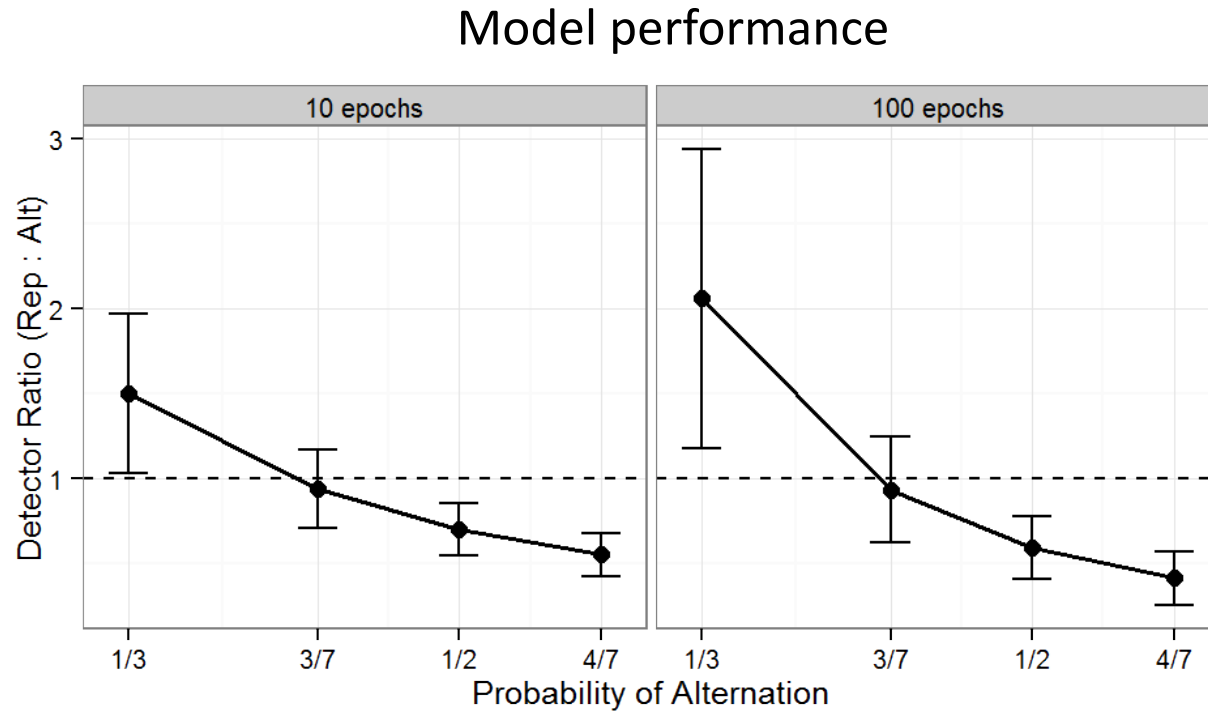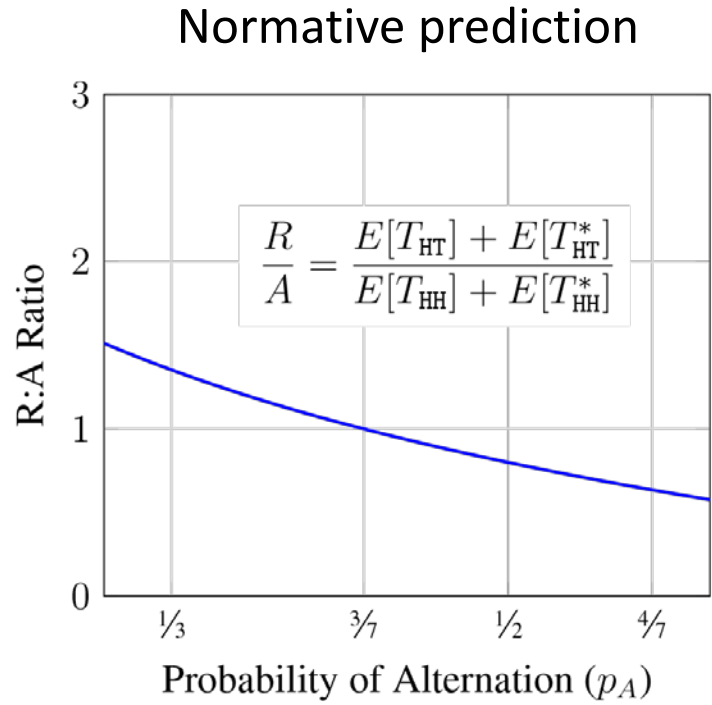# Summary

