

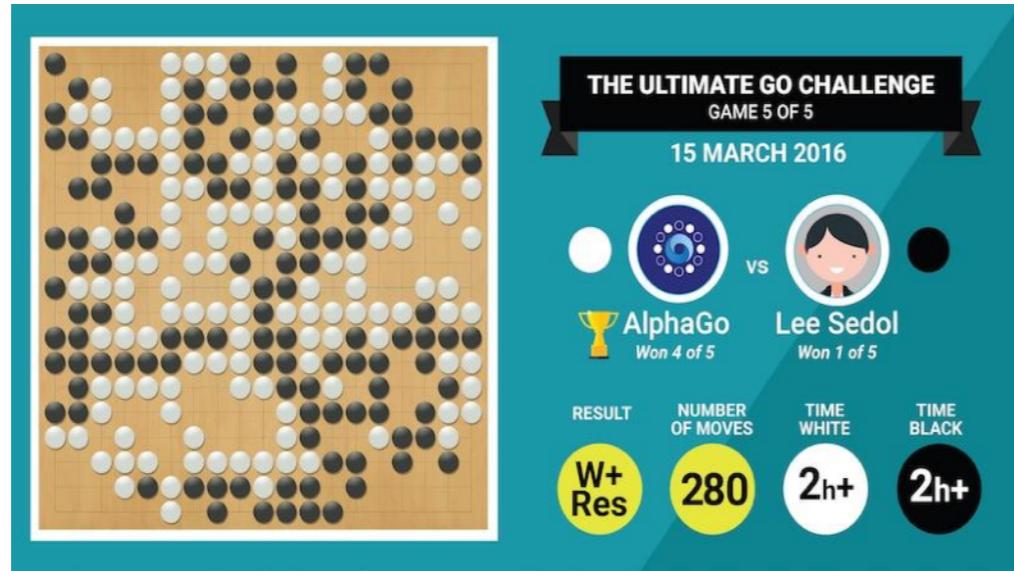
MODEL-BASED AND MODEL-FREE DECISION-MAKING IN A COMPLEX PLANNING TASK

Ionatan Kuperwajs

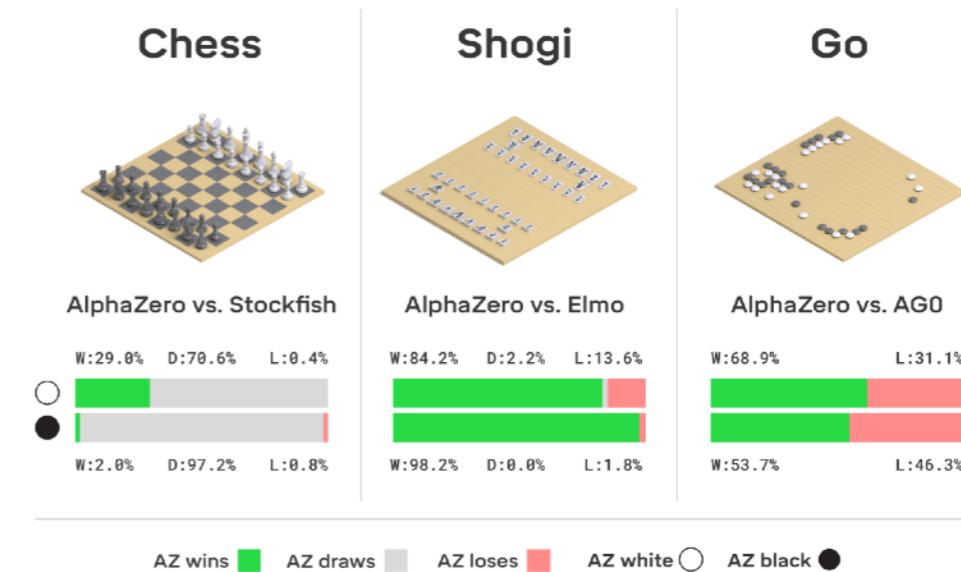
Ma Lab
Center for Neural Science
New York University



COMPLEX PLANNING IN GAMES



Silver et al., *Nature*, 2016



Silver et al., *Science*, 2017

Sequence of decisions

- 1st move: 361 options
- 2nd move: 360 options
- 3rd move: 359 options
- ⋮

Complex planning problem

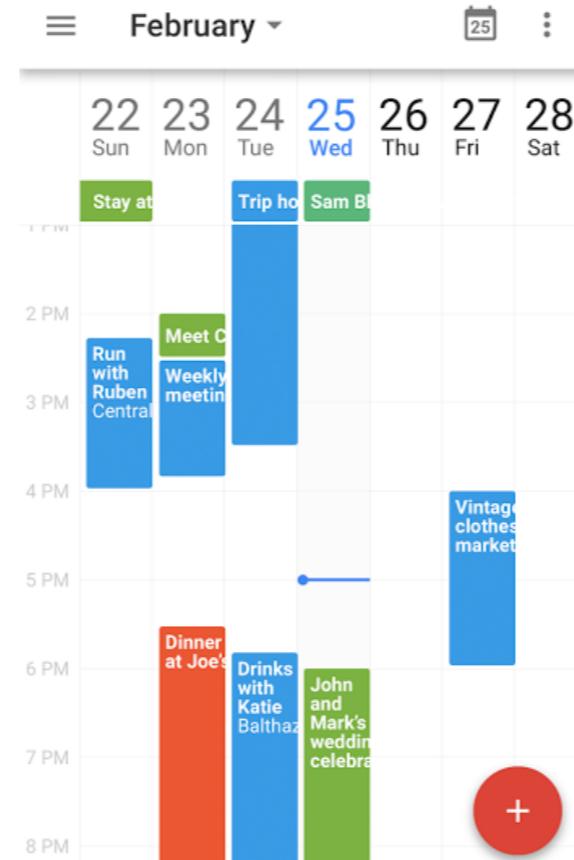


Mental simulation of possible futures

Are complex planning problems limited to games?



Foraging



Organizational
strategy



Spatial navigation

COMPLEX PLANNING

Complex planning problems:

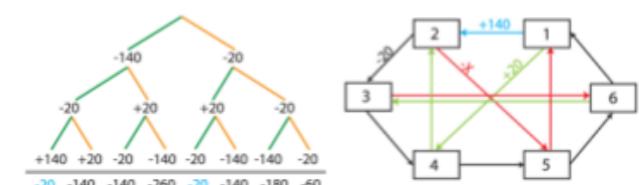
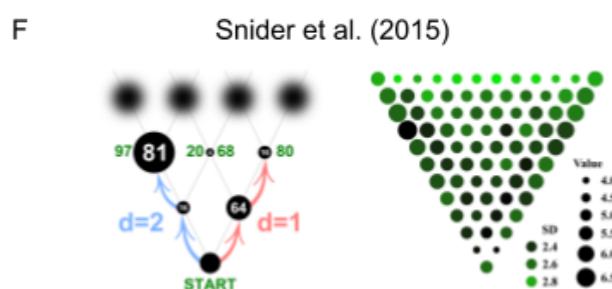
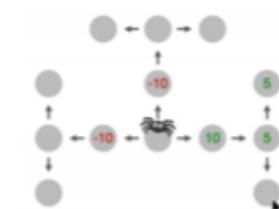
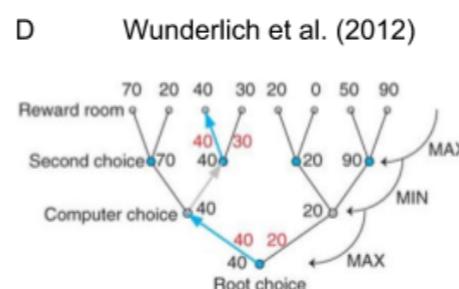
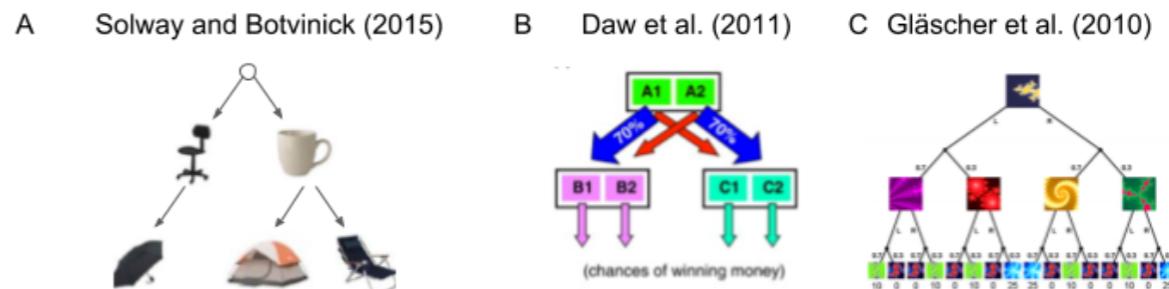
- Incredibly challenging for agents
 - Common in natural behavior

How do humans solve such problems?

Are these solutions similar to AI algorithms?

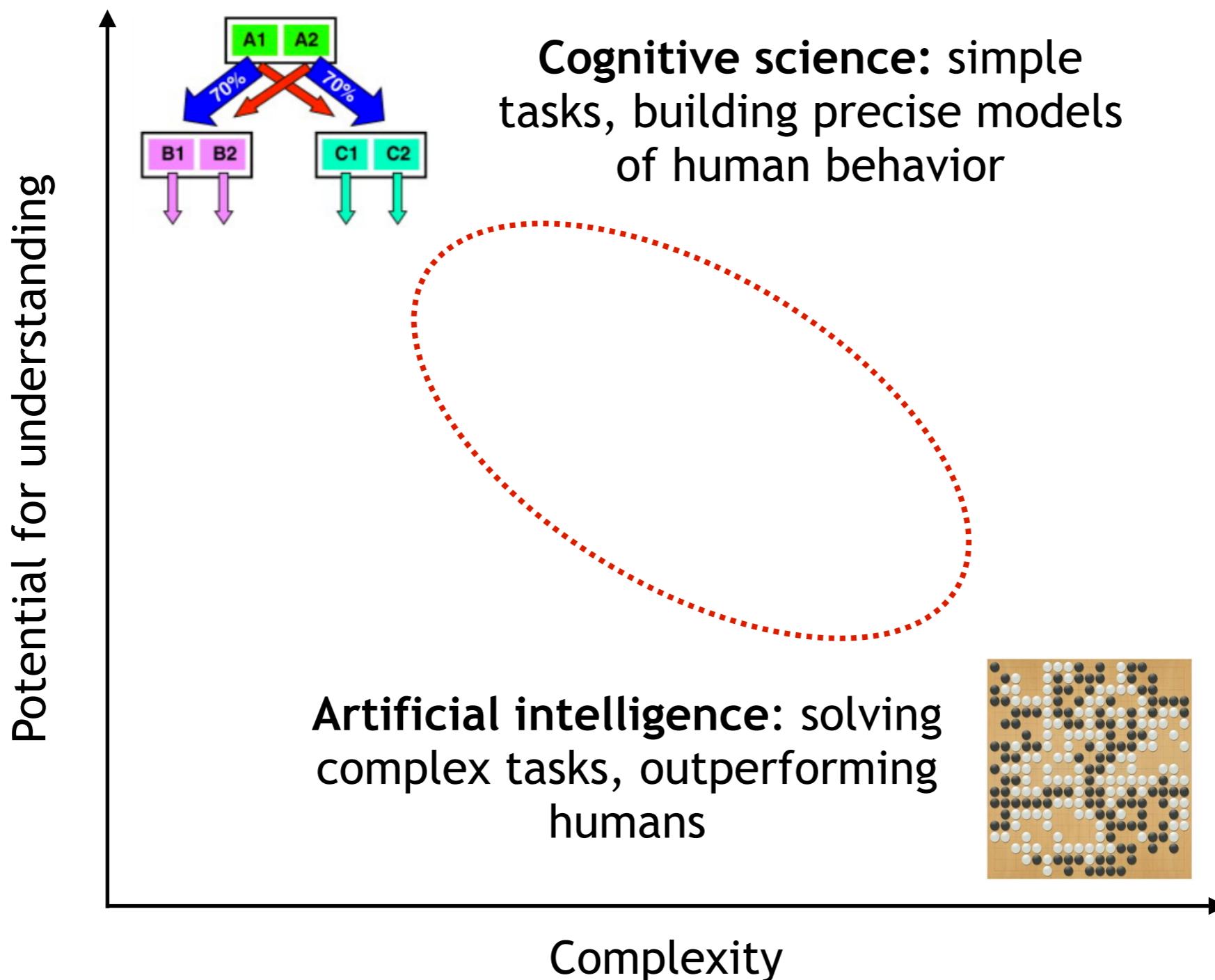
What tasks do we use to study these questions?

