UC Davis
UC Davis Previously Published Works

Title

Using Multilevel Models to Estimate Variation in Foraging Returns: Effects of Failure Rate, Harvest Size, Age, and Individual Heterogeneity

Permalink

https://escholarship.org/uc/item/2qk1v7rk

Authors

McElreath, Richard
Koster, Jeremy

Publication Date

2013-11-26

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Human Nature manuscript No.
(will be inserted by the editor)

Using Multilevel Models to Estimate Variation in Foraging Returns

Effects of Failure Rate, Harvest Size, Age, and Individual Heterogeneity

Richard McElreath · Jeremy Koster

In press manuscript, prior to copy editing

Abstract Distributions of human foraging success across age has implications for many aspects of human evolution. Estimating the distribution of foraging returns is complicated by (1) the zero-inflated nature of hunting returns, as many if not most trips fail, and (2) the substantial variation among hunters, independent of age. We develop a multilevel mixture analysis of human foraging data to address these difficulties. Using a previously published 20 year record of hunts by 147 individual Aché hunters in Eastern Paraguay, we estimate returns-by-age functions for both hunting failures and the size of harvests, while also estimating the heterogeneity among hunters. Consistent with previous analyses, we find that most hunters peak around 40 years of age. We can also show, however, that much more of the variation among Aché hunters arises from heterogeneity in failure rates (zero returns), not harvest sizes. We also introduce a new R package, `glmer2stan`, to assist in defining and fitting similar multilevel mixture models.

Keywords human behavioral ecology · foraging · multilevel modeling · life history

1 Introduction

Foraging returns vary for many reasons. Prey vary in value and cost. Individual foragers may be differently skilled, due to differences in endurance or strength, knowledge or memory. Across the lifespan, foraging returns vary, as young hunters can be substantially less productive than the old. Foraging returns also entail some risk, in that many hunting trips fail to produce any returns and others vary in yield, producing considerable trip-to-trip variation. In empirical data, all of these sources of variation—individual, age, and trip-to-trip

R. McElreath
Department of Anthropology
One Shields Avenue
UC Davis
Davis, CA 95616
E-mail: mcelreath@ucdavis.edu

J. Koster
Department of Anthropology
University of Cincinnati
Cincinnati, OH 45221

variation—are mixed together. Furthermore, as Jones et al. (forthcoming) have recently emphasized, it may be essential to carefully estimate variation, since foraging returns are non-linearly related to fitness, and therefore expected values may not lead to correct inferences about the value of prey items.

These aspects of the data make foraging records challenging to analyze. We examine two general difficulties. First, the outcome measure is a mixture of zeros and positive values, unlike any convenient statistical distribution. No effective transformation exists for such "zero-inflated" data. The simple statistical solution of averaging across trips, for example using rates of return, obscures trip-to-trip variation and discards information. Second, data of this sort tend to exhibit imbalance in sample size across individuals. Since hunters may vary in skill, both within and across ages, ignoring the unbalanced repeat measures can lead to misleading estimates. In some cases, such as estimating the effect of age on foraging returns, since no individual presents data from all ages, a failure to account for heterogeneity can reduce the precision of age estimates as well.

Because of these difficulties, previous analyses of hunting returns have made pragmatic compromises. Alvard (1995) and Koster (2008), for example, addressed the problematic outcome variable by using non-parametric comparisons of trip-by-trip returns, calculated as kilograms per hour. And while cognizant of hunter-level variation, they did not attempt to model it. Bliege Bird and Bird (2002) addressed the zero-inflation problem by discarding information about harvest size, using logistic regression to model the failure rates of foraging bouts, while using fixed effects to account for forager-level variation. Koster and Tankersley (2012) implemented a multilevel model with varying intercepts to account for hunter-level variation, but transformed the returns from hunting trips by adding an arbitrary constant to zeros and taking the logarithm. Walker et al. (2002) and Gurven et al. (2006) also used varying intercepts for individual hunters while modeling aspects of hunting that impact daily returns, including encounters per trip and the success of initiated pursuits, but they relied on locally estimated smoothers to examine the effects of age on return rates.

All of these studies are informative. But if these analyses were now to be repeated, it would be practical to fit unified foraging models that disentangle variation in failure rate from variation in the size of returns, while controlling for and modeling heterogeneity among individual foragers or foraging households. We provide an example of such an analysis, predicting trip-by-trip returns while allowing individual hunters to vary, as well as a software package to make such modeling more accessible. The data we use comprise about 15-thousand individual trips by 147 different hunters among the Aché of Paraguay (Hill and Kintigh 2009). Hunter ages in the data range from 19 to 74 years, with a mean of 48. We simultaneously address the problems of zero-inflated outcomes and imbalanced sampling in human foraging data, using a mixture distribution to model both failures (zero returns) and the size of returns (non-zeros) as functions of age and individual variation. Our estimates allow for zero and non-zero harvests to vary independently or to be correlated within individual hunters. Each component of returns is allowed to change across age in independent or correlated ways.

To cope with imbalance in sample size and heterogeneity among individuals, we use a multilevel modeling approach. Multilevel models, also known as hierarchical models or mixed effects models, can cope with imbalanced samples (for an accessible introduction, see Gelman and Hill 2007). Such models assume that there are clusters of observations within the data, such as all observations from individual hunters, that have unique coefficients. For example, each hunter can have his own parameter for average kilograms of meat returned to camp. These unique coefficients are related to one another by simultaneously

modeling the population of clusters. The estimates for each hunter take advantage of the information in the total sample, providing better estimates than by assuming either that each hunter provides no information about other hunters (zero pooling) or that all hunters are the same (complete pooling). The multilevel approach has considerable advantage over "fixed effects" using dummy variables for forager identity, because fixed effects will overfit on the data available to each forager and potentially be unidentified.[1] Additionally, if we wish to make inferences about the population of foragers, then we must model the population, and only a multilevel model does that explicitly.

We estimate and interpret the model fits within a Bayesian framework (Gelman et al. 2004). The model fit itself is Bayesian, because it comprises samples from the joint posterior density of the parameters. The interpretation is Bayesian, because we use the probabilistic interpretation of these parameters to simulate predictions and assess the quality of model fit to data. When parameters are less certain, predictions are less certain, and the Bayesian approach makes easy the propagation of uncertainty at the level of parameters to the level of predictions. Readers unfamiliar with Bayesian data analysis will be surprised to find no significance tests in our results. Instead, relative confidence in different models and parameter values is assessed by means and confidence (sometimes called *credible*) intervals and information criteria, as well as visual inspection of model predictions.

Our major goal is methodological. We hope to encourage other researchers to adopt statistical modeling approaches that (1) meet the difficult nature of the outcome variable and (2) assume and estimate individual level variation. Even a decade ago, such models would have been difficult or impossible to fit to data. But desktop computing and statistical algorithms have advanced rapidly, and now even average desktop computers are capable of fitting complex multilevel models. While the most complex models presented in this paper took as many as 8 hours to fit, it took many more hours—indeed, years—to collect the data. The cost in estimation time is worth the benefits of estimating the models the theory recommends rather than merely the models that are convenient. Winterhalder et al. (2010) make a similar argument for the use of Bayesian model fitting in the analysis of archaeological settlement data.

