Planning In Humans And Machines

by

Xinlei Lin

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Department of Neuroscience

New York University

May, 2025

Dr. Wei Ji Ma

Dedication

To my future self: Remember that you did the hard thing when you're facing your next impossible challenge.

Acknowledgements

This dissertation is more than the research contained in its pages; it's the culmination of a decade-long journey shaped by mentors, colleagues, friends, family and two cats who cheered me on. These acknowledgments attempt the impossible: to adequately thank those who transformed an intimidating academic marathon into a memorable journey

First and foremost, I owe my deepest gratitude to my advisor, Weiji. Your intellectual rigor and warmth taught me that brilliant science and genuine kindness are not mutually exclusive. Thank you for giving me the freedom to chase ideas that make me excited, for encouraging my often overly optimistic plans (and giving me reality checks when needed), and for reminding me regularly that karaoke and escape rooms are essential forms of self-care. I couldn't imagine crossing this finish line without your support, and this thesis wouldn't have explored so many fascinating paths without your guidance.

To my committee – Christine Constantinople, Cate Hartley, and Todd Gureckis – thank you for sharpening my arguments, encouraging my progress, and making every meeting an inspiring conversation. To my collaborators Sam, Jake, Victor, Shucheng and Brenden – many chapters of this dissertation simply wouldn't exist without your contributions.

To my cats, Boba and Milk: thank you for relocating across the country to New York with me, sacrificing your sprinting space for my academic pursuits. Your purrs during late-night writing sessions and judgment-free cuddles during moments of self-doubt were therapeutic. Without you, I would have surely spiraled into a pandemic-induced existential crisis during those first

two isolated years of my PhD. I hope you are excited about moving back to California.

To my lab family – Jordan, Jeroen, and Nastaran: walking into our fifth-floor office and seeing your friendly faces transformed even the most frustrating days into bearable ones. Our random lunch conversations were often the highlight of my day. Thanks to Xiang, Ionatan, and Heiko for the warmest welcome when I first joined the lab. Thanks to everyone I've overlapped with in lab – Dongjae, Hsin, Yotam, Jenn, Aspen, Peipei, Sixuan, Shucheng, Yiran, Jieyu, Yichen – for making our lab feel like a second home; And Yanli: even though our lab timelines didn't overlap, our shared stories and laughter about navigating this field as women always made me feel less lonely.

Thanks to the fifth floor higher cognition suite that makes a workspace feels like a community, and Pat for enthusiastically advertising my chili crisp.

A special acknowledgment to Sweetgreen Astor Place – your overpriced yet strangely addictive salads kept my blood sugar stable enough to avoid the afternoon crash, even if your portion size is significantly smaller compared to other Sweetgreens.

To friends since day-one: Miao, Yue, Shiyun, Xu and Jiadai, who've known me since before I could pronounce "computational neuroscience" – thanks for sticking around and never asking "what is your research about?". To Ning, Chun, Megan, and Ray – I'm honored to be the godmother to your adorable tiny humans. To Penny and Bolong, thanks for being the witnesses to many milestones of my life and for all those delicious meals we shared that nourished my soul. To Wanwan and Kedi: who knew that a stranger sliding into my Instagram DMs for wedding tips would turn into a special friendship? To the girl squad – Jenny, Penny, Shishi, Jingjin, Meimei, April, Haihai, Wenli – our brunches were sanity-saving oases in a field where female colleagues were sometimes as rare as statistically significant p values. Lixing, thank you for the memorable snowboarding adventures, and Panpan, you truly were the best flower dude. To Ruby, Mengdi, for always infusing our gatherings with vibrant energy. To my cohort – Keelin, Aaron, and Heejae – for those yummy butter chicken cooking nights. I'm grateful for everyone's persistence

in dragging me out for weekend nights when all I wanted was solitude – yet somehow always leaving me refreshed and reconnected to the world outside my research.

To my undergraduate mentor, Takaki: You saw potential in me when I knew nothing about academic research. I wouldn't have discovered my passion for this field – let alone found the courage to pursue a PhD – without the wonderful experience in your lab.

To Mom and Dad, thanks for giving me a childhood free from constraints, allowing my curiosity to run wild. I wouldn't be as optimistic and resilient as I am now without your unconditional love. And to the extended family – grandparents, aunts, uncles, parents-in-law, sister-in-law, and Zhouzhou, my lovely little cousin with whom I shared my childhood – thank you for teaching me that "home" transcends geography.

Nate, my life co-author: thank you for believing in me through failed experiments, buggy codes, and midnight crises when I was convinced my research was heading nowhere. Your confidence in me has been more reliable than any statistical significance I've ever achieved. This dissertation bears my name, but your fingerprints are on every page.

Finally, to my past self: I know you often questioned why you picked such challenging topics, but thank you for taking the leap.

Abstract

Planning – the ability to imagine future possibilities, evaluate options, and choose actions to reach desired goals – shapes decisions as simple as organizing one's day or as complex as guiding a spacecraft to Mars. Cognitive science has traditionally studied planning through small state-space tasks, while artificial intelligence research pushes algorithms to master highly games like Go or Chess. This dissertation bridges this gap by using Four-in-a-Row – a game complex enough to challenge both humans and machines yet still computationally tractable. Across four studies, we ask (i) what cognitive components underlie planning? (ii) How do humans construct plans moment-by-moment? (iii) How well can transformer-based sequence models that condition on long move histories predict human actions? and (iv) what the state-of-the-art planning agent – AlphaZero – learns and misses when mastering the game?

To answer these questions, I combine large-scale behavioral experiments, think-aloud protocol analyses, transformer-based behavioral modeling, and deep reinforcement learning tools, aiming to deepen our understanding of how humans and machines approach complex planning tasks. The resulting picture sketches a virtuous loop between cognitive science and AI: empirical insights from human cognition help reveal blind spots in AI algorithms, while machine learning tools set new ceilings and analytic probes for refining cognitive models and theories of human planning.

Contents

Dedication iii

Acknowledgments iv

Abstract vii

List of Figures xi

1 Introduction 2

1.1 Motivation and Significance of Planning 2

1.1.1 Planning as a Fundamental Ability 2

1.1.2 A Brief Historical Perspective 2

1.1.3 Planning matters for humans 3

1.1.4 Planning matters for AI 4

1.1.5 Planning Bridges Minds and Machines 4

1.2 Planning in Humans: Theories and Paradigms 5

1.2.1 Planning and Problem-Solving 5

1.2.2 Foundational Frameworks 5

1.2.3 Model-Based vs. Model-Free 6

1.2.4 Tree-Search 7

1.2.5 Paradigms for Studying Human Planning 8

1.2.6 A Case Study: Four-in-a-Row . . . 12
1.2.7 Methods for Investigating Planning . . . 15
1.2.8 Cognitive Abilities . . . 18
1.3 Planning in Machines: Models, Searches, and Learning . . . 20
1.3.1 Classical Symbolic Planning . . . 20
1.3.2 Adversarial Game-Tree Search . . . 20
1.3.3 Sampling–Based Search: Monte-Carlo Tree Search . . . 21
1.3.4 Learning to Guide Search. . . . 22
1.3.5 Transformer . . . 23
1.4 Overview of This Dissertation . . . 25

2 What are the cognitive components of planning? 28
2.1 Introduction . . . 28
2.2 Methods . . . 30
2.3 Results . . . 36
2.4 Discussion . . . 47

3 Do humans think in trees? Lesson from think-aloud protocal 52
3.1 Introduction . . . 52
3.2 Methods . . . 55
3.3 Results . . . 63
3.4 Discussion . . . 70

4 Are humans Markovian planners? 73
4.1 Introduction . . . 73
4.2 Methods . . . 75
4.3 Results . . . 79

4.4 Discussion 83

5 What machines can learn from humans? Lessons from AlphaZero 87

5.1 Introduction 87

5.2 Method 88

5.3 Results 92

5.4 Discussion 99

6 Conclusion 107

Supplementary Materials 112

Bibliography 120

List of Figures

1.1 Example board of Four-in-a-Row. 12

2.1 Task battery used in the online experiment 37

3.1 Example puzzle in think-aloud experiment 57
3.2 All Four-in-a-Row puzzles 58
3.3 Correlation between subjective difficulty and planning metrics 65
3.4 Relationships between verbalization metrics and Elo rating 66
3.5 Verbalized depth progression within puzzles 68

4.1 Tokenization scheme 76
4.2 GPT-4IAR architecture 78
4.3 GPT-4IAR training and validation loss 80
4.4 GPT-4IAR Action accuracy vs. context length 82
4.5 Example distributions of predicted moves in GPT-4IAR 83
4.6 GPT-4IAR reaction time prediction vs. context length 84
4.7 Reaction time RMSE in GPT-4IAR 85

5.1 AlphaZero Network Architecture 89
5.2 Elo and planning metric comparison between AlphaZero and human. 102
5.3 Policy quality and policy entropy. 103

5.4 Mediation analysis: illustration and results. 103
5.5 The effect of N_{MCTS} manipulation on depth and Elo. 104
5.6 Elo ratings as a result of the value or policy function manipulation. 104
5.7 Feature Probing Analysis 105
5.8 Visualization of NMF for selected factors. 105
5.9 AlphaZero's puzzle failures 106

1 Example spreadsheet for think-aloud protocol analysis showing Articulated Tree rows (yellow), Feature rows (green), and Qualitative Description rows (red). . . . 116

List of Tables

2.1 Correlation table of task metrics 40

2.2 Factor Loadings for Three-Factor Solution 45

2.3 Lasso regression results for planning tasks 46

2.4 Cross-task cross-half correlations 47

3.1 Verbalization-derived metrics and their computation 62

3.2 Descriptive statistics of verbalized planning behaviors (N=34). 64

3.3 Correlation Matrix of Metrics 67

4.1 Fixed hyperparameters used for training. 79

4.2 Comparison between the fully connected model and GPT-4IAR 81

1 Correlation table of task metrics with p values 113

2 Comparison of Split-half Reliability Estimates for Each Task 114

3 Varimax-Rotated Factor Loadings for Three-Factor and Two-Factor Solutions. Bolded loadings exceed 0.30. 115

1 Introduction

1.1 Motivation and Significance of Planning

1.1.1 Planning as a Fundamental Ability

Whether it is a traveler stringing together metro lines to reach the airport or a robot navigating clutter to fetch a glass, intelligent agents – biological or artificial – succeed by anticipating future states before acting. Planning is the process of generating, evaluating, and selecting sequences of actions that transform a present state into a desired one. Everyday activities such as scheduling meetings or furnishing an apartment showcase this foresight. In artificial systems the very same faculty powers route–finding, task–and–motion control, and the reasoning in large language models (LLM) [37, 176, 207, 229, 234]. As Mattar and Lengyel [125] put it, planning is "a fundamental component of intelligent behavior in both biological and artificial agents".