Tasks for studying human planning



However, these tasks are low in complexity and not fully combinatorial

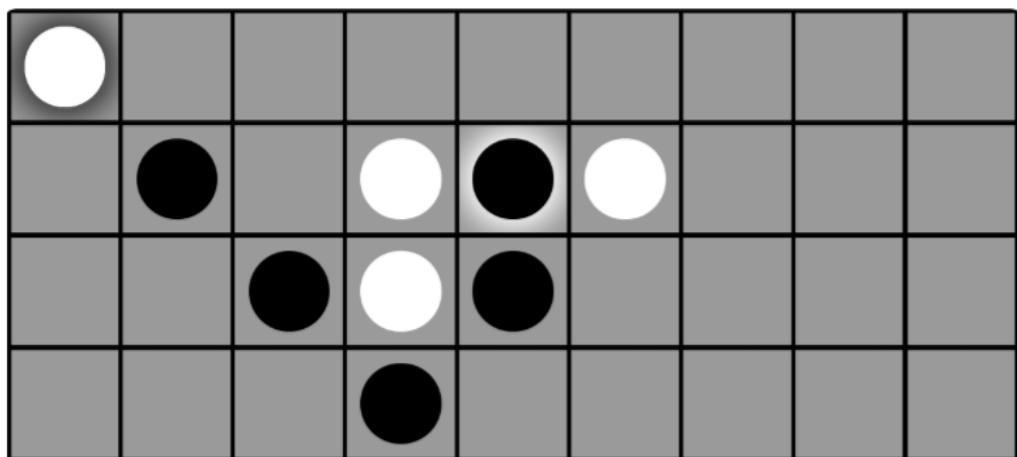
SPACE OF PLANNING TASKS



Needed: a combinatorial planning task of intermediate complexity with high potential for modeling human behavior

FOUR-IN-A-ROW

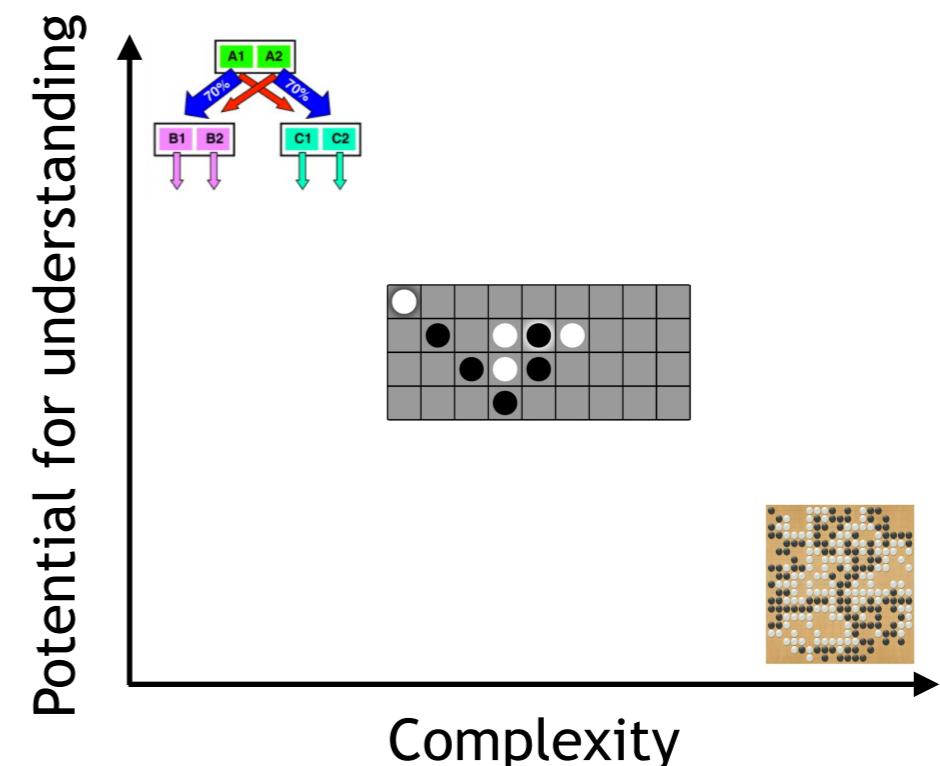
- 4-by-9 board
- 2 players, black plays first
- Alternate placing pieces on the board, trying to complete four in a row in any orientation



$1.2 \cdot 10^{16}$ possible states

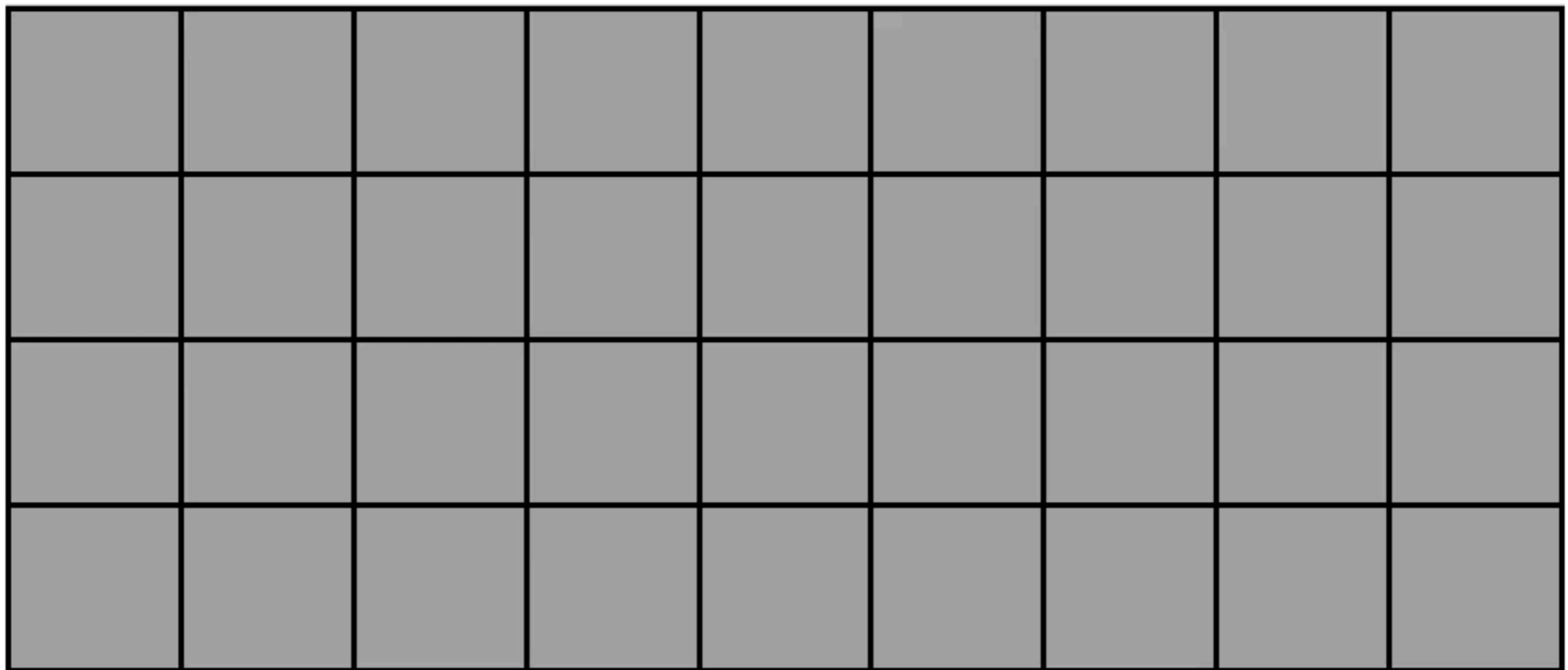
State space comparison

Task	State space complexity	Human cognition studies	AI studies
Solway and Botvinick task	3	[Solway and Botvinick, 2015]	
Daw et al. "two-step" task	3	[Daw et al., 2011]	
Gläscher et al. task	5	[Gläscher et al., 2010]	
Wunderlich et al. task	7	[Wunderlich et al., 2012a]	
Callaway et al. task	$\lesssim 20$	[Callaway et al., 2017]	
Snider et al. task	66	[Snider et al., 2015]	
Huys et al. task	2 to 128	[Huys et al., 2012] [Huys et al., 2015]	
Four-in-a-row	$1.2 \cdot 10^{16}$	[van Opheusden et al., 2017]	Beck, 2008
Chess	$\sim 10^{40}$	[de Groot, 1946] [Chase and Simon, 1973] [Holding, 1985]	[Shannon, 1950] [Campbell et al., 2002] [Silver et al., 2017]
Go	$2.1 \cdot 10^{170}$		[Silver et al., 2016]

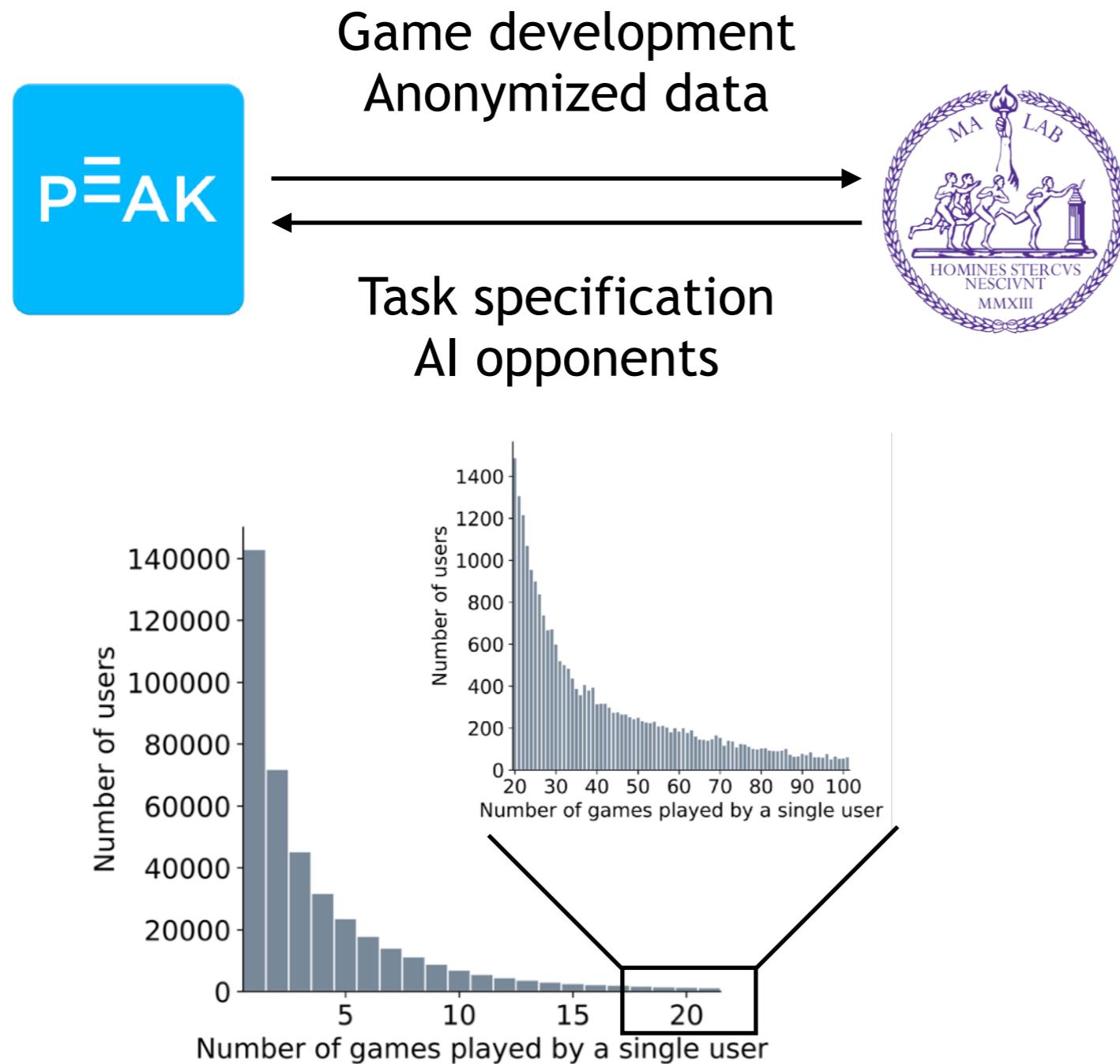


GAMEPLAY

A game between two novice players
Opheusden, Galbiati, Bnaya, Li, Ma, *CogSci*, 2017



LARGE-SCALE DATA



- Human vs. AI games
- User always plays first
- ~3.2 million games
- ~430,000 unique users
- ~1.5 million new games/month

OUTLINE

Given this task of intermediate complexity, we want to:

- Build an AI-inspired prospective model of human play
- Fit the model to data, and explore signatures of retrospective decision-making in the large data set
- Develop models for integrating prospection and retrospection (ongoing work)

OUTLINE

Given this task of intermediate complexity, we want to:

- **Build an AI-inspired prospective model of human play**
- Fit the model to data, and explore signatures of retrospective decision-making in the large data set
- Develop models for integrating prospection and retrospection (ongoing work)

THE AI OF GAMES



Philosophical Magazine, Ser.7, Vol. 41, No. 314 - March 1950.

