Supplemental materials for: Demonstrating the Utility of Egocentric Relational Event Modelling Using Focal Follow Data from Congolese BaYaka Children and Adolescents Engaging in Work and Play

French Abstract: Les aspects de la répartition du temps des enfants et des adolescents dans divers contextes culturels ont été difficiles à modéliser à l'aide des techniques statistiques conventionnelles. Une nouvelle approche statistique, la modélisation d'événements relationnels égocentriques (EREM), permet de modéliser simultanément la fréquence, la durée et l'enchaînement des activités. Dans cet article le EREM est appliqué à un ensemble de données de suivi focal des activités de jeu et de travail des enfants et des adolescents BaYaka de la Republique du Congo. Les résultats montrent qu'en vieillissant, les enfants s'adonnent à des jeux moins fréquents et plus longs et à des travaux plus fréquents et plus longs. La fréquence et la durée des jeux sont une mesure plus sensible des différences entre les sexes à un âge précoce que l'allocation de temps globale. Les modèles séquentiels de travail et de jeu suggèrent que ces activités constituent des compromis énergétiques à court terme. Cet article démontre que le EREM peut révéler à la fois des schémas stables et variables dans le développement de l'enfant.

Sequence forms unpacked

In creating the s-form, it is important to include all the subsequences that match the pattern of interest in the definition. For interval timing models an s-form must be defined for both the onset and termination of the pattern. For example, the s-form for the onset of a work to play transition would be defined by the set of all subsequences where work is followed by the start of a play event (e.g., collecting|START → collecting|STOP → pretense play|START). The s-form for termination of the transition is therefore identical except that it includes the appropriate STOP event for each type of play (e.g., collecting|START → collecting|STOP → pretense play|START → pretense play|STOP). Creating s-forms is made easier because the *relevent* package represents each event type by a default alphabetic code “a” through to “zz” in the order they first occur in the data set (for a maximum of 52 event types). In addition, it is possible to use regular expression (“regex”) such as “ab” to represent a → b or (a|b)c to represent a → c OR b → c.

Consider a simplified example where there are only two types of events, one type of work event and one type of play event. The first event, a work event, would be represented by the default codes “a” and “b” representing the onset and termination of the work event. Let us assume that the second work event is coded “c” and “d” and the pairs “e’ and “f” and “g” and “h” represent other activities and play respectively. In this simple case, we could create the onset of the work to play transition s-form by defining the following subsequences that exhaust all possible work to play transitions with or without an interruption by the other activities event type:

Onset subsequence	Regular expression (regex)
a → b → g	abg
c → d → g	cdg
a → b → e → f→ g	abefg
c → d → e → f→ g	cdefg

Termination subsequence	Regular expression (regex)
a → b → g	abgh
c → d → g	cdgh

a → b → e → f→ g abefgh
c → d → e → f→ g cdefgh

For interval timing data a termination subsequence differs only by the addition of the termination of the final event in the sequence (e.g., "h" or "play|STOP" in this example). Thus, the onset subsequences identify the patterns of event that are of interest and enable the s-form to estimate their relative frequency of the final activity, while the termination sequence enables the relative duration of the final activity to be estimated. While the preceding example enumerated all the possible subsequences exhaustively, regular expressions syntax does not require this. For example (ab|cd)g is a disjunction that represents the sequences abg and cdg within a single regular expression (see Table 1 of Marcum and Butts 2015).

The four sets of s-forms included in the analysis were added to the model by modifying the statslist to incorporate the appropriate dummy code whenever an event history matches one of the subsequences defined by the s-form.[1] The model was then refitted to include the new statslist to obtain parameter estimates for the onset and duration of the work to play and play to work transitions we defined.

Granularity of event coding and handling repetition of events

A key consideration in modeling event data is the granularity at which to model behavior. This is partly constrained by the form in which the raw data are collected and informed by theoretical considerations. Here we use focal follow data taken at one-minute intervals on a relatively small number of children. Coding activity at a very high level of granularity produces a very large number of event types and results in a very sparse data set. Coding at a higher level of granularity allows us to capture types of activity that are of particular interest (e.g., different types of work and play) and fit a more parsimonious model. In addition, the resolution of focal follow data will not always support precise timing of events if high granularity of coding is employed. One consequence of coding types of activity at this level of granularity with focal follow data is that there are no repeated events. If there are repeated events in the data set it is important to capture the persistence of these behaviors. This can be done by including s-forms to estimate activity specific repetition effects (i.e., a → a, b → b etc.) or a global omnibus repetition effect (x → x) to control for this phenomenon.

