A. Derivations

The marginal likelihood, also known as Bayes factor, is the expectation of the likelihood under the prior distribution. The logarithm of this quantity (LML) is

$$LML = \log \mathbb{E}_{\mathbf{p}} \left[\Pr(\mathbf{N} \mid \mathbf{p}) \right]$$
$$= \log \mathbb{E}_{\mathbf{p}} \left(\prod_{\mathbf{x}} \prod_{m=1}^{M} p_{\mathbf{x},m}^{N_{\mathbf{x},m}} \right)$$
$$= \sum_{\mathbf{x}} \log \left(\frac{B(\mathbf{N}_{\mathbf{x}} + \alpha)}{B(\alpha)} \right).$$
(A 1)

The log predictive density (LPD) given a model defined by an inferred posterior distribution $\boldsymbol{p}|\boldsymbol{N}$ may be computed

$$LPD = \log \mathbb{E}_{\mathbf{p}|\mathbf{N}} [Pr(\mathbf{N} | \mathbf{p})]$$

= $\log \mathbb{E}_{\mathbf{p}|\mathbf{N}} \left(\prod_{\mathbf{x}} \prod_{m=1}^{M} p_{\mathbf{x},m}^{N_{\mathbf{x},m}} \right)$
= $\sum_{\mathbf{x}} \log \left(\frac{B(2\mathbf{N}_{\mathbf{x}} + \alpha)}{B(\mathbf{N}_{\mathbf{x}} + \alpha)} \right).$ (A 2)

The log pointwise predictive density (LPPD), requires partition of data into disjoint "points." Treating trajectories as points yields

$$LPPD = \sum_{j} \sum_{\mathbf{x}} \log \mathbb{E}_{\mathbf{p}_{\mathbf{x}} | \mathbf{N}_{\mathbf{x}}} \left[\Pr \left(\mathbf{N}_{\mathbf{x}}^{(j)} \mid \mathbf{p}_{\mathbf{x}} \right) \right]$$
$$= \sum_{j} \sum_{\mathbf{x}} \log \mathbb{E}_{\mathbf{p}_{\mathbf{x}} | \mathbf{N}_{\mathbf{x}}} \left(\prod_{m=1}^{M} p_{\mathbf{x},m}^{N_{\mathbf{x},m}^{(j)}} \right)$$
$$= \sum_{j} \sum_{\mathbf{x}} \log \left(\frac{B(\mathbf{N}_{\mathbf{x}} + \mathbf{N}_{\mathbf{x}}^{(j)} + \alpha)}{B(\mathbf{N}_{\mathbf{x}} + \alpha)} \right).$$
(A 3)

The LOO, as defined in this manuscript, is similar to the LPPD. For the LOO, each of the pointwise posterior distributions is computed after leaving out the corresponding trajectory. Hence,

$$LOO = -2\sum_{j}\sum_{\mathbf{x}} \log \mathbb{E}_{\mathbf{p}_{\mathbf{x}}|\mathbf{N}_{\mathbf{x}}\setminus\mathbf{N}_{\mathbf{x}}^{(j)}} \left[\Pr\left(\mathbf{N}_{\mathbf{x}}^{(j)} \mid \mathbf{p}_{\mathbf{x}}\right) \right]$$
$$= -2\sum_{j}\sum_{\mathbf{x}} \log \mathbb{E}_{\mathbf{p}_{\mathbf{x}}|\mathbf{N}_{\mathbf{x}}\setminus\mathbf{N}_{\mathbf{x}}^{(j)}} \left(\prod_{m=1}^{M} p_{\mathbf{x},m}^{N_{\mathbf{x},m}^{(j)}}\right)$$
$$= -2\sum_{j}\sum_{\mathbf{x}} \log \left(\frac{B(\mathbf{N}_{\mathbf{x}} - \mathbf{N}_{\mathbf{x}}^{(j)} + \mathbf{N}_{\mathbf{x}}^{(j)} + \alpha)}{B(\mathbf{N}_{\mathbf{x}} - \mathbf{N}_{\mathbf{x}}^{(j)} + \alpha)}\right).$$
(A 4)