1.1.2 A Brief Historical Perspective

In psychology, Miller, Galanter, and Pribram [130] argued that humans organize behavior around internal "plans" – hierarchical building blocks of behavior. In their book, *Plans and the Structure of Behavior*, they proposed that much of human cognition involves formulating internal representations of intended actions before execution. Around the same time, AI researchers sought algorithmic realizations of the same idea. STRIPS (Stanford Research Institute Problem

Solver) encoded actions as symbolic operators and solved navigation tasks for Shakey the robot [62]. Newell and Simon's General Problem Solver (GPS) modeled human problem solving via means–ends analysis: repeatedly identifying the difference between current and goal states and selecting operators to reduce that gap (1959). Despite arising in separate disciplines, these efforts shared the conviction that planning is inherently computational and necessarily heuristic: exhaustive enumeration is infeasible, so both brains and machines must rely on abstractions and shortcuts to navigate complex problems.

1.1.3 Planning matters for humans

Planning is a hallmark of human intelligence because it lets us rehearse possible futures, score their outcomes, and commit to the most promising actions, all without acting in the world. Non-human primates and corvids show rudimentary foresight (e.g., tool caching or delayed gratification) [138, 164], but only humans routinely orchestrate multi-step, cross-domain plans that may unfold over hours, months, or years [203]. For example, graduate students need to schedule coursework, experiments, and dissertation milestones across a multi-year program. Intuitively, effective planning requires a working memory to hold intermediate states in mind, the ability to recognize patterns or analogies to past experiences, mental simulation of future states, inhibitory control to suppress habitual responses that derail goals, and logical reasoning to select the best actions. Lesions in the prefrontal cortex leave perception intact, but shatter the ability to coordinate multi-step tasks such as cooking or budgeting[186]. In contrast, the maturation of prefrontal circuitry in childhood parallels the gains in everyday problem solving[32]. Understanding the cognitive architecture of planning therefore promises both theoretical insight and clinical leverage.

1.1.4 Planning matters for AI

Since 1950s chess programs to AlphaZero and modern tool-calling LLMs, headline AI systems succeed through planning: Deep Blue's victory over world champion Kasparov [29]; AlphaZero's superhuman play in chess, Go, and shogi [190]. More recently, iterative prompting frameworks (ReAct, Tree-of-Thoughts) harness planning to turn large language models into multi-step reasoners. Across these milestones, planning provides the bridge between perception, prediction, and action. [176]

1.1.5 Planning Bridges Minds and Machines

Planning represents a natural domain for interdisciplinary inquiry, with both cognitive science and AI investigating how agents navigate spaces of possible futures to achieve their goals. Cognitive scientists borrow AI formalisms such as state representations and heuristic search to build algorithms of human planning [20, 27, 97, 148, 162]. Conversely, empirical discoveries about how people plan such as chunking, hierarchical abstraction, and resource-rational allocation – inspire elements in modern AI systems [19, 34, 119, 194]. Recognizing this reciprocity is crucial: progress in one field continuously reshapes the hypothesis space of the other. The present thesis leverages that synergy, using AI models both as explanatory tools for human data and as beneficiaries of insights drawn from cognitive experiments. The cross-fertilization that began in the 1960s remains a guiding theme for this dissertation.

Despite six decades of progress, gaps remain. Human planning demonstrates remarkable flexibility in new or uncertain environments while operating under working-memory constraints. In contrast, AI systems achieve success through biologically implausible computational resources yet often lack the flexibility and generalizability that humans exhibit, exemplified by chess engines that can calculate millions of positions per second but struggle to transfer their strategies to even slightly modified game variants. Recent developments in LLMs further highlight these con-

trasts: they scale impressively in language but frequently falter on multi-step reasoning problems that are effortless for most humans [215]. The very limitations of LLMs have rekindled interest in planning within AI despite the headline triumphs of AlphaGo and AlphaZero in 2017–2018. At the same time, cognitive scientists are moving toward richer tasks that better capture the complexity of human planning[148]. By uniting these efforts – exploring how people plan and applying that knowledge to inform AI, while also leveraging AI models to test cognitive theories – we can progress toward a more comprehensive science of planning.

1.2 Planning in Humans: Theories and Paradigms

1.2.1 Planning and Problem-Solving

In cognitive psychology, problem-solving is the broader umbrella: it begins with problem identification, spans plan construction and execution, and ends with outcome evaluation [75]. This leads some authors to note that while all planning scenarios can be framed as problem-solving, not all problem-solving tasks necessarily involve explicit planning [75, 174]. For instance, some insight problems resolve through sudden re-conceptualization – the “aha!" moment.

1.2.2 Foundational Frameworks

Newell and Simon cast planning as search through a problem space [141]. Under this framework, a person encodes the current situation as a state, defines a goal state, and mentally explores possible actions that transform states until the goal is reached. Because exhaustive search is typically infeasible for complex tasks, Newell and Simon highlighted the use of subgoals and heuristics – mental shortcuts or rule of thumb that prune tree of possibilities. Their General Problem Solver (GPS), a model that emulated human-like problem-solving strategies, operationalized these ideas: it keeps internal representations of the current state, the goal state, and a library

of operators, then applies means–ends analysis to generate intermediate subgoals that steadily close the gap.

A parallel line of work asked why experts plan better than novices. De Groot [51] showed chess grandmasters do not consider more moves or search deeper than weaker players when choosing a move. Instead, their advantage comes from superior pattern recognition and memory for game configurations. De Groot observed that experts recognize promising moves almost immediately, relying on "chunks" of familiar piece patterns rather than brute-force search. Subsequent work by Chase and Simon [34] estimated that chess masters store tens of thousands of such patterns, enabling them to identify good moves without exhaustively enumerating many possibilities. Together these findings reframed human planning as limited-depth search guided by richly learned representations.

1.2.3 Model-Based vs. Model-Free

Decision makers are often described as relying on two complementary systems. Model-based control uses an internal model of how actions change the world to prospectively evaluate futures. Model-free control caches action values from past reward and selects habits quickly but myopically.

Mathematically, the two systems are specializations of a *Markov Decision Process (MDP)* $\langle S, A, P, R, \gamma \rangle$. A model-based agent solves or searches the Bellman equations:

$$V^*(s) = \max_a \sum_{s'} P(s'|s,a)\,[R(s,a,s') + \gamma V^*(s')]$$

whereas a model-free learner updates cached Q values via temporal-difference rules:

$$Q(s,a) \leftarrow Q(s,a) + \alpha\big[r + \gamma \max_{a'} Q(s',a') - Q(s,a)\big].$$

where $\alpha \in (0, 1]$ is a learning-rate and $\gamma \in [0, 1)$ the discount factor.

Evidence from cognitive science shows that humans blend model-based and model-free control [50]. In artificial agents the same division yields Deep Q-Networks on the model-free side [136] and search-based planners on the model-based side[190]. Hybrid schemes such as DYNA interleave model-free updates with simulated model-based rollouts, blending the two systems[204]

1.2.4 Tree-Search

Model-based planners treat an MDP's *state space* – the set of all legally reachable configurations of the environment – as a search tree whose nodes are states and whose edges are actions; different algorithms explore that tree in different ways, all sharing the core idea of "simulate futures before acting."

Depth– and Breadth-First Search. The simplest strategies explore either the deepest unexplored node first (DFS) or all nodes at a given depth before going deeper (BFS). Both guarantee reaching a goal in finite trees but explode combinatorially and ignore action cost.

Heuristic Best-First and A*. Introducing a domain heuristic $h(s)$ that estimates the distance-to-goal yields *best-first search*. A* refines this by selecting the child that minimizes

$$f(s) = g(s) + h(s),$$

where $g(s)$ is path cost so far. If heuristic h is admissible, meaning it never over-estimates the true cheapest remaining cost for all states, the first goal node popped by A* is guaranteed optimal [80].

Adversarial Minimax with α–β Pruning. Two-player games replace path cost with a value function that one player maximizes and the other minimizes. Minimax evaluates leaves and backs

up values; the α–β bounds skip branches that cannot influence the final decision, turning exponential search into something closer to $b^{d/2}$ effective branching.

