using least-squares curve fitting analysis. We assumed that the function with the lowest degrees of freedom-adjusted r-squared value provided the best fit to the data. We considered linear, logarithmic, inverse, quadratic, logistic, growth, sigmoid and exponential functions. Curves were fitted for the 12 trials where at least four individuals solved the task (to avoid redundancy between curve functions). The shape of the diffusion data from these

Table 1. Parameters definitions: Each time step (t, $t+1$) represents an increment of 10 seconds. At each time step, both stimulus enhancement and observational learning correct for the influence of the appropriate asocial factor within the same trial. The values of each of these correction factors, appeared to be consistent between tasks within groups (Paired samples test: $\overline{m}$: $t=0.073$, $df=8$, $P=0.994$ and $\overline{a}$: $t=0.25$, $df=7$, $P=0.81$).

Parameter	Calculation per time step, t	Calculating the Mean Value per Trial
Intrinsic Movement to the Task ($m_{N,towards}$)[1]	The intrinsic movement coefficient $$m^t_{N,towards} = \frac{n_{ND^t \rightarrow P^{t+1}}}{n^t_{ND}}, \qquad (1)$$ is the number of naïve individuals at a distance from the task at time t, n^t_{ND}, that move to be at proximity to the task at time $t+1$, $n_{ND^t \rightarrow P^{t+1}}$. This parameter is calculated at each time step for which $n^t_{ND} > 0$ and there are no successful manipulations of the task during the previous time step, $n^t_{dem} = 0$.	The mean value of m^t, calculated across the total number of time steps, $T_{(n_{dem}=0)}$ for which the intrinsic movement parameter is calculated, is given by $$\overline{m}_{N,towards} = \frac{1}{T_{(n_{dem}=0)}} \sum_{t=1}^{T_{(n_{dem}=0)}} m^t .(2)$$
Stimulus Enhancement (s)	The stimulus enhancement coefficient, calculated at each time step for which $n^t_{ND} > 0$ and there is a successful demonstration of the foraging behaviour during the previous time step (i.e. $n^t_{dem} > 0$), is given by $$s^t = \frac{1}{n^t_{dem}} \left(\frac{n_{ND^t \rightarrow P^{t+1}}}{n^t_{ND}} - \overline{m} \right). \qquad (3)$$ The coefficient accounts for the mean probability of moving to the task by intrinsic movement for that trial, $\overline{m}$, and is scaled by the number of successful manipulations of the task[2] during the previous time step, n^t_{dem}, so that it represents the effect of a single demonstration.	The mean value of s^t, calculated across the $T_{(n_{dem}>0)}$ time steps for which the stimulus enhancement coefficient is calculated, is given by $$\overline{s} = \frac{1}{T_{(n_{dem}>0)}} \sum_{t=1}^{T_{(n_{dem}>0)}} s^t. \qquad (4)$$
Asocial Learning (a)	The asocial learning coefficient, $$a^t = \frac{n_{NP^t \rightarrow L^{t+1}}}{n^t_{NP}}, \qquad (5)$$ is the number of naïve individuals at proximity to the task at time step t, n^t_{NP}, that manipulate the task successfully by time $t+1$, $n_{NP^t \rightarrow L^{t+1}}$. The parameter is calculated when $n^t_{NP} > 0$ and there are no successful demonstrations during the previous time step (i.e. $n^t_{dem} = 0$).	The mean value of a^t, calculated across the $T_{(n_{dem}=0)}$ time steps for which the asocial learning coefficient is calculated, is given by $$\overline{a} = \frac{1}{T_{(n_{dem}=0)}} \sum_{t=1}^{T_{(n_{dem}=0)}} a^t. \qquad (6)$$
Observational Learning (b)	The observational learning coefficient, calculated at time steps where $n^t_{NP} > 0$ and demonstrations take place during the previous time step (i.e. $n^t_{dem} > 0$), is given by $$b^t = \frac{1}{n^t_{dem}} \left(\frac{n_{NP^t \rightarrow L^{t+1}}}{n^t_{NP}} - \overline{a} \right). \qquad (7)$$ The coefficient accounts for the mean probability of learning asocially within the trial ($\overline{a}$) and is scaled by the number of demonstrations during the previous time step.	The mean value of b^t, calculated across the $T_{(n_{dem}>0)}$ time steps for which the observational learning coefficient is calculated, is given by $$\overline{b} = \frac{1}{T_{(n_{dem}>0)}} \sum_{t=1}^{T_{(n_{dem}>0)}} b^t. \qquad (8)$$

[1] Movement of naïve and informed individuals away from the task, and movement of informed individuals towards the task were calculated in an analogous fashion, however, they were calculated for time steps independent of the number of demonstrations (i.e. $n^t_{dem} \geq 0$).

[2] The alternative assumption, that stimulus enhancement is a function of the number of individuals at proximity to the task (n_P), made no significant difference to the results.