Model comparison

Fit can assessed be with the likelihood ratio statistic (the change in deviance from a null model with no predictors) and information criteria (AIC and BIC). While, for the most part, we focus on parameter estimation rather than model comparison, we will report changes in AIC and BIC to demonstrate the use of information criteria in model comparison. The parameter estimates are 95% posterior probability interval—an interval in which there is a 95% probability that the true parameter falls assuming the model is correct. Comparing information criteria demonstrates that the model that includes sex as a covariate was an improvement on the baseline model, with both deviance and AIC substantially lower, though BIC (which penalizes more markedly the large number of additional parameters) is slightly higher. The model including age as a covariate has as substantially better fit in terms of lower residual deviance as well as both AIC and BIC relative to the baseline model. This provides evidence that the frequency or duration of event types varies between younger and older children. Finally, including s-forms into the model substantially improves the overall fit of the model—with residual deviance, AIC and BIC all considerably lower than for the baseline model.

Calculating overall time allocation
By multiplying bout duration by bout frequency, we can calculate model predictions regarding overall time allocation and compare these predictions with the mean proportion of observations spent in each activity from the raw data. We retained the missing values to account for time allocation to these missing activities. As can be seen in Table S1, the predicted values are consistent with the raw mean overall time allocation and demonstrate that the bulk of the children's time (52% of known events) is spent on other activities—and hence neither work nor play.

Table S1. Percent (%) predicted and observed overall time allocation for activities of interest.

Event type	Predicted overall time allocation	Observed overall time allocation
Household	5.79	5.78
Collecting	3.97	3.97
Tuber digging	2.57	2.58
Hunting	0.59	0.59
Honey collecting	0.39	0.38
Pretense play	5.54	5.56
Structured games	2.17	2.17
Other play	8.97	8.94
Other activities	32.88	32.89
Unknown	37.13	37.15

Table S2. Parameter estimates (posterior mean and SD with 95% posterior probability interval for the mean) for the intercept-only baseline egocentric relational event model.

	Intercepts for Activity Onset Frequency (START)			
Event type	Posterior *M* (*b*)	Posterior *SD*	Q2.5%	Q97.5%
Household	7.133	0.059	7.02	7.25
Collecting	6.237	0.091	6.05	6.41
Tuber digging	5.319	0.147	5.02	5.60
Hunting	3.934	0.296	3.32	4.47
Honey collecting	3.735	0.329	3.08	4.34
Pretense play	6.677	0.074	6.53	6.82
Structured games	5.646	0.125	5.38	5.88

Other play	7.737	0.046	7.65	7.82
Other activities	8.320	0.032	8.26	8.38
Unknown	6.605	0.078	6.45	6.75

Intercepts for Activity Duration (STOP)				
Event type	Posterior *M* (*b*)	Posterior *SD*	Q2.5%	Q97.5%
Household	-1.171	0.061	-1.29	-1.06
Collecting	-1.691	0.093	-1.88	-1.52
Tuber digging	-2.177	0.147	-2.47	-1.90
Hunting	-2.082	0.298	-2.69	-1.53
Honey collecting	-1.860	0.328	-2.58	-1.28
Pretense play	-1.589	0.074	-1.74	-1.45
Structured games	-1.678	0.074	-1.74	-1.45
Other play	-1.004	0.044	-1.09	-0.92
Other activities	-1.723	0.033	-1.787	-1.657
Unknown	-3.560	0.079	-3.720	-3.406

Notes. Null deviance (*df* = 4612) = 20496.9, residual deviance (*df* = 4592) =-17950.4, χ^2(*df* = 20) = 38447.3, AIC: -17910.4, BIC: -17781.7.

Table S3. Parameter estimates (posterior mean and *SD* with 95% posterior probability interval for the mean) including sex as a covariate. Covariate effects are estimated as male sex differences relative to the (female) intercept.

	Intercepts (female) for Activity Onset frequency (START)				
	Posterior *M* (*b*)	Posterior *SD*	Q2.5%	Q97.5%	
Event type					
Household	7.296	0.074	7.16	7.45	
Collecting	6.628	0.101	6.45	6.81	
Tuber digging	5.700	0.153	5.41	5.97	
Hunting	1.642	1.077	-0.24	3.04	
Honey collecting	1.707	1.074	-0.75	3.01	
Pretense play	6.405	0.092	6.23	6.58	
Structured games	5.759	0.152	5.43	6.02	
Other play	7.518	0.060	7.40	7.62	
Other activities	8.341	0.042	8.25	8.43	
Unknown	6.605	0.055	6.50	6.70	
Coefficients for sex differences (male = 1) for Activity Onset Frequency (START)					
	Posterior *M* (*b*)	Posterior *SD*	Q2.5%	Q97.5%	*Pr*
Event type					
Household	-0.326	0.060	-0.45	-0.22	< .0001
Collecting	-0.862	0.101	-1.05	-0.68	< .0001
Tuber digging	-0.836	0.183	-1.14	-0.51	< .0001
Hunting	2.242	0.224	1.80	2.69	< .0001
Honey collecting	2.020	0.282	1.56	2.62	< .0001