- Detector ratio R/A: the ratio between the numbers of repetition (HH or TT) and alternation (HT or TH) detectors:
  - R/A decreases as p_A increases. R-A asymmetry: $R/A < 1$ at $p_A = 1/2$ (i.e., representativeness bias and gambler's fallacy).
  - "Equilibrium" $R/A \approx 1$ at $p_A = 3/7$, where $E[T_{\mathrm{HH}}] + E[T_{\mathrm{HH}}^*] = E[T_{\mathrm{HT}}] + E[T_{\mathrm{HT}}^*]$, indicating that mean time and waiting time make equal contributions to pattern dissociation.
  - Robust results when network parameters vary over fairly wide ranges.

- Training on indicator variables ("null model"):
  - Effectively keeps the effect of mean time but removes the effect of waiting time.
  - $R/A \approx 1$ at $p_A = 1/2$ (i.e., no R-A asymmetry), indicating that the self-overlapping property of local patterns, namely, waiting time, is responsible for the R-A asymmetry at $p_A = 1/2$.

- A new Bayesian model ("LM") to fit human data (Goodfellow, 1938):
  - The model is an extension of Griffiths and Tenenbaum (2001)'s Bayesian model of random sequence production.
  - With the results from the neural model (naturally emerged p_A' and λ), plus one additional log-likelihood function for the second-order pattern event (applying the same λ), the new Bayesian model outperformed the model by G&T 2001.

R-A Dissociation & Parameter Sweeps
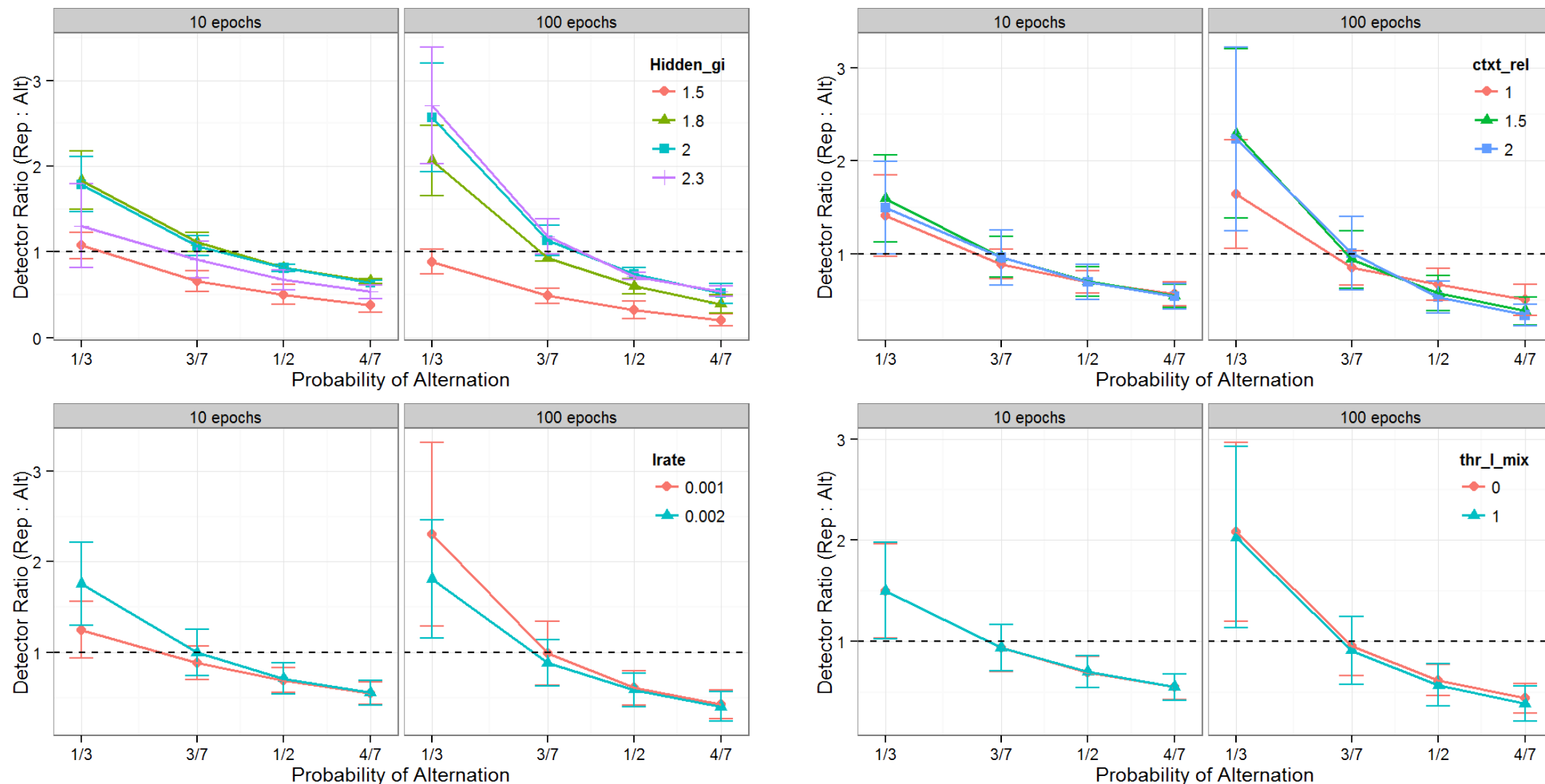
- R-A ratio is monotonically sensitive to the training p_A.
- Robust result: R-A asymmetry $R/A \approx 0.6{\sim}0.8$ at $p_A = 1/2$ and equilibrium $R/A \approx 1$ at $p_A = 3/7$.
- Hidden layer inhibition (**Hidden_gi**) is the most significant factor. At $p_A = 1/2$, it is the only factor.