For the WAIC, the two variants of complexity parameters are

$$k_{\text{WAIC1}} = 2\text{LPPD} - 2\sum_{j}\sum_{\mathbf{x}} \mathbb{E}_{\mathbf{p}_{\mathbf{x}}|\mathbf{N}} \left[\log \mathbf{p}_{\mathbf{x}}^{N_{\mathbf{x}}^{(j)}}\right]$$

$$= 2\text{LPPD} - \sum_{j}\sum_{\mathbf{x}}\sum_{m=1}^{M} N_{\mathbf{x},m}^{(j)} \mathbb{E}_{\mathbf{p}_{\mathbf{x}}|\mathbf{N}_{\mathbf{x}}} \left(\log p_{\mathbf{x},m}\right)$$

$$= 2\text{LPPD} - 2\sum_{j}\sum_{\mathbf{x}}\sum_{m=1}^{M} N_{\mathbf{x},m}^{(j)} \left[\psi(N_{\mathbf{x},m} + \alpha_m) - \psi\left(N_{\mathbf{x}} + \sum_{m} \alpha_m\right)\right]$$

$$= 2\text{LPPD} - 2\sum_{\mathbf{x}}\sum_{m=1}^{M} N_{\mathbf{x},m} \left[\psi(N_{\mathbf{x},m} + \alpha_m) - \psi\left(N_{\mathbf{x}} + \sum_{m} \alpha_m\right)\right], \quad (A 5)$$

and

$$k_{\text{WAIC2}} = \sum_{j} \sum_{\mathbf{x}} \operatorname{var}_{\mathbf{p}_{\mathbf{x}}} \left[\log \Pr \left(\mathbf{N}_{\mathbf{x}}^{(j)} \mid \mathbf{p}_{\mathbf{x}} \right) \right]$$

$$= \sum_{j} \sum_{\mathbf{x}} \operatorname{var}_{\mathbf{p}_{\mathbf{x}}} \left\{ \log \left(\prod_{m=1}^{M} p_{\mathbf{x},m}^{(j)} \right) \right\}$$

$$= \sum_{j} \sum_{\mathbf{x}} \operatorname{var}_{\mathbf{p}_{\mathbf{x}}} \left[\sum_{m=1}^{M} N_{\mathbf{x},m}^{(j)} \log p_{\mathbf{x},m} \right]$$

$$= \sum_{j} \sum_{\mathbf{x}} \sum_{m=1}^{M} \sum_{n=1}^{M} N_{\mathbf{x},m}^{(j)} N_{\mathbf{x},n}^{(j)} \operatorname{cov} \left(\log p_{\mathbf{x},m}, \log p_{\mathbf{x},n} \right)$$

$$= \sum_{j} \sum_{\mathbf{x}} \sum_{m=1}^{M} \sum_{n=1}^{M} N_{\mathbf{x},m}^{(j)} N_{\mathbf{x},n}^{(j)} \left[\psi' \left(\alpha_{n} + N_{\mathbf{x},n} \right) \delta_{nm} - \psi' \left(\sum_{m} \alpha_{m} + N_{\mathbf{x}} \right) \right]$$

$$= \sum_{j} \sum_{\mathbf{x}} \sum_{m=1}^{M} \left[\sum_{m=1}^{M} [N_{\mathbf{x},m}^{(j)}]^{2} \psi' \left(\alpha_{m} + N_{\mathbf{x},m} \right) - [N_{\mathbf{x}}^{(j)}]^{2} \psi' \left(\sum_{m} \alpha_{m} + N_{\mathbf{x}} \right) \right].$$
(A 6)