Pretense play	0.445	0.056	0.34	0.56	< .0001
Structured games	-0.242	0.121	-0.49	0.02	.0320
Other play	0.369	0.035	0.31	0.43	< .0001
Other activities	-0.031	0.032	-0.10	0.03	.1810

Intercepts (female) for Activity Duration (STOP)

Event type	Posterior *M* (*b*)	Posterior *SD*	Q2.5%	Q97.5%
Household	-1.470	0.070	-1.61	-1.34
Collecting	-1.640	0.094	-1.81	-1.45
Tuber digging	-1.978	0.182	-2.28	-1.68
Hunting	-1.435	1.172	-5.00	0.00
Honey collecting	-2.497	1.153	-4.73	-0.89
Pretense play	-1.786	0.099	-1.99	-1.60
Structured games	-1.592	0.149	-1.92	-1.31
Other play	-0.967	0.059	-1.08	-0.86
Other activities	-1.683	0.044	-1.77	-1.61
Unknown	-3.549	0.057	-3.65	-3.45

Coefficients For Sex Differences (male = 1) for Activity Duration (STOP)

Event type	Posterior *M* (*b*)	Posterior *SD*	Q2.5%	Q97.5%	*Pr*[†]
Household	0.804	0.060	0.70	0.92	< .0001
Collecting	-0.153	0.111	-0.38	0.05	.0867
Tuber digging	-0.512	0.195	-0.82	-0.17	.0013
Hunting	-1.283	0.252	-1.72	-0.67	< .0001
Honey collecting	0.049	0.242	-0.38	0.54	.4357
Pretense play	0.308	0.059	0.18	0.42	< .0001
Structured games	-0.173	0.134	-0.44	0.10	< .0001
Other play	-0.053	0.036	-0.12	0.02	< .0001
Other activities	-0.059	0.027	-0.11	0.00	.0167

Notes. Null deviance (df = 4612) = 20496.9, residual deviance (df = 4592) = -18067.7, $\chi^2(df = 38)$ = 38564.6, AIC: -17991.7, BIC: -17747.1. Change in model fit relative to the baseline model: $\Delta\chi^2(18)$ = 117.3, $p < .0001$, ΔAIC = -81.3, ΔBIC = 34.6.

[†] Bayesian '*p* value' obtained from twice the posterior probability that the effect is in the opposite direction to the observed effect.

Table S4. Parameter estimates (posterior mean and *SD* with 95% posterior probability interval for the mean) including age in years as a covariate. Covariate effects are estimated relative to the intercept for a child of mean age (10.6 years).

	Intercepts (for Child of Mean Age) for Activity Onset Frequency (START)				
Event type	Posterior *M* (*b*)	Posterior *SD*	Q2.5%	Q97.5%	
Household	7.151	0.044	7.08	7.23	
Collecting	6.238	0.065	6.11	6.37	
Tuber digging	4.960	0.137	4.70	5.22	
Hunting	2.998	0.336	2.14	3.60	
Honey collecting	3.597	0.279	3.06	4.17	
Pretense play	6.528	0.062	6.41	6.64	
Structured games	5.654	0.086	5.46	5.80	
Other play	7.562	0.037	7.49	7.63	
Other activities	8.316	0.022	8.28	8.36	
Unknown	6.608	0.050	6.50	6.70	
Coefficient of Age with Respect to Activity Onset Frequency (START)					
Event type	Posterior *M* (*b*)	Posterior *SD*	Q2.5%	Q97.5%	*p*
Household	0.085	0.009	0.07	0.10	<.0001
Collecting	0.101	0.014	0.08	0.13	<.0001
Tuber digging	0.274	0.023	0.23	0.31	<.0001
Hunting	0.408	0.038	0.33	0.48	<.0001
Honey collecting	0.206	0.056	0.12	0.34	<.0001
Pretense play	-0.089	0.010	-0.11	-0.07	<.0001
Structured games	0.018	0.020	-0.02	0.06	.1816
Other play	-0.096	0.006	-0.11	-0.08	<.0001
Other activities	0.000	0.005	-0.01	0.01	.5334
Intercepts (for Child of Mean Age) for Activity Duration (STOP)					
Event type	Posterior *M* (*b*)	Posterior *SD*	Q2.5%	Q97.5%	
Household	-1.140	0.046	-1.25	-1.06	
Collecting	-1.562	0.066	-1.68	-1.44	
Tuber digging	-1.968	0.163	-2.30	-1.64	
Hunting	-6.954	0.722	-8.05	-4.92	
Honey collecting	-1.682	0.564	-3.21	-0.85	
Pretense play	-1.696	0.058	-1.81	-1.60	
Structured games	-1.632	0.081	-1.78	-1.49	
Other play	-1.016	0.034	-1.08	-0.94	
Other activities	-1.751	0.023	-1.80	-1.71	
Unknown	-3.549	0.058	-3.67	-3.46	
	Coefficient of Age with Respect to Activity Duration (STOP)				
Event type	Posterior *M* (*b*)	Posterior *SD*	Q2.5%	Q97.5%	*Pr*