Equilibrium at $p_A = 3/7$

Normative prediction

$$\frac{R}{A} = \frac{E[T_{\text{HT}}] + E[T_{\text{HT}}^*]}{E[T_{\text{HH}}] + E[T_{\text{HH}}^*]}$$

R:A Ratio — Probability of Alternation (p_A)

Model performance

Detector Ratio (Rep : Alt) — Probability of Alternation

- Repetition (HH or TT) and alternation (HT or TH) patterns have the same mean time $E[T]$ at $p_A = 1/2$, and the same waiting time $E[T^*]$ at $p_A = 1/3$.
- An "equilibrium" at $p_A = 3/7$ can be derived where $E[T_{\text{HH}}] + E[T_{\text{HH}}^*] = E[T_{\text{HT}}] + E[T_{\text{HT}}^*]$.
- We predict that the neural model would show R-A indifference at $p_A = 3/7$. The model did exactly that!
- Note that more trainings (e.g., 100 epochs, 100 trials each) may change the detector ratio at other p_A levels, but not at $p_A = 3/7$, which is a desired property of equilibrium.
- This result indicates that mean time and waiting time make equal contributions in pattern dissociation.

Subjective Probability of Alternation p'_A

A major focus of the current study is to account for human "biases" in randomness perception, e.g.,

- people's avoidance of repeating patterns,

- subjective probability of alternation p'_A different from actual $p_A = 1/2$ in truly random sequences.

Average $R/A \approx 0.70$ at $p_A = 1/2$:

$$p'_A = \frac{A}{A+R} = \frac{1}{1 + R/A} = \frac{1}{1 + 0.70} \approx 0.59$$

- Naturally emerging: $R/A < 1$, namely, under-representation or avoidance of repeating patterns. This amounts to the representativeness bias and the gambler's fallacy (Kahneman & Tversky, 1972).

- Subjective probability of alternation $p'_A \approx 0.59$ is consistent with existing empirical findings on human random sequence production (e.g., Falk & Konold, 1997), and consistent with theoretical analyses of waiting time statistics (Sun & Wang, 2010, 2012).

Training on Indicator Variables ("IV_TRAIN") as a "Null Model"

original sequence of flips:

H	T	H	H	T	H	H	H	T	T	H	H

sequence of indicator variables:

|	A	A	R	A	A	R	R	A	R	A	R

- In the original sequence, pattern events (repetition HH, TT, or alternation HT, TH) may have the same mean time but different waiting time (particularly at $p_A = 1/2$) due to the temporal overlaps of pattern elements (e.g., HH occurs twice in sequence HHH).

- In the sequence of indicator variables, events (repetition R or alternation A) have the same mean time as the corresponding pattern events in the original sequence. However, because now there are no temporal overlaps, events R and A have the same mean time and waiting time at each p_A level.