XXII. Programming a Computer for Playing Chess¹

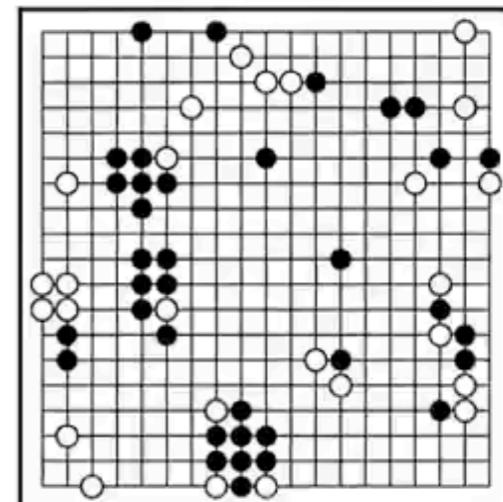
By CLAUDE E. SHANNON

Claude Shannon

Visualizing selective search
Silver et al., *Nature*, 2016

Two key ingredients:

- (1) **Evaluation function:** quantified value of how good a position is
- (2) **Selective search strategy:** look-ahead when a move seems promising according to the evaluation function



EVALUATION FUNCTION

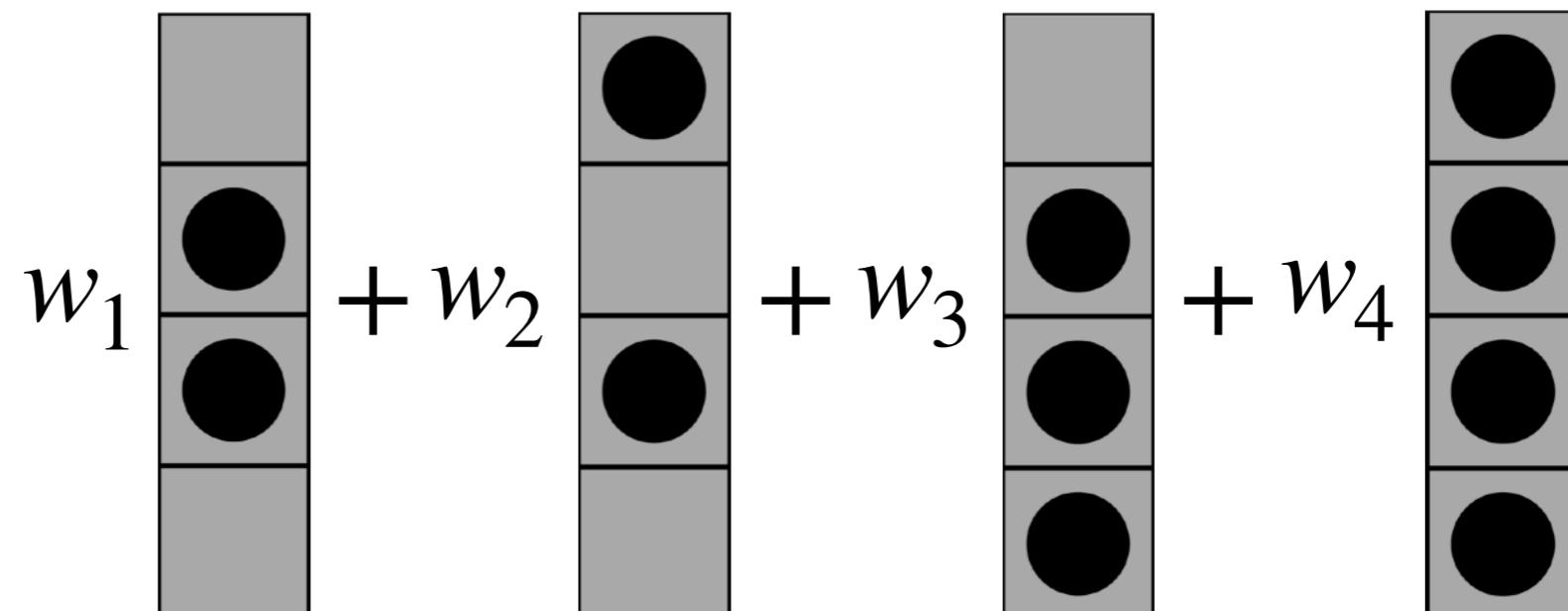
Key ingredient #1

Positional
features

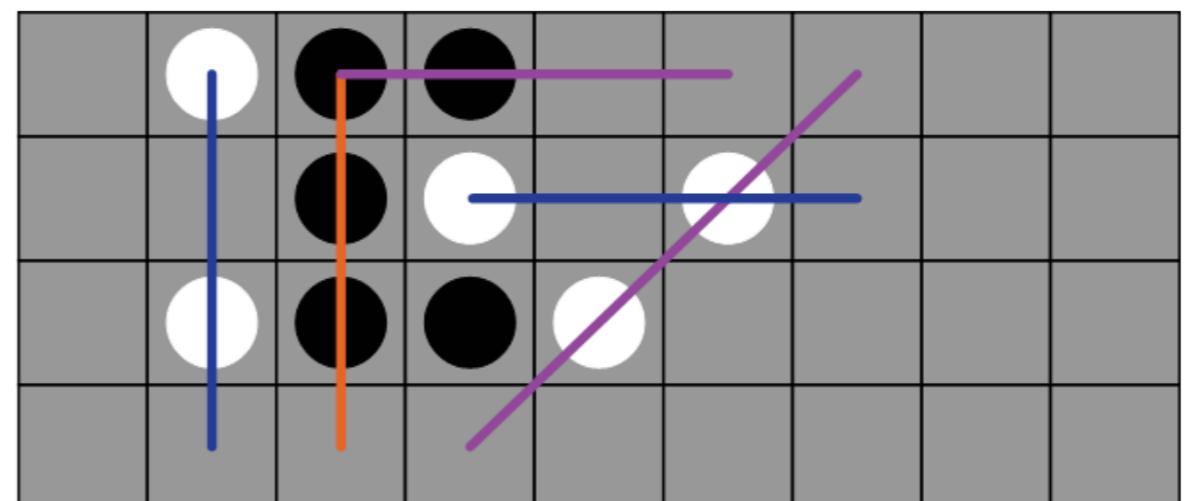
Learned
weights

Scaling
constant

Opponent

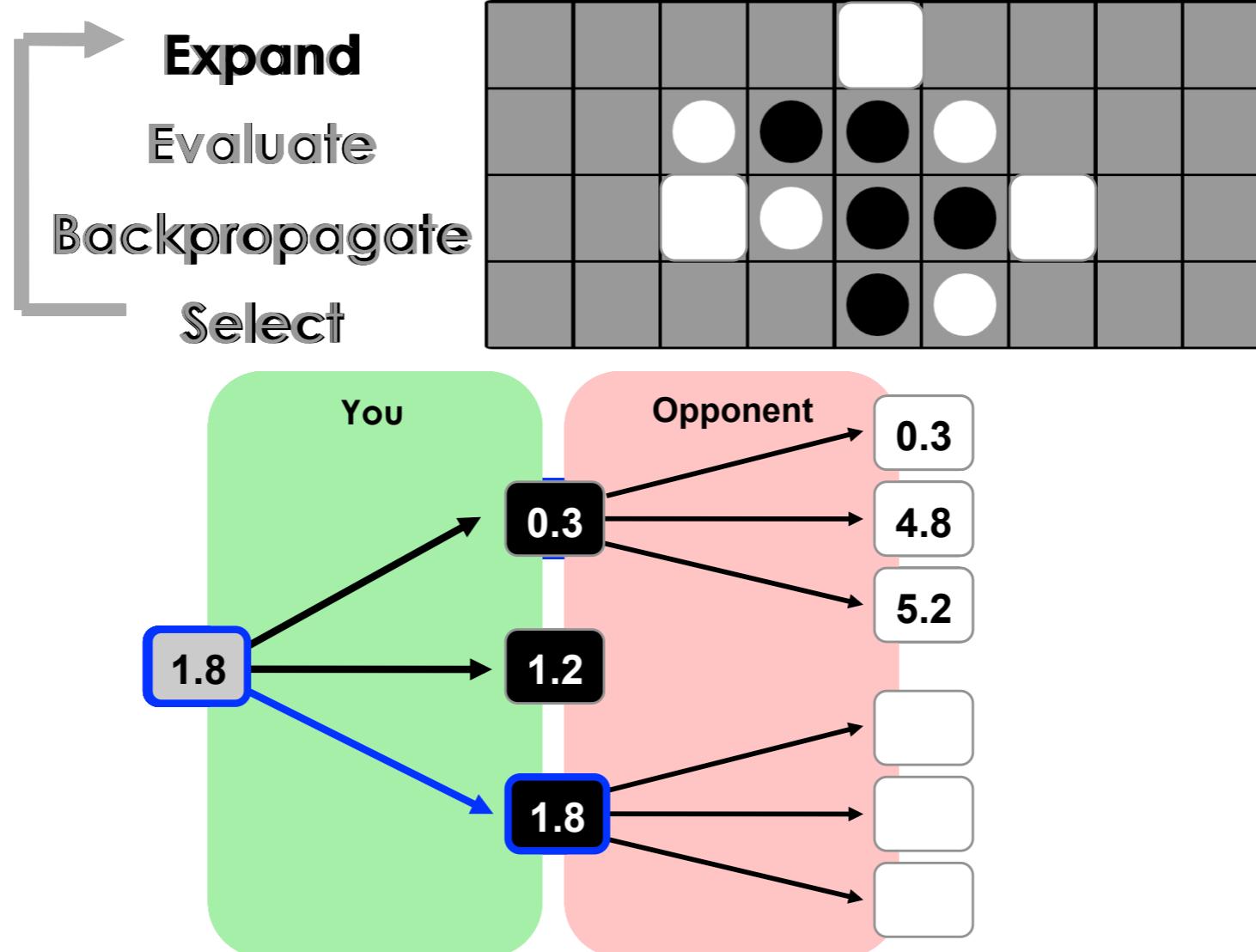


$$V(s) = c_{self} \sum_{i=0}^4 w_i f_i(s, self) - c_{opp} \sum_{i=0}^4 w_i f_i(s, opp)$$



DECISION TREE SEARCH

Key ingredient #2



OTHER COMPONENTS OF THE MODEL

Pruning

- Nodes with a value below that of the best move minus a threshold are removed from the tree (θ)

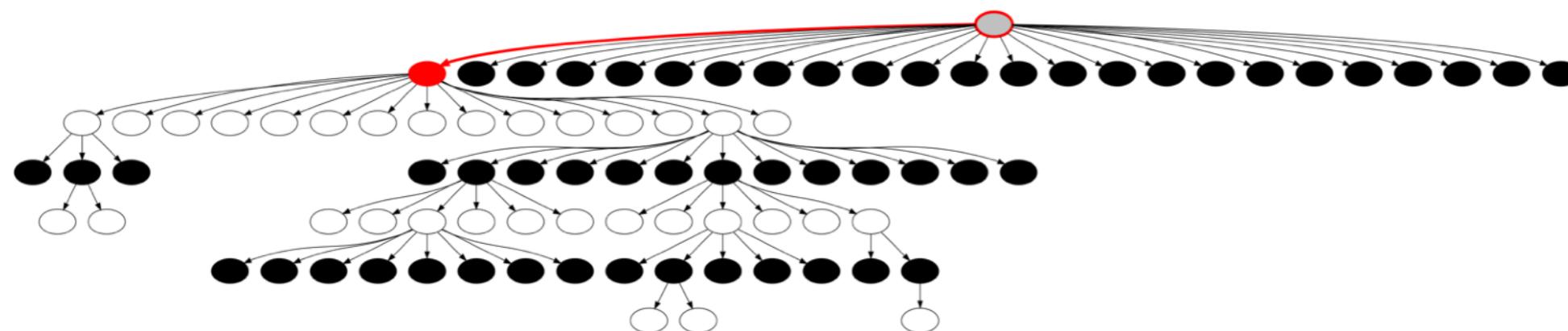
"Human-like" modes of failure

- Gaussian noise added to $V(s)$
- Feature dropping (δ): overlook features when calculating $V(s)$
- Lapse rate (λ): make a random move

Stopping criterion to terminate search

- Randomly with a small probability (γ)
- When the preferred move hasn't changed for a number of iterations
OR the difference between the best and second best moves is above a threshold (N)

A SINGLE MODEL SIMULATION



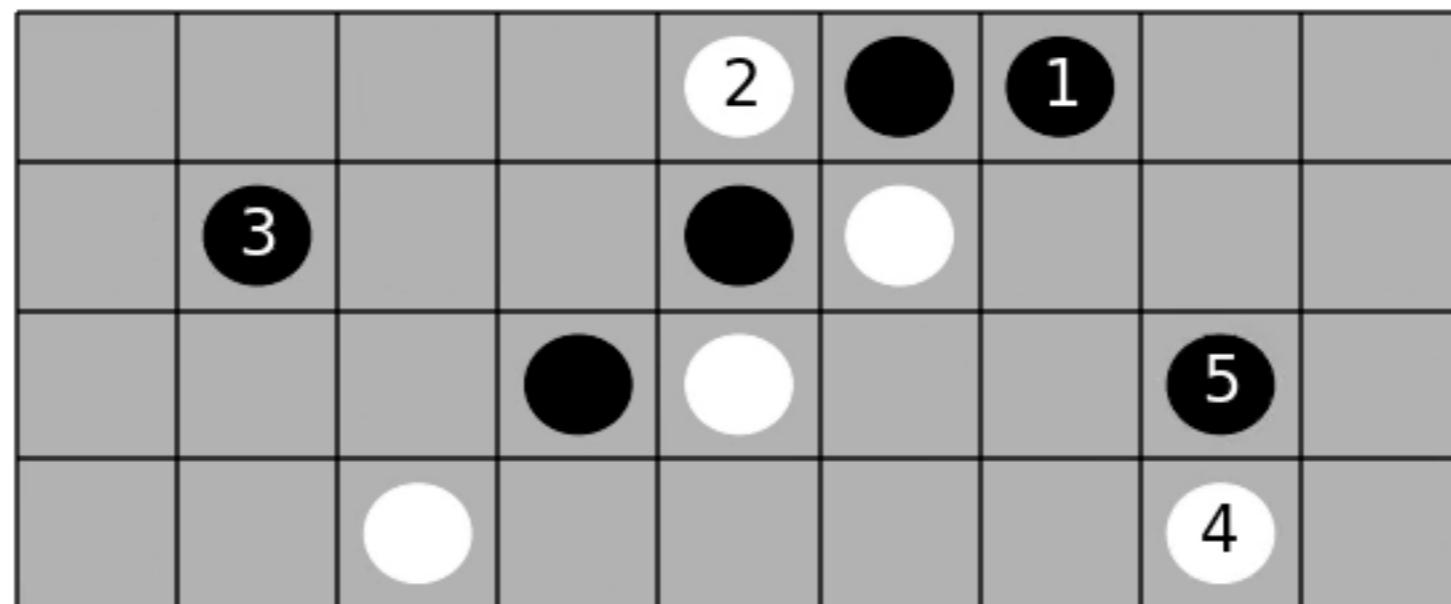
Principal variation 1

Principal variation 2

Principal variation 3

Principal variation 4

Decision



OUTLINE

Given this task of intermediate complexity, we want to:

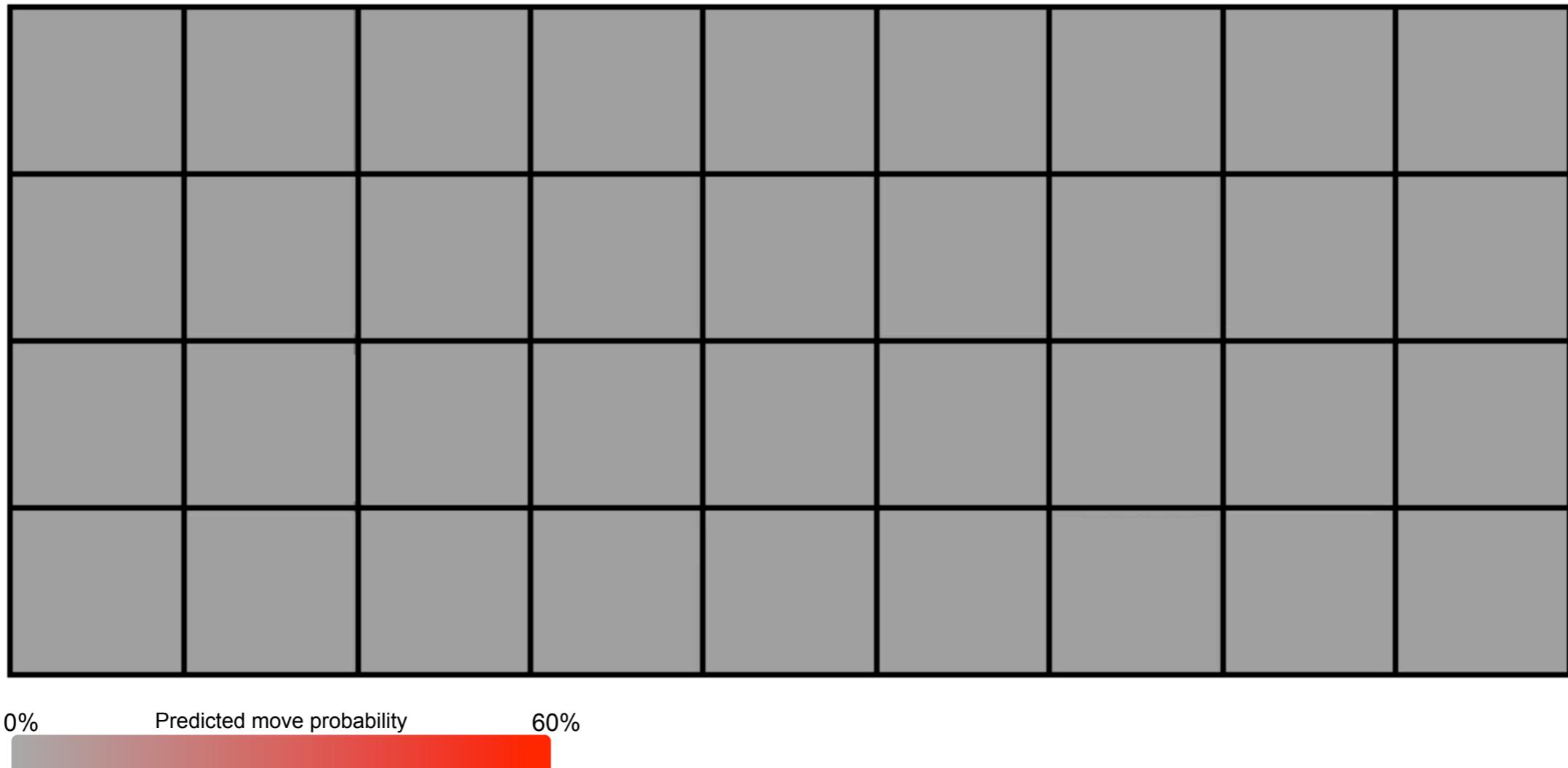
- Build an AI-inspired prospective model of human play
- Fit the model to data, and explore signatures of retrospective decision-making in the large data set
- Develop models for integrating prospection and retrospection (ongoing work)

OUTLINE

Given this task of intermediate complexity, we want to:

- Build an AI-inspired prospective model of human play
- Fit the model to data, and explore signatures of retrospective decision-making in the large data set
- Develop models for integrating prospection and retrospection (ongoing work)

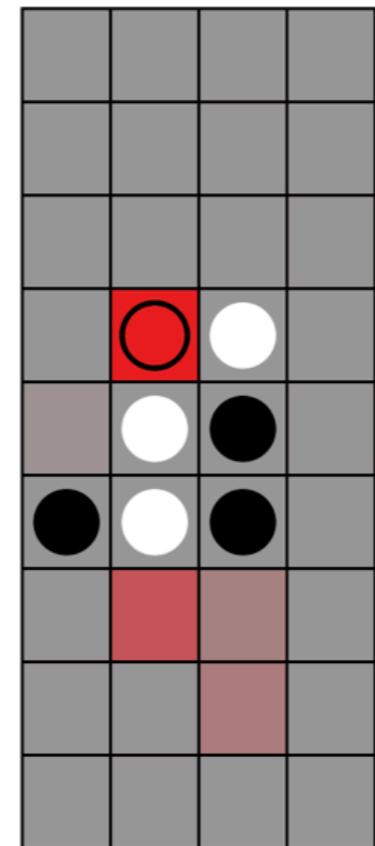
FITTING THE MODEL TO HUMAN PLAY



Model is simulated many times in each position, returning the probability that the subject moves on a given square

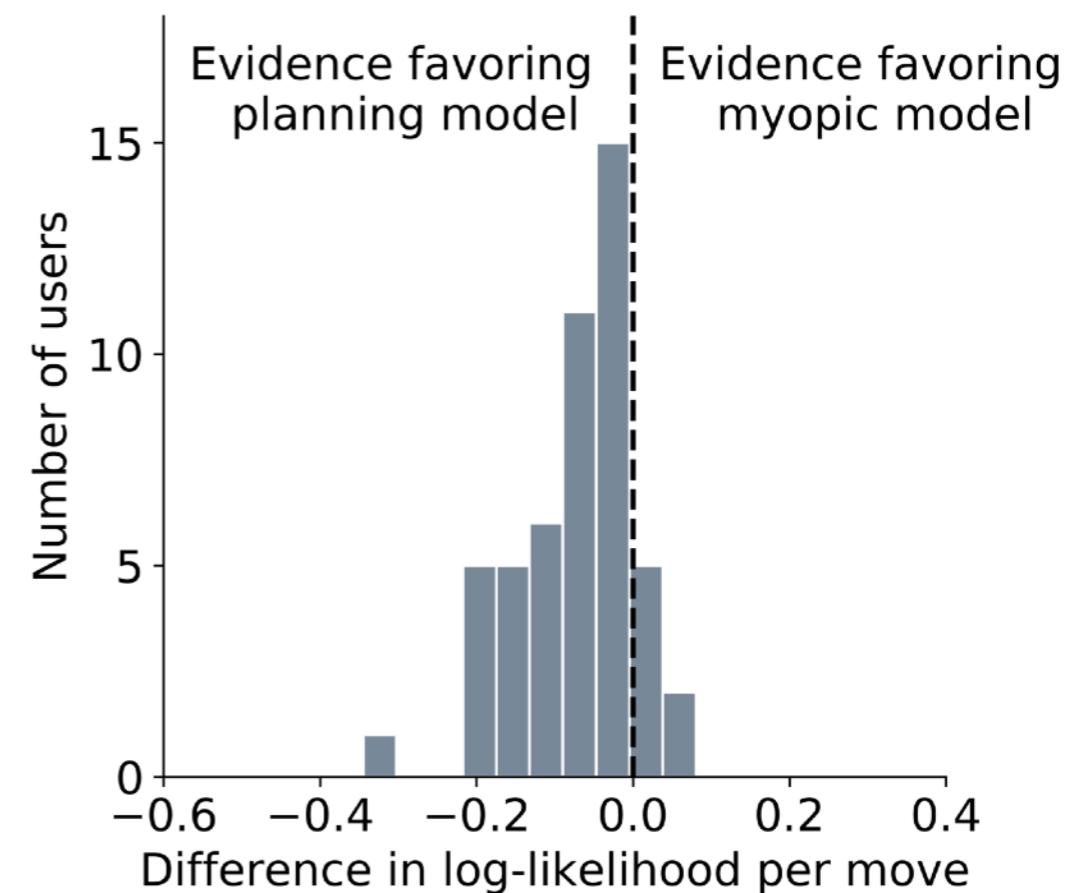
MODEL PERFORMANCE

50 pseudo-randomly selected users



Move predictions

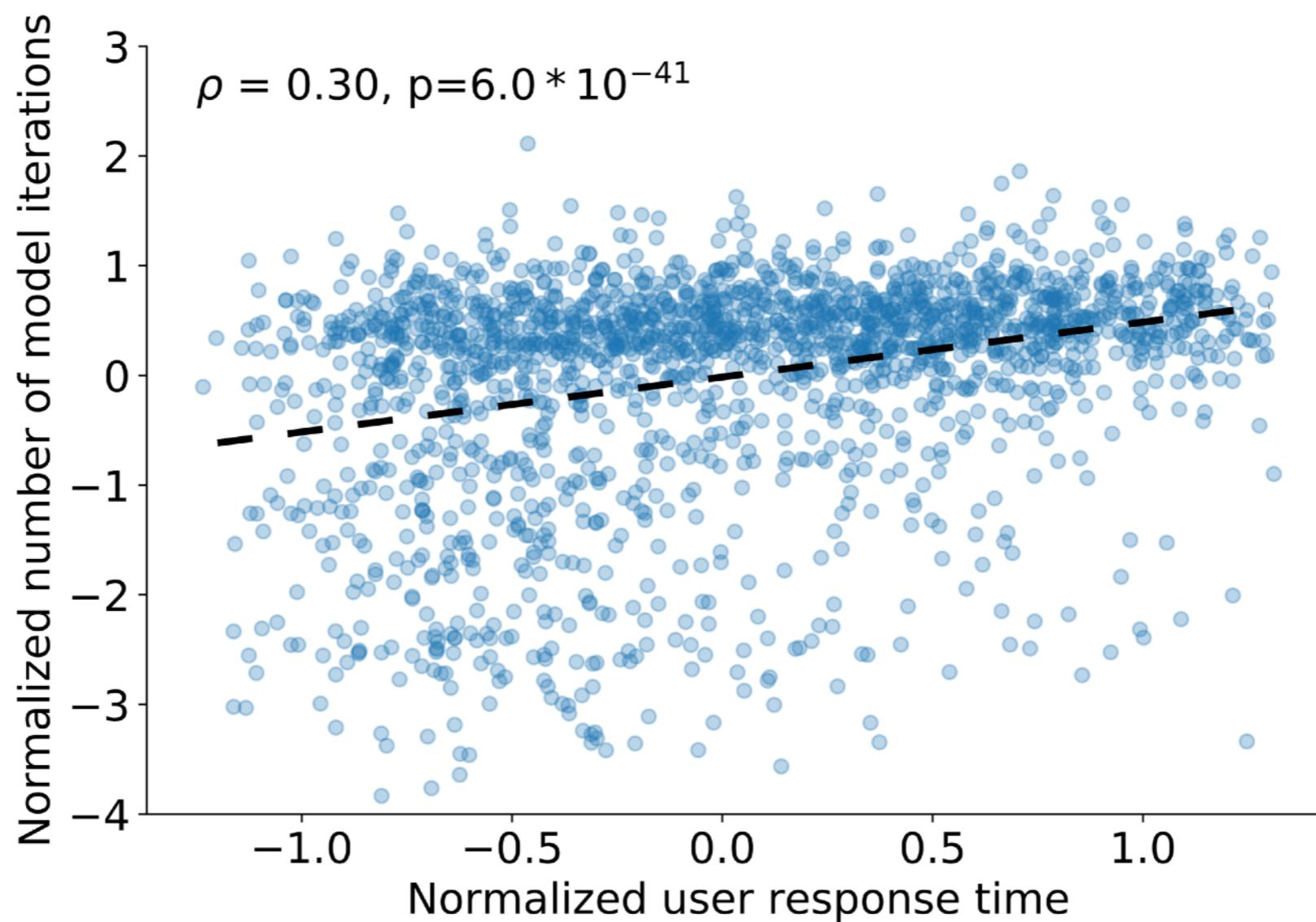
~25% correct
High compared to chance (4%)