Box 1. Modelling the diffusion of a novel foraging behaviour.

We model the frequency of individuals that are naïve and at a distance from the task, f_{ND}, the frequency that are naïve and at proximity to the task, f_{NP}, the frequency that have learned the novel behaviour and at proximity to the task, f_{LP}, and the frequency that have learned the novel behaviour and are at a distance from the task, f_{LD}. Between sequential time steps, t and $t+1$, we assume that individuals may move and then learn to solve the task. Naïve individuals move towards the task as a result of their rate of intrinsic movement, $m_{N,towards}$ and the effect of stimulus enhancement, sf_{LP}^t. Informed individuals move towards the task also move away from the task with a probability $m_{N,away}$ and $m_{L,away}$, respectively. Following movement, naïve individuals at the task learn the novel food-extraction behaviour as a result of asocial learning, with a probability a, and observational learning, with a probability bf_{LP}^t. The effect of both social learning processes, stimulus enhancement and observational learning, are assumed to be affected by the frequency of demonstrators, that is individuals that have learned the novel behaviour and are proximal to the task, f_{LP}, and weighted by their respective coefficients, s and b. We found that an alternative assumption, that stimulus enhancement was dependent on the frequency of individuals at proximity to the task $\left(f_P = f_{LP} + f_{NP} \right)$, made no significant difference to the results.

The frequency of the four states of individuals at time $t+1$ can be expressed as the following set of recursions:

$$f_{ND}^{t+1} = f_{ND}^t \left[1 - \left(m_{N,towards} + sf_{LP}^t \right) \right] + f_{NP}^t m_{N,away} ,$$

$$f_{NP}^{t+1} = \left[f_{NP}^t \left(1 - m_{N,away} \right) + f_{NA}^t \left(m_{N,towards} + sf_{LP}^t \right) \right] \left[1 - \left(a + bf_{LP}^t \right) \right] ,$$

$$f_{LP}^{t+1} = \left[f_{NP}^t \left(1 - m_{N,away} \right) + f_{NA}^t \left(m_{N,towards} + sf_{LP}^t \right) \right] \left(a + bf_{LP}^t \right) + f_{LP}^t \left(1 - m_{L,away} \right) + f_{LA}^t \left(m_{L,towards} \right) ,$$

$$f_{LD}^{t+1} = 1 - \left(f_{ND}^{t+1} + f_{NP}^{t+1} + f_{LP}^{t+1} \right) \text{ and the starting condition is } f_{ND}^{t=0} = 1.$$

Model Variants under Selection

We compare the following model variants:
(i) ($m_{N,towards}$, a) > 0, all other parameters set to zero.
(ii) ($m_{N,towards}$, $m_{N,away}$, a) > 0, all other parameters set to zero.
(iii) ($m_{N,towards}$, s, a, b) > 0, all other parameters set to zero.
(iv) ($m_{N,towards}$, $m_{N,away}$, s, a, b) > 0, all other parameters set to zero.
(v) ($m_{N,towards}$, $m_{N,away}$, s, a, b, $m_{L,towards}$, $m_{L,away}$) > 0.

trials is illustrated using a single standardized diffusion plot. We assessed whether the 'best model' accurately predicted the shape of observed diffusion curves.

Where multiple comparisons or tests were made, the family-wise error rate was controlled by reducing the significance level of alpha, designated in the text as α^*. For each family of comparisons or tests $\alpha^*= \alpha/c$, where $\alpha =0.05$ and c corresponds to the number of comparisons or tests. Data were treated as independent and equivalent between both trials and tasks (see Appendix). The statistical power of the tests below was based on $\alpha =0.05$ and Cohen's (1988) medium effect size (Howell, 1997). Any parametric statistics were performed after testing for normality of residuals (Kolmogorov-Smirnov test) and homogeneity of variance (Levene's test). Where indicated, data were log transformed to meet these criteria.

RESULTS

Descriptive Statistics

The median latency for an individual to first contact a task was approximately 300 seconds, and the median latency for an individual to solve a task was 550 seconds. On average, each subject performed 6.64 successful manipulations of a task per trial ($\pm SE = 1.11$).

Parameter Values and the Observed Latency to Solve the Task

The mean parameter values across trials for intrinsic movement, stimulus enhancement and asocial learning, but not observational learning, were significantly greater than zero (m: $t= 3.556$, $df= 28$, $P< 0.001$; s: $t= 4.580$, $df= 21$, $P< 0.001$; a: $t= 4.0280$, $df= 26$, $P< 0.001$; b: $t= 0.183$, $df= 20$, $P= 0.428$). There were no significant differences between the mean movement parameter values, m and s (Paired t test: $t= 0.408$, $df= 21$, $P= 0.688$; power≈ 0.76) or the mean learning parameter values, a and b (Paired t test: $t= 1.502$, $df= 20$, $P= 0.149$; power$= 0.74$; see Figure 1).