Monte-Carlo Tree Search (MCTS). MCTS merges simulation with sampling. Each simulation performs *selection*, *expansion*, *roll-out*, and *back-propagation*. During selection a child a of state s maximizes

$$UCB(s,a) = Q(s,a) + c\sqrt{\frac{\ln N(s)}{N(s,a)}},$$

with Q the mean return and N visit counts. Repeating this loop concentrates search on promising branches while ensuring exploration, producing an anytime estimate of the optimal action – crucial for high-branching domains such as Go. We revisit neural network guided MCTS variants in (§1.3.3).

All four algorithms implement the same template –simulate, rank, and back-up – but differ in how they prioritize nodes. Later chapters, we will see how different algorithms exploit these differences.

1.2.5 Paradigms for Studying Human Planning

Researchers have investigated human planning through a variety of experimental paradigms ranging from highly controlled laboratory tasks to messy real-world scenarios. This section provides an overview of several widely used tasks.

Two-step Task One widely cited paradigm is the Two-Step Task, originally introduced to distinguish between model-based (goal-directed) and model-free (habit-driven) decision-making strategies [50]. This task requires minimal "look-ahead". On each trial, the participant makes a first-stage choice between two options. This leads, with a fixed transition probability (typically 70% for a common transition and 30% for a rare transition), to one of two second-stage states. At the second stage, the participant chooses again between two options, each of which is associated

with a certain probability of yielding a reward (often a binary reward like a coin). The second-stage reward probabilities drift slowly over time (e.g. via a random walk), encouraging participants to continually evaluate and update their decision strategy. A purely model-free learner, which simply reinforces actions that were rewarded, will tend to repeat any first-stage choice that was followed by reward. In contrast, a model-based planner understands the task's transition structure: after a rare transition, the model-based strategist would realize that a repeat of the same first-stage choice is less likely to lead back to the previously rewarded second-stage state, and might switch choices. By examining participants' tendency to stay with or switch first-stage choices as a function of the previous trial's outcome and transition type, one can quantify the degree to which their behavior reflects model-based versus model-free. Empirically, healthy humans exhibit a mixture of model-based and model-free control on this task. In terms of complexity, the two-step task is deliberately simple with state space complexity of three. This simplicity allows researchers to fit parameters to hybrid RL models to recover individual model-based weights. The two-step task exemplifies how even a minimal sequential decision problem illuminates multi-step decision-making.

Towers Task The Towers tasks is a common way to operationalize planning for researchers. They are both puzzle tasks that require formulating a series of moves in advance to reach a goal state.

Tower of Hanoi (TOH). Invented by Lucas in 1883, TOH asks participants to move a stack of n disks to a new peg while (i) moving only one disk at a time and (ii) never placing a larger disk atop a smaller. The optimal solution length is 2^n–1 moves and the legal state space grows as 3^n (27 states for n=3, 243 for n=5). The state space of the Tower of Hanoi consists of all legal configurations of n disks on the pegs. This state-space size grows exponentially as 3^n. For instance, with 5 disks there are $3^5 = 243$ distinct legal states. The Tower of Hanoi has been extensively used in cognitive psychology to analyze problem-solving strategies and learning. Anzai and Simon's [9]

famous study had a participant think aloud while repeatedly solving a 5-disk Tower of Hanoi; the analysis of her verbal protocols revealed how she formed subgoals and transformed her strategy with practice. Because the optimal path is known, TOH has served both as a classic cognitive psychology paradigm – Anzai & Simon's think-aloud study famously traced subgoal formation.

Tower of London (TOL). Shallice [184] variant replaces disks with three colored balls and pegs of capacities 3–2–1, yielding 36 reachable configurations. Unlike the Tower of Hanoi, there is no restriction about larger vs. smaller since all balls are the same size; any ball can be placed on any peg as long as there is space. Typically, Tower of London problems are given with a minimum moves criterion (e.g. "solve this in the fewest moves possible"), and problems can vary in difficulty (from requiring 2 moves up to 7 or more moves in the hardest versions used clinically). TOL has become a standard test of executive planning in neuropsychology. Performance is often measured by the number of optimal moves. Patients with frontal lobe lesions, for example, tend to perform worse on the TOL, taking more moves and more time, indicative of planning deficits [184].

Board Games Games are attractive scientific stimuli because they combine formal structure (clear rules, fully observable states) with rich combinatorial complexity, offering a sweet-spot between sterile laboratory puzzles and messy real-world situations [7]. Their intrinsic fun also motivates millions of people to generate openly available play logs, giving researchers an unprecedented window on large-scale planning behavior.

In cognitive psychology and artificial intelligence alike, chess has been called the "Drosophila" in AI, serving as a model organism for understanding the mind (a phrase inspired by de Groot's pioneering chess studies [51] and later popularized by others). The game of chess is exceedingly complex – far more so than the puzzle tasks discussed above – yet humans can become exceptionally skilled at it, which naturally invites the question of how they manage to plan and decide on moves in such a vast search space.

Chess is played on a 8×8 board and has an estimated state-space complexity on the order

of 10^{43} to 10^{47} possible legal positions. [187] approximated about 10^{43} positions and a game-tree complexity of 10^{120} for chess. Clearly, exhaustive search or brute-force planning over this space is impossible for human minds (and even for computers, beyond limited depths). Yet, human chess players can look ahead a number of moves and devise effective strategies.

From a state-space standpoint, chess epitomizes planning under combinatorial explosion: at each turn, a player faces numerous legal moves, leading to exponential growth in possible future positions [187]. This complexity necessitates the use of heuristics and chunking strategies. Heuristics such as "develop pieces toward the center" or "attack the king's position" guide players in decision-making [51]. Chunking involves grouping pieces into meaningful patterns, allowing players to process complex positions efficiently [35].

These cognitive strategies are developed through extensive experience and instruction. Cognitive task analyses have demonstrated that strong players internalize high-level plans, such as controlling key squares and coordinating piece activity, which guide their move choices [89]. Unlike puzzles like the Tower of Hanoi, where an optimal sequence can be computed, chess requires setting intermediate goals (e.g., gaining material advantage or improving pawn structure) and continual re-planning in response to the opponent's moves [193]. Studies in chess reveal phases of situation assessment, where players evaluate the position using learned principles, followed by progressive deepening search, where they explore a move, analyze opponent replies, then backtrack and refine their plan. Notably, much of what we understand about human chess planning has informed AI as well – early chess programs were designed to mimic human-like selective search as described by De Groot [51], and the concept of a heuristic evaluation function (assigning a value to a board position) is analogous to human intuitive judgment of a position's merits [141]. Eye-tracking studies show that expert players' gaze fixates more on important pieces and key squares compared to novices, reflecting their internal guidance on relevance [189]. Experts also exhibit a larger visual span, allowing them to perceive configurations with fewer glances, and they recognize familiar patterns of piece relations at a glance [168]. These perceptual and

memory-driven advantages illustrate that planning in chess is not a uniform tree search but a knowledge-guided process [35, 71].

1.2.6 A Case Study: Four-in-a-Row

Why Four-in-a-Row: This is a tic-tac-toe variant on a 4×9 grid. Four-in-a-Row has a large state space: approximately 10^{16} possible positions over all game progressions, allowing us to combine the rich behavioral phenomena of real games with computational tractability for modeling. Opheusden et al. [148] developed a computational model that predicts human moves reliably using sophisticated log-likelihood estimation technique and optimization [1, 216]

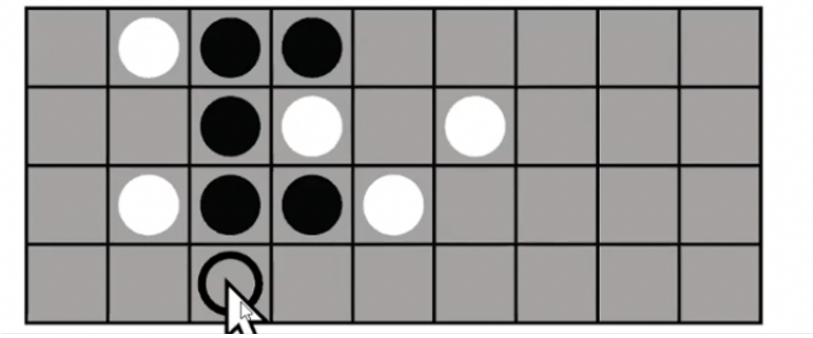


Figure 1.1: Example board of Four-in-a-Row.

The heuristic search model: Opheusden et al. [148] posit that each move arises from a stochastic, heuristic–guided best-first search. Two components specify the model:

(1) **Value function.** A leaf state s is scored by a linear weighted sum of features

$$V_\theta(s) \;=\; \sum_{i=1}^{5} w_i\, \phi_i(s, \text{self}) \;-\; C \sum_{i=1}^{5} w_i\, \phi_i(s, \text{opponent}), \tag{1.1}$$

where the feature vector

$\boldsymbol{\phi}(s, \text{player}) = \big[\text{centrality},\ \text{connected–2},\ \text{unconnected–2},\ \text{connected–3},\ \text{4-in-a-row}\big]$ counts

the number of features in the current state.[1] $\theta = \{w_1, \ldots, w_5, \kappa\}$ are free parameters fitted to data.

(2) **Search and termination.** Starting from the current board, the algorithm iterates: *(a) selection* – choose the leaf node with maximal V_θ; *(b) expansion* – generate all legal children and evaluate each with (1.1); *(c) back-propagation* – replace the parent's value by the maximum value of its children. After every expansion the process halts with probability $\gamma \in (0, 1)$.

