Testing reinforcement learning models for the Iowa Gambling Task

THE OHIO STATE UNIVERSITY

haines.175@osu.edu Nathaniel Haines[1], Woo-Young Ahn[1] ahn.280@osu.edu

[1]Department of Psychology, The Ohio State University

The Iowa Gambling Task

- The Iowa Gambling Task (IGT) is widely used to study decision making in both healthy and clinical populations. The IGT involves a complex interplay between multiple decision making processes but cognitive models may be used to better understand the multiple facets of choice behavior.
- Past modeling work with the IGT has focused on many types of models including: heuristic models [1][2], Bayesian updating models [1], expectancy-valence models [1][3], and hybrid models [4]. None of these models, however, show good performance for both short (e.g. one-step-ahead) and long-term (e.g. simulation) prediction accuracy [3; 5]. This suggests that we are still in need of a better model to explain choice behavior on the IGT.
- Our goal was to develop a new model for the IGT that shows excellent performance for both short and long-term prediction. We developed a new model using post-hoc fits, simulation, and parameter recovery. The new model addresses concerns made by previous studies by explicitly modeling frequency of outcomes in combination with expected value and a perseverance strategy.
- To accurately assess the new model's performance relative to existing models, we compared it with: (1) the Prospect-Valence Learning model with Delta rule (PVL Delta), and (2) the Value-Plus-Perseverance model (VPP). We chose these models because a previous study [6] showed that the PVL Delta performs excellent for simulation and parameter recovery while the VPP performs excellent for post-hic fit.

Structure of the Iowa Gambling Task

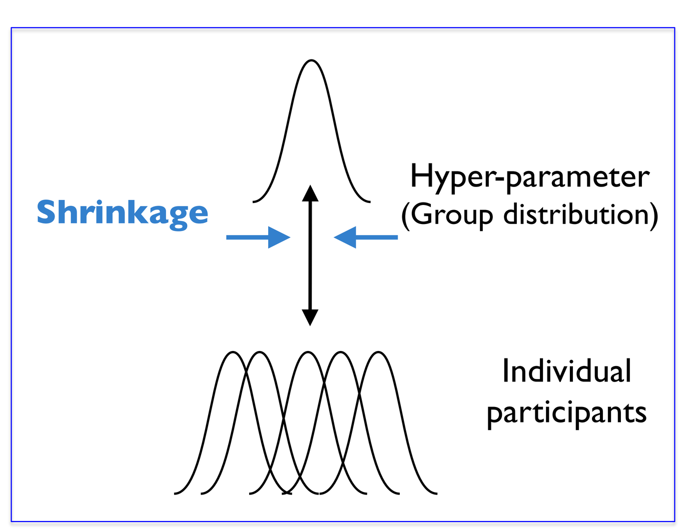
Modified Iowa Gambling Task (IGT) Bechara et al. (2001)

	"Bad"		"Good"	
Deck:	**A**	**B**	**C**	**D**
Average Gain:	100	100	50	50
Likelihood of loss:	50%	10%	50%	10%
Average Loss:	-150 to -350	-1250	-25 to -75	-250
Expected value:	**-250**	**-250**	**250**	**250**

- Instruction: Select cards from decks in a way that would maximize earnings.
- Each selection yields a gain of some amount, but some selections also involve a loss.
- Two decks (C and D) are "good" or "advantageous" and two other decks (A and B) are "bad" or "disadvantageous". Participants are not told which decks are "good" and which are "bad", and that they should learn this by trial-and-error as the task progresses.

Methods

- Data from N=48 healthy controls from Ahn et al., 2014 were used to test all models.
- We used Hierarchical Bayesian Analysis (HBA) and the hBayesDM R package (development version) [8] for parameter estimation. The hBayesDM package uses the Stan software (mc-stan.org) for Markov-Chain Monte-Carlo sampling.