Our results are consistent with but also extend previous analyses. We find that failures (zero returns) contribute much more to variation in expected hunting returns, across all ages, than do the sizes of harvests. Most hunters peak around age 40 in both success and size of returns. Success and the size of returns are correlated, as hunters with lower failure rates (zeros) also experience higher mean non-zero returns. Success rates decline rapidly late in age, but the size of non-zero returns shows no similarly rapid decline. By decomposing hunting returns, more detailed questions can be addressed about the nature of human foraging strategies and life history evolution.

2 The data

The data we analyze here derive from a 1980–2007 study of the Aché of eastern Paraguay (Hill and Kintigh 2009). The outcome variable we will be concerned with, because it will

[1] For example, suppose a particular hunter is sampled 10 times and is never observed to succeed on a hunt. In a typical fixed effects model, it is not possible to produce a reasonable estimate unique to this hunter, because the maximum likelihood estimate of the log-odds of hunting success would be $-\infty$ with an enormous standard error. In a multilevel context, Bayes theorem is used to augment these 10 observations with the inferences from other hunters, moving the estimate off of zero probability of success.

be available in many other samples, is the kilograms of game returned to camp. For each recorded hunt, the data provide an associated hunter identity (as a number), the hunter's age at the time of the hunt, and the calendar date of the hunt. In some cases, the duration of the hunt is also recorded. See Hill and Kintigh (2009) for more details of data collection.

Out of 14,358 recorded returns, 7,502 (52%) are zero returns where nothing was brought back to camp. We call these zero returns *failures*. The remaining non-zero returns are widely dispersed, ranging from 0.04 kilograms to 191 kilograms, with a median of 5 kilograms. We call these non-zero returns *harvest sizes*, to indicate that they may comprise multiple animal packages.

One aspect of these data that is likely to be common in much field data of this kind is the extraordinary imbalance in sampling across individuals. The number of hunts for each unique hunter ranges from 1 to 1049. The median number of hunts is 26. The 72 individuals with more than the median 26 hunts contribute all together 13,639 hunts out of a total 14,358, 95% of the entire sample. If individual hunters vary much in skill, then a statistical model that ignores this variation may construct misleading estimates by substantially overweighting some individuals. This estimation hazard extends to regression slopes as well, such as estimates of the effect of age on returns.

3 Model definition

We model hunting returns, which are a mix of zero and positive real values. Observations that mix zeros with a large range of positive values arise in many fields, and are called both *zero-inflated* and *zero-augmented*. We adopt a mixture of bernoulli (zero/one) and gamma values to model this outcome variable. The probability density $f(y)$ for the returns y is:

$$f(y) = \begin{cases} \pi & \text{if } y = 0 \\ (1-\pi)\text{Gamma}(k,\theta) & \text{if } y > 0 \end{cases}$$

where π is the probability of a failure ($y = 0$), and k and θ define a gamma density with mean $k\theta^{-1}$ and rate (inverse scale) θ. For clarity, let $\text{ZGamma}(\pi,\mu,\theta)$ indicate the zero-inflated gamma distribution with probability of a zero π and non-zero outcomes defined by the mean μ and rate θ. See McCullagh and Nelder (1989), Chapter 8, for a thorough mathematical treatment of gamma GLM's. Jones et al. (forthcoming) independently arrived at a very similar solution to the same problem, using a zero-inflated log-normal distribution.[2]

This mixture distribution is capable of modeling the observed returns. But the model is descriptive—the true distribution of returns depends upon a complex stew of encounter rates, handling times, prey choice decisions, and heterogenous risks (Stephens and Krebs 1986). Given enough data, one might be able to build a process-oriented model of hunting returns along those lines. But it remains useful to pursue a descriptive model that can be applied in many contexts.

The details of our model framework comprise the remainder of this section. For those eager to jump ahead to the results, we summarize the approach as assuming that hunters vary in both baseline skill and how skill changes with age. The models estimate the extent of that variation while constructing individual hunter estimates. We also consider several,

[2] We considered a log-normal density for the non-zero values, as is sometimes used to model similar mixtures (Tooze et al. 2002, Fletcher et al. 2005), but found that for our data a Q-Q plot indicated substantial non-normality in the tails of the logged non-zero values. The gamma distribution has the same number of parameters as the log-normal, but provides additional flexibility.

potentially study-specific, covariates: Julian date, unique day effects, and hunt duration. These help illustrate the generality of our approach, how it can incorporate study-specific variables. The fitted multilevel model yields predictions at the level of the population of hunters, as well as predictions for each individual hunter. Perhaps most important, we compare models of differing structure, in order to learn both about statistical issues such as overfitting as well as anthropological issues such as how predictions vary as variables and assumptions are added and deleted.

3.1 Modeling age

We model the effect of age using a cubic polynomial, in both the zeros (π) and the non-zeros (μ). The advantages of the cubic polynomial include familiarity to many scientists, flexibility, and ability to accommodate differential rates of improvement and then decline with age (unlike a quadratic). The cubic is relatively easy to estimate, being an additive function. Higher-order polynomials are possible, but we found (using DIC, explained later) that fourth-order polynomials added nothing to the estimates, in this case.

The cubic model's flexibility is also a potential disadvantage. In principle, it can predict impossible hunter life histories with peak hunting skill at birth or death. Polynomials are also inconvenient because their parameters rarely have meaning that is easy to interpret. To navigate between the benefits and costs of the cubic function, we present results for the cubic polynomial in the main text, and we check the robustness of our inferences by estimating another, biologically-motivated and constrained function. We present the alternative age function in the ESM, showing that general inference remains the same under both models.

We wish to allow for heterogeneity among hunters in every parameter of these age functions. Varying slopes can accommodate this assumption. In the varying slopes approach, each cluster (hunter) in the data has its own parameter values. The ensemble of parameter estimates come from a population, and the parameters which describe the population's shape are estimated from the data. It is the population assumption that allows multilevel models to share statistical power among clusters, using the total information in the sample to improve individual estimates (for further explanation, see Efron and Morris 1975).

The varying slopes model structure is given by:

$$
\begin{aligned}
y_{ij} &\sim \mathrm{ZGamma}(\pi_{ij}, \mu_{ij}, \theta_j), \\
\log \frac{\pi_{ij}}{1-\pi_{ij}} &= s_{j1} + s_{j2}\mathrm{A}_{ij} + s_{j3}\mathrm{A}_{ij}^2 + s_{j4}\mathrm{A}_{ij}^3 + \beta_\pi \mathbf{x}_i, \\
\log \mu_{ij} &= s_{j5} + s_{j6}\mathrm{A}_{ij} + s_{j7}\mathrm{A}_{ij}^2 + s_{j8}\mathrm{A}_{ij}^3 + \beta_\mu \mathbf{x}_i, \\
\log \theta_j &= s_{j9}, \\
\mathbf{s}_j &\sim \mathrm{N}_9(\mathbf{s}, \Sigma),
\end{aligned}
$$

where i is an index referring to case, j an index referring to hunter, and A_{ij} is the age (standardized) of the j-th hunter at case i. Each coefficient s_{jk} is a parameter unique to hunter j. $\mathbf{s}_j$ is a vector of these individual hunter varying effects, sampled from a 9-dimensional multi-variate normal. $\mathbf{s}$ is a vector of the means of each effect. The 9-by-9 matrix Σ is the variance-covariance matrix of the varying effects. Additional covariates $\mathbf{x}$ (such as hunt duration and Julian date) are estimated by the β_π and β_μ vectors of standard regression coefficients. This allows a covariate, such as hunt duration, to have different effects on the

different components of the outcome. See the model code in the ESM for exact details of implementation.