- Training on the indicator variables effectively keeps the effect of mean time but removes the effect of waiting time.

Training on Indicator Variables as a "Null Model"
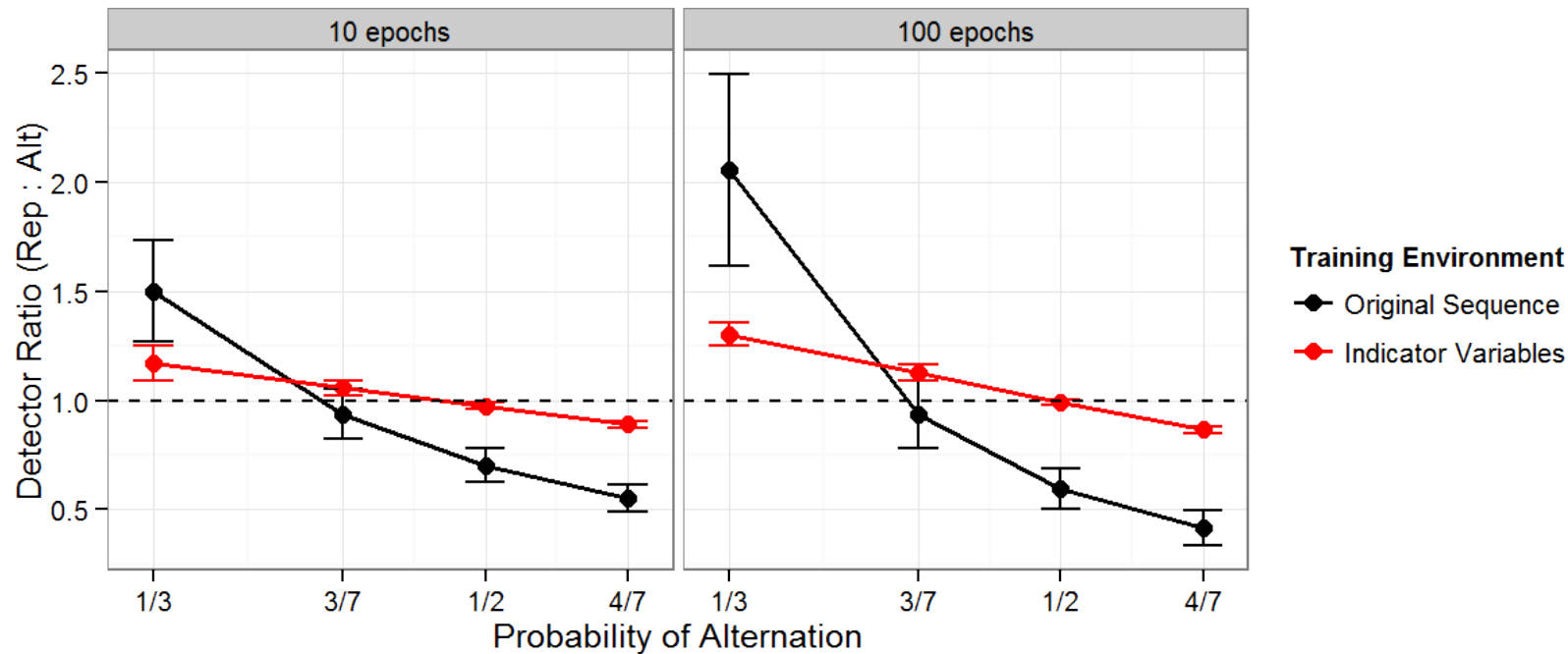
- When training on the indicator variables (without the effect of waiting time): detector ratio R/A is still monotonically responsive to p_A due to different training frequencies (i.e., the effect of mean time), however, with a different equilibrium $R/A \approx 1$ at $p_A = 1/2$.

- This finding confirms that the temporal overlap of local patterns in the original sequences, namely, the waiting time, is responsible for the R-A asymmetry at $p_A = 1/2$.

- Note that when training on the original sequences, detector ratio R/A is more sensitive to the underlying p_A, consistent with our previous argument that pattern waiting times can facilitate signal detection (Sun & Wang, 2010).