The commonly used Deviance Information Criterion (DIC)

$$DIC = -2\sum_{\mathbf{x}} \log p\left(\mathbf{N}_{\mathbf{x}} \mid \mathbf{p}_{\mathbf{x}} = \mathbb{E}_{\mathbf{p}_{\mathbf{x}} \mid \mathbf{N}_{\mathbf{x}}} \mathbf{p}_{\mathbf{x}}\right) + 2k_{\text{DIC}}$$
(A 7)

also resembles the WAIC, consisting of two variants in the computation of model complexity,

$$\begin{split} k_{\text{DIC1}} &= -2\left\{\sum_{\mathbf{x}}\sum_{m=1}^{M}N_{\mathbf{x},m}\log\left(\frac{N_{\mathbf{x},m}+\alpha_{m}}{N_{\mathbf{x}}+\sum_{m}\alpha_{m}}\right) - \sum_{j}\sum_{\mathbf{x}}\mathbb{E}_{\mathbf{p}_{\mathbf{x}}|\mathbf{N}}\log\mathbf{p}_{\mathbf{x}}^{\mathbf{N}_{\mathbf{x}}^{(j)}}\right\} \\ &= 2\left\{\sum_{\mathbf{x}}\sum_{m=1}^{M}N_{\mathbf{x},m}\log\left(\frac{N_{\mathbf{x},m}+\alpha_{m}}{N_{\mathbf{x}}+\sum_{m}\alpha_{m}}\right) - \sum_{j}\sum_{\mathbf{x}}\sum_{m=1}^{M}\mathbf{N}_{\mathbf{x},m}^{(j)}\left[\psi(\alpha_{m}+N_{\mathbf{x},m})-\psi\left(\sum_{m}\alpha_{m}+N_{\mathbf{x}}\right)\right]\right\} \\ &= 2\left\{\sum_{\mathbf{x}}\sum_{m=1}^{M}N_{\mathbf{x},m}\log\left(\frac{N_{\mathbf{x},m}+\alpha_{m}}{N_{\mathbf{x}}+\sum_{m}\alpha_{m}}\right) - \sum_{\mathbf{x}}\sum_{m=1}^{M}N_{\mathbf{x},m}\left[\psi(\alpha_{m}+N_{\mathbf{x},m})-\psi\left(\sum_{m}\alpha_{m}+N_{\mathbf{x}}\right)\right]\right\}, \end{split}$$

$$(A 8)$$

rsos.royalsocietypublishing.org R. Soc. open sci. 0000000

and $k_{\text{DIC2}} = 2 \text{var}_{\mathbf{p} | \mathbf{N}} [\log \Pr(\mathbf{N} | \mathbf{p})]$, which may be computed

$$\begin{aligned} k_{\text{DIC2}} &= 2 \text{var}_{\mathbf{p}_{\mathbf{x}}} \left[\sum_{\mathbf{x}} \sum_{m} N_{\mathbf{x},m} \log p_{\mathbf{x},m} \right] \\ &= 2 \sum_{\mathbf{x}} \text{var}_{\mathbf{p}_{\mathbf{x}}} \left(\sum_{m} N_{\mathbf{x},m} \log p_{\mathbf{x},m} \right) \\ &= 2 \sum_{\mathbf{x}} \sum_{m} \sum_{n} N_{x,m} N_{x,n} \text{cov}(\log p_{\mathbf{x},m}, \log p_{\mathbf{x},n}) \\ &= 2 \sum_{\mathbf{x}} \sum_{m} \sum_{n} N_{x,m} N_{x,n} \left[\psi'(\alpha_{m} + N_{\mathbf{x},m}) \delta_{mn} - \psi'\left(\sum_{m} \alpha_{m} + N_{\mathbf{x}}\right) \right] \\ &= 2 \sum_{\mathbf{x}} \left(\sum_{m} N_{\mathbf{x},m}^{2} \psi'(\alpha_{m} + N_{\mathbf{x},m}) - (N_{\mathbf{x}})^{2} \psi'\left(\sum_{m} \alpha_{m} + N_{\mathbf{x}}\right) \right), \end{aligned}$$
(A 9)

where δ_{mn} refers to the Kronecker delta function.