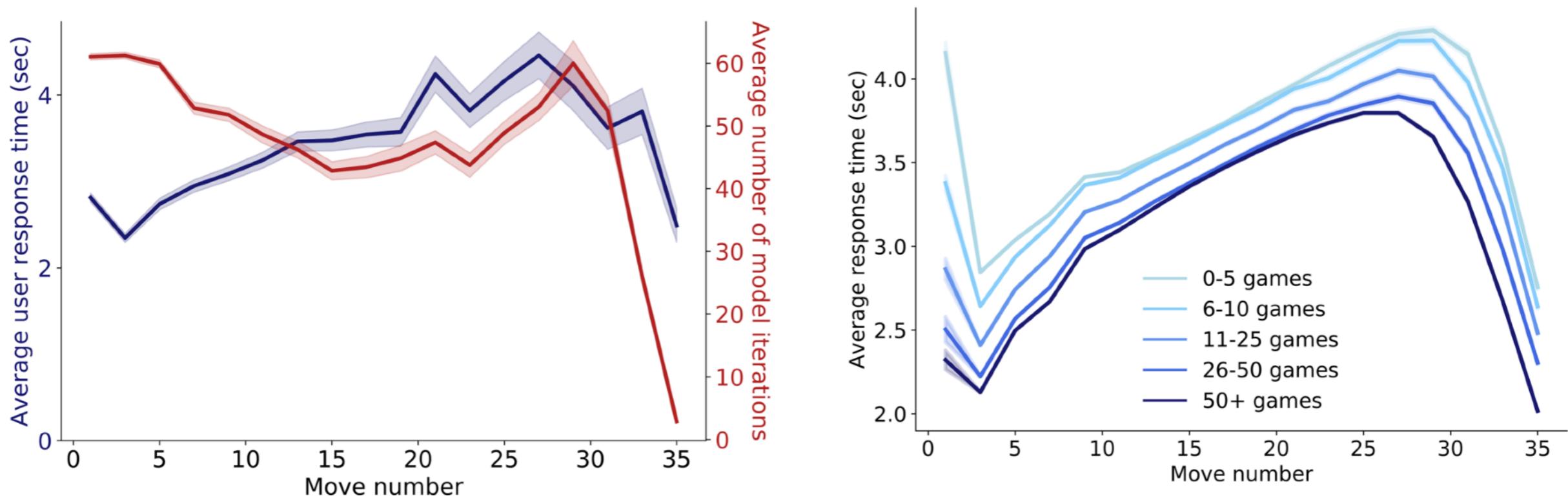
Model comparison

RESPONSE TIME PREDICTIONS

Prediction: when the model builds a larger tree, people's **response times increase**



RESPONSE TIME PREDICTIONS

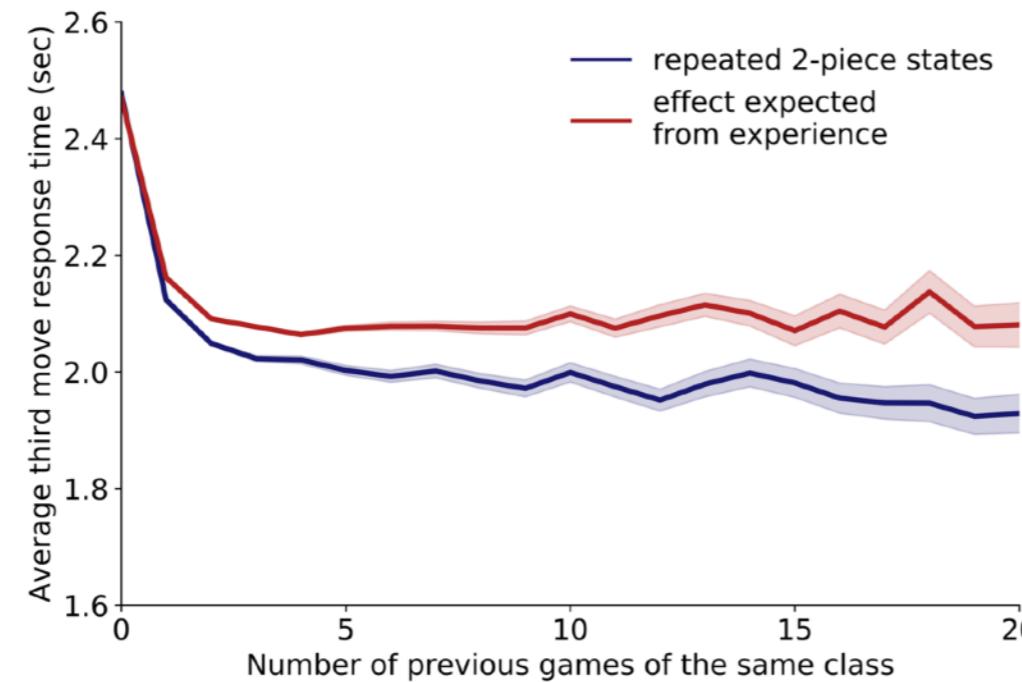
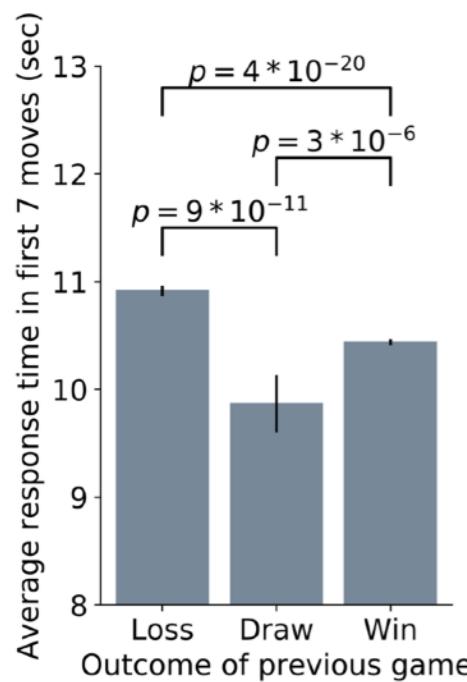


Model fails to predict human response times trends at the individual move level irrespective of experience level

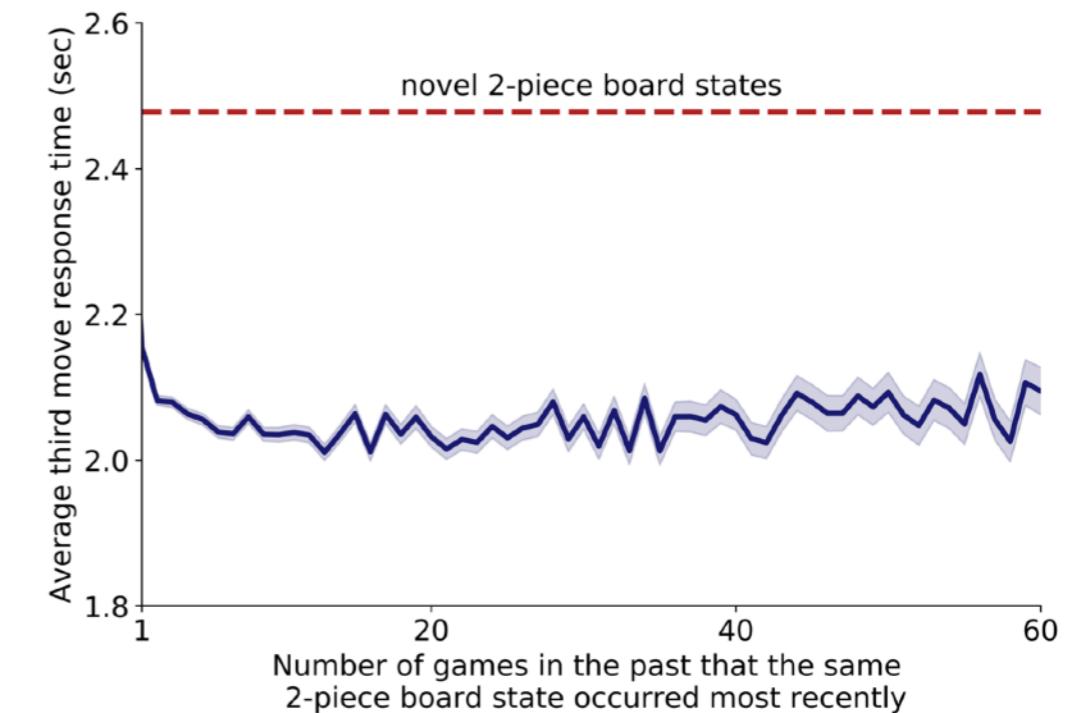
Possible explanation: **model-free RL or retrospective decision-making**

MODEL-FREE RESPONSE TIMES

RL prediction #1: response times increase after a loss



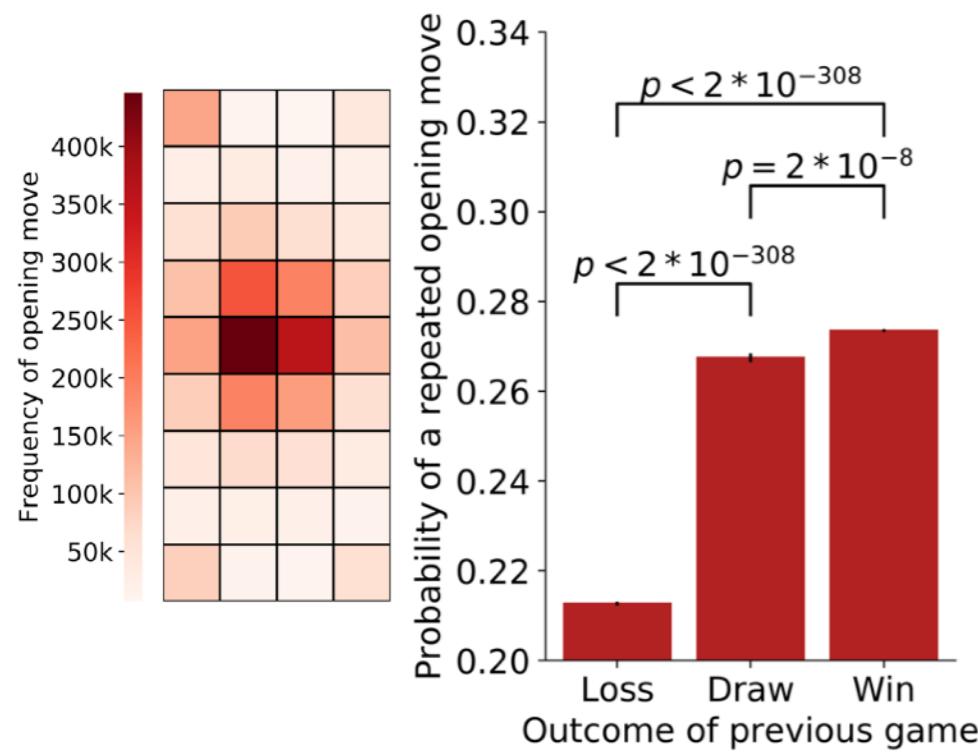
Control for experience



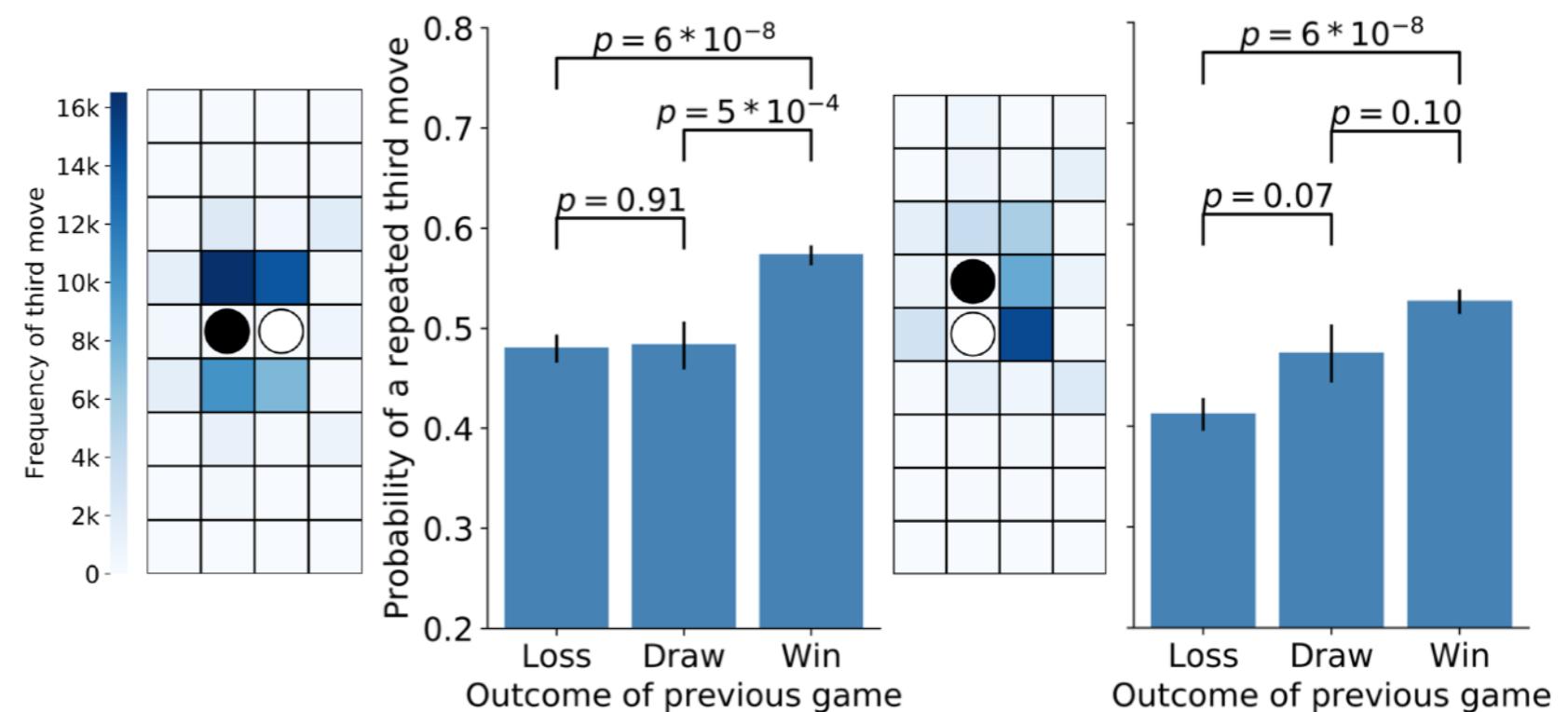
Control for memory

MODEL-FREE ACTION SELECTION

RL prediction #2: win-stay-lose-shift in early game positions



Opening move



3rd move

OUTLINE

Given this task of intermediate complexity, we want to:

- Build an AI-inspired prospective model of human play
- Fit the model to data, and explore unexplained signatures of retrospective decision-making in the large data set
- Develop models for integrating prospection and retrospection (ongoing work)

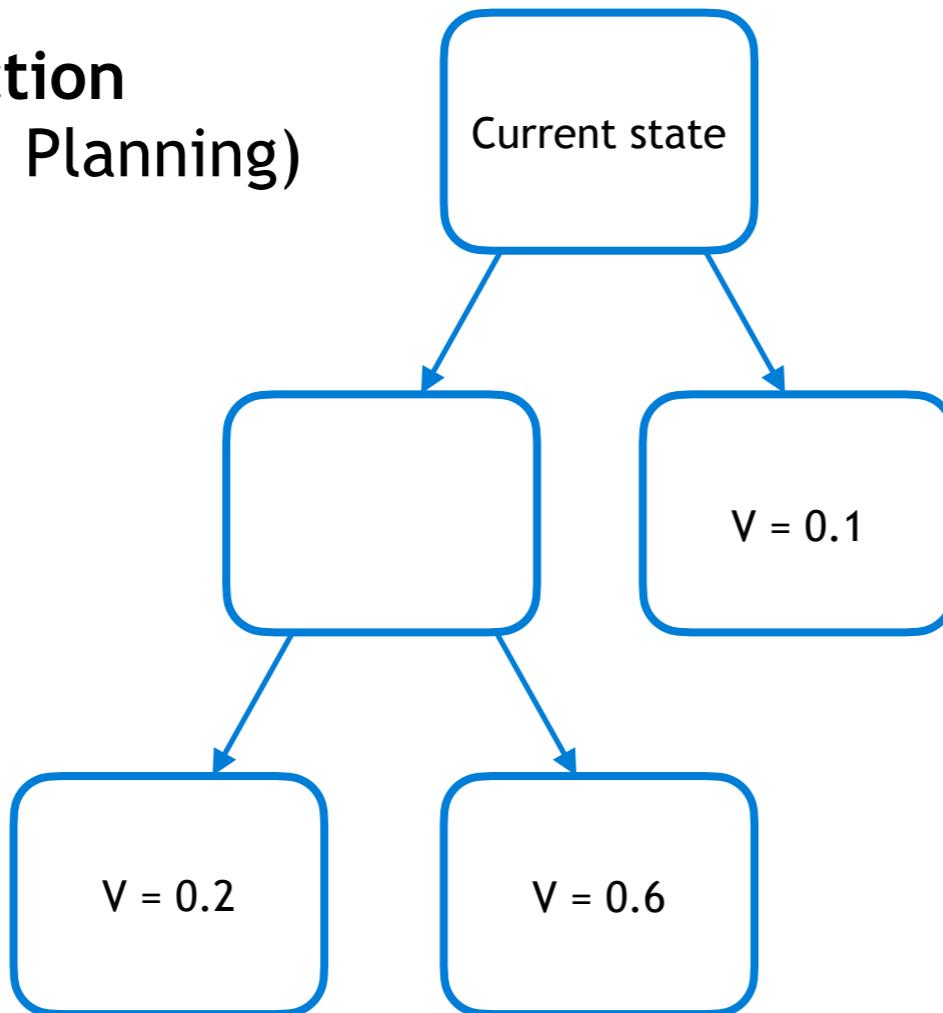
OUTLINE

Given this task of intermediate complexity, we want to:

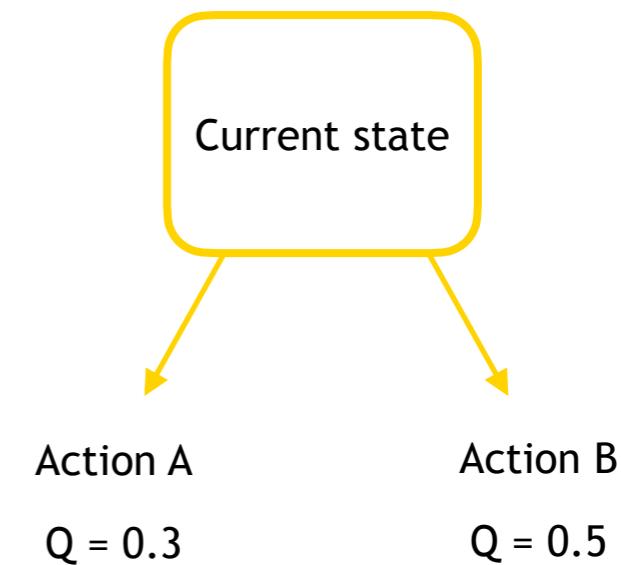
- Build an AI-inspired prospective model of human play
- Fit the model to data, and explore unexplained signatures of retrospective decision-making in the large data set
- **Develop models for integrating prospection and retrospection (ongoing work)**

DUAL SYSTEMS FOR SEQUENTIAL DECISION-MAKING

Prospection
(Model-Based, Planning)



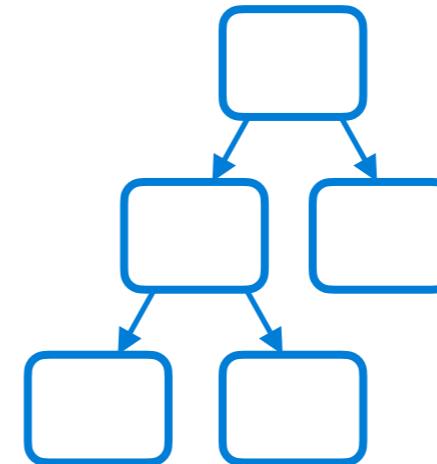
Retrospection
(Model-Free, Habitual)



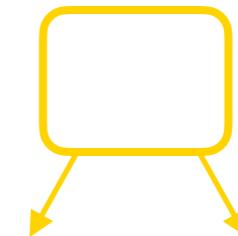
What algorithms do humans use to **integrate prospective and retrospective decision-making** in a more complex task?

MIXTURE MODEL: THEORY SKETCH

Goal: compute a measure of uncertainty for both systems over p_{win} , and run the one with lower uncertainty



Planning model



Q-learning

How do we calculate uncertainty?

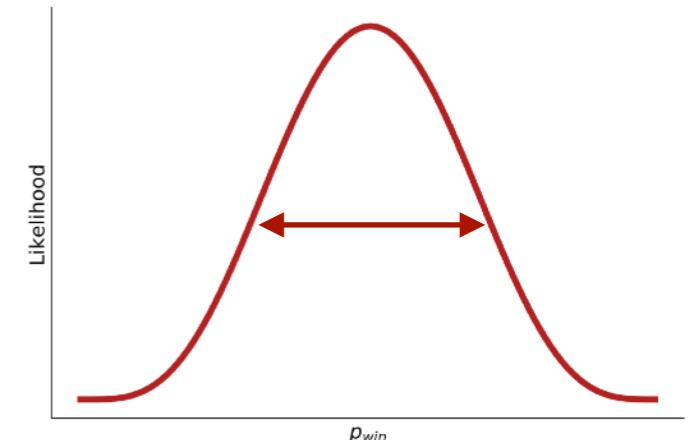


Generative model

$$P(n_{wins} | n_{past}, p_{win}) \sim \text{Binomial}$$

Inference

$$P(p_{win} | n_{win}, n_{past}) \sim \text{Beta}$$



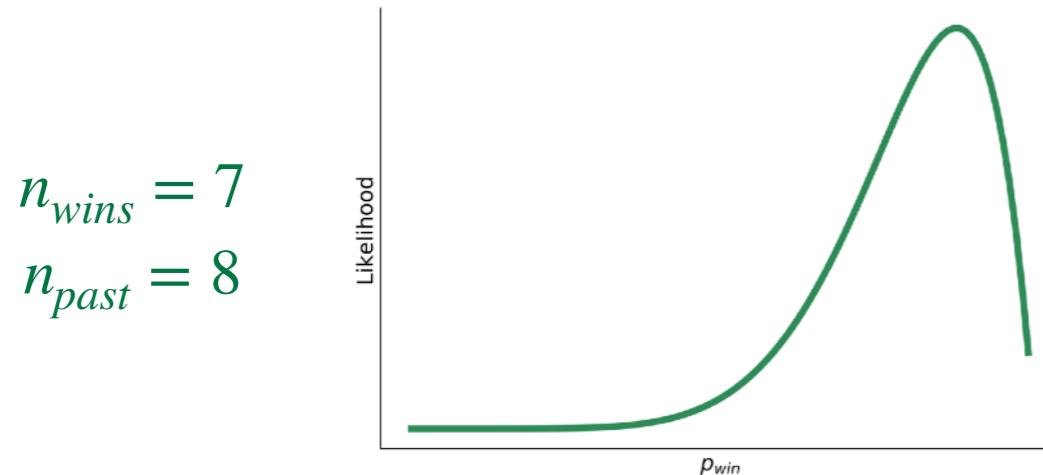
Typically, uncertainty is the **variance** of the best action's posterior distribution
However, we care about **comparing with alternative actions**

MIXTURE MODEL: THEORY SKETCH

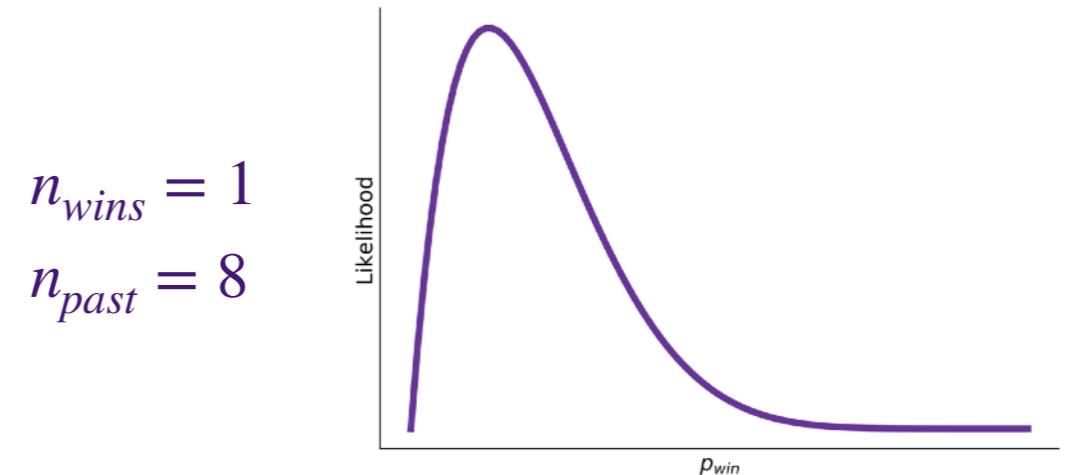


Toy example for **expected regret** using Q-learning
and two possible actions

Posterior for the **chosen action**



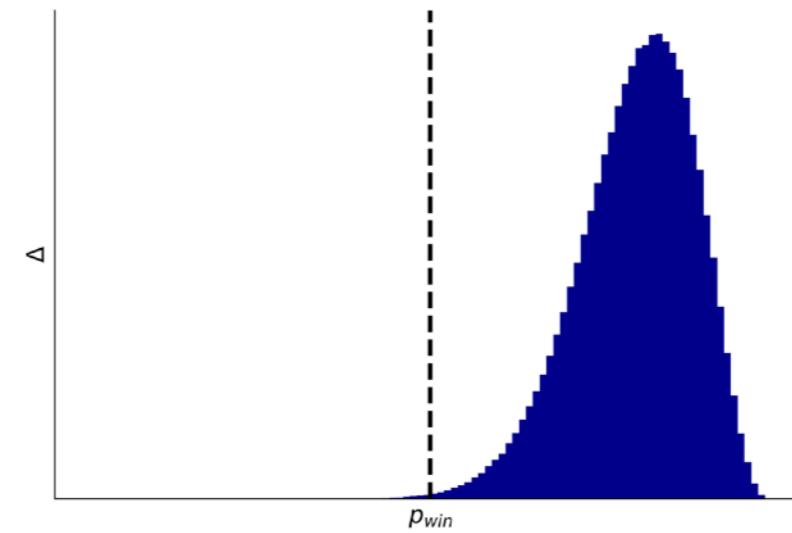
Posterior for the **alternative action**