3.2 Julian date

To consider the possibility that average failure rates and harvest sizes have changed over time in these data, potentially distorting our age estimates, we fit models containing Julian date (standardized) as a covariate in both the zero and non-zero components of the model.

3.3 Hunting duration and Bayesian imputation

The duration of a hunt may strongly covary with both success rate and the size of returns. To model this possibility, we allow hours hunted as simple fixed effect on both components of returns. Our interest is in controlling for hours hunted, so as to produce better estimates of the age functions.

A complication is that most (12,044 out of 14,358 cases, 84%) of the recorded hunts in the Aché data are missing hours hunted values. In order to use the 2,314 records that include hours hunted, we impute the missing values from a Gaussian duration distribution with mean and variance estimated from the data. We use a Gaussian, because the empirical distribution of observed hunt durations is approximately Gaussian, with almost no skew. We perform this imputation within a Bayesian framework, integrating over uncertainty in each imputed value h_{ij} for duration for hunt i of hunter j, while computing the uncertainty in the other parameters. This approach allows us to use the duration information where it is present to estimate any relationship between duration and returns, without having to discard the majority of cases where it is missing. Unlike *multiple imputation*, the Bayesian method of imputation uses a hyper-parameter for each missing value, simultaneously estimating the missing values and the impact of the variable (hunt duration) on outcomes. See Chapter 25 of Gelman and Hill (2007) for an overview of imputation terminology and methods.

Since we have good reason to think that duration entries are missing completely at random (Hill and Kintigh 2009), we adopt that approach (Rubin 1976). We considered the possibility that hunt duration is also a function of age, requiring a more complex imputation model. However, the correlation between hours hunted and age is weak, whether modeled as a linear, quadratic, or cubic effect.

3.4 Correlated returns by day

While returns are assigned to individual hunters, failures and harvest sizes may be correlated across hunts on the same day. First, there may be exogenous factors such as weather that affect all hunters, even if they hunt completely independently. Second, while the Ache mainly hunt individually, they do move into the forest in groups and sometimes recruit help in pursuing some prey (Hill 2002). Therefore correlations may arise because of cooperative aspects of hunting.

As a minimal approach to modeling these correlations, we consider a varying effect on each of the 2,525 unique days in the sample. Number of hunts per day range from 1 to 29, with a median of 4 and mean of 5.7. We introduce two varying effects on each day, centered on zero, that allow for an increase or decrease in log-odds of failure and mean harvest

size, respectively. These effects apply to all hunts on the same day, creating correlations in outcomes. They are independent of the varying effects on individual hunters, making this a cross-classified multilevel model with both varying effects on age, clustered by hunter, as well as varying effects on failure and harvest size, clustered by day.

3.5 Priors

Different priors produce different models, just like different likelihoods and link functions produce different models. We adopt "uninformative" priors for all parameters, resulting in estimates that echo maximum likelihood inferences. Precise prior specifications are available in the model definitions in the supplemental. However, in general we believe that the best priors are rarely uninformative, because the analyst always knows something *a priori* that can improve model-based inference. See for discussion and example Gelman et al. (2008).

More generally, the term "noninformative" appears to be a Whorfian trick of language. Just as an "empty" gasoline drum is more dangerous than a "full" one (Whorf 1941), a noninformative prior may distort estimates more than would an informative prior. Better to regard all priors as informative, just as all likelihood functions are informative.

In that spirit, we checked that inference is robust to changes in variance-covariance priors for the varying effects. We estimate the models using both the common inverse-Wishart prior and a Cholesky decomposition that provides for a uniform prior over positive definite matrices (Lewandowski et al. 2009, as implemented in Stan 1.0.2). Moreover, we provide all of our model fitting code so that others can easily adjust the priors (as well as likelihoods and links) and see how changes alter inferences.

3.6 Nested models

The full model contains 9 varying effects clustered on individual hunters, 2 varying effects clustered on day, and four simple fixed effects. There is also one parameter for each missing hunt duration value to be imputed, as well as two parameters to estimate for the distribution of the missing values. This implies a total of 63 traditional (fixed effect) parameters and 15,892 hyper-parameters (varying effects and missing values).

Many simpler nested models can be constructed from this full model. In order to guard against overfitting, we fit 10 different models to the Aché hunting data. The simplest model, Fixed, uses cubic models of age for both failures (zeros) and harvests (non-zeros), but contains no varying effects of any kind. This model also omits Julian date, hunt duration, and varying effects on days. This is a poor model for the unbalanced, heterogenous data considered here, but comparing its fit to the other models provides information about how much better prediction is after considering varying effects and the remaining covariates. We then add in varying effects on components of the age model, from Vary_2 (varying intercepts only) to Vary_9 (varying slopes on every component). Combinations of hunt duration (with imputation), Julian date, and varying effects on day allow for judging the robustness of age inferences to inclusion of these covariates.

3.7 Software

Models of this kind, with non-standard outcome distributions and cross-classified varying effects and missing data imputation, are most easily estimated with Markov chain Monte Carlo (MCMC) techniques. We fit the models using both JAGS 3.1.0 (Plummer 2003) and Stan 1.0.2 (Stan Development Team 2012) to draw samples from the joint posterior density of the parameters. The results we present are based on the Stan samples, which produce the same inferences as the JAGS results, but with more rapid convergence and better mixing. The estimates here derive from 10-thousand samples of each parameter, after 5-thousand samples for adaptation. Sampling was very efficient within Stan, which uses a variant of Hamiltonian Monte Carlo (see Neal 2011, for lucid explanation). We present example trace plots from two chains in the supplemental, to illustrate Stan's rapid convergence and efficient mixing.

We have also developed an open source R package, `glmer2stan` (McElreath 2012), that allows the user to define and fit a large variety of mixed-outcome multilevel models, using standard R mixed model formula notation. Further details of this package are available in the ESM or by contacting the corresponding author.

4 Results

4.1 Model comparison

Table 1 presents formal model comparison statistics for ten models that recombine different varying effect structures and covariates. For each of the ten models, the table presents (left to right) the number of fixed parameters, the number of varying hyper-parameters, the deviance, the effective number of parameters p_D, the information criterion DIC, the difference between each DIC and the smallest DIC (ΔDIC), and finally the approximate posterior model weight of each model.

The metric DIC (Deviance Information Criterion: Plummer 2008, Spiegelhalter et al. 2002) provides a multilevel analogue of AIC (Burnham and Anderson 2002, Claeskens and Hjort 2008, Lunn et al. 2013). In the absence of varying effects, DIC reduces directly to AIC. When one is not interested in generalizing to the same clusters, then AIC may provide useful information, while DIC is more relevant for generalization of varying effects estimates (Fang 2011, Vaida and Blanchard 2005). Like AIC, smaller values of DIC indicate better expected generalization. The model weights are analogous to Akaike weights (Burnham and Anderson 2002). The "weight" of a model is an estimate of the posterior probability of the model family, providing a standardized metric of comparison.