Intuitively, there are either asocial or social routes to approaching the task (through intrinsic movement or stimulus enhancement) and learning the solution to the task. Hence we might expect negative correlations between m and s and also a and b, respectively. Our prediction was confirmed for the two learning parameters, a and b (Pearson coefficient$= -0.591$, $N= 21$, $P= 0.005$, $\alpha^*=0.017$), and the negative correlation between the two movement parameters, m and s, approached significance (Pearson coefficient$= -0.413$, $N= 22$, $P= 0.056$, $\alpha^*= 0.017$). In the discussion we raise the possibility that these correlations may be artifactual.

We might also expect that individuals drawn to approach the task asocially will solve the task asocially, while those that approach the task for social reasons may solve the task through social learning. We did find that intrinsic movement was positively correlated with asocial learning (Pearson coefficient$= 0.665$, $N= 28$, $P< 0.001$, $\alpha^*=0.017$), but no significant correlation was found between the magnitude of stimulus enhancement and observational learning parameters across trials (Pearson coefficient$= -0.159$, $N= 18$,
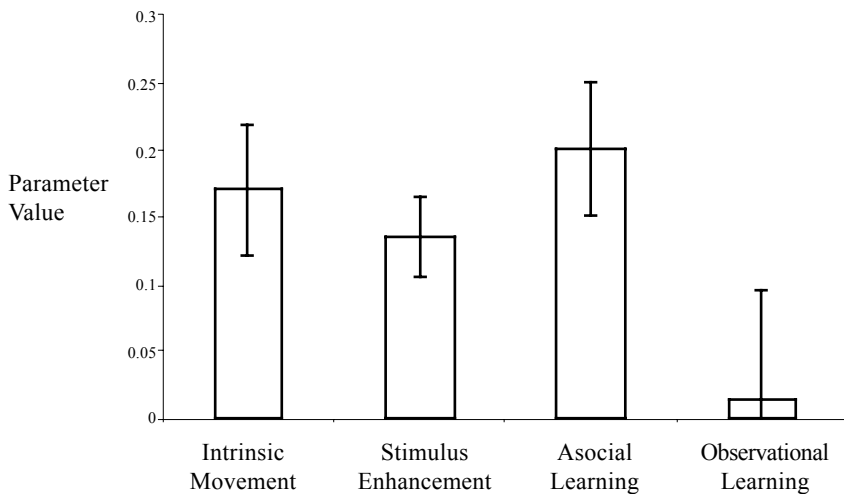
Figure 1. The mean value for each parameter across trials (±SE).

$P= 0.529$).

If any of the four parameters were critical to task solution, we would expect the latency to solve the task to be negatively correlated with its magnitude. It is important to note that testing for a correlation between latency and the magnitude of either of the social processes provides evidence, in a social context, that is either consistent, or not, with social learning. This test can potentially help to identify traditions in natural populations. Here, we found that the latency to solve the task was negatively correlated with each of the two non-social parameters, intrinsic movement (Pearson correlation coefficient= -0.525, $N= 29$, $P= 0.003$) and asocial learning (Pearson correlation coefficient= -0.479, $N= 28$, $P= 0.01$). There was no significant relationship, however, between the latency and either stimulus enhancement or observational learning.