- Three model comparison methods were used to test model performance:
  1. Post-hoc model fit indices (Watanabe-Akaike information criterion (WAIC) and Leave One Out information criterion (LOOIC))
  2. Simulation performance (refer to [6])
  3. Parameter recovery performance (refer to [6])

Models

➢ PVL Delta and VPP

Utility function:
$$u(t) = \begin{cases} x(t)^{\alpha}, & \text{if } x(t) \geq 0 \\ -\lambda |x(t)|^{\alpha}, & \text{if } x(t) < 0 \end{cases}$$

Value Updating:
$$E_j(t+1) = E_j(t) + A \cdot (u(t) - E(t))$$

Perseverance: (VPP only)
$$P_j(t+1) = \begin{cases} k \cdot P_j(t) + \epsilon_p, & \text{if } x(t) \geq 0 \\ k \cdot P_j(t) + \epsilon_n, & \text{if } x(t) < 0 \end{cases}$$

Total Value: (VPP only)
$$V_j(t+1) = \omega \cdot E_j(t+1) + (1-\omega) \cdot P_j(t+1)$$

Choice Sensitivity:
$$\theta = 3^c - 1$$

Action Selection: Softmax with θ (inverse temperature) and $E_j(t+1)$ for PVL, or $V_j(t+1)$ for VPP.

* Both models are available in the hBayesDM R package

➢ Proposed Model

Value Updating:
$$E_j(t+1) = \begin{cases} E_j(t) + A_{rew} \cdot (x(t) - E_j(t)), & \text{if } x(t) > 0 \\ E_j(t) + A_{pun} \cdot (x(t) - E_j(t)), & \text{if } x(t) \leq 0 \end{cases}$$

Frequency Updating:
$$F_j(t+1) = \begin{cases} F_j(t) + A_{rew} \cdot (\text{sign}(x(t)) - F_j(t)), & \text{if } x(t) > 0 \\ F_j(t) + A_{pun} \cdot (\text{sign}(x(t)) - F_j(t)), & \text{if } x(t) \leq 0 \end{cases}$$

Fictive Frequency Updating:
$$F_{j'}(t+1) = \begin{cases} F_{j'}(t) + A_{pun} \cdot (-\text{sign}(x(t)) - F_{j'}(t)), & \text{if } x(t) > 0 \\ F_{j'}(t) + A_{rew} \cdot (-\text{sign}(x(t)) - F_{j'}(t)), & \text{if } x(t) \leq 0 \end{cases}$$

Perseverance:
$$P_j(t+1) = 1 \quad \text{, and}$$
$$P_{j'}(t+1) = k \cdot P_{j'}(t)$$

Total Value:
$$V_j(t+1) = E_j(t+1) + \beta_F \cdot F_j(t+1) + \beta_P \cdot P_j(t+1)$$