We detail calculation of DIC and DIC weights in the supplemental. It is important to keep in mind that DIC is computed from samples from a Markov chain, so there can be simulation uncertainty in its value (Gelman and Hill 2007, page 525). This means that we should not base strong inferences on small differences in DIC between models. Additionally, the proper use of model comparison metrics like DIC is not to "select" a single best model. First, both theory and simulation suggest that averaging predictions using relative DIC weights produces better results than choosing a single model (Burnham and Anderson 2002, Claeskens and Hjort 2008). Second, much can be learned from all models, including those with very large DIC values. Later in the paper, and in the ESM, we discuss compar-

isons of predictions among models, which provides further information not available from concise metrics like DIC.

The DIC rankings and model weights show that models that include both 9 varying effects on age ("Vary_9") and varying effects on day ("v") are expected to generalize best to new data from the same population of hunters. Removing varying effects on day from a model results in a one-thousand point increase in DIC, and removing varying slopes on age results in another one-thousand point increase. Other factors have smaller effects. Considering Julian date adds very little to the model, as models that differ only by inclusion of Julian date ("D") differ very little in both deviance and DIC. Models that include hunt duration ("h") out-perform models that do not, but the improvement in DIC is only about 40 points of deviance. Together, these results imply that varying intercepts and slopes on age and day-specific intercepts are both very important to successful prediction, while the other effects are less important.

Table 2 presents the expected values and standard deviations of the samples of each parameter, across the four top-ranking models from Table 1. In the ESM, we provide estimates for the six lower-ranked models. While we reference this table of estimates in the following sections, it will be easier to appreciate the nature of the estimates by plotting their implied predictions. In the following sections, we take each component of the top-ranked model, Vary_9hDv, and plot its implied predictions. In the ESM, we show predictions for other models. In each case, the implied predictions incorporate the uncertainty embodied in the joint posterior. As a result, highly uncertain estimates produce highly uncertain predictions.

4.2 Failure rates

Figure 1 displays predictions from model Vary_9hDv, using the sample-average length of hunts, 6.8 hours (the empirical mean duration), in all plots unless otherwise noted. Figure 1 is complex, but focus for now on only the top row. Panel (a) displays the predicted average hunter's probability of failure (zero return) across the lifespan. The youngest hunters are expected to fail more than 80% of the time. By age 40, the average hunter is expected to fail only 50% of the time. Later in life, failure rates rise again, but the predictions are much less precise, as seen by the broadening of the 95% confidence interval (dashed curves).

The circles in Figure 1(a) indicate observed average failure proportions for each age, averaged across trips for all hunters. Comparing predictions to the raw data can be useful both for validating the estimates and understanding their implications. The size of each circle indicates the relative sample size at each age. Notice that the estimates of the predicted average trend across age does not pass through the center of the observed data averages. Instead, the predicted trend for failures passes above the mean failure rates across hunters. Why is this? The best hunters hunt more often, and so they bias population estimates, unless variation among hunters is explicitly modeled. Thus the apparent lack of fit between the model predictions (the curves) and the raw data (the circles) is a feature of the multilevel approach, not a flaw. Predictions based on fixed effects models, shown in the ESM, do indeed pass through the empirical averages at each age, but are nevertheless not better estimates of the performance of an average hunter, because they overweight the data from the best hunters.

Figure 1(b) illustrates variation across individual hunters. Predictions for the average hunter are useful, but it is also useful to characterize the variation among hunters at each

Table 1 Model comparison information. In model names, "Vary$_x$" indicates x varying effects on age, "D" indicates Julian date, "h" indicates hours hunted, and "v" indicates varying effects on day. ΔDIC is the difference between a model's DIC and the smallest DIC in the set.

Model	parameters	hyper-parameters	deviance	p_D	DIC	ΔDIC	weight
Vary$_9$hDv	63	15892	50006	2126	54258	0	0.73
Vary$_9$hv	61	15892	49979	2140	54260	2	0.27
Vary$_9$v	57	3848	50431	1932	54294	36	< 0.01
Vary$_9$Dv	59	3848	50448	1928	54304	46	< 0.01
Vary$_9$h	58	13367	54094	631	55355	1097	< 0.01
Vary$_9$hD	60	13367	54097	630	55357	1099	< 0.01
Vary$_9$	54	1323	54625	393	55411	1153	< 0.01
Vary$_9$D	56	1323	54635	390	55415	1157	< 0.01
Vary$_2$	12	294	56297	156	56608	2350	< 0.01
Fixed	9	0	57287	9.03	57305	3047	< 0.01

Table 2 Parameter sample expectations and standard deviations (in parentheses) for the top four models. Variance components reported as variances, with standard deviation of the variance in parentheses. In model names, "Vary$_9$" indicates 9 varying effects on age, "D" indicates Julian date, "h" indicates hours hunted, and "v" indicates varying effects on day.

	Vary$_9$v	Vary$_9$Dv	Vary$_9$hv	Vary$_9$hDv
Fixed effects				
Zeros				
Intercept	0.04(0.10)	0.01(0.10)	1.07(0.25)	0.97(0.21)
Age	0.14(0.12)	0.02(0.12)	0.15(0.12)	0.01(0.12)
Age^2	0.17(0.10)	0.23(0.10)	0.18(0.10)	0.24(0.10)
Age^3	-0.11(0.06)	-0.08(0.06)	-0.11(0.06)	-0.08(0.06)
Julian date		0.17(0.06)		0.17(0.05)
Hours			-0.15(0.03)	-0.14(0.03)
Non-zeros				
Intercept	1.83(0.04)	1.84(0.04)	1.79(0.06)	1.79(0.06)
Age	-0.06(0.05)	-0.06(0.05)	-0.06(0.05)	-0.06(0.05)
Age^2	-0.02(0.05)	-0.02(0.05)	-0.02(0.05)	-0.03(0.05)
Age^3	0.05(0.03)	0.05(0.03)	0.05(0.03)	0.05(0.03)
Julian date		0.00(0.02)		0.00(0.02)
Hours			0.01(0.01)	0.01(0.01)
log θ	-0.64(0.11)	-0.66(0.11)	-0.64(0.10)	-0.66(0.11)
Variance				
Zeros				
Intercept	0.55(0.14)	0.46(0.12)	0.57(0.15)	0.45(0.12)
Age	0.33(0.13)	0.27(0.11)	0.33(0.13)	0.29(0.11)
Age^2	0.19(0.09)	0.16(0.07)	0.19(0.08)	0.16(0.08)
Age^3	0.09(0.03)	0.08(0.03)	0.09(0.03)	0.09(0.03)
Day	0.85(0.07)	0.85(0.07)	0.87(0.07)	0.86(0.07)
Non-zeros				
Intercept	0.08(0.02)	0.08(0.02)	0.08(0.02)	0.08(0.02)
Age	0.05(0.01)	0.06(0.01)	0.06(0.01)	0.06(0.02)
Age^2	0.05(0.01)	0.05(0.01)	0.05(0.01)	0.05(0.01)
Age^3	0.03(0.01)	0.03(0.01)	0.03(0.01)	0.03(0.01)
log θ	0.91(0.17)	0.88(0.17)	0.90(0.17)	0.89(0.17)
Day	0.02(0.004)	0.02(0.004)	0.02(0.004)	0.02(0.004)