B. Supplemental results

In Fig. 6, it is apparent that models where h = 0 and h = 1 have comparable predictive power. As a form of permutation test, we consider resamplings without replacement of James' 2016–2017 free throws where within each game the order of his shot outcomes are scrambled. The results here (Fig. S1) are similar to those found in Fig. 6, where the statistical power of the information criteria are evaluated using simulated data.

Fig. S2 presents a version of the same tests performed in the main manuscript (where an eight state system is used) on a four-state system. In comparing Fig. 2 to Fig. S2, one finds consistent results. Note that on average the simulated trajectory lengths are shorter in the four state system relative to the eight state system due to the fact that there are fewer interior states relative to the number of absorbing states.

3

		J = 4	J = 16	J = 64	J = 256	J = 1024	J = 4096	J = 16384	
e = 0	WAIC ₁	63% 35%	71%	65%	86%	86%	93%	100%	WAIC ₁
	$WAIC_2$	84%	86%	86%	100%	86%	93%	100%	WAIC ₂
	LOO	91%	93%	93%	100%	86%	93%	100%	-L00
	LHO	84%	85%	79%	100%	93%	86%	100%	LHO
	DIC ₁	63%	71%	65%	86%	86%	93%	100%	DIC ₁
$i_{\rm tru}$	DIC_2	93%	100%	100%	100%	93%	93%	100%	DIC ₂
	AIC	100%	100%	100%	100%	86%	93%	100%	AIC
	BIC	100%	100%	100%	100%	100%	100%	100%	BIC
	-2LML	55%	78%	84%	77%	100%	100%	100%	-2LML
	WAIC	50%	43% 35%	93%	93%	100%	93%	100%	WAIC
	WAICa	68%	71%	03%	100%	100%	100%	100%	WAIC ₂
	100	61%	71%	100%	100%	100%	100%	100%	100
H	LHO	68%	71%	93%	86%	100%	100%	93%	1.00
	DIC	46%	36% 35%	93%	03%	100%	100%	100%	
true	DIC	36% 64%	100%	100%	100%	100%	100%	100%	
ч	AIC.	78%	03%	100%	100%	100%	100%	100%	<u>-</u>
	BIC	75%	100%	100%	100%	100%	100%	100%	RIC
	-21 ML	35%	50%	86%	100%	100%	03%	100%	-21 MI
	-201410	3370	3070	0070	10070	10070	5570	10070	-261416
	WAIC ₁	56%	79%	93%	100%	100%	100%	100%	WAIC ₁
	WAIC ₂	56%	93%	100%	100%	100%	100%	100%	WAIC ₂
~	LOO	56%	100%	100%	100%	100%	100%	100%	-L00
	LHO	44%	72%	98%	100%	100%	100%	100%	LHO
. en	DIC ₁	50%	65%	93%	100%	100%	100%	100%	DIC1
$h_{ m tr}$	DIC ₂	69% 31%	43% 57%	100%	100%	100%	100%	100%	DIC ₂
	AIC	75%	42% 58%	100%	100%	100%	100%	100%	AIC
	BIC	88%	99%	58% 42%	100%	100%	100%	100%	BIC
	-2LML		85%	100%	100%	100%	100%	100%	-2LML
	WAIC ₁	36%	42%	56%	100%	100%	100%	100%	WAIC ₁
	WAIC ₂	35%	42%	72%	100%	100%	100%	100%	WAIC ₂
	LOO	35%	56%	72%	100%	100%	100%	100%	-L00
3	LHO	35%	43%	79%	100%	100%	100%	100%	LHO
1	DIC ₁		35%	58%	100%	100%	100%	100%	DIC ₁
k true	DIC ₂	64% 36%	58% 42%	49% 37%	100%	100%	100%	100%	DIC ₂
1	AIC	49% 51%	79%	51% 35%	100%	100%	100%	100%	AIC
	BIC	86%	86%	100%	65%	93%	100%	100%	BIC
	-2LML		50%	86%	100%	100%	100%	100%	-2LML
	MAIC	260/ 260/	269/	700/	000/	1009/	1008/	1009/	MAIC
	WAIC1		30%	70%	90%	100%	100%	100%	WAIC1
	WAIC ₂	43% 35%	30%	11%	100%	100%	100%	100%	- WAIC2
4	LUO	- 30% 30%	30%	0470 709/	100%	100%	100%	100%	-LOO
	LHO	- 30%	44%	70%	91%	100%	100%	100%	LHU
rue	DIC1	43% 30%	30%	70%	98%	100%	100%	100%	
$h_{ m t}$	DIC ₂		70%	30% 42%	04% 70%	93%	100%	100%	-DIC ₂
	AIC	02%	10%	100%	70%	93%	100%	100%	AIC
	BIC	92%	93%	100%	44% 50%	00%	00%	100%	BIC
	-2LIVIL	3370	43%	100 /0	100 /0	100 /0	100 /0	100 %	-2LIVIL
	WAIC ₁	50%	42%	44% 49%	86%	100%	100%	100%	WAIC ₁
	WAIC ₂	43%		44% 42%	93%	100%	100%	100%	WAIC ₂
	L00	43%	42%	58%	100%	100%	100%	100%	-L00
10	LHO	- 50%		44%	72%	100%	100%	100%	LHO
 •	DIC_1	64%	42%	37% 56%	93%	100%	100%	100%	DIC ₁
2 tru	DIC ₂	63% 37%	50% 50%	42%	72%	44% 56%	93%	100%	DIC ₂
-	AIC	63% 37%	72%	35% 58%	86%	65% <mark>35%</mark>	93%	100%	AIC
	BIC	93%	71%	93%	51% 49%	58%	42% 51%	79%	BIC
	-21 ML	50%	56%	100%	100%	100%	100%	100%	- 21 MI