Action Selection: Softmax with $V_j(t+1)$

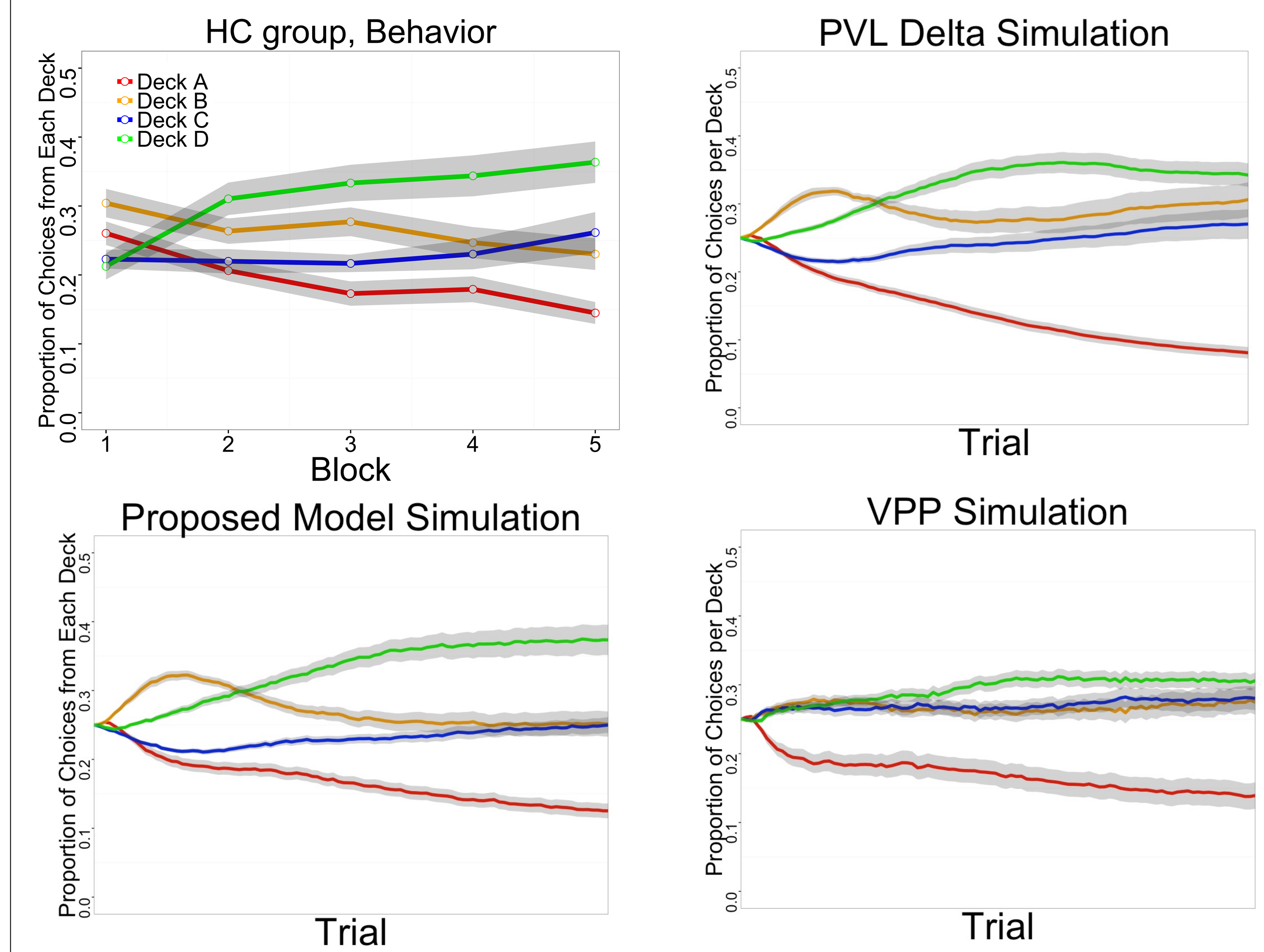
* Will be made available in the hBayesDM R package

Results: WAIC & LOOIC

	WAIC	LOOIC	Number of Parameters
PVL Delta	12319	12357	4
VPP	11543	**11581**	8
Proposed Model	**11532**	11604	5