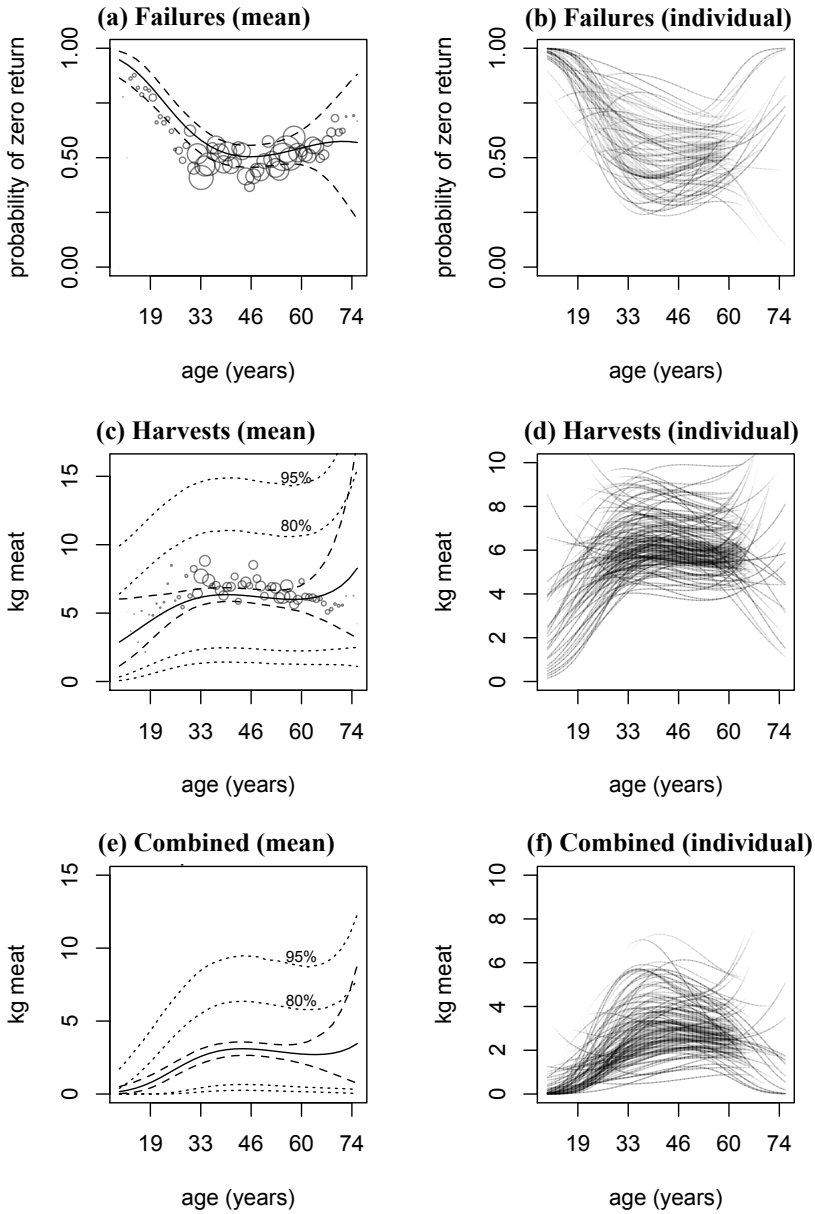

Fig. 1 Predicted relationships between age and failures (top row), non-zero harvests (middle row), and their combined expectation (bottom row). Left column: Means (solid curve) and 95% confidence intervals (dashed) of the means. Dotted curves represent prediction intervals at 80% and 95% of the realized harvest size distribution. The circles show the observed average, across all hunters, at each age, with size indicating sample size. Right column: Individual hunter estimated functions. Each curve is the mean prediction for one of 147 hunters, with transparency indicating width of confidence interval at each age. See ESM for each individual curve plotted seperately.

age. Each curve in this plot is an individual hunter's average predicted relationship between probability of failure and age. Transparency of the curves is proportional to the width of the confidence interval at each age. Notice that there is less variation at low ages, with the curves bundled more tightly, but more variation at middle ages, with some hunters keeping very low failure rates for most of their lives. The best hunters enjoy a low 30% failure rate in their 30's or 40's, while the average hunter does no better than a 50% failure rate.

As age increases, failure rates increase. Above 60 years old, the estimates are very scattered and uncertain, as indicated by the flaring confidence bounds in the top-left plot and the frequent transparency of individual curves in the lower-left. This is a result of paucity of data at high ages and the flexibility of the cubic polynomial. As we noted earlier, the cubic age function is very flexible. However, when there is little data for an individual hunter, the mean prediction for the hunter can look odd, even predicting best performance (lowest failure rates) at the youngest and oldest ages in a small number of cases. The corresponding uncertainty for such curves is huge, and therefore they are transparent in the figure. We show confidence bounds for all individual hunters in the ESM, with no transparency, to provide a detailed sense of this phenomenon.

There are several strategies that could improve the estimates at the level of individual hunter. We could adopt informative priors on the age function that preclude success at extreme ages exceeding success at central ages. Another strategy is to use another, less flexible age function. We provide in the ESM individual hunter curves for an alternative age function that is more constrained and has interpretable parameters. These alternative estimates show that the general inferences about pattern and variation are not sensitive to the function we've chosen. However, the high uncertainty at late ages is partly a result of the flexibility of the cubic polynomial, as the alternative age function produces more confident trends at both the population and individual level.

4.3 Harvest size

The middle row in Figure 1 plots predicted average non-zero returns (c) and individual hunter estimates for non-zero returns (d). These plots illustrate the distribution of returns from successful hunts only. The dotted curves in panel (c) show the predicted density of the actual returns, not just the mean returns indicated by the solid and dashed curves. So for example the 80% dotted contour shows the boundary within which 80% of predicted returns for an average hunter are expected to fall.

The pattern for the size of non-zero returns (harvest sizes) demonstrates a rapid increase early in life, with an average hunter in his 40's returning twice as many kilograms of meat per trip as the youngest hunter. However, there is more diversity at early ages, compared to failures, and less evidence of a decline in performance late in age. Estimates late in life remain highly uncertain.

The individual hunter predictions, Figure 1(d), demonstrate variation at all ages, with the best hunters appearing to enjoy a larger average return for most of their lives. In the flat region between ages 35 to 60, the best hunters on average return twice as much meat as the worst hunters. Still, realized returns overlap substantially, making it difficult to identify the best hunters (Hill and Kintigh 2009).