Figure S1. Permuting LeBron James' 16'-17' free throws. (a) Distributions of $\Delta Criterion(h) = Criterion(h) - Criterion(h = 0)$ Evaluations of information criterion relative to h = 0 for resamplings without replacement of James' 16'-17' free throws. (b) Frequency of choosing h = 0: black, 1: red, 2: blue, 3: green.

4

		J = 4	J = 16	J = 64	J = 256	J = 1024	J = 4096	J = 16384	
e = 0	WAIC ₁	63% 35%	71%	65%	86%	86%	93%	100%	- WAIC ₁
	WAIC ₂	84%	86%	86%	100%	86%	93%	100%	- WAIC ₂
	L00 ·	91%	93%	93%	100%	86%	93%	100%	-L00
	LHO ·	84%	85%	79%	100%	93%	86%	100%	LHO
	DIC ₁	63%	71%	65%	86%	86%	93%	100%	- DIC ₁
tru	DIC ₂	93%	100%	100%	100%	93%	93%	100%	DIC ₂
1	AIC	100%	100%	100%	100%	86%	93%	100%	AIC
	BIC	100%	100%	100%	100%	100%	100%	100%	BIC
	LML	55%	78%	84%	77%	100%	100%	100%	LML
	MAIC	F09/	420/ 250/	0.20/	0.29/	1009/	0.20/	100%	MAIC
	WAIC1	- 30% - 60%	43% 33%	93%	93%	100%	93%	100%	WAIC1
	WAIC ₂	619/	71%	93%	100%	100%	100%	100%	- WAIC ₂
-	LOO	60%	71%	100%	100%	100%	100%	100%	-100
	LHU	00%		93%	00%	100%	100%	93%	- LHO
rue	DIC1.	269/ 649/	30% 33%	93%	93%	100%	100%	100%	
\mathbf{p}_{f}	DIC ₂ .	30% 04%	100%	100%	100%	100%	100%	100%	DIC ₂
	AIC	70%	93%	100%	100%	100%	100%	100%	AIC
	BIC	/5%	100%	100%	100%	100%	100%	100%	BIC
	LML	35%	50%	80%	100%	100%	93%	100%	LML
	WAIC ₁	56%	79%	93%	100%	100%	100%	100%	- WAIC ₁
	WAIC ₂	56%	93%	100%	100%	100%	100%	100%	- WAIC ₂
~	L00 ·	56%	100%	100%	100%	100%	100%	100%	-L00
1	LHO ·	- 44%	72%	98%	100%	100%	100%	100%	LHO
e	DIC ₁	50%	65%	93%	100%	100%	100%	100%	- DIC ₁
$h_{\rm tr}$	DIC ₂	69% 31%	43% 57%	100%	100%	100%	100%	100%	- DIC ₂
	AIC	- 75%	42% 58%	100%	100%	100%	100%	100%	- AIC
	BIC ·	88%	99%	58% 42%	100%	100%	100%	100%	BIC
	LML		85%	100%	100%	100%	100%	100%	- LML
	WAIC ₁	36%	42%	56%	100%	100%	100%	100%	- WAIC ₁
	WAIC ₂	35%	42%	72%	100%	100%	100%	100%	- WAIC ₂
	LOO ·	35%	56%	72%	100%	100%	100%	100%	-L00
3	LHO ·	35%	43%	79%	100%	100%	100%	100%	LHO