The Model Latency to Solve the task

The simulated median latencies to solve a task, generated from the full model (see Box 1) using mean parameter values calculated from each trial, correlated positively with the observed median latencies to solve the task across 23 trials (Pearson Correlation coefficient, $r= 0.632$, $N= 23$, $P= 0.001$; Figure 2a). If the model and observed latencies had correlated perfectly, there would be an intercept of zero and a gradient of one (dashed line). In fact, the best-fit line intercepted the observed latency axis significantly above zero ($t= 5.098$, $df= 22$, $P< 0.001$) and the gradient (= 0.39, SE= 0.104) was significantly shallower than a gradient of one ($t= 5.865$, $df= 21$, $P< 0.001$). A paired t-test showed no significant difference between the (log transformed) median observed and model latencies across trials (Figure 2b; $t=-0.932$, $df= 22$, $P= 0.362$). Overall, the simulated latencies were consistent with the observed data, but there was a tendency for
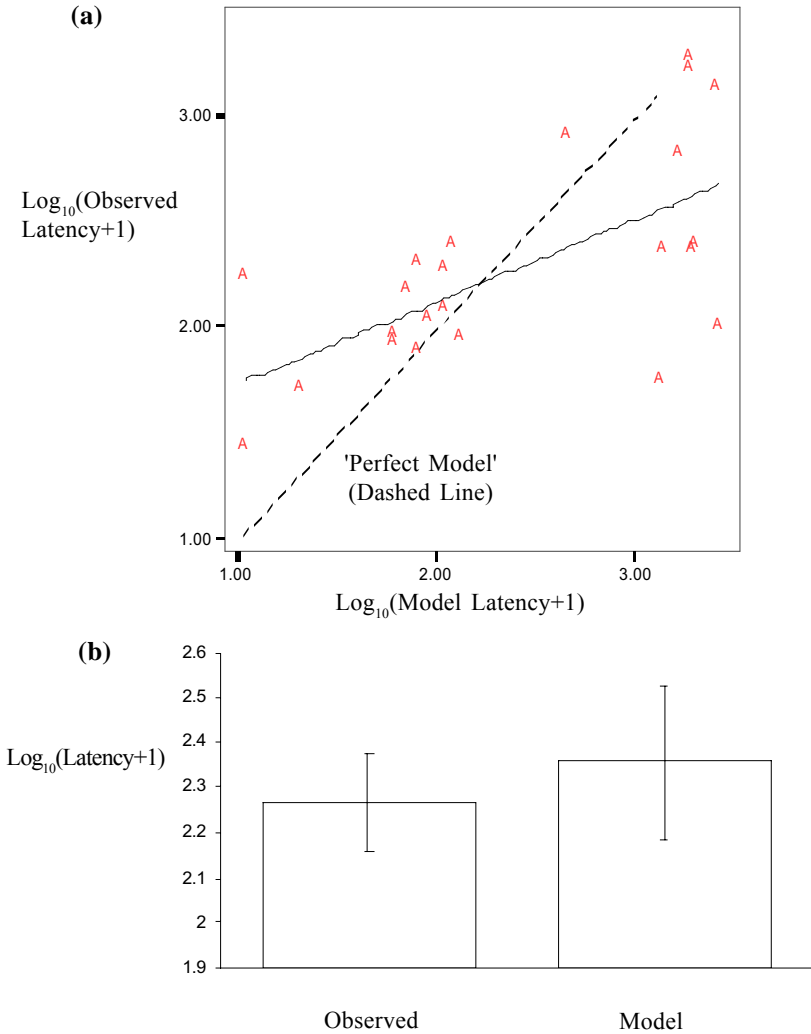
Figure 2. The relationship between the model and observed median latencies to solve the task per trial (each point represented by 'A'. The linear best-fit line (solid) is given by *y*= 0.39x+1.34 (where y represents the observed latency and x represents the model latency). The correlation coefficient, *r*= 0.632. Part (b) shows the model and observed mean of the median latencies across trials (±SE), to solve the task. In both parts, the model estimate of the median latency was generated using mean parameter values calculated for each trial.

the model to underestimate the latency (i.e. too fast) in trials when the task was solved quickly and overestimate the latency (i.e. too slow) when the task was solved slowly.

   Note that the preceding results cannot be used to assess the predictive power of

the model as the observed and simulated latencies for each trial were not independent: the simulations used parameter values gathered from the same trials used to collect observed latency data. The absence of sufficient data to split the dataset, using half to sample latencies and the other half to estimate parameters (Table 1) used in the model
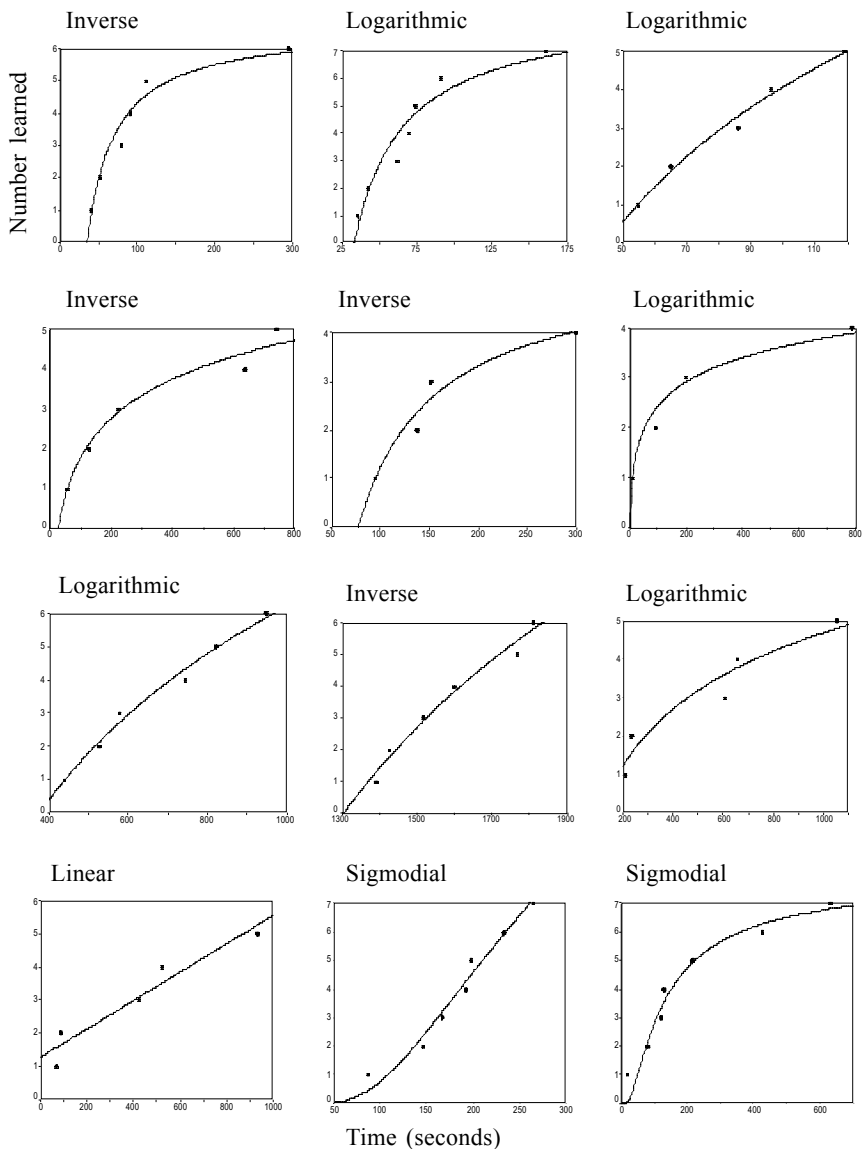