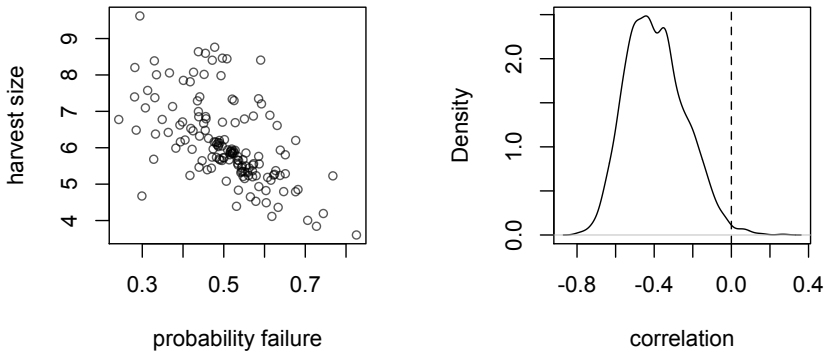

Fig. 2 Correlation between failure and size of returns (in kilograms), evaluated at age 46. Lefthand plot: Each point is an individual hunter. The estimated average harvest size is plotted against probability of failure. Righthand plot: Marginal posterior density of the correlation between failure and harvest size.

4.4 Correlations across failure rate and harvest size

Before explaining the bottom row in Figure 1, it will help to present estimates of the correlations between components of the individual hunter varying effects. Recall that this model assumes that each hunter has unique parameter values. Per-hunter parameters are drawn from a population, however, and the model estimates the shape of this population. Part of the structure of the population lies in correlations along different dimensions of hunting performance. For example, do hunters with lower average failure rates also get larger average harvests?

Figure 2, lefthand plot, shows the mean estimates for each hunter, evaluated at age 46, for the probability of failure and harvest size. The clear negative trend suggests that hunters with high failure probabilities also have low mean harvest sizes. The righthand plot displays the posterior density for the correlation between log-odds of failure at age 46 and log harvest size at age 46. The density has a mean of -0.40 and a 95% confidence interval from -0.67 to -0.12. Better hunters are better at both getting prey and in getting more and larger prey. There is substantial uncertainty about the magnitude of this correlation, although it is unlikely to be positive.

4.5 Combined returns

The expected returns from a hunter are computed by multiplying his estimated average success probability by the estimated mean harvest size. The bottom row in Figure 1 plots these expectations, for both an average hunter (e) and for individual hunters in the sample (f). These are the returns experienced by the people in the hunter's community, in the long run.

Being able to analyze the zero and non-zero components of the outcome allows us to identify how variation in each contributes to the overall pattern of variation. By far

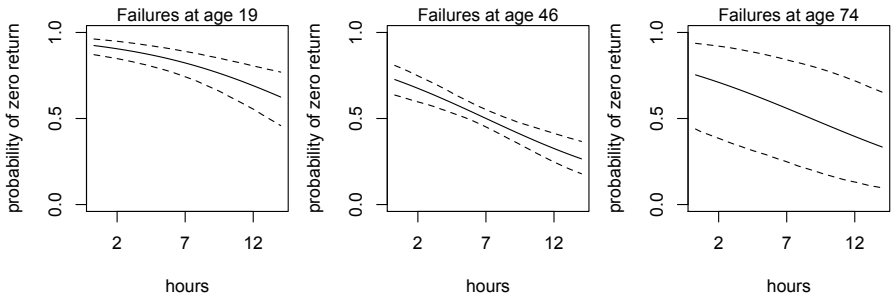

Fig. 3 Predicted impact of hours hunted on failure. Predictions computed for an average hunter at age 19 (left), age 46 (middle), and age 74 (right). Dashed lines are 95% confidence intervals.

the most important component is failure rate. Since harvest size is irrelevant when a hunt fails, hunting success multiplies the distribution of harvest sizes and has a dominant role in determining variation in combined returns.

Some counter-factual calculation will help reveal this fact. The standard deviation in expected returns at age 46 is 0.76 kg. Now consider two thought experiments. First, imagine a world in which all hunters have identical failure rates at the same age but retain their estimated difference in non-zero returns. Under this thought experiment, the standard deviation of expected returns falls 65%, from 0.76 kg to 0.27 kg. In the second thought experiment, suppose instead that all hunters have identical harvest size distributions at the same age but variable failure rates. Now the standard deviation of expected returns falls only by 30%, to 0.54 kg. Both components are important to combined expected returns, especially given the correlation between them, but failure rates contribute much more than does harvest size.

4.6 Day varying effects

Next to age, the most important effect in these models is varying effects clustered on day. These estimates express correlations among all trips on the same calendar date, and no model without these effects comes close to either the deviance or DIC of the top four models in Table 1. Estimates of the variation in log-odds of failure across days is consistently near 0.86 (95% interval 0.72–0.99). This is nearly twice as much variation as is estimated to exist among hunters of the same age, 0.45 (95% interval 0.25–0.69). Day effects, whatever their cause, structure per-trip failure and success in powerful ways.

In contrast to failure rates, there is not much evidence that day effects impact harvest sizes. The estimated variation in log kilograms of meat across days is 0.023 (95% interval 0.015–0.031). This is a third of the variation estimated to exist among hunters of the same age (0.08 with 95% interval 0.05–0.12).

4.7 Hunt duration

Hunt duration is an important covariate of failure rates, but not harvest sizes. The models that include hours hunted (see Table 2) dominate the other models in both fit (deviance) and expected out-of-sample accuracy (DIC). The estimate for the impact of each additional hour hunted on log-odds of hunt failure is -0.14 (95% confidence interval: -0.20 to -0.08). In terms of proportional odds, each additional hour of hunting is expected to reduce the odds of failure by about 13% (95% interval: 8% to 18%). Figure 3 plots the impact of this estimate on expected failure rates, at three different ages. The wide confidence intervals at high ages is due to uncertainty in baseline failure rate at high ages.

In contrast to failures, the estimate for the impact of hours hunted on harvest size is essentially zero—an expected additional 0.01 log-kg of meat per additional hour hunted (95% interval: -0.01 to 0.03). If hunt duration has any important effect on the expected size of non-zero returns, we cannot detect it with these data and these models.

4.8 Julian date

We consider Julian date as a predictor in order to account for secular trends in prey density or other factors that increase or reduce return rates across the entire population through time. The models provide weak support for a small increase in failure rates through time. However, models with Julian date incorporated as a predictor fit almost identically to those without it, and the coefficient estimates for Julian date are very small (Table 2). The mean coefficient estimate for the effect of date on log-odds of failure in model Vary_9hDv (the best model that includes Julian date as a predictor) is 0.17, with a 95% confidence interval of 0.07 to 0.27. Since date was standardized before fitting, this estimate says that an increase of one standard deviation of Julian date (2400 days, or 6.5 years) increases log-odds of failure by 0.17. This is a small effect, although maybe a very real one (Hill et al. 2003). There is also little evidence of a trend in harvest size (mean coefficient < 0.01, 95% interval -0.04 to 0.04).

Note that because all of the models include hunter age, this estimate for secular trend in failure rate controls for changes in the age structure of the hunter population. This is important, because a reduction of the average hunter age, for example, could produce an apparent increase in mean failure rate, without any change in the prey population.

5 Discussion

5.1 Summary and advantages of the approach

This paper presents a unified Bayesian analysis of variation in human foraging returns. Variation in these data arise from differences in age, skill, and hunt duration, as well as many unmeasured and un-modeled factors. Instead of coercing the outcome measure, kilograms of meat returned to camp, into a convenient distribution, we modeled these returns using a two process zero-inflated gamma mixture model. The benefit of this additional complexity is that we are able to discuss risk as well as average returns within the same model. When analyses focus on only one part of the mixture, whether zeros or non-zeros, or average across trips to blend zeros and non-zeros together, information is lost. In addition, the models estimate the correlation between the components of hunting returns, a superior

approach to running separate regression models on zero and non-zero outcomes, because it allows information about failures to inform estimates about harvests, and visa versa.