Modeling Random Sequence Production: A Bayesian Model
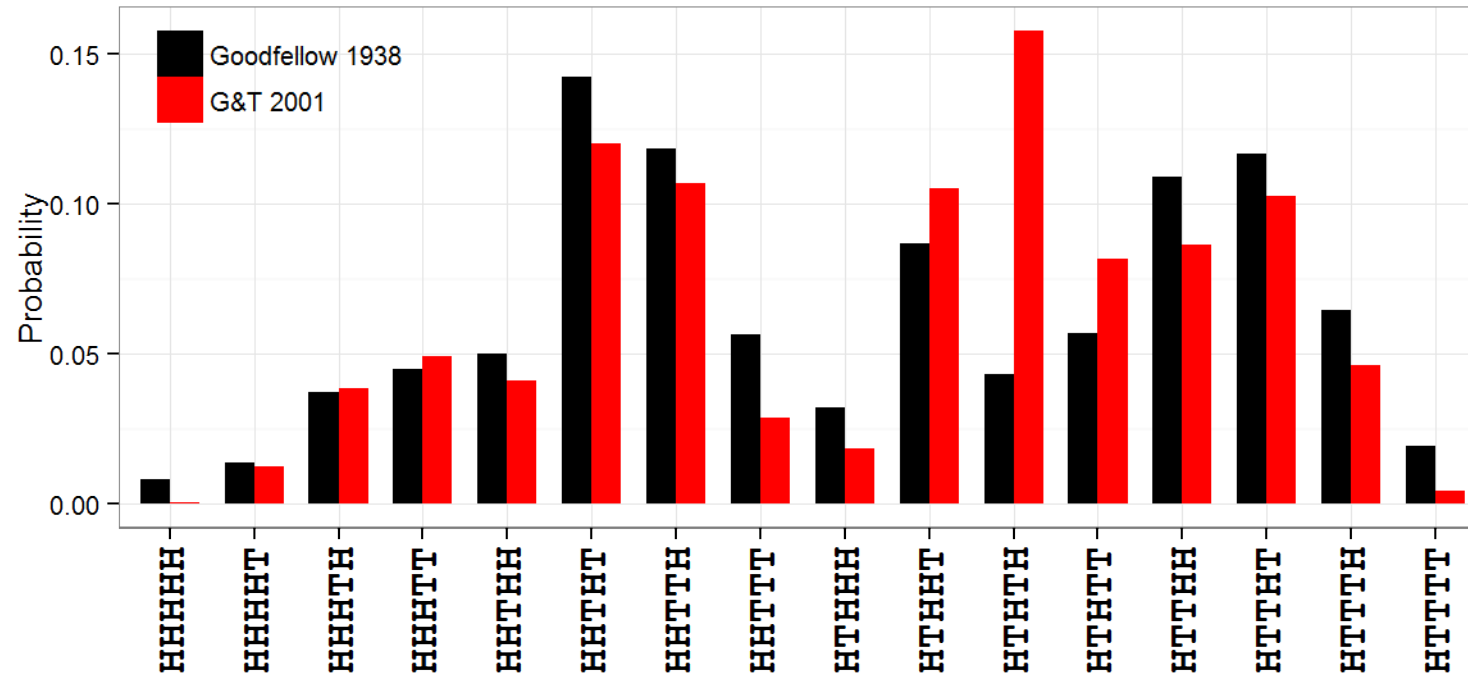
Griffiths & Tenenbaum (2001) propose a Bayesian model to account for the representativeness bias. Except apparent discrepancies at some data points (e.g., HTHTH), the model provides a fairly good fit to human data (Goodfellow, 1938, "Zenith radio data", 20,099 participants' preference to random sequences).

Modeling Random Sequence Production: G&T 2001

Subjective randomness:

$$\text{random}(X) = \log\frac{P(X|\text{random})}{P(X|\text{regular})} = \log\binom{H+T}{H} + f(H+T)$$

Local representativeness:

$$L_k = \sum_{i=1}^{k-1} \text{random}(H_i + 1, T_i) - \text{random}(H_i, T_i + 1) = \log\prod_{i=1}^{k-1}\left(\frac{T_i + 1}{H_i + 1}\right)$$

Probability of responses:

$$P(R_k = \text{H}) = \frac{1}{1 + e^{-\lambda L}}$$

Built-in bias:

$$p_A' = \frac{1}{1 + 2^{-\lambda}} \qquad p_A' > 0.5 \text{ for all } \lambda > 0$$

- The model by Griffiths & Tenenbaum (2001) has a built-in bias towards alternating patterns (HT or TH), and a free scaling parameter λ.