Difference distribution obtained via sampling

$$\Delta = p_{chosen} - p_{alternative}$$

$$Expected\ regret = \int_{-\infty}^0 \Delta f(\Delta) d\Delta$$



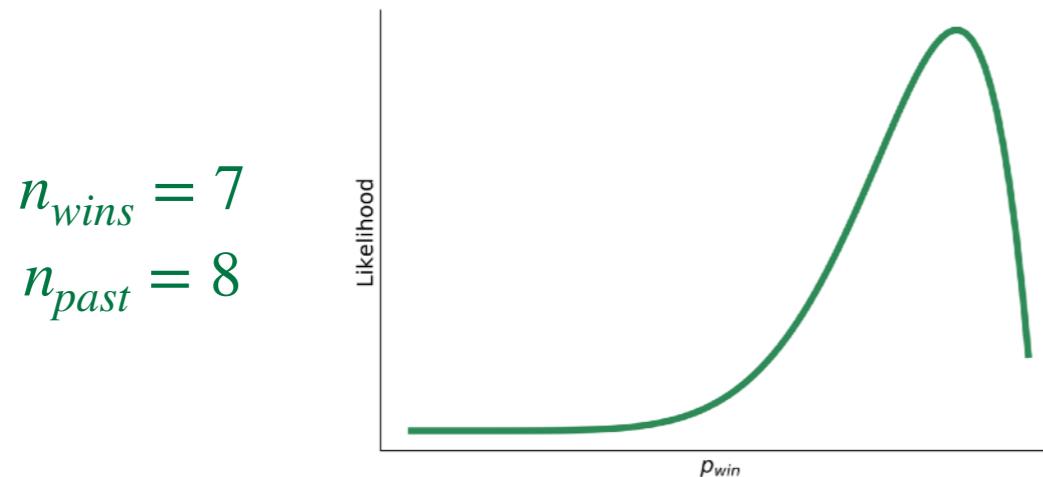
Low expected regret

MIXTURE MODEL: THEORY SKETCH

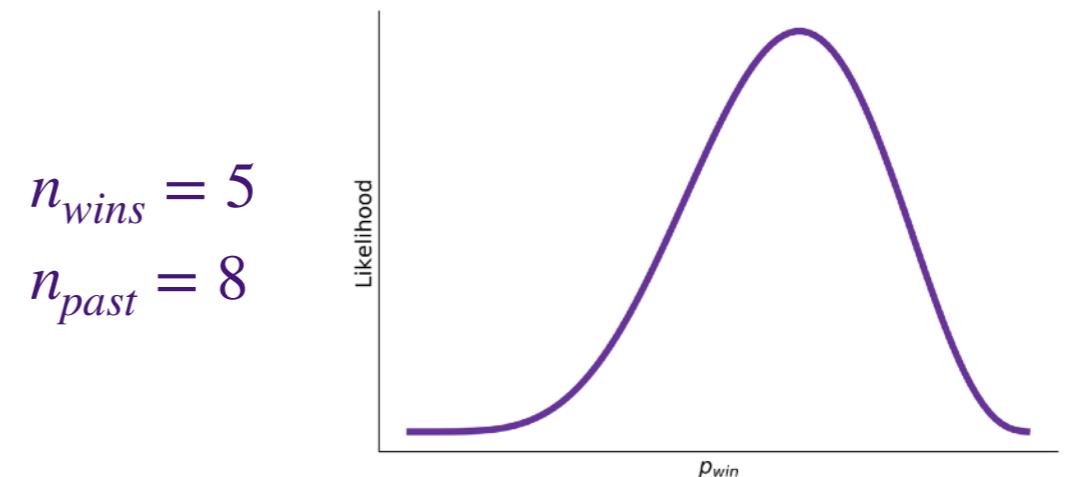


Toy example for **expected regret** using Q-learning
and two possible actions

Posterior for the **chosen action**



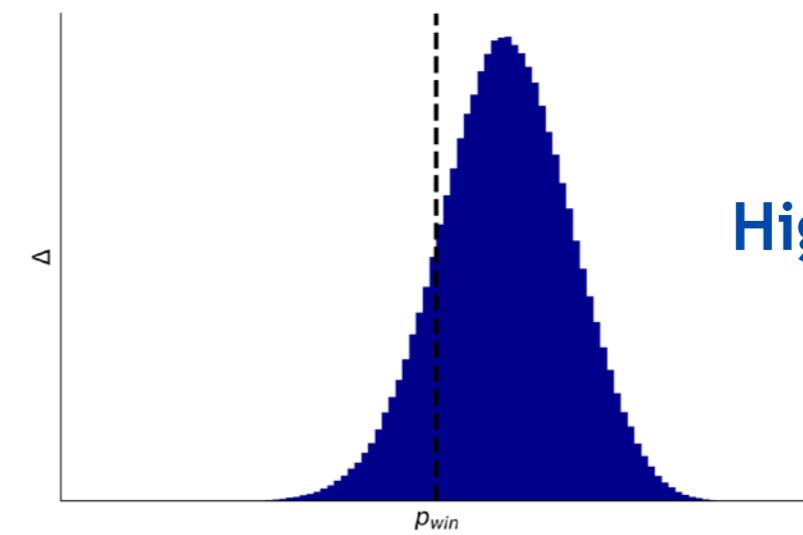
Posterior for the **alternative action**



Difference distribution obtained via sampling

$$\Delta = p_{chosen} - p_{alternative}$$

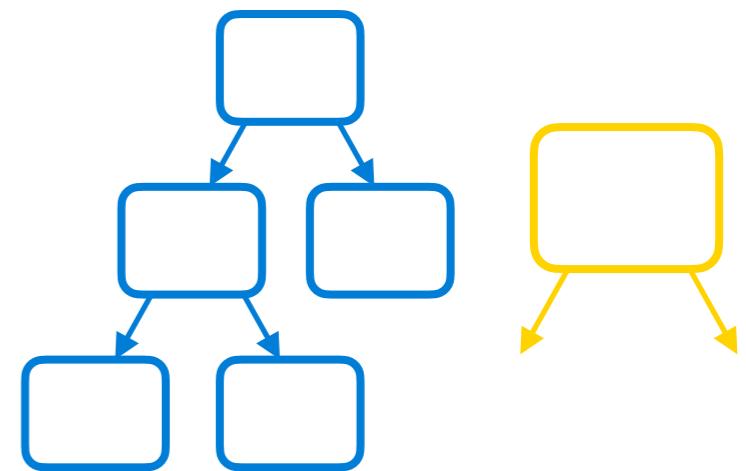
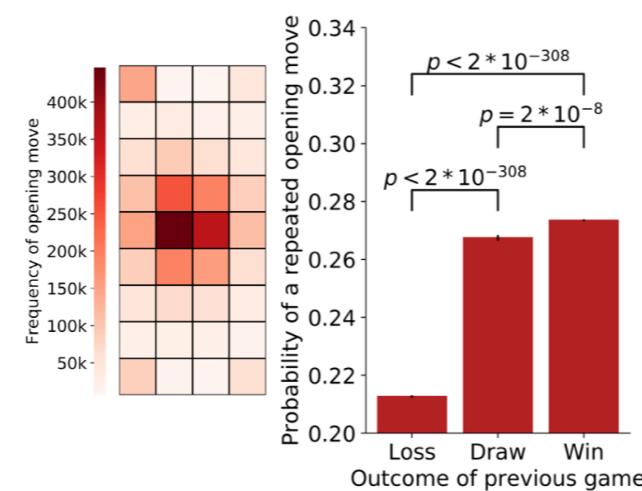
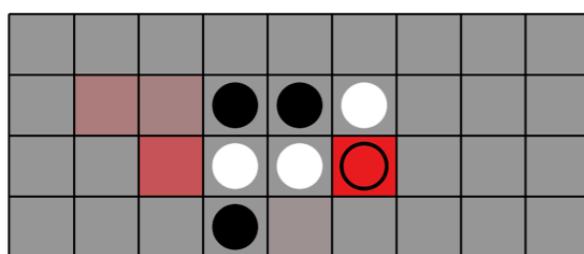
$$Expected\ regret = \int_{-\infty}^0 \Delta f(\Delta) d\Delta$$



SUMMARY

Computational modeling of human decision-making in complex planning tasks is tractable

- Built a computational model inspired by concepts in classical artificial intelligence
- Fit the planning model to human play, providing evidence for decision tree search
- Found signatures of model-free decision-making early in gameplay using a large data set
- Started investigating the integration of model-based and model-free algorithms



NEURAL BASIS OF COMPLEX PLANNING

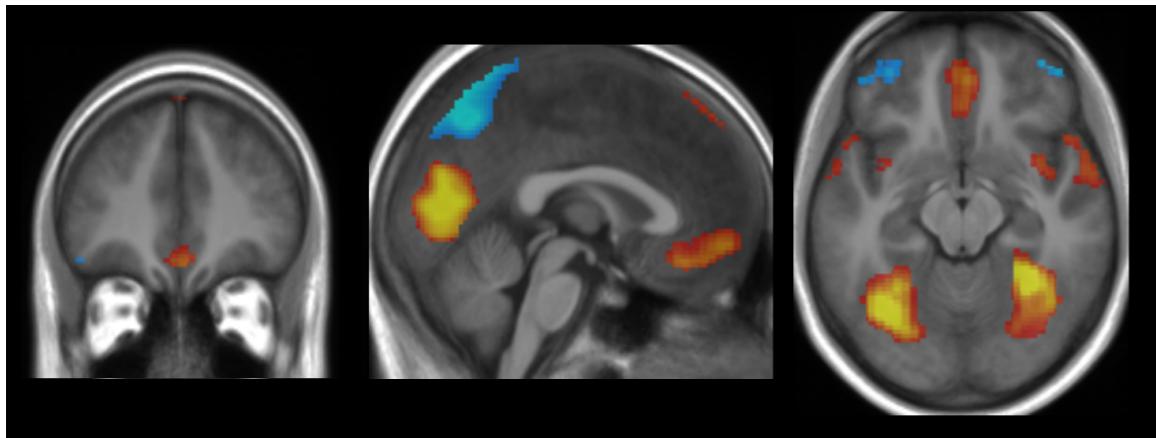
Humans



Nathaniel Daw



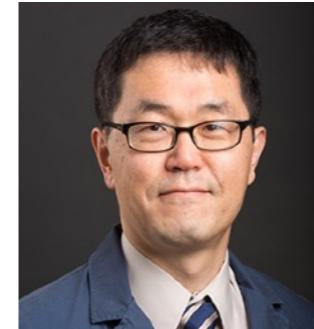
Marcelo Mattar



n=35

Are there fMRI signatures of board states and values in human players?

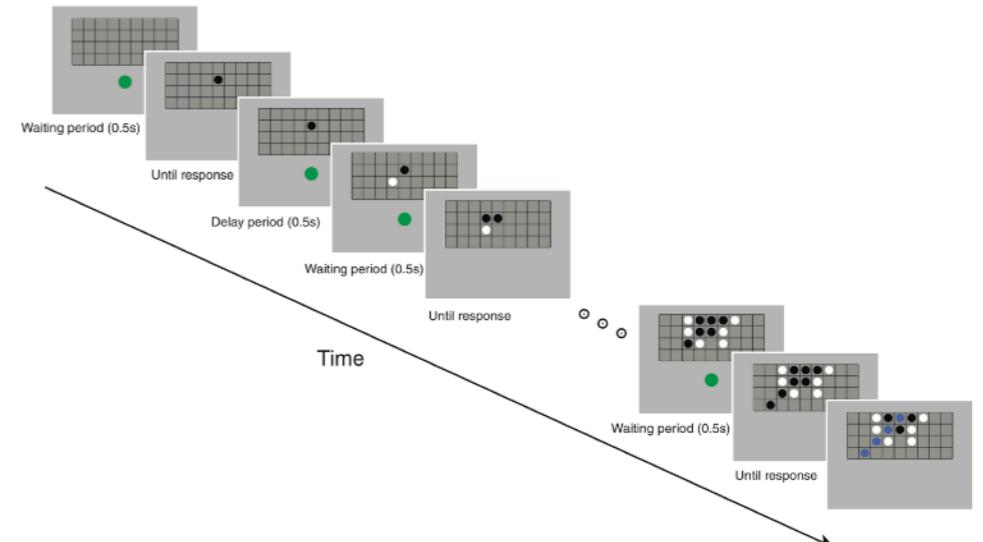
Non-human primates



Daeyeol Lee

Neural correlates of strategic reasoning during competitive games

Hyojung Seo,^{1,*} Xinying Cai,^{1,†} Christopher H. Donahue,^{1,‡} Daeyeol Lee^{1,2,3*}



Can non-human primates learn complex planning tasks? If so, what neural substrates support planning versus other forms of decision-making?