If the algorithm stops, each legal move a from the root has an associated backed-up value $Q_\theta(s, a)$. Action selection follows a noisy soft-max rule. A "lapse" parameter mixes in uniform random choice to capture attentional oversight.

Parameter inference. Because the search tree and stopping time are stochastic, the exact likelihood is intractable. The authors therefore maximize an unbiased simulation likelihood estimator using Inverse Binomial Sampling (IBS; [216]) and the Bayesian Adaptive Direct Search optimizer (BADS; [1]).

Full parameter set.

$$\theta = \left\{ w_{\text{center}},\ w_{2\text{–conn}},\ w_{2\text{–uncon}},\ w_3,\ w_4,\ C,\ \vartheta,\ \gamma,\ p_{\text{fd}},\ p_{\text{lap}} \right\}.$$

The first five elements are the feature weights already introduced in Eq. (1.1). The remaining five parameters are:

- **Active–passive scaling factor** C: multiplies all opponent feature weights, reflecting the tendency to weight one's own features relative to the opponent's.
- **Pruning threshold** ϑ: any node whose backed-up value falls below ϑ is discarded, limiting search for clearly losing lines.

[1] $C \in [0, 1]$ scales how strongly the player discounts the opponent's threats.

- **Stopping probability** γ: after each expansion step the algorithm stops with probability γ.
- **Feature-drop rate** $p_{\mathbf{fd}}$: with probability p_{fd} a random feature is omitted during evaluation, introducing stochastic variability that captures the imperfection of human play.
- **Lapse rate** $p_{\mathbf{lap}}$: with probability p_{lap} the final move is chosen uniformly at random, accounting for attentional lapses.

This model provides the analytical backbone for the work that follows and will be revisited throughout the remainder of the dissertation.

Real-world Planning Tasks Laboratory puzzles omit the uncertainty, interruptions, and multiple goal streams of daily life. We frequently engage in planning in far messier real-world contexts where the "rules" are not fully explicit, the environment can change unpredictably, and there may be multiple goals to juggle at once. One example is the Multiple Errands Test (MET) introduced by [186]: participants receive several simple errands (buy bread, mail a letter, meet the experimenter at 11:45) plus rules (enter shops only to purchase, keep track of time). None of the subtasks is hard, yet patients with frontal-lobe damage – who perform normally on tower puzzles – forget errands, break rules, or become sidetracked, exposing deficits in planning, scheduling, and prospective memory.

Variants such as the *Six Elements* and *Hotel* tests require juggling 5–6 easy subtasks within a deadline, again taxing the ability to switch goals, monitor progress, and recover from interruptions. Because the state space of everyday planning is effectively unbounded, success is measured qualitatively: completion rate, time efficiency, and rule adherence.

In conclusion, MET-style tasks bridge laboratory and life: they reveal cognitive demands—multi-goal coordination, time monitoring, error recovery—that simplified puzzles obscure. While the present dissertation concentrates on controlled games that allow precise modeling, any full theory of human planning must ultimately account for performance in such ecologically valid set-

tings.

1.2.7 Methods for Investigating Planning

Planning is an inherently covert process: much of the look-ahead and evaluation occurs without direct behavioral markers. Accordingly, researchers have employed a wide range of methods to capture the cognitive processes that underlie planning and decision-making.

Behavioral data. Behavioral data – final choices, number of moves, reaction times, error rates – provide a scalable way to quantify performance. For instance, on the Tower of London, one can measure planning time (how long a person examines the problem before making the first move) and solution optimality (how close they came to the minimum moves possible). Behavioral data is usually the first layer of analysis and can be statistically compared across groups (e.g., patients vs. controls, experts vs. novices, children vs. adults), but it often lacks fine-grained insight into how people plan.

Process tracing Researchers sometimes supplement these measures with process-tracing methods such as eye-tracking and Neuroimaging techniques (e.g., fMRI, EEG). Eye-tracking provides a window onto the process of planning rather than just the outcome. By monitoring where participants look on a problem display, researchers can infer what elements are currently under consideration and in what sequence.[26, 50, 54, 61, 73, 148, 169, 170]. For example, in chess, eye-tracking studies have shown that expert players fixate on the most relevant pieces and potential move locations in a chess position, essentially zooming in on the critical aspects of the problem. Novices, by contrast, often look in a more diffused pattern and inspect pieces more randomly, reflecting their lack of a clear plan [169, 170]. In the Tower-of-Hanoi, fixations shift to the smallest movable disk or the goal peg just before a subgoal move, exposing the plan's intermediate structure [82]. Neuroimaging methods also allow us to observe the brain in action while people plan.

For example, Balaguer et al. [12] and Lally et al. [114] used fMRI to identify brain regions that correlate with the depth of lookahead or the weighting of prospective outcomes.[50] reported that neural prediction error signals in the ventral striatum were modulated by whether subjects' behavior was model-based or model-free, and the lateral prefrontal cortex showed greater activity in individuals with more model-based choice patterns. Similarly, Vikbladh, Russek, and Burgess [219] reported that specific neural signatures correlate with model-based computations in sequential decision-making. Although these methods can be resource-intensive and require specialized equipment, they enable researchers to link cognitive planning stages to underlying neural mechanisms.

Besides those additional process-tracing measures, two ways to study planning mechanism using behavioral measures alone include enriching the data with verbalization, and using modern modeling methodologies such as log-likelihood estimation and optimization to recover parameters and predict behavior.

Think-aloud. The think-aloud method has a long history in cognitive psychology, tracing back to pioneering work by Newell and Simon [141] and De Groot [51]. Participants verbalize their ongoing thoughts while solving a problem. The resulting verbal protocols are then analyzed to infer the sequence of mental operations.[57]. In [9], the single participant's transcript was parsed line by line and matched to problem states. The researchers could identify points where she set subgoals (e.g., "I need to free up the largest disk") and points where she discovered more efficient methods. They noted when she made evaluative statements (like "that move was a mistake, backtrack") versus planning statements ("first move this, then that"). From such data, one can infer the structure of the internal plan. Think-aloud protocols in chess ([51]'s method) revealed that strong players often verbalize general ideas ("the knight on f5 is strong; maybe sacrifice on g7") rather than enumerating moves, whereas weaker players talk more about specific moves ("maybe move the bishop here, then... no, that loses the rook"), showing that expert advantage arises not

simply from deeper search but from more efficient heuristics. Though think-aloud data offer qualitative insight into the mechanisms of planning, criticisms were raised about the validity of introspection and verbalization (need citation). Subsequent research confirmed that think-aloud protocols can be "nonreactive," meaning they do not necessarily alter performance outcomes with careful instruction [65, 179]. However, its main drawback is cost: coding thousands of utterances is labor-intensive and subject to inter-rater variability.

Computational Modeling. Computational modeling offers a rigorous way to turn informal descriptions of planning into algorithms. The earliest cognitive-inspired planner, the *General Problem Solver* (GPS; [140]), instantiated means–ends analysis decades ago. Although GPS reproduced human step-by-step solutions on problems such as the Tower of Hanoi, there was no principled optimization or parameter recovery. A major methodological leap came with model-fitting techniques. For example, RL models that dissociate model-free from model-based control in the two-step task, a weighting parameter $w \in [0, 1]$ quantifies each participant's reliance on prospective evaluation [50, 52]. For combinatorial tasks such as chess endgames or Four-in-a-Row, human move sequences are often fit with stochastic tree-search models whose parameters (search depth, breadth, pruning threshold) are recovered by maximum-likelihood estimation [148]. A complementary line casts planning as Bayesian inference: the optimal policy is treated as a latent variable, recovered with message passing or sampling in a graphical model [20]. Hierarchical RL further assumes that agents can learn or select temporally extended "options," providing a quantitative account of chunking and abstraction [196]. Recent evidence shows that people do not merely search faster – they strategically simplify the problem they search over. In a series of large behavioral studies, Ho et al. found that humans dynamically compress task representations to reduce planning cost, balancing representational simplicity against policy quality [88]. Resource-rational models make the optimization problem explicit: an ideal planner trades off expected utility against computational cost, allowing fits that reveal when and why humans satisfice

rather than search exhaustively [118].

Fitting such models links algorithmic assumptions—depth of search, pruning thresholds, planning costs—to observable behavior, providing process-level explanations. In Chapter 2, we will explore the extent to which these assumptions capture the true mechanisms of human planning.

1.2.8 Cognitive Abilities

What cognitive machinery allows people to orchestrate multi-step behavior? Consider organizing a multi-day trip across an unfamiliar city: to decide on the optimal schedule, one must mentally simulate potential outcomes (e.g., train schedules, museum hours), juggle multiple constraints in memory, and resist impulsive decisions that might seem convenient in the moment but conflict with overall goals. Previous work has examined the contributions of several cognitive abilities, such as working memory, inhibitory control, and fluid intelligence, to planning behavior using the individual differences methodology[159, 212, 223, 235].Individual-differences methodology examines how variation between people on one measure can be statistically explained by their variation on other measures. The typical approach is to administer a planning task and a battery of tests to the same participants, then correlate (or regress) each participant's score on one task with their scores on other tasks.

Working Memory. Working memory is the capacity to temporarily maintain and manipulate information [11]. It plays a critical role in multi-step planning because each step in a plan depends on retaining various goals, rules, and intermediate states. In classic planning tasks such as the TOL, performance correlates significantly with measures of working memory span [56, 68, 102, 152, 206, 212, 223], suggesting that individuals with higher working memory capacity can more effectively track evolving task states while holding strategic subgoals in mind.