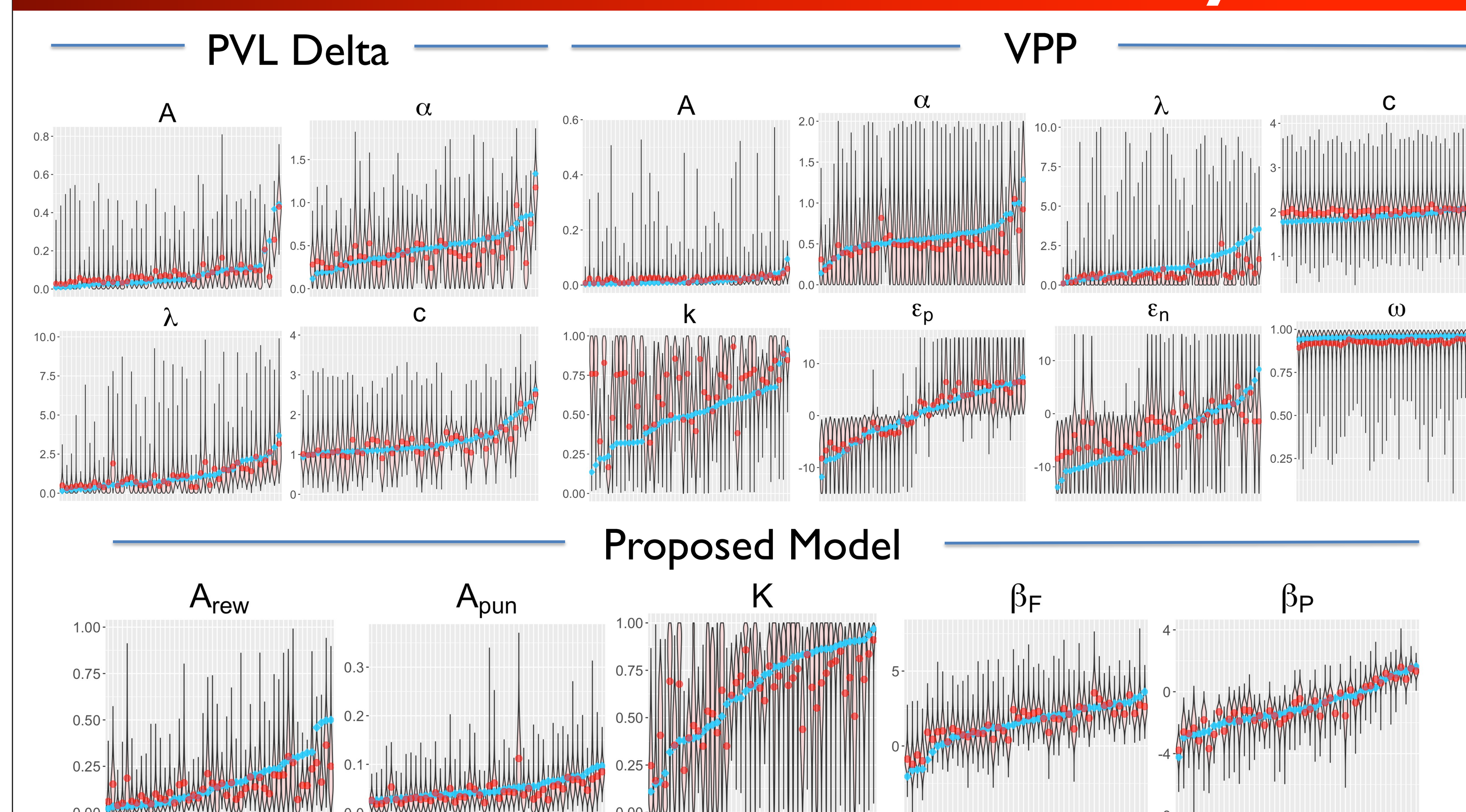
- Lower WAIC and LOOIC indices indicate better fits.
- The Proposed model and the VPP perform similarly in terms of post-hoc fits (one-step-ahead prediction accuracy).

Results: Simulation



- The proposed model showed the best simulation performance

Results: Parameter Recovery



- Blue diamonds: true parameters used to simulate choice behavior. Red circles: Recovered posterior means. The full densities of the recovered parameters are shaded red.
- The proposed model shows comparable parameter recovery to the PVL Delta. Both PVL Delta and the proposed model perform better than the VPP.

Conclusions

Our study investigated a novel reinforcement learning model for the Iowa Gambling Task (IGT) that measures expected value, expected outcome frequency, and perseverance. We used post-hoc model fits, simulation, and parameter recovery to assess how well the new model explains data collected from healthy subjects. Our results show that the new model performs very well in all three indices. These findings suggest that long-run average, outcome frequency, and perseverance all play important roles in choice behavior in the IGT. Future studies should use the new model to investigate differences between healthy and clinical populations of interest.

References

1. Busemeyer, J. R., & Stout, J. C. *Psychological assessment,* **14(3)**, 253 (2002).
2. Worthy, D.A., & Maddox, W.T. Journal of mathematical psychology, **59**, 41-49 (2014).
3. Ahn, W.Y., Busemeyer, J. R., & Wagenmakers, E. J. *Cognitive Science*, **32(8)**, 1376–1402 (2008).
4. Worthy, D.A., Pang, B., & Byrne, K.A. *Frontiers in psychology*, **4**, 640 (2013).
5. Steingroever, H., Wetzels, R., & Wagenmakers, E. J. *The Journal of Problem Solving,* **5(2)**, 2 (2013).
6. Ahn, W.-Y., Vasilev, G., Lee, S.-H., Busemeyer, J. R., Kruschke, J. K., Bechara, A., & Vassileva, J. *Frontiers in Psychology,* **5**, 1376 (2014).
7. Bechara, A., Dolan, S., Denburg, N., Hindes, A., Anderson, S.W., & Nathan, P. E. *Neuropsychologia*, **39,** 376-389 (2001).
8. Ahn, W.Y., Haines, N., & Zhang, L. *bioRxiv*, **064287** (2016).