Figure 3. The best-fit curve functions for the cumulative number of individuals to solve the task over time, for 12 trials. In each case, the best-fit curve function is indicated above the curve.

to simulate latencies, is likely to be a widespread problem, particularly in studies of natural populations. The proceeding section compares the predictive power of variations in the model by treating the parameters independently of the monkey data (allowing their values to vary freely) and assesses the fit of the simulated and observed latencies between model variants.

Model Selection

We compared the model variants outlined in Box 1. For each model, we searched parameter space to find (i) a single estimate of the median latency that minimized the residual sum of squares (RSS) across all trials (i.e. parameter values held constant between trials) and (ii) separate estimates of the median latencies for each trial that, when summed over trials, minimized the RSS (i.e. parameter values allowed to vary between trials). In both cases, we found that the purely asocial model variant ($m > 0$, $a > 0$, all other parameters held at zero) was the 'best model' (case (i) AICc = 38.235, Akaike weighting, $w_{m,a} > 0 = 0.749$; case (ii) AICc = 155.175, $w_{m,a} > 0 > 0.999$). These results support the finding from the parameter calculations that the asocial processes had more effect on the latency to solve the task than the social processes.

The Shape of Diffusion Curves

Figure 3 shows that deceleratory functions (logarithmic and inverse) provided the best fit for most (9 out of 12) of the diffusion curves. The solid line in figure 4 summarizes this finding, showing that standardized diffusion data (see legend) from all 12 trials follows a deceleratory trajectory.

We were interested to know whether the parameter values that generated the best fit between the model and the monkey latencies might also predict the shape of the monkey diffusion curves. Figure 4 shows simulated diffusion curves from the 'best' model variant ($m > 0$, $a > 0$, all other parameters held at zero), using the mean parameter values[3] that minimized the RSS, where parameters were either allowed to vary between trials or were held constant. The two cases generated very similar-shaped diffusion curves that were largely deceleratory. (Note that initially, there was a slight acceleratory period, despite the absence of social processes). Overall, the simulated curves predicted the deceleratory shape of the observed data but overestimated the normalized rate of spread.

DISCUSSION

We have attempted to quantify social and asocial processes that may influence the spread of a novel behavior in animal populations. Parameters representing these processes were estimated from actual data and applied to a diffusion model, allowing us to tease apart the effects of asocial and social processes on the diffusion dynamics. Potentially, this approach of measuring the magnitude of social processes, testing for their correlation with the latency to solve the task, and selecting between asocial and
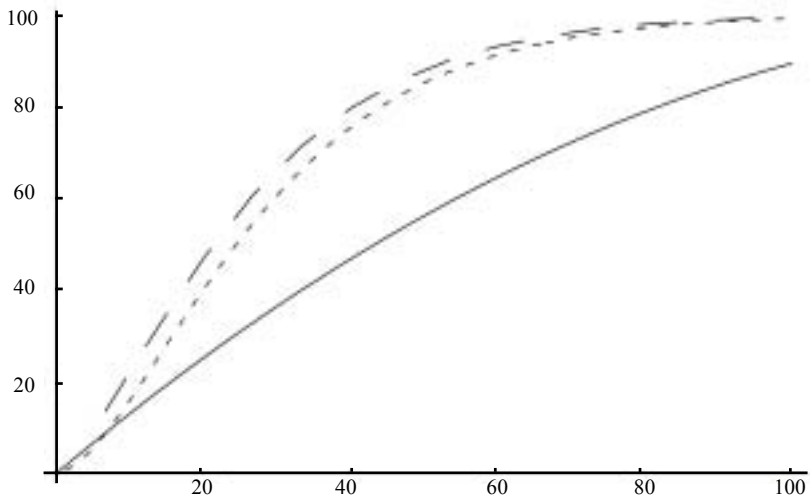
Figure 4. Observed (solid line) and model (dashed lines) diffusion curves. The solid line shows the best-fit function across all the (67) data points shown in figure 3, where for each of the 12 trials, the data points have been scaled to the proportion of individuals that solved the task (vertical axis), and the proportion of time taken between the start of the trial and the last individual to solve the task (horizontal axis). Diffusion curves predicted by the 'best model' ($m,a> 0$, all other parameters held at zero), using parameters that minimized the RSS, where parameters were (large dashes; $m= a= 0.367$) and were not (small dashes; $m= a= 0.085$) allowed to vary between trials. We stopped the simulation (i.e. 100% time) when $f_{LP}> 0.99$. (The difference in gradient between the observed and model curves is affected largely by the proportion of informed individuals at which we stop the simulation –e.g. the difference in gradient would have been steeper if the simulation were stopped at $f_{LP}> 0.999$).