Inhibitory Control. Inhibitory control is the ability to suppress impulsive or habitual actions that conflict with a current goal [53, 135]. In tasks like the TOL, participants must resist seemingly intuitive moves that would ultimately hinder reaching the goal configuration in the fewest steps [122, 206, 223, 235]. These data imply that inhibitory control functions as a gate that prunes implausible actions before they enter deeper search.

Fluid Intelligence. Fluid intelligence is the ability to reason over novel relations and generate abstract representations [33]. Within the context of planning tasks, fluid intelligence often helps infer which action sequence best satisfies multiple constraints. Individuals who excel in standard tests of fluid intelligence also tend to perform better on multi-step puzzles such as the TOL and Two-Step Task [212, 235]. This suggests that abstract reasoning supports the ability to identify efficient strategies and integrate multiple information sources into a coherent action plan.

Spatial Abilities Mental rotation [188] – the capacity to mentally manipulate spatial representations – has been theoretically linked to planning tasks that require envisioning future states [36]. Similarly, pattern detection – the rapid extraction of recurring regularities in spatial or temporal arrays – may support the identification of advantageous configurations or sequences in spatially complex planning scenarios [34, 51, 173, 205].

To date, much of the evidence for these cognitive underpinnings comes from relatively simple paradigms with limited state spaces – the Tower of London and the Two-Step Task [50, 184]. Many scenarios like strategic board games or scheduling itineraries involve massive state spaces where possible action sequences grow exponentially. Whether the same cognitive components that underpin simple planning tasks also support more complex, combinatorially rich forms of planning remains an open question. Chapter 1 will address this question.

1.3 Planning in Machines: Models, Searches, and Learning

In artificial intelligence, an agent must compute a sequence of actions that maximizes rewards in an environment. The environment is typically formalized as a Markov decision process (MDP) or, when observability is partial, a POMDP. Given a transition model , planning becomes a search or optimization problem over the (often enormous) state–action graph. Three questions organize the algorithmic landscape: How are future states simulated? How is the next action selected? How is computation apportioned under resource limits? The remainder of the section reviews the major families of answers, moving from classical symbolic methods to modern neural and language-based planners.

1.3.1 Classical Symbolic Planning

Early AI assumed deterministic, fully observable worlds. STRIPS represents each action by logical pre-conditions and add/delete effects; planning reduces to finding a sequence of operators that transforms the initial predicate set into the goal set [63]. Generic graph-search algorithms—depth-first, breadth-first, Dijkstra, A*—supply the control, while domain-independent heuristics (e.g., relaxed-plan and delete-relaxation estimates) scale A* to thousands of steps. The Planning Domain Definition Language (PDDL) and planners such as GraphPlan, SATPlan, and Fast-Forward formalised competitions that steadily improved heuristic design.

1.3.2 Adversarial Game-Tree Search

Two-player, perfect-information games introduce an adversary. Here, the planner (player) must account for an opponent's counteractions. Minimax search is the foundational algorithm, where one alternates between a "max" player trying to maximize the evaluation score and a "min" player trying to minimize it. However, minimax by itself is expensive if the game tree is large.

Alpha-beta pruning improves efficiency by skipping over branches that cannot affect the final minimax decision [106], thereby reducing the number of expanded nodes. This pruning works by maintaining upper and lower bounds (α and β) on the possible outcomes. If a branch's best achievable score is worse than the current α or β, further exploration of that branch is cut off. IBM's *Deep Blue* scaled the same recipe to2×10^8 positions per move and defeated the world champion in 1997 [28].

1.3.3 Sampling–Based Search: Monte-Carlo Tree Search

Monte-Carlo Tree Search (MCTS) trades exhaustive enumeration for *selective sampling*. Instead of expanding every child of every node, it grows the tree where past simulations suggest value is highest, while still reserving some probability for exploration. This character makes MCTS the planner of choice for domains whose branching factor explodes.

UCT: multi-armed bandits in a tree. At each internal node, the uct rule treats every legal action as a bandit arm and chooses

$$a^\star(s) \;=\; \arg\max_a \Big[\underbrace{\bar{Q}(s,a)}_{\text{exploitation}} + \underbrace{c\sqrt{\frac{\ln N(s)}{N(s,a)}}}_{\text{exploration}} \Big],$$

where $N(s)$ is the visit count of state s and $\bar{Q}$ the empirical mean return [107]. The square-root bonus guarantees that the regret of picking sub-optimal moves grows only $O(\log N)$ with the number of simulations, and the tree can be queried after *any* number of iterations for its current best action.

Modern refinements. Go engines and general RL agents boost vanilla UCT with two key ideas:

- **PUCT.** Replace the exploration bonus by a learned policy prior $\pi_\theta(a \mid s)$, focusing roll-outs on moves a neural network already believes are promising [191].
- **Progressive widening.** In huge or continuous action spaces, add children gradually rather than all at once [47].

Coupled with deep networks that supply *policy* priors and *value* estimates, these refinements powered AlphaGo/Zero and MuZero to superhuman play while expanding $\sim 10^2$–10^3 times fewer nodes than alpha–beta search. Chapter 4 dissects how much of this performance gain comes from the network versus the search.

1.3.4 Learning to Guide Search.

Deep networks can supply the prior policy and state evaluations that MCTS needs, replacing naïve roll-outs with *neural roll-outs*. *AlphaGo* first demonstrated the power of this marriage, beating human professionals in Go [191]. *AlphaGo Zero* and the more general *AlphaZero* removed the human data entirely: self-play games are analyzed by MCTS, and the resulting move counts and outcomes supervise the next network update. Repeating this policy-iteration loop yields super-human play in Go, chess, and shogi [191, 192]. This synergy of learning and search is commonly viewed as a *model-based* approach, since the agent uses an internal simulator (MCTS) to plan moves. Yet after training, the network alone can play quite strongly even without deep tree expansions [78], suggesting a partial shift toward "model-free" execution once knowledge is internalized.

These achievements invite questions about how AlphaZero's learning compares to human learning: What makes AlphaZero a superhuman model? Later, we will explore the interplay of knowledge and lookahead in both humans and AI.

1.3.5 Transformer

The rise of the transformer reframed planning as sequence modeling. Decision Transformer treats return-conditioned trajectories as text and performs autoregressive hindsight planning [38]. Gato unifies language, vision, and control tokens in a single foundation model, while Latent Plan Transformer abstracts trajectories into high-level plans before decoding low-level actions. Diffusion and energy-based planners (e.g., Diffuser, Trajectory Transformer) generate smooth action sequences in continuous domains. Transformer architectures underpin today's most powerful language and vision models, yet they also provide a useful mental model for *how* an agent might integrate information over extended sequences when planning.

Self-Attention. A transformer processes an input sequence $x_{1:T}$ using stacked layers of *multi-head self-attention*. Each token first projects to *query* (Q), *key* (K), and *value* (V) vectors. Attention weights are computed as

$$\text{Attn}(Q, K, V) \;=\; \text{softmax}\Big(\frac{QK^\top}{\sqrt{d_k}}\Big)\, V$$

allowing every position to attend to every other, modulated by content similarity rather than fixed distance. Because this operation is repeated in parallel heads, the network can capture diverse dependency patterns—short-range syntax and long-range semantic links—within a single layer.

Positional Encoding. Unlike recurrent networks, transformers lack an inherent notion of order. Positional encodings—either sinusoidal functions or learned vectors—are therefore added to the token embeddings so that the model can tell "first move" from "last move."

Autoregressive Decoding. In a *decoder-only* transformer (e.g. GPT-style models) the network is trained to predict the next token given all previous tokens, subject to a causal mask that prevents peeking into the future. At inference time, this supports step-by-step generation: feed the context, sample a next token, append, and repeat.

Why Transformers for Planning. Two properties make transformers attractive cognitive and algorithmic models of planning:

1. **Long-range integration.** Self-attention connects every time-step to every other, letting the model ground each decision in arbitrarily distant context without the vanishing-gradient problems that hamper RNNs

2. **Parallel evaluation.** Because attention acts on the entire trajectory in a single pass, a transformer can simultaneously weigh alternative continuations while conditioning on the complete action history.

Chapter 3 exploits these properties by training *GPT-4IAR*, a transformer that predicts human Four-in-a-Row moves from extended histories.

Planning in LLMs. Although not the focus of this dissertation, it is worth noting that Transformer-based LLMs excel at text generation yet lack explicit look-ahead mechanisms. Trained on massive corpora, these models display impressive language understanding and fluency, but do not inherently plan in the sense of searching through explicit state–action sequences. Prompting strategies—chain-of-thought [222], self-consistency [221], and ReAct [228]—coax them to reveal intermediate reasoning. External control structures such as Tree-of-Thoughts guide breadth-first or MCTS-style exploration in the LLM's latent space. Other work incorporates explicit search procedures, such as Monte Carlo-like branching in a "Tree of Thoughts" [229] or partial world models [79]. While these approaches are still in development, they show promise in allowing LLMs to reason more systematically about future states, bridging purely reactive text generation with the deliberative search found in classical planners.

The parallel developments in cognitive and AI approaches to planning invite a convergence: Can insights from human planning improve AI planners, and can AI techniques, in turn, illuminate the intricacies of human behavior and refine cognitive theories? This dissertation sits at that intersection.