Household	-0.035	0.011	-0.05	-0.01	.0010
Collecting	-0.091	0.016	-0.12	-0.06	<.0001
Tuber digging	-0.062	0.027	-0.12	-0.02	.0052
Hunting	0.644	0.046	0.57	0.74	<.0001
Honey collecting	-0.121	0.074	-0.26	0.03	.0482
Pretense play	-0.053	0.012	-0.08	-0.03	.0002
Structured games	-0.126	0.019	-0.16	-0.09	<.0001
Other play	-0.005	0.008	-0.02	0.01	.2134
Other activities	-0.047	0.005	-0.06	-0.04	<.0001

Table S5. Parameter estimates (posterior mean and *SD* with 95% posterior probability interval for the mean) for s-forms including play to work and work to play transitions.

	Intercept—Activity Onset Frequency (START)				
	Posterior *M* (*b*)	Posterior *SD*	Q2.5%	Q97.5%	
Event type					
Household	7.492	0.097	7.30	7.68	
Collecting	6.778	0.120	6.54	7.01	
Tuber digging	5.857	0.152	5.55	6.14	
Hunting	4.494	0.287	3.90	5.02	
Honey collecting	4.296	0.314	3.65	4.87	
Pretense play	6.884	0.078	6.72	7.03	
Structured games	5.854	0.124	5.60	6.08	
Other play	7.942	0.066	7.81	8.07	
Other activities	8.320	0.031	8.26	8.38	
Unknown	6.606	0.075	6.45	6.75	
	S-forms for Activity Onset Frequency (START)				
	Posterior *M* (*b*)	Posterior *SD*	Q2.5%	Q97.5%	*p*
Event type					
Household to play	-0.567	0.086	-0.74	-0.40	< .0001
Food production to play	-1.079	0.144	-1.36	-0.80	< .0001
Play to household	-0.780	0.095	-0.96	-0.60	< .0001
Play to food production	-1.387	0.143	-1.67	-1.11	< .0001
	Intercept—Activity Duration (STOP)				
	Posterior *M* (*b*)	Posterior *SD*	Q2.5%	Q97.5%	
Event type					
Household	-1.348	0.100	-1.55	-1.16	
Collecting	-1.732	0.127	-1.98	-1.50	
Tuber digging	-2.216	0.150	-2.52	-1.92	
Hunting	-2.077	0.277	-2.66	-1.56	
Honey collecting	-1.873	0.312	-2.53	-1.31	
Pretense play	-1.596	0.076	-1.75	-1.45	
Structured games	-1.678	0.125	-1.93	-1.44	
Other play	-1.013	0.066	-1.14	-0.89	
Other activities	-1.722	0.032	-1.79	-1.66	
Unknown	-3.558	0.074	-3.71	-3.42	
	S-forms for Activity Duration (STOP)				
	Posterior *M* (*b*)	Posterior *SD*	Q2.5%	Q97.5%	*Pr*
Event type					
Household to play	-0.040	0.086	-0.21	0.13	.3197
Food production to play	0.149	0.145	-0.14	0.43	.1517
Play to household	0.545	0.097	0.36	0.74	< .0001
Play to food production	0.121	0.147	-0.17	0.40	.2103

Notes. Null deviance (df = 4612) = 20496.9, residual deviance (df = 4592) = -18234.4, $\chi^2(df = 38)$ = 38731.4, AIC: -18158.4, BIC: -17913.9. Change in model fit relative to the baseline model: $\Delta\chi^2(18)$ = 284.0, p < .0001, ΔAIC = -248.0, ΔBIC = -132.2.

Notes. Null deviance (*df* = 4612) = 20608.9, residual deviance (*df* = 4584) = -18178.3, χ^2(*df* = 28) = 38675.2, AIC: -18122.3, BIC: -17942.1. Change in model fit relative to the baseline model: $\Delta\chi^2(8) = 227.9$, $p < .0001$, ΔAIC = -211.9, ΔBIC = -160.4.

Note

1. The statslist is the list of statistics, effectively a list of contrast matrices, defining the sufficient statistics to be estimated for the model.