social model variants can be used to distinguish socially contingent diffusions, traditions, or culture, from unlearned or asocially learned behavior.

Social learning theory predicts that information may often spread faster within groups that use social learning than within groups that do not (Giraldeau *et al.*, 1994). The results of this study were not consistent with this prediction as there was no relationship between the social parameter values and the latency to extract food from the novel tasks. Here, the social learning processes appeared to have less influence on the rate of adoption of the novel behavior than the asocial processes, perhaps because the novel tasks were not so difficult to solve as to require anything more than asocial learning (Lefebvre & Palameta, 1988; Zentall, 2004). Further analysis (submitted) suggests that social learning of a preference to use one of the two options (doors) to extract food from the task (rather than learning more quickly to use either option) may occur, and is more likely for difficult tasks than for easy tasks. Overall, social learning may not be preferred over asocial learning unless, for example, the task is too complex or costly

to solve through asocial learning alone (Boyd & Richerson, 1988; Byrne & Russon, 1998; Kendal *et al.*, 2005b). Furthermore, there may be variation in the capacity for social learning between species or genera.

Nonetheless, paradoxically, we found a positive mean value for the stimulus enhancement parameter. This relatively simple social learning process may be particularly widespread in animal populations (e.g. Hosey *et al.*, 1997; Fritz *et al.*, 2000; Reader *et al.*, 2003; Worden & Papaj, 2005). In our study, stimulus enhancement could have been dependent on either the frequency of task extractions or simply the frequency of individuals at the task (Sherry & Galef, 1991). It is perhaps curious that the incidence stimulus enhancement did not co-vary with the rate of task solution. This may reflect a non-linear effect of the number of demonstrations (or proportion of individuals at the task) per unit time (n^t_{dem}) on the stimulus enhancement coefficient, *s*. For instance, an individual's attention may be drawn away from the task when most group members are not interacting with it, and only drawn to the task when a substantive proportion of the group interacts with the task. This kind of nonlinear social influence on learning, characterised as conformity, has been reported in fish (Day *et al.*, 2001). There may also be a trade-off between reliance on stimulus enhancement and neophilia, as suggested by the negative correlation of *s* and *m*. Conceivably, stimulus enhancement may only enhance rates of learning in relatively neophobic individuals, which would perhaps explain why it generates no overall enhancement in learning rates.

The positive correlation found between intrinsic movement and asocial learning may provide support for a suggested association between neophilia and innovation (Greenberg, 1990). Further analysis of the callitrichid data is required, however, to confirm that the measures of intrinsic movement and asocial learning related specifically to novel rather than familiar objects. In birds, the association between neophilia and innovation may be found particularly in opportunistic generalist species (Greenberg, 1990; Lefebvre, 2000), for which exploration of novel environments encourages innovation (Greenberg, 2003). For example, Webster and Lefebvre (2001) found a correlation between neophilia and innovation across species of Barbadian passeriformes and columbiformes. Also, Sol *et al*. (2002) found that the invasion success of avian species, introduced into New Zealand, positively correlated with innovation frequency in their original homeland.

The negative relationship between the non-social parameters and the latency to extract food from the task supports the notion that high levels of intrinsic movement (perhaps in combination with neophilia) and asocial learning (perhaps resulting in innovation) may facilitate the discovery and extraction of novel sources of food.

There was some evidence that the social (*s* and *b*) and non-social (*m* and *a*) parameters were inversely related to one another. However, this result may have been an artefact of the parameter calculation methods as computation of each social parameter included a negative correction for the effect of the relevant non-social parameter. Also, the social parameters may have been less likely than the non-social parameters to be calculated from subjects that were fastest to approach and extract food from the task. Further, if demonstrations were rare, the mean social parameter values may include error from a low sample size. Independent calculation of the social and non-social

parameters is required to determine whether the inverse relationship is an artefact of the calculation method and also to remove any effect on the within-parameter variance that may have resulted from non-independent calculation. Also, the parameter calculations could be developed by accounting for the effect of previous interactions with the task and of previous observations of task manipulations by conspecifics.