- The model shows consistency between the pattern preferences from Goodfellow (1938) and the bias $p_A' > 0.5$ (Falk & Konold, 1997), but did not explain the origin of the bias.

Fitting Human Data: Neural + Bayesian

$$p_A' = \frac{1}{1 + R/A} \approx 0.59$$

- $p_A' \approx 0.59$ naturally emerges from the neural model.

$$\lambda = -\log_2(R/A) \approx 0.515$$

- The naturally emerged p_A' imposes a constraint on the scaling parameter λ.

$$M_k = \log\left(\frac{O_\mathrm{T} + 1}{O_\mathrm{H} + 1}\right)$$

- $p_A' \approx 0.59$ means that alternation outweighs repetition. We can apply the same tendency to the second-order pattern events $(O_\mathrm{H}, O_\mathrm{T})$ and define another log-likelihood function M, to avoid repeating the alternation event itself.

$$P(R_k = \mathrm{H}) = \frac{1}{1 + e^{-\lambda(L+M)}}$$

- The same scaling parameter λ is applied to both log-likelihood functions.

Fitting Human Data: Neural + Bayesian
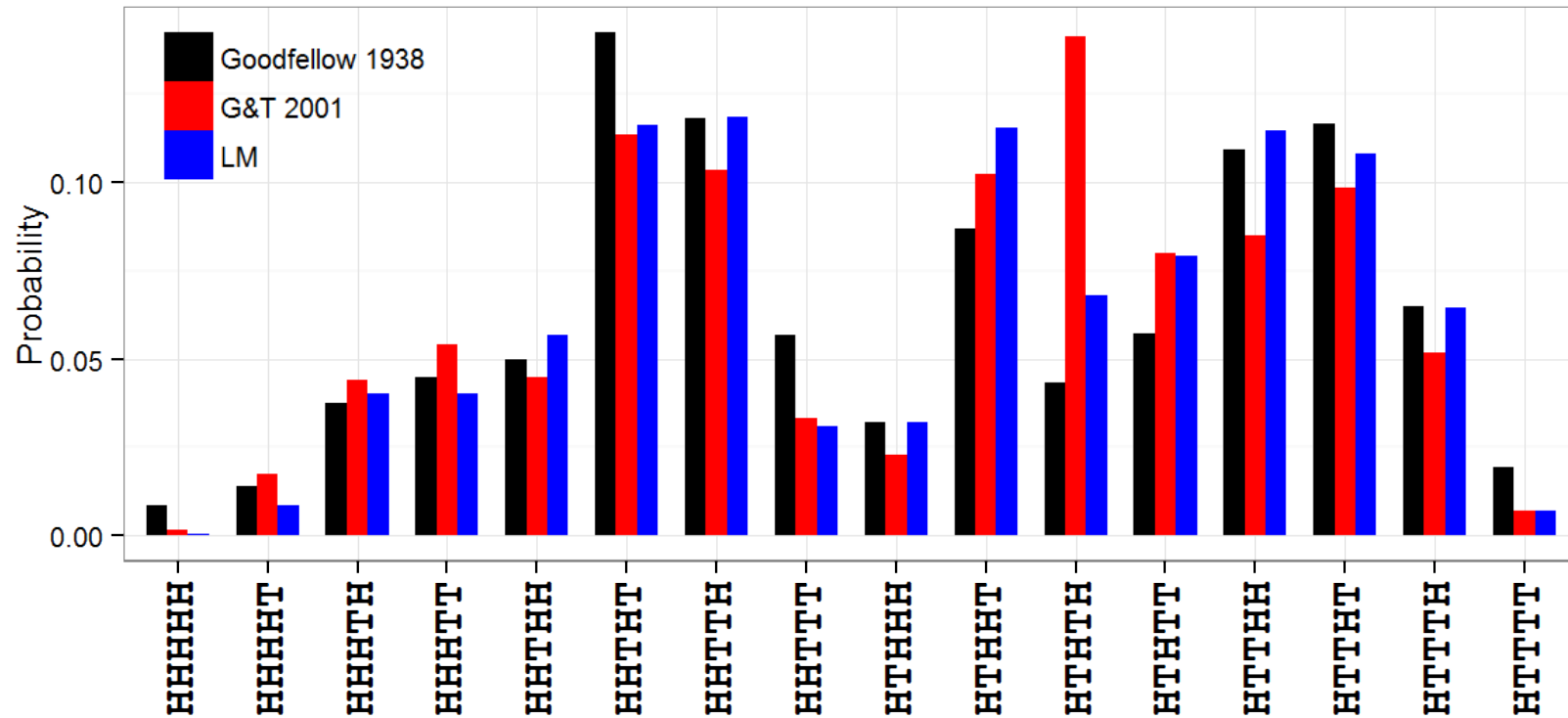
$\lambda = .516$, LM ($r = .929, R^2 = .863$), G&T 2001 ($r = .736, R^2 = .542$)
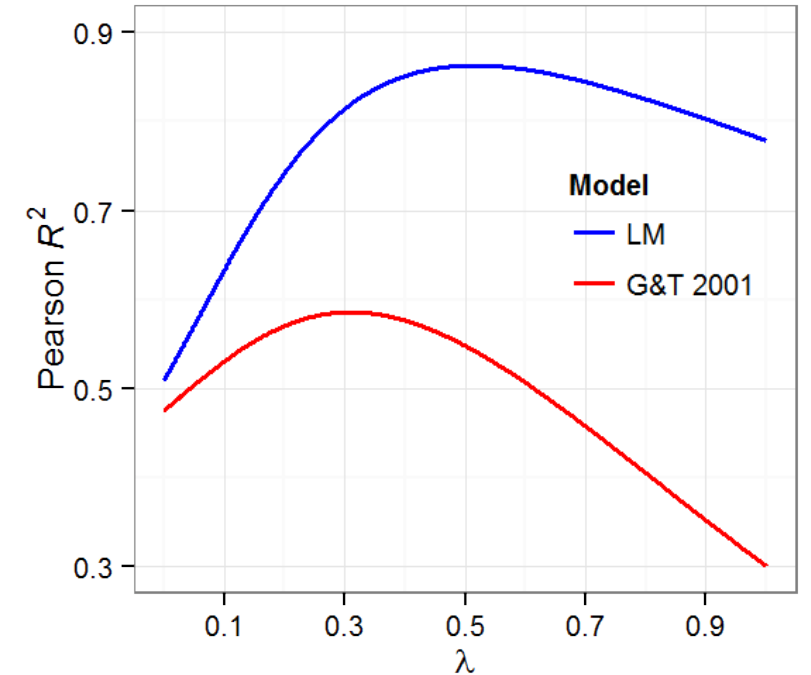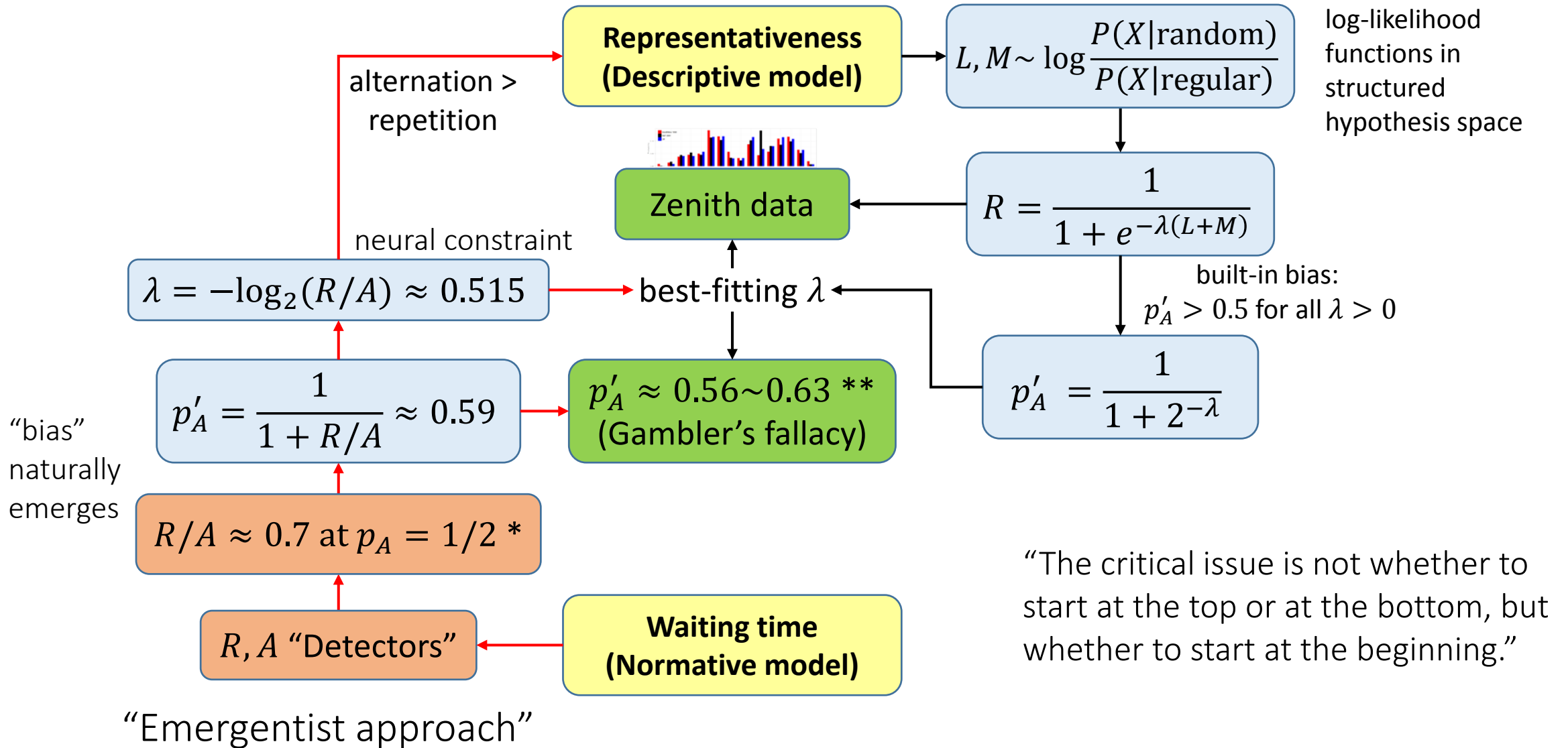
Best-fitting λ for the new model is indeed around 0.515.