 	DIC ₁		35%	58%	100%	100%	100%	100%	DIC1
tra	DIC ₂	64% 36%	58% 42%	49% 37%	100%	100%	100%	100%	- DIC ₂
4	AIC	49% 51%	79%	51% 35%	100%	100%	100%	100%	AIC
	BIC	86%	86%	100%	65%	93%	100%	100%	BIC
	LML		50%	86%	100%	100%	100%	100%	- LML
	WAIC	260/ 260/	260/	70%	00%	100%	100%	100%	WAIC
	WAIC-	130/0 30/0 130/ 3E0/	36%	77%	100%	100%	100%	100%	WAIC
	100	259/ 259/	26%	040/	100%	100%	100%	100%	100
4	100.	35/6 35/6	44%	70%	01%	100%	100%	100%	
	DIC	13% 36%	36%	70%	08%	100%	100%	100%	DIC.
true	DIC	63% 37%	78%	58% 42%	84%	03%	100%	100%	
h_{i}	AIC	56% 44%	78%	77%	70%	93%	100%	100%	- AIC
	BIC	92%	93%	100%	44% 56%	86%	86%	100%	BIC
	LMI	35%	43%	100%	100%	100%	100%	100%	LML
	WAIC ₁	50%	42%	44% 49%	86%	100%	100%	100%	- WAIC1
	WAIC ₂	43%		44% 42%	93%	100%	100%	100%	- WAIC ₂
10	100 ·	43%	42%	58%	100%	100%	100%	100%	- LOO
$h_{ m true} = 1$	LHO	50%		44%	/2%	100%	100%	100%	LHO
		04%	42%	3/% 50%	93%	100%	100%	100%	
	DIC2 -	63% 37%	50% 50%	42%	72%	44% 56%	93%	100%	
	AIC	03% 37%	72%	<u>35%</u> 58%	519/ 409/	05% <u>35%</u>	93%	100%	AIC
	BIC	93%	71% 56%	93%	100%	100%	42% 51%	100%	BIC
	1 1/11 -	310770	20%	10070	10070	11/10/20	10070	10070	

Figure S2. Chosen degree of memory h for M = 4 system in simulations for varying true degrees of memory h_{true} and number of observed trajectories J. Choices made on basis of model with lowest value of given criterion. Rows correspond to model selection under a given degree of memory. Columns correspond to the number of trajectories. Depicted are the percent of simulations in which each degree of memory is selected using the different model evaluation criteria (percents of at least 20 are labeled). Colors coded based on degree of memory: (0: black, 1: red, 2: blue, 3: green, 4: purple, 5: orange). Example: For $h_{true} = 1$ and J = 4, the WAIC₁ criteria selected h = 1 approximately 68% of the time.

5

rsos.royalsocietypublishing.org R. Soc. open sci. 0000000