1.4 Overview of This Dissertation

This dissertation explores how humans and machines plan, using *Four-in-a-Row* as a tractable yet sufficiently complex model domain. The chapters build cumulatively: from identifying what cognitive components constitute planning, through analyzing how plans are verbalized, to modeling long-range dependencies in human planning with transformers, and finally to dissecting what state-of-the-art RL agents – AlphaZero – learns and misses during planning. The concluding chapter knits the strands together.

- **Chapter 1: Cognitive Components of Human Planning** – We begin by asking what constitutes planning. To uncover the cognitive components of planning, Chapter 1 reports an individual-differences study (N = 476) using three planning paradigms spanning fifteen orders of magnitude in state-space complexity – Two-Step, Tower-of-London, and Four-in-a-Row – against a battery of basic cognitive ability tasks. Factor and LASSO analyses revealed three-factor structure: (i) visuospatial processing, (ii) working memory, and (iii) inhibitory control. Different planning tasks load onto different factors, reinforcing that human planning is not a unitary construct but rather an orchestration of partially dissociable abilities that shift with complexity.

- **Chapter 2: Do Humans "Think in Trees"?** – Having established what abilities underpin planning, this chapter examines how plans are formed moment to moment with think-aloud data. Thirty-four participants solved Four-in-a-Row puzzles while verbalizing their thought processes. More than 4,000 coded statements reveal a heterogeneous mix of strategies – ranging from canonical tree-like searches to shallow heuristics. The results suggest 1) strategic diversity: not all participants rely on deep or systematic search; 2) weak alignment between verbalized depth and model-inferred depth: a single fixed search template may miss nuances of human planning, motivating richer behavioral proxies for

modeling planning in addition to final choices or reaction times.

- **Chapter 4: Learning human gameplay with action history** – Shifting focus from process to prediction, Chapter 4 evaluates whether human planning strategies can be approximated by a sequence model conditioned on move history. Training a transformer (GPT-4IAR) on ten million human Four-in-a-Row games, we find that conditioning on up to 90 prior moves yields +7% improvement in move prediction over state-of-the-art Markovian models, and a 3.6% reduction in log-likelihood per move. These gains challenge the assumption that humans plan solely from the *current* board state. Instead, long-horizon context substantially shapes upcoming moves. GPT-4IAR therefore sets a new behavioral ceiling: any process-level cognitive model that ignores long-term strategies now has a tangible gap to close.

- **Chapter 5: What AlphaZero Learns and Misses about Planning** – Finally, Chapter 5 reverses the lens to dissect how a superhuman RL planner, AlphaZero, acquires and applies expertise in Four-in-a-Row. Our experiments showed that (i) Policy quality drives playing strength gains more than value quality or increased search depth. (ii) the network spontaneously recovers human-interpretable features such as 3-in-a-row but not weaker precursors such as 2-in-a-row; and (iii) the agent fails spectacularly on forced-win puzzles that demand reasoning chains. Augmenting AlphaZero's value head with human-inspired features partially mitigates these failures. The results suggest directions for improving deep RL models with insights from human planning while generating hypothesis for improving cognitive modeling with policy prior.

- **Chapter 6: Conclusion** – The final chapter synthesizes insights across levels of analysis. It (i) maps where human and machine planning converge and diverge, (ii) spells out theoretical implications for multi component models of planning, (iii) reflects on methodological contributions—from large-scale individual-differences designs to transformer-based behav-

ioral emulators, (iv) discusses practical limitations and open questions, and (v) outlines a bi-directional research agenda in which cognitive data inform AI algorithms and AI failures inspire new cognitive experiments. By continuing to learn from each other – cognitive science learning from AI successes, and AI adopting cognitive insights – we move closer to artificial planners that not only match human performance in narrow domains but also approach the versatility and robustness of human planning in the real world.

2 What are the cognitive components of planning?

2.1 Introduction

Understanding human intelligence requires identifying the cognitive abilities that underlie complex, goal-directed behaviors. One form of intelligence is planning. Planning pervades everyday activities, from solving multi-step math problems, furnishing a space, and organizing events, to strategizing during board games. Even writing, which involves creating sequential statements that lead to effective transmission of a message, can be thought of as a planning problem [64, 81]. Effective planning requires the integration of several cognitive processes such as working memory, mental simulation, inhibitory control, and abstract reasoning[16, 68, 159]. Consider, for example, planning a wedding: one must mentally simulate the overall theme and venue, manage timelines and budgets, hold logistical details (e.g., guest lists, catering menus) in working memory, and revise decisions in response to new constraints such as vendor availability or weather forecasts. Such real-world examples illustrate that planning often requires the coordinated use of diverse cognitive processes.

Despite the ubiquity of planning, the cognitive processes that underlie planning ability have been studied only to a limited extent. One limitation has been the field's reliance on simple tasks to study planning, which implicitly treats planning as a unitary construct. Commonly used

paradigms include the Tower of London (TOL) and the Two-Step Task. The TOL involves rearranging items to match a predetermined configuration and is widely utilized in clinical populations to assess planning deficits[41, 117, 213]. Studies using the TOL highlight contributions from working memory, inhibitory control, and fluid intelligence [68, 151, 158, 212, 235]. The Two-Step Task is a minimalistic sequential decision-making paradigm designed to reveal the interplay between goal-directed (model-based) and habitual (model-free) decision strategies [50]. Model-based behavior has been associated with planning [60, 131], although this characterization remains debated [4, 44]. Studies using this paradigm have connected model-based strategies with cognitive abilities such as working memory and fluid intelligence. [56, 151, 161, 236].

These paradigms typically involve small state spaces with limited sequential complexity. As a result, it is unknown whether the cognitive mechanisms at play in simple forms of planning generalize to more complex situations with "combinatorial complexity", where the number of possible outcomes grows exponentially with each decision. Such complexity quickly surpasses the capacity for exhaustive mental exploration and is typical in tasks that challenge artificial intelligence systems, including strategic games like chess and Go [8, 177, 187]. Common activities such as planning travel itineraries also quickly become combinationally complex. Each decision—such as choosing destinations, times, or transportation—exponentially expands the subsequent possibilities. By contrast, how biological intelligence tackles tasks with combinatorial complexity is poorly understood. Recent efforts have thus begun exploring human planning through larger state-space board games. One example is the game Four-in-a-Row –a two-player strategic game analogous to tic-tac-toe on a 4x9 board – which has been used to demonstrate how depth of planning changes with expertise, and to understand how memory improves with planning [92] and to characterize clinical deficits [90]

The current work uses an individual-differences approach to clarify whether planning should indeed be considered a unitary construct or rather a collection of different cognitive components. Burgess et al. [25] have argued that planning is best understood as comprising multiple subpro-

cesses (e.g., mental simulation, retrospective and prospective memory, and inhibition) rather than a unitary construct. Building on this perspective, we used a combinatorial complex planning task to examine which basic cognitive abilities underpin complex planning, and to what extent do different planning paradigms share common cognitive mechanisms? Addressing these questions is essential for clarifying how cognitive abilities coordinate to support complex, goal-directed behavior. We hypothesize that as complexity increases, individuals may require more extensive mental simulations, enhanced working memory capacity for maintaining extended action sequences, and potentially greater reliance on domain-specific heuristics.

To investigate these questions, we administered to the same participants the above-mentioned planning tasks as well as basic cognitive tasks measuring working memory, inhibitory control, mental rotation, pattern detection, and fluid intelligence. This allows us to examine relationships among planning abilities as well as between planning abilities and basic cognitive abilities.

2.2 Methods

Participants

We recruited 568 participants via Prolific, requiring them to be at least 18 years old and fluent in English. To ensure robust power for factor analyses, we aimed for a sample of at least 300 [123], with a target of 480 participants that can complete all sessions. Of the 568 participants, 88 failed to complete all four sessions, and 4 additional participants reporting technical issues that interrupted their experiments were excluded. The final sample included 476 participants.

Experiment

The experiment consisted of four sessions conducted online through Prolific. The initial session includes demographic information and surveys, including the Future Orientation Scale [199],

and the Barratt Impulsiveness Scale [13]. Cognitive tasks were spread across three roughly equal-length subsequent sessions. Each session included the same set of tasks, but the order of the sessions, as well as the order of tasks within each session, were randomized.

Tasks

Four-in-a-Row: Four-in-a-Row is a two-player game similar to Tic-Tac-Toe. It is played on a 4×9 board. Players alternate turns placing pieces, attempting to form an uninterrupted row of four (horizontally, vertically, or diagonally). This paradigm features a significantly larger state space (approximately $1.2 \cdot 10^{16}$) than TOL or the Two-Step Task [148]. Participants played 40 games. We used Elo rating as performance metric (See Elo Estimates).

Tower of London (TOL): Participants moved colored balls on three pegs from an initial arrangement to match a specified target configuration in the minimum number of moves. Each arrangement has an optimal number of moves, and the balls can only be moved one at a time from peg to peg. Our 25 puzzles have optimal move ranging from 3 to 7 (5 puzzles for each level). We used number of optimally solved puzzles as performance metric.

Two-step Task: This is a sequential decision-making paradigm designed to disentangle model-based from model-free strategies in humans [50]. Following the implementation of Decker et al. [52] and Nussenbaum et al. [146], participants chose between two "spaceships" (first step), leading probabilistically to one of two "planets" (second step), where participants selected between two "monsters" for a potential reward. Participants completed 80 trials, separated into four blocks of 20 trials. We used model-based weight as performance metric by fitting a reinforcement learning model described in Daw et al. [50], Nussenbaum et al. [146], and Otto et al. [150].