Ma Lab
Wei Ji Ma
Heiko Schütt
Hsin-Hung Li
Ili Ma
Aspen Yoo
Carolina Di Tella
Jenn Lee
Peiyuan Zhang
Xiang Li
Donqi Bao
Yichen Li
Qixiu Fu



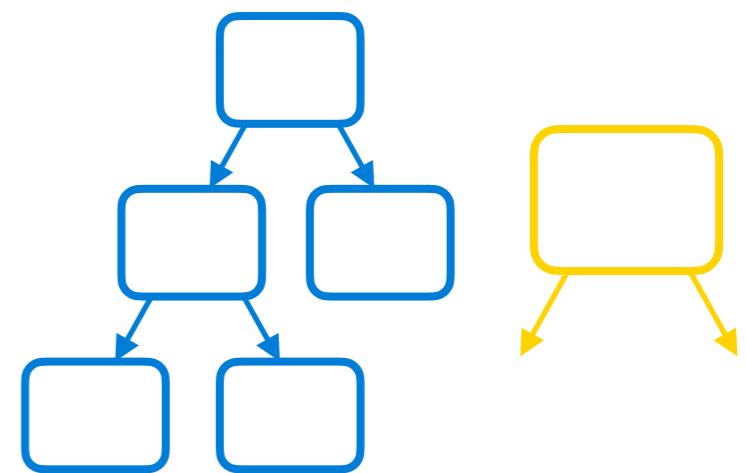
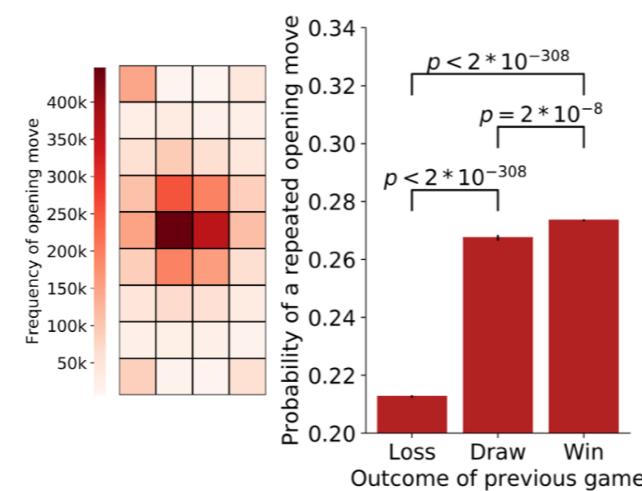
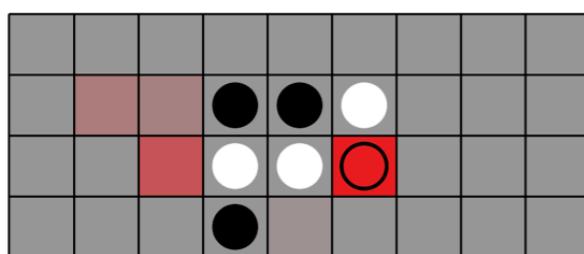
Collaborators
Bas van Opheusden
Daeyeol Lee
Nathaniel Daw
Marcelo Mattar



SUMMARY

Computational modeling of human decision-making in complex planning tasks is tractable

- Built a computational model inspired by concepts in classical artificial intelligence
- Fit the planning model to human play, providing evidence for decision tree search
- Found signatures of model-free decision-making early in gameplay using a large data set
- Started investigating the integration of model-based and model-free algorithms



Extra slides

PLANNING IN NEUROSCIENCE

Rats consider possible futures

Johnson and Redish, *J Neurosci*, 2007



White: rat's current position
Color: position in hippocampal place cell activity

2011

Neural Correlates of Forward Planning in a Spatial Decision Task in Humans

Dylan Alexander Simon¹ and Nathaniel D. Daw^{1,2}

2013

Neural Computations Underlying Arbitration between Model-Based and Model-free Learning

Sang Wan Lee,^{1,2,3,*} Shinsuke Shimojo,^{1,2,4} and John P. O'Doherty^{1,2,3}

2019

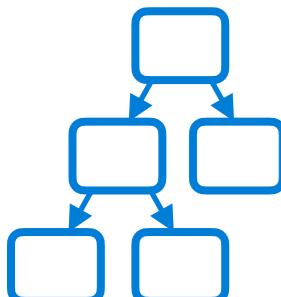
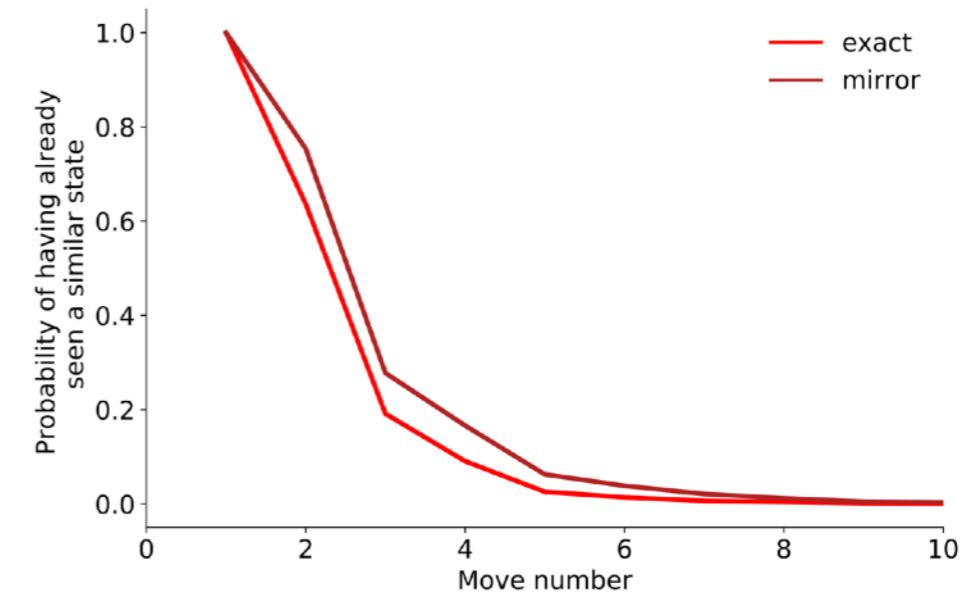
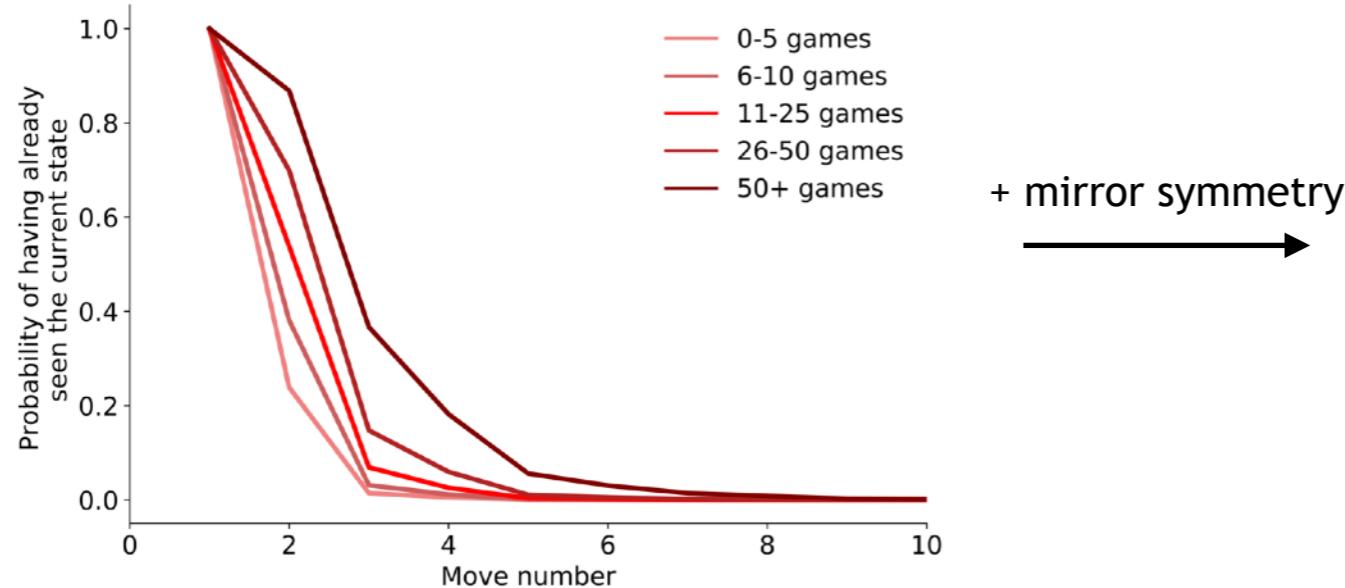
Hippocampal Contributions to Model-Based Planning and Spatial Memory

Oliver M. Vikbladh^{1, 14, 15}  , Michael R. Meager^{2, 3}, John King⁴, Karen Blackmon⁵, Orrin Devinsky^{3, 6, 7}, Daphna Shohamy^{8, 9}, Neil Burgess^{10, 11}, Nathaniel D. Daw^{12, 13}  

MIXTURE MODEL: THEORY SKETCH

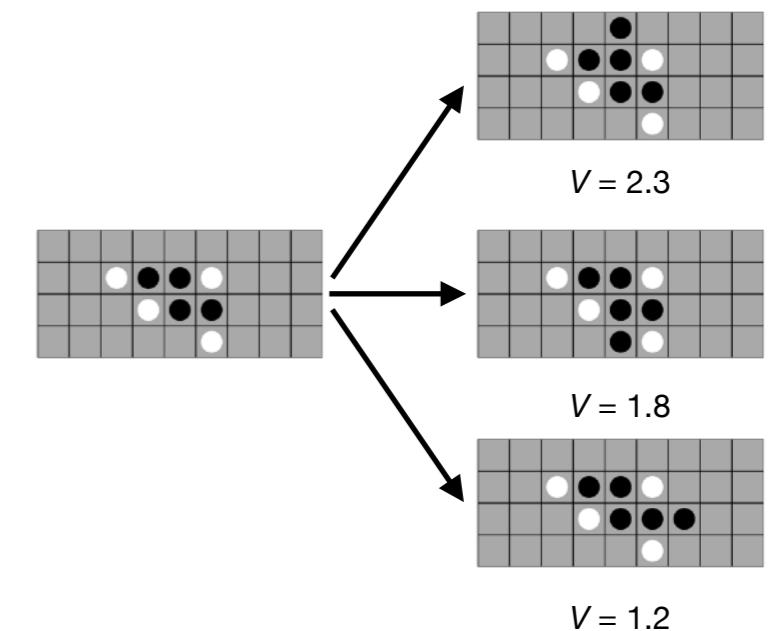


Defining a similarity metric



What about the planning model?

- Compute expected regret without forward search and in terms of win probability
- Myopic evaluations correlate with win probability

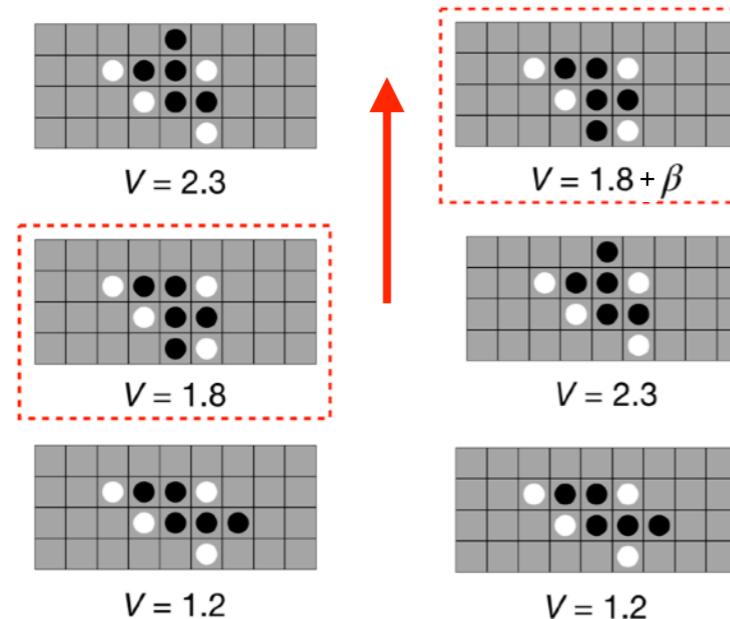


OTHER MODEL CONSIDERATIONS

Value updating

Add a retrospective component to the prospective model

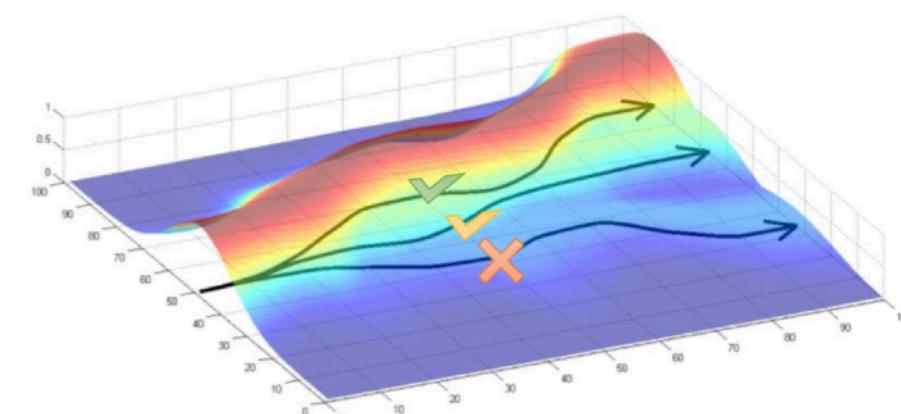
$$V(s) = c_{\text{self}} \sum_{i=0}^4 w_i f_i(s, \text{self}) + \beta(s)$$



Credit assignment

In specific states, which features led to certain outcomes?

$$V(s) = c_{\text{self}} \sum_{i=0}^4 \textcolor{red}{w}_i f_i(s, \text{self})$$



What does it mean for an integration model to be successful in this task?

Response time trends in the opening

- Resolving the mismatch between the data and the original planning model

Move prediction at two scales

- Within-game: early moves are better predicted by retrospection, later moves are better predicted by prospectation
- Between games: more experience transfers control of decision-making from prospectation to retrospection

FUTURE DIRECTIONS

Train constrained convolutional neural network on human play

- Theoretical upper bound on how well any model can fit the data
- Powerful way to learn features

Normative optimizations

- Resource rationality: maximize game outcome while minimizing computational costs
- Per iteration: is expanding one more node worthwhile?
- Per move: is prospective search worthwhile at all?

Working memory limitations

- Exponential decay of values in tabular RL component of the model
- Optimal strategy for memory refresh: when is it most beneficial to revisit states in your decision tree?