Bridging the Gaps:

- between neurons (low-level neural circuits and constraints) and behavior (high-level cognition);
- between descriptive models (intuition and biases) and normative models (probability theory);
- between the "dynamical emergentist approach" and the "structured probability approach".

Bridging the Gaps (red arrows)

"Structured probability approach"

Representativeness (Descriptive model)

$L, M \sim \log \dfrac{P(X|\text{random})}{P(X|\text{regular})}$

log-likelihood functions in structured hypothesis space

alternation > repetition

Zenith data

$R = \dfrac{1}{1 + e^{-\lambda(L+M)}}$

neural constraint

$\lambda = -\log_2(R/A) \approx 0.515$

best-fitting λ

built-in bias: $p_A' > 0.5$ for all $\lambda > 0$

$p_A' = \dfrac{1}{1 + R/A} \approx 0.59$

$p_A' \approx 0.56 \sim 0.63$ ** (Gambler's fallacy)

$p_A' = \dfrac{1}{1 + 2^{-\lambda}}$

"bias" naturally emerges

$R/A \approx 0.7$ at $p_A = 1/2$ *

"The critical issue is not whether to start at the top or at the bottom, but whether to start at the beginning."

R, A "Detectors"

Waiting time (Normative model)

"Emergentist approach"

*: Neural model robustly produced $R/A \approx 0.6 \sim 0.8$ at $p_A = 1/2$.
**: Empirical p_A' in randomness generation task (Falk & Konold, 1997, Table 1).

Manuscript Abstract (draft)

Perceiving patterns from randomness: Emergence of structures and cognitive biases in time

The concept of randomness is a cornerstone in probability theory and foundational to human mathematical learning, but people's intuition of randomness and systematic biases remain controversial. Here, we provide a novel account for the computational bases of subjective randomness through a combination of biologically-based neural network and structured Bayesian inference models. We show that by simply observing random sequences over time, neurons in the neural model naturally differentiated according to the waiting time of local patterns, giving rise to the representativeness bias and the gambler's fallacy. Applying emergent neural properties to the Bayesian model, we obtain an accurate fit to human data in random sequence production. Our models establish that neurons' ability to integrate information over time is critical in bridging the gap between low-level sensory processing and high-level structured probability inference. In addition, we demonstrate that in order to reconcile people's intuition of randomness with probability theory, the statistical structures in the learning environment must be taken into consideration.