Corsi Block-Tapping Task (Corsi): This task is a standard measure of spatial working memory. Participants viewed an arrangement of blocks on the screen, which highlighted a sequence of positions. Their task was to reproduce each sequence in the same order, with sequence length ranging from 2 to 9 [14, 45]. Participants completed 2 trials for each sequence length. We used

Corsi score, total number of correctly reproduced sequences, as performance metric.

Change Detection Task (CDT): Adapted from Brady and Tenenbaum [21], participants briefly viewed an array of objects for 750 ms, which then disappeared. After a 1000 ms delay, either the same array or a slightly altered one reappeared for 750 ms, and participants indicated whether a change had occurred. This task measures how many visual items can be simultaneously maintained with sufficient precision to detect changes, while the Corsi task measures sequential processing aspects of visuo-spatial working memory, requiring participants to reproduce increasingly lengthy spatial sequences. Participants completed 48 trials. We quantified performance using d-prime (d'), a signal detection theory metric calculated as d' = z(hit rate) - z(false alarm rate) [124]. This measure of sensitivity provides a more robust performance index than accuracy alone, with higher values indicating better discrimination between changed and unchanged arrays [198]

Wisconsin Card Sorting Task (WCST): A measure of cognitive flexibility and inhibitory control, requiring participants to sort cards by color, shape, or number without explicit instructions. After a run of correct responses, the sorting rule is changed without warning, and the participant must infer the new rule based on feedback (correct vs. incorrect). Participants completed 64 trials. Performance was quantified using perseverative errors, which occur when participants continue to sort according to a previously correct rule despite negative feedback. This metric is widely used as an indicator of cognitive flexibility and executive function, with fewer perseverative errors indicating better performance [15, 85]. We transformed the perseverative error count by negation (multiplying by -1) to create a metric where higher values represent better performance, consistent with our other cognitive measures.

Mental Rotation Task: This task assesses an individual's ability to mentally manipulate objects in space [188, 217]. During each trial, participants viewed two three-dimensional figures and decided, via a two-alternative forced choice (2AFC), whether the figures were identical apart from rotation or if one was a mirror image of the other. We used stimuli from Ganis and Kievit

[67]. Participants completed 96 trials in total. Performance is measured through d'[74].

Standard Progressive Matrix (SPM): This task measures abstract reasoning and fluid intelligence. Participants completed 2D pattern matrices by selecting the missing segment from multiple choices. Each matrix follows a pattern or a rule, which the participant must reason to identify the missing piece. Participants completed 60 questions in total. We used SPM score, the number of correctly solved problems, as performance metric [166, 167]

Pattern Detection Task In this newly developed measure of visual pattern detection, participants inspect a 10×10 board with black and white pieces placed randomly. They must decide whether any four identically colored pieces form a continuous row (horizontally, vertically, or diagonally) through 2AFC. Each trial features 12 to 70 pieces, with no guarantee of color balance. Half of the trials contain a valid four-in-a-Row configuration, while the other half do not. Participants complete 78 trials in total, and performance is quantified using d'.

Elo Estimates

To evaluate a player's playing strength based on their performance against computer agent opponents, we used the Elo rating system using the open-source software Bayeselo[55, 94, 232]. To improve the estimate, we conducted a tournament among computer agents, where each agent played 230 games against every other agent, including itself. For efficiency, we grouped the original 200 computer agents in Opheusden et al. [148] into bins of 10, with each bin consisting of agents having similar Elo ratings. This approach reduced the total number of games required without compromising the accuracy of our evaluations. All games involving human players versus computer agents, as well as computer versus computer games, were combined into a single dataset as input to Bayeselo.

Split-half Reliability

We employed split-half reliability analysis to assess the internal consistency of cognitive measures. For all tasks, trials were divided into odd-numbered and even-numbered trials, creating two equivalent test halves. For each metric, we calculated the Pearson correlation coefficient between performance on odd and even trials across participants. The resulting correlation represents the reliability of a half-length test. To estimate full-test reliability, we applied the Spearman-Brown prophecy formula [23, 197]. This correction compensates for the reliability reduction inherent in test shortening, providing a more accurate estimate of the measure's true reliability.

Correlation Analysis

To investigate correlations among task performance measures while controlling for multiple comparisons, we employed a permutation testing approach with 10,000 iterations, which is commonly employed in fMRI but less standard in individual differences research. For each permutation, we randomly shuffled the data along the subject dimension to break any true correlations between variables while preserving the distribution of each variable. We then calculated correlations between all pairs of variables and recorded the maximum correlation value for that permutation. This process was repeated 10,000 times to generate an empirical null distribution of maximum correlations that could occur by chance. This approach avoids overly conservative corrections like Bonferroni and provides robust control of the familywise error rate [58, 143].

Factor Analysis

We conducted exploratory factor analysis to identify the latent structure underlying performance across cognitive tasks. Prior to analysis, we standardized all variables to have a mean of 0 and standard deviation of 1. To verify the adequacy of our data for factor analysis, we calculated the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy and performed Bartlett's

test of sphericity. A KMO value closer to 1.0 indicates a more favorable factorable dataset [103], and Bartlett's test assesses whether correlations among variables are sufficiently large for factor analysis [13].

The number of factors to retain was determined using multiple criteria, including the Kaiser criterion (eigenvalues $>$ 1) and the cumulative percentage of variance explained (aiming for 50–60% coverage), and interpretability[77, 157, 224]. Although parallel analysis is often recommended as a robust method for determining the number of factors [91], methodological research indicates that with small variable number (fewer than 10-20 variables), all statistical retention methods become less reliable [59, 163]. Specifically, with only 9 variables as in our dataset, parallel analysis can become overly conservative, potentially underestimating smaller but meaningful factors [175]. As Preacher and MacCallum [163] note, when the number of variables is small, researchers should rely more on theory and interpretability rather than automated retention methods. Following these recommendations, we prioritized variance-explained thresholds (50-60%) and factor interpretability as the primary criteria for factor retention, which is consistent with best practices for exploratory factor analysis with small variable sets [224, 226, 237].

Factor loadings were evaluated according to the guidelines in Hair et al. [77], where loadings of approximately ±0.30 are considered minimal, ±0.40 are important, and ±0.50 or higher indicate practical significance. This approach ensured that retained factors were both statistically robust and interpretable in light of the literature.

Regression

For each planning measure, we employed LASSO regression with cross-validation to identify significant cognitive predictors while controlling for multicollinearity. All variables were standardized prior to analysis to facilitate direct comparisons and enhance model interpretability. We regressed each planning measure on the full set of cognitive predictors (detection d', rotation d', WCST perseverative errors, Corsi span, pattern detection d', and matrix reasoning) using LASSO

regression with cross-validated alpha parameter selection. The LASSO approach automatically performs feature selection by shrinking less important predictor coefficients to zero, effectively removing them from the model. For robust determination of the regularization strength, we implemented a 5-fold cross-validation procedure using a logarithmic grid of potential alpha values ranging from $10^{-}4$ to 10^{1}. The alpha value that minimized the mean squared error across validation folds was selected for the final model.

Additionally, we examined whether there remained any stable, systematic variance in planning performance unaccounted for by our cognitive measures, following the approach of Mitko and Fischer [133]. We conducted split-half reliability analyses on regression residuals by dividing each participant's task performance into odd and even trials, creating two separate measures for each planning task. We then regressed our cognitive predictors on each half-measure separately and computed correlations between the resulting residuals to assess the reliability of variance not explained by our cognitive measures. To further investigate whether the unexplained variance represented the same construct across different planning tasks, we conducted cross-task residual correlation analyses. For each pair of planning tasks, we examined the correlation between Half 1 residuals of Task A with Half 2 residuals of Task B, and vice versa. Strong cross-task residual correlations would suggest a common underlying planning ability not captured by our cognitive measures.

2.3 Results

To characterize the cognitive components underlying human planning and clarify the relationships between planning abilities in tasks of varying complexity, we administered a battery of planning and basic cognitive tasks to 476 participants recruited via Prolific. Participants' ages ranged from 18 to over 64 years, with a balanced gender distribution: 48.7% identified as male, 48.6% as female, 2.5% as non-binary, and 0.1% preferred not to disclose their gender.