The best-fit asocial model variant predicted the finding that the majority of observed diffusion curves were deceleratory. In contrast, Lefebvre (1995) found that a sigmoidal pattern of diffusion was most common for putative culturally-transmitted feeding behaviors in primates, although acceleratory functions were more common for diffusions over long, rather than short, time periods (years rather than minutes/days). Also in contrast to Lefebvre (1995), most of the diffusion curves in our study were asymptotic; all populations were 'closed' and typically, a high proportion of the population completed the task.

We regard this analysis as a tentative first step towards using modeling to draw inferences about underlying psychological processes in animal populations. Clearly, the model selection procedure could be improved upon. Future development might include a stochastic model that explicitly accounts for some error. Model selection can then proceed by generating likelihood functions for competing model variants and comparing AIC values. Nonetheless, the analysis is sufficient to illustrate that the approach has considerable potential. Quantifying asocial and social learning processes in animal populations, and selecting between the fit of asocial and social model variants to the observed data, should help researchers test the validity of claims of 'culture' and 'tradition' in animal populations.

NOTES

1. Solving the task is assumed to result from learning although information retention was not tested.
2. The parameter values were also calculated using a 60 second time scale but this proved an unreliable measure. Demonstrations were likely to influence movement over a faster time scale than 60s as it was not uncommon for both demonstrators and naïve subjects to move towards and away from the task multiple times within 60s. In addition, a subject often solved the task within 60 seconds of a demonstration. Hence, observational learning was likely to take place over a time scale that was faster than 60 seconds.
3. The mean value for each parameter was taken as there was redundancy in the effect of parameter values on the simulated median latency.

References

Boyd R & Richerson PJ (1988). An evolutionary model of social learning: The effects of spatial and temporal variation. In TR Zentall & BG Galef Jr. (Eds.) *Social Learning: Psychological and Biological Perspectives* (pp. 29-48). Erlbaum, New Jersey.

Byrne RW & Russon AE (1998). Learning by imitation: a hierarchical approach. *Behavior and Brain Science, 21,* 667-721.

Cohen J (1988). *Statistical power analysis for the behavioral sciences.* Hillsdale, New Jersey: Erlbaum.

Day R, MacDonald T, Brown C, Laland KN & Reader SM (2001). Interactions between shoal size and conformity in guppy social foraging. *Animal Behaviour, 62,* 917-925.

Fritz J, Bisenberger A & Kotrschal K (2000). Stimulus enhancement in greylag geese: socially mediated

learning of an operant task. *Animal Behaviour, 57,* 1119-1125.

Galef BG Jr. (1988). Imitation in animals: History, definition, and interpretation of data from the psychological laboratory. In TR Zentall & BG Galef Jr. (Eds.) *Social Learning: Psychological and Biological Perspectives* (pp. 3-28). Erlbaum, New Jersey.

Galef BG Jr. (2004). Approaches to the study of traditional behaviours of free-living animals. *Learning and Behavior, 32,* 53-61.

Galef BG Jr. & Allen C (1995). A new model system for studying behaviour traditions in animals. *Animal Behaviour, 49,* 1325-1334.

Galef BG Jr. & White DJ (1998). Mate-choice copying in Japanesee quail, *Coturnix coturnix japonica*. *Animal Behaviour, 55,* 545-552.

Giraldeau L-A, Caraco T & Valone TJ (1994). Social foraging: Individual learning and cultural transmission of innovations. *Behavioural Ecology, 5,* 35-43.

Greenberg R (1990). Feeding neophobia and ecological plasticity: a test of the hypothesis with captive sparrows. *Animal Behaviour, 39,* 375-379.

Greenberg R (2003). The role of neophobia and noephilia in the development of innovative behaviour of birds. In SM Reader & KN Laland (Eds.). *Animal Innovation* (pp. 175-196). Oxford: Oxford University Press, Oxford,.

Heyes CM, Jaldow E, Nokes T & Dawson GR (1994). Imitation in rats: conditions of occurrence in a bidirectional control paradigm. *Behavioral Processes, 32,* 173-182.

Hosey GR, Jacques M & Pitts A (1997). Drinking from tails: social learning of a novel behaviour in a group of ring-tailed lemurs (*Lemur catta*). *Primates, 38,* 415-422.

Howell DC (1997). *Statistical methods for psychology, fourth edition.* Belmont: Duxbury Press.

Johnson J.B., & Omland, K.S. (2004). Model selection in ecology and evolution. *Trends in Ecology and Evolution, 19,* 101-108.

Kendal RL, Coe RL & Laland KN (2005a). Age differences in neophilia, exploration and innovation in family groups of callitrichid monkeys. *American Journal of Primatology, 66,* 167-188.