In addition to structuring returns using a two-process model, the multilevel structure of the analysis assumes that hunters vary and allow the data to tell us how much and in which ways. The model allows for individual hunters to differ not only on average, but also to differ in how their performance changes with age. In principle, because foraging data are nearly always unbalanced, presenting more data from certain individuals, a multilevel model is needed to produce accurate estimates that do not give excessive influence to the most frequent hunters. But beyond this basic need to deal with unbalanced repeat measures, the multilevel structure also facilitates inferences about variation in different aspects of foraging at different points in the lifespan.

The model produces estimates about how foraging success changes with age by partially pooling data across individual hunters. This is important, because rarely does any individual hunter present a complete record of hunting across all ages. Instead we must estimate the relationship between age and performance by piecing together information from many hunters. Attempting this kind of estimate outside of a multilevel framework can bias estimates in many ways. The multilevel framework uses information from all hunters to improve estimates for all ages, but it does not create illusionary confidence where there is none. This conservatism of the approach can be seen in the wide confidence intervals around pieces of the individual hunter estimates.

The advantages of multilevel models are numerous, making them useful as a default form of regression analysis. Field data have naturally occurring clustering and usually unbalanced sampling. An additional advantage of the multilevel framework is that it allows for clusters, such as hunters, to have unique intercepts while still using cluster-level predictors like age (Gelman and Hill 2007, page 269). Software such as the `glmer2stan` package (McElreath 2012, described in the ESM) provides a friendlier interface to Bayesian model fitting, which will help researchers incorporate such models into their standard procedures.

It is important to note that the multilevel approach is not confined to large-data situations, like this one. Whenever there is clustering or repeat sampling in the data, multilevel estimates are likely to outperform traditional fixed effects. This is especially true when there is imbalance in sampling, such that some individuals, households, or populations are sampled more. In another analysis (House et al. forthcoming), similar age-dependent models (fit with `glmer2stan`) were usefully fit to much less data, because the data were similarly clustered and imbalanced. Of course the amount of data does impose limits on model complexity, but the relationship between data and parameters in a multilevel model is not of the same kind as in simple, fixed effect models. That is why measures like DIC are needed. More generally, one always learns more by comparing models of differing structure and complexity than by examining a single model, simple or complex.

Because it is now possible to fit the models we wish to, rather than just the models we are able to, it is also possible to work towards a more satisfying match between statistical and conceptual models of human foraging. The framework we have outlined is flexible, allowing researchers to propose and compare different models, using additional information about prey type, hunting strategies, and community composition. For example, it is not necessary to treat returns as products of single hunters. Instead, multiple membership models (Browne et al. 2001) allow for returns to be modeled as products of cooperative hunts, while still disentangling evidence of individual hunter ability. The only requirement is that individuals do not always hunt with the same partners. But even if partners were un-

changing, estimates could be made for teams, and the success of teams could be modeled as a function of individual characteristics of its members.

5.2 Risk-sensitive foraging and reciprocity

There are many substantive issues to unpack with respect to how our estimates reflect on ongoing debates within evolutionary anthropology, from signaling (Hawkes and Bliege Bird 2002) to human life history (Kaplan et al. 2000, 2009) to optimal strategies under different stochastic distributions (Stephens and Charnov 1982). We have focused on methodology. But this approach makes it possible to ask direct statistical questions about topics such as risk, both within and across individuals, in precise new ways. We conclude the discussion by relating our results to one of these traditional human behavioral ecological topics, risk and risk reduction through sharing.

An immediate benefit of modeling both zero and non-zero returns within the same model is that it becomes possible to discuss their relative importance in determining the empirical distribution of foraging returns. We found evidence of substantial variation in both failures (zeros) and harvest sizes (non-zeros). We also found that success and the size of returns are correlated across hunters. Nevertheless, the estimates strongly suggest that variation in failures contributes much more to average returns than does variation in harvest sizes. This pattern is unlikely to be true in all foraging contexts, especially those with greater variation in prey mass—there are no giraffes in Paraguay. But in the Aché sample, this finding echoes the general recognition that foraging is inherently risky (see chapters in Cashdan 1990) and that sharing does sometimes result facultatively from increased risk (Kaplan et al. 2012). Individual hunters can expect long stretches of failure, with no returns to show for their efforts, even when the community at large enjoys high returns. Our estimates further suggest that some hunters must expect longer periods of total failure. For example, the best hunters in our estimates have failure probabilities around 0.3 (at peak age), while the worst have probabilities around 0.7. Translating these probabilities into the probability of realizing 5 failed hunts in a row results in estimates of $0.7^5 \approx 0.17$ for the worst hunters and $0.3^5 \approx 0.002$ for the best hunters.

This result sharpens the puzzle of food sharing, because it suggests that the benefits and costs of sharing are highly asymmetric within at least some foraging groups. Sharing can still evolve under such conditions (Boyd 1992, Gurven et al. 2000, Gurven 2006), but it is a poorly studied problem. Additional complexity arises from risk on the production side, as some individuals will be unable to share, for entirely stochastic reasons. There is some work on reciprocity under noisy conditions (Green and Porter 1984, Le and Boyd 2007), but to our knowledge reciprocity has never been formally modeled under the peculiar conditions we find here, with both highly stochastic production of goods for exchange and large asymmetries in costs and benefits. On top of these factors, we also found evidence that independent hunts on the same day have correlated failure probabilities. The impact of correlations on the production side of sharing reduces the risk reduction benefits of sharing and therefore incentives for reciprocity (Winterhalder 1990). Among other advantages, the use of multilevel statistical models can inform future theorizing by helping to narrow and parameterize modeling effort.

Statistical tools sometimes spur new theorizing (Gigerenzer 1991). We hope that the ability to effectively model and estimate individual variation at multiple scales (individual, family, society) will encourage new questions, and eventually new answers, about the

strategic nature and evolutionary history of human foraging behavior, in relation to both ecology and culture.

Acknowledgements Thanks to Mark Grote, Jamie Holland Jones, Bruce Winterhalder, members of the UC Davis Cultural Evolution and Human Behavioral Ecology labs, and two anonymous reviewers for advice and comments. Bob Carpenter and the members of the Stan Development Team helped us improve the efficiency of our Stan code. Keith O'Rourke suggested the analogy between Whorf's gasoline drums and uninformative priors. Nicholas Blurton-Jones, back in 1997, encouraged RM to find a solution to modeling imbalanced field data, leading him eventually to Bayesian data analysis.

References

Alvard, M. (1995). Shotguns and sustainable hunting in the neotropics. *Oryx*, 29:58–66.

Bliege Bird, R. and Bird, D. W. (2002). Constraints of knowing or constraints of growing? Fishing and collecting by the children of Mer. *Human Nature*, 13(2):239–267.

Boyd, R. (1992). The evolution of reciprocity when conditions vary. In Harcourt, A. and DeWaal, F., editors, *Coalitions and Alliances in Humans and Other Animals*. Oxford University Press, New York.