Kendal RL, Coolen I, van Bergen Y & Laland KN (2005b). Trade-offs in the adaptive use of social and asocial learning. *Advances in the Study of Behavior, 35,* 333-379.

Krützen M, Mann J, Heithaus MR, Connor RC, Bejder L & Sherwin WB (2005). Cultural transmission of tool use in bottlenose dolphins. *Proceedings of the National Acadamy of Science, 102,* 8939-8943.

Laland, KN & Plotkin HC (1993). Social transmission of food preferences amongst Norway rats by marking of food sites, and by gustatory contact. *Animal Learning and Behavior, 21,* 35-41.

Lefebvre L & Palameta B (1988). Mechanisms, ecology and population diffusion of socially learned, food-finding behaviour in feral pigeons. In TR Zentall & BG Galef Jr. (Eds.) *Social Learning: Psychological and Biological Perspectives* (pp. 141-163). Erlbaum, New Jersey.

Lefebvre L (1995). Culturally-transmitted feeding behaviour in primates: evidence for accelerating learning rates. *Primates, 36,* 227-239.

Lefebvre L (2000). The origin and cultural transmission of feeding innovations in birds. In CM Heyes & L Huber (Eds.) *The evolution of Cognition* (pp. 311-328). Cambridge MA: MIT Press.

McGrew WC (1998). Culture in nonhuman primates. *Annual Review of Anthropology, 27,* 301-328.

Perry S, Baker M, Fedigan L, Gros-Louis J, Jack K, MacKinnon KC, Manson JH, Panger M, Pyle K & Rose L (2003). Social conventions in wild white-faced capuchin monkeys. *Current Anthropology, 44,* 241-268.

Reader SM, Kendal JR & Laland KN (2003). Social learning of foraging sites and escape routes in wild Trinidadian guppies. *Animal Behaviour, 66,* 729-739.

Reader SM (2004). Distinguishing social and asocial learning using diffusion dynamics. *Learning and Behavior, 32,* 90-104.

Rendell L & Whitehead H (2001). Culture in whales and dolphins. *Behavior and Brain Science, 24*, 309-324.

Sherry DF & Galef BG Jr. (1991). Social learning without imitation: more about milk bottle opening by birds. *Animal Behaviour, 40,* 987-989.

Sol D, Timmermans S & Lefebvre L (2002). Behavioural flexibility and invasion success in birds. *Animal Behaviour, 63,* 495-502.

van Schaik CP, Ancrenaz M, Borgen G, Galdikas B, Knott CD, Singleton I, Suzuki A, Utami, SS & Merrill M (2003). Orangutan cultures and the evolution of material culture. *Science, 299,* 102-105.

Webster SJ & Lefebvre L (2001). Problem solving and neophobia in a columbiform-passeriform assemblage in Barbados. *Animal Behaviour, 62,* 23-32.

Whiten A & Ham R (1992). On the nature and evolution of imitation in the animal kingdom: Reappraisal of a century of research. *Advances in the Study of Behavior, 21,* 239-283.

Whiten A, Goodall J, McGrew WC, Nishida T, Reynolds V, Sugiyama Y, Tutin CEG, Wrangham RW & Boesch C (1999). Cultures in chimpanzees. *Nature, 399,* 682-685.

Worden BD & Papaj DR (2005). Flower choice copying in bumblebees. *Biology Letters, 1,* 504-507.

Zentall TR (1996). An analysis of imitative learning in animals. In CM Heyes & BG Galef Jr. *Social Learning in Animals: The Roots of Culture* (pp. 221-244). San Diego: Academic Press.

Zentall TR (2004). Action imitation in birds. *Learning and Behavior, 32,* 15-23.

APPENDIX

Between-trial data were treated as independent and equivalent. This assumption followed the findings that the parameter values and latency to extract food were unaffected by the number of tasks or the type of task to which the group was exposed.

First, there was no difference in any of the parameters or the latency to extract food across trials between the 9 groups of monkeys exposed to both tasks and the 11 groups exposed to only one task (e.g. for Latency: $F_{1,27}= 0.699$, $P= 0.410$).

Second, repeated measures ANOVAs were performed on the values of the four parameters and the latencies, comparing groups that had been tested over two trials with a random set of pairings made up from all the remaining groups that had each been tested in one trial only. The repeated measures ANOVA test was repeated ten times for each of the four parameters and the latency, using different sets of random pairings in each repeat. There were no significant interactions between the type of pairing (i.e. 'two trial pairing or random pairing). And the repeated measure for any of the parameters of the latency to manipulate the task successfully (having adjusted for family-wise error rate). The results provided evidence that there was no difference in the variation between independent groups that were tested once and within groups that were tested twice.

Third, the 29 trials were treated as independent of the task presented as there was no effect of task on any of the four parameters or the latency to manipulate the task successfully (e.g. for Latency: $F_{1,27}= 2.164$, $P= 0.153$), although we could rule out a type II error.