Browne, W. J., Goldstein, H., and Rasbash, J. (2001). Multiple membership multiple classification (MMMC) models. *Statistical Modelling*, 1:103–124.

Burnham, K. and Anderson, D. (2002). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. Springer-Verlag, 2nd edition.

Cashdan, E., editor (1990). *Risk and uncertainty in tribal and peasant economies*. Westview Press, Boulder, CO.

Claeskens, G. and Hjort, N. (2008). *Model Selection and Model Averaging*. Cambridge University Press.

Efron, B. and Morris, C. (1975). Data analysis using Stein's estimator and its generalizations. *Journal of the American Statistical Association*, 70:311–319.

Fang, Y. (2011). Asymptotic equivalence between cross-validations and Akaike information criteria in mixed-effects models. *Journal of Data Science*, 9:15–21.

Fletcher, D., Mackenzie, D., and Villouta, E. (2005). Modelling skewed data with many zeros: A simple approach to combining ordinary and logistic regression. *Environmental and Ecological Statistics*, 12:45–54.

Gelman, A., Carlin, J. C., Stern, H. S., and Rubin, D. B. (2004). *Bayesian Data Analysis*. Chapman & Hall/CRC, 2nd edition.

Gelman, A. and Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press.

Gelman, A., Jakulin, A., Pittau, M. G., and Su, T.-S. (2008). A weakly informative default prior distribution for logistic and other regression models. *The Annals of Applied Statistics*, 2(4):1360–1383.

Gigerenzer, G. (1991). From tools to theories: A heuristic of discovery in cognitive psychology. *Psychological Review*, 98(2):254–267.

Green, E. J. and Porter, R. H. (1984). Noncooperative collusion under imperfect price information. *Econometrica*, 52(1):87–100.

Gurven, M. (2006). The evolution of contingent cooperation. *Current Anthropology*, 47:185–192.

Gurven, M., Hill, K., Kaplan, H., Hurtado, A., and Lyles, R. (2000). Food transfers among Hiwi foragers of Venezuela: Tests of reciprocity. *Human Ecology*, 28(2):171–217.

Gurven, M., Kaplan, H., and Gutierrez, M. (2006). How long does it take to become a proficient hunter? Implications for the evolution of delayed growth. *Journal of Human Evolution*, 51:454–470.

Hawkes, K. and Bliege Bird, R. (2002). Showing off, handicap signaling, and the evolution of men's work. *Evolutionary Anthropology*, 11:58–67.

Hill, K. (2002). Altruistic cooperation during foraging by the Ache, and the evolved human predisposition to cooperate. *Human Nature*, 13(1):105–128.

Hill, K. and Kintigh, K. (2009). Can anthropologists distinguish good and poor hunters? implications for hunting hypotheses, sharing conventions, and cultural transmission. *Current Anthropology*, 50(3):369–377.

Hill, K., McMillan, G., and na, R. F. (2003). Hunting-related changes in game encounter rates from 1994 to 2001 in the Mbaracayu reserve, Paraguay. *Conservation Biology*, 17(5):1312–1323.

House, B. R., Silk, J. B., Henrich, J., Barrett, H. C., Scelza, B., Boyette, A., Hewlett, B., McElreath, R., and Laurence, S. (forthcoming). The ontogeny of prosocial behaviour across diverse societies. *Proceedings of the National Academy of Sciences, USA*.

Jones, J., Bird, R., and Bird, D. (forthcoming). To kill a kangaroo: Understanding the decision to pursue high-risk/high-gain resources. *Proceedings of the Royal Society of London, B*.

Kaplan, H. S., Hill, K. R., Lancaster, J. B., and Hurtado, A. M. (2000). A theory of human life history evolution: Diet, intelligence, and longevity. *Evolutionary Anthropology*, 9:156–185.

Kaplan, H. S., Hooper, P. L., and Gurven, M. (2009). The evolutionary and ecological roots of human social organization. *Philosophical Transactions of the Royal Society B*, 364:3289–3299.

Kaplan, H. S., Schniter, E., Smith, V. L., and Wilson, B. J. (2012). Risk and the evolution of human exchange. *Proceedings of the Royal Society B*, 279(1740):2930–2935.

Koster, J. (2008). Hunting with dogs in Nicaragua: An optimal foraging approach. *Current Anthropology*, 49:935–944.

Koster, J. M. and Tankersley, K. B. (2012). Heterogeneity of hunting ability and nutritional status among domestic dogs in lowland nicaragua. *PNAS*, 109:E463–E470.

Le, S. and Boyd, R. (2007). Evolutionary dynamics of the continuous iterated Prisoner's dilemma. *Journal of Theoretical Biology*, 245:258–267.

Lewandowski, D., Kurowicka, D., and Joe, H. (2009). Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis*, 100:1989–2001.

Lunn, D., Jackson, C., Best, N., Thomas, A., and Spiegelhalter, D. (2013). *The BUGS Book: A practical introduction to Bayesian analysis*. CRC Press, Boca Raton, Florida.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*. Chapman and Hall/CRC Press, 2nd edition.

McElreath, R. (2012). glmer2stan: Rstan models defined by glmer formulas, version 0.96. http://xcelab.net/rm/.

Neal, R. (2011). MCMC using Hamiltonian dynamics. In Brooks, Gelman, Jones, and Meng, editors, *Handbook of Markov Chain Monte Carlo*. Chapman and Hall/CRC Press.

Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In *Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003)*.

Plummer, M. (2008). Penalized loss functions for Bayesian model comparison. *Biostatistics*, 9(3):523–539.

Rubin, D. B. (1976). Inference and missing data (with discussion). *Biometrika*, 63:581–592.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and va der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society B*, 64:583–639.

Stan Development Team (2012). Stan: A C++ library for probability and sampling, version 1.0.2. http://mc-stan.org/.

Stephens, D. W. and Charnov, E. L. (1982). Optimal foraging: Some simple stochastic models. *Behavioral Ecology and Sociobiology*, 10:251–263.

Stephens, D. W. and Krebs, J. R. (1986). *Foraging Theory*. Princeton University Press, Princeton.

Tooze, J. A., Grunwald, G. K., and Jones, R. H. (2002). Analysis of repeated measures data with clumping at zero. *Statistical Methods in Medical Research*, 11:341–355.

Vaida, F. and Blanchard, S. (2005). Conditional Akaike information for mixed-effects models. *Biometrika*, 92(2):351–370.

Walker, R., Hill, K., Kaplan, H., and McMillan, G. (2002). Age-dependency in hunting ability among the Ache of eastern Paraguay. *Journal of Human Evolution*, 42:63900657.

Whorf, B. (1941). The relation of habitual thought and behavior to language. In Spier, L., editor, *Language, culture, and personality, essays in memory of Edward Sapir*, pages 75–93. Sapir Memorial Publication Fund, Menasha, Wisconsin.

Winterhalder, B. (1990). Open field, common pot: Harvest variability and risk avoidance in agricultural and foraging societies. In Cashdan, E. A., editor, *Risk and Uncertainty in Tribal and Peasant Economies*, pages 67–87. Westview Press, Boulder, CO:.

Winterhalder, B., Kennett, D. J., Grote, M. N., and Bartruff, J. (2010). Ideal free settlement of California's northern Channel Islands. *Journal of Anthropological Archaeology*, 29:469–490.