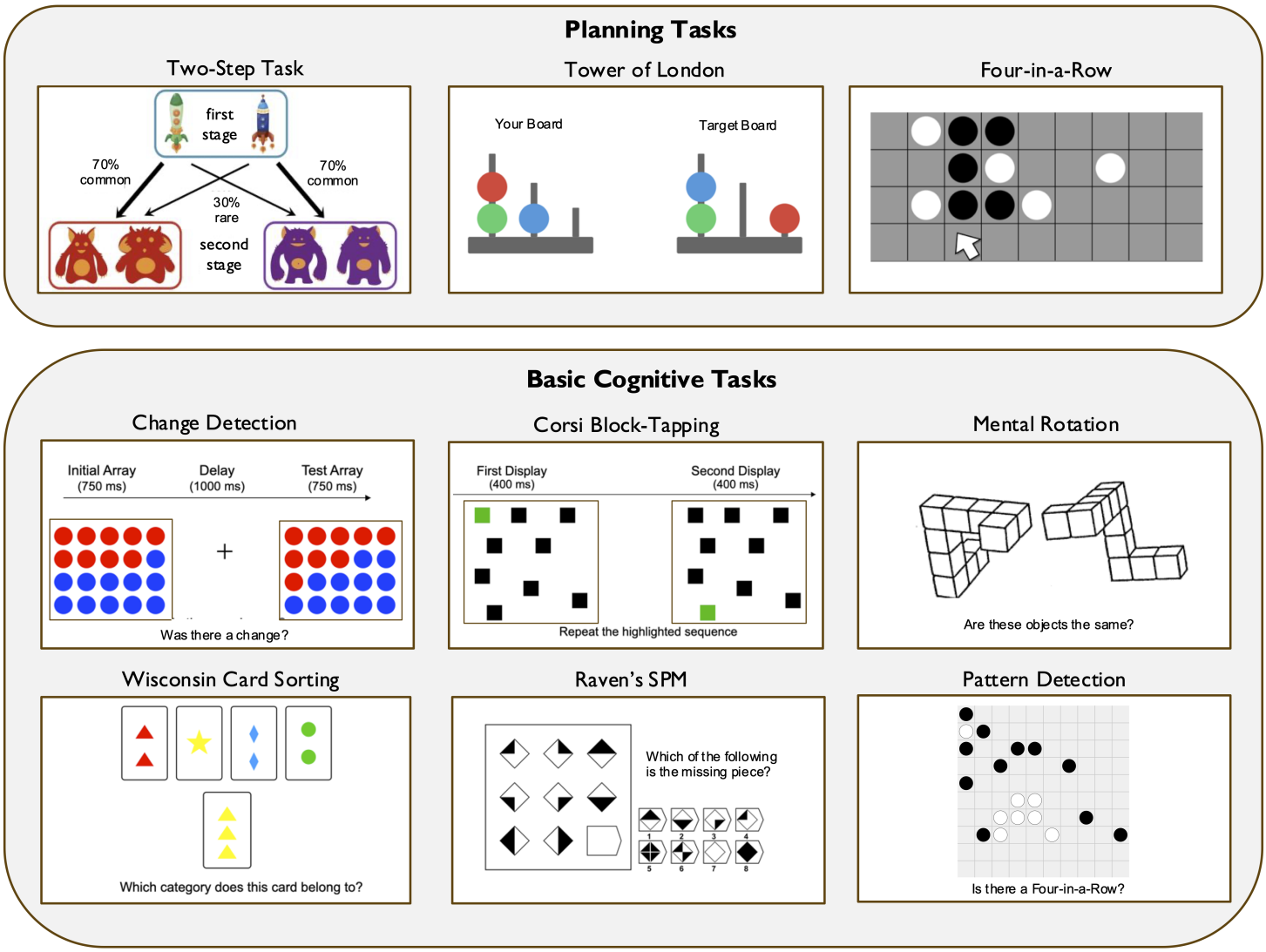


Figure 2.1: The figure displays nine cognitive tasks organized into two categories: Planning Tasks (top) and Basic Cognitive Tasks (bottom). Planning Tasks include the Two-Step Task, Tower of London, and Four-in-a-Row. Basic Cognitive Tasks include Change Detection, Corsi Block-Tapping, Mental Rotation, Wisconsin Card Sorting, Raven's Standard Progressive Matrices, and Pattern Detection.

Participants completed three planning tasks: the Two-Step Task, Tower of London (TOL), and Four-in-a-Row (FIAR). In the Two-Step Task, participants navigated two sequential stages through binary choices to accumulate rewards. This task differentiates between model-based and model-free decision-making strategies [50, 52, 146]. Model-based strategies involve actions guided by an internal cognitive model of the environment, characteristic of planning behavior [60, 131], but some researchers argue it may just reflect sophisticated habit formation [4, 44]. We quantified performance by computing a model-based weight through reinforcement learning model [150]. Participants demonstrated a mean model-based weight of 2.17 (SEM = 0.046). The Tower of London task involved rearranging colored balls placed on three pegs from an initial configuration to match a target arrangement [185]. We computed a weighted performance score by multiplying each solved puzzle by the number of minimal steps required and summing up these trial scores, yielding an average weighted score of 56.85 (SEM = 0.83). Four-in-a-Row, a combinationally complex game similar to tic-tac-toe played on a 4-by-9 grid, involved players aiming to form four uninterrupted pieces horizontally, vertically, or diagonally [148]. We used Elo ratings to quantify performance (See Methods). Participants achieved an average Elo rating of 77.09 (SEM = 5.2).

Additionally, participants completed six basic cognitive tasks hypothesized to support planning processes. The Corsi Block-Tapping Task evaluated spatial sequential working memory, where participants reproduced sequences of visually highlighted blocks [14, 45], with performance quantified as the Corsi span multiplied by the number of correctly solved trials. Participants achieved a mean Corsi score of 53.54 (SEM = 1.1). The Change Detection Task (CDT), adapted from Brady and Tenenbaum [21], assessed visual working memory precision for detecting subtle changes in visual arrays, yielding a mean d' of 1.80 (SEM = 0.037). [74, 198]. We evaluated inhibitory control with the Wisconsin Card Sorting Test (WCST), which required participants to flexibly adapt sorting rules [15, 85]. We quantified performance using the number of perseverative errors, which we negated for consistency (higher values indicating better perfor-

mance). Participants showed a mean negated perseverative error score of -2.45 (SEM = 0.17). We measured participants' ability to mentally simulate and manipulate three-dimensional objects using the Mental Rotation Task [188, 217], producing a mean d' of 2.16 (SEM = 0.055). Participants achieved a mean d' of 2.78 (SEM = 0.031) in the newly introduced Pattern Detection Task designed to measure visual pattern identification ability. Lastly, we assessed fluid intelligence using the Standard Progressive Matrices (SPM) which involves abstract reasoning to identify missing elements in patterned displays [166, 167]. which yielded a mean correct score of 46.09 (SEM = 0.38). Together, these tasks cover cognitive processes potentially underpinning planning performance.

To ensure our planning and cognitive tasks reliably measure individual differences, we first assessed the internal consistency of task performance. We computed split-half correlations between even and odd trial subsets for each task. Reliability values ranged from 0.50 to 0.86, with the exception of the Two-Step Task ($r = 0.30$). Reliability values for tasks that have been used in the literature were comparable. The relatively lower reliability of the Two-Step Task is consistent with the previously reported low model-based weight in Two-Step Task across individuals[22]. Full reliability table in 6. Collectively, these findings suggest that the most tasks we measure are sufficiently stable for subsequent individual differences analyses.[40].

Correlation Analysis

To investigate the inter-task relationships, we computed Pearson correlations among all tasks. Self-report surveys (Future Orientation Scale; Barratt Impulsiveness Scale) were excluded because they were uncorrelated with every performance measure (See 6

To control for multiple comparisons, we generated 10,000 permutations by randomly shuffling participant labels, recomputed the full correlation matrix on each iteration, and built the empirical null distribution from the maximum absolute coefficient in every permutation [144, 225].This produced permuted correlations representing chance-level associations. We determined statisti-

cal significance when an observed coefficient's absolute value exceeded the 95th percentile of this null. Table 2.1 displays pairwise correlations among the nine tasks.See 6 for exact p values.

Table 2.1: Correlation table of task metrics

	Non-Planning Tasks						Planning Tasks		
	SPM	Corsi	Rotation	WCST	CDT	Pattern	TOL	Two-Step	FIAR
Raven's SPM		0.325***	0.591***	0.295***	0.323***	0.409***	0.394***	0.233***	0.344***
Corsi			0.269***	0.156*	0.416***	0.242***	0.215***	0.141	0.355***
Mental Rotation				0.252***	0.324***	0.325***	0.392***	0.184**	0.213***
WCST					0.181**	0.136	0.221***	0.179**	0.187**
Change Detection						0.234***	0.249***	0.153*	0.321***
Pattern Detection							0.230***	0.136	0.328***
Tower of London								0.166*	0.280***
Two-Step Task									0.185**
Four-in-a-Row									

Significance levels:

* $p \leqslant 0.05$ ($|r| \geqslant 0.146$) ** $p \leqslant 0.01$ ($|r| \geqslant 0.167$) *** $p \leqslant 0.001$ ($|r| \geqslant 0.195$)

Between non-planning tasks Fluid intelligence (Raven's SPM) correlated positively with every other cognitive measure in our battery. The correlations with visuospatial working memory tasks were moderate – Corsi: $r = .325$; Change–Detection: $r = .323$ – and sit close to the meta-analytic mean for simple span measures ($r \approx .28$) [2]. Primary studies span a wide range: $r \approx .27$–$.34$ for forward and backward digit-span tasks in school-age samples [100], and about $r = .40$ for backward digit span in undergraduates [42]. Thus, our observed correlations fall squarely within the expected bandwidth for expected correlation between fluid intelligence and simple-storage WM tasks.

SPM showed a strong association with Mental Rotation ($r = .591$), slightly exceeding the average G_v–G_f correlation reported in a recent meta-analysis ($r \approx .52$) [24] and consistent with earlier individual study estimates (e.g., [99] – what are the numbers). The moderate correlation between SPM and Pattern Detection ($r = 0.409$) aligns with prior evidence rapid recognition of novel visual patterns correlate significantly with fluid-intelligence test scores(e.g., mental-rotation slope

$r = -.29$, inspection time $r = -.34$ with Raven's APM) [205]. Finally, SPM scores and (negated) WCST perseverative errors were modestly related ($r = 0.295$), in line with the literature linking set-shifting to fluid intelligence at around $r \approx 0.27–0.32$ [108].

Working-memory tasks also correlated with Mental Rotation (Corsi: $r = .269$; CDT: $r = .324$). Adult studies report comparable or stronger links, spanning $r \approx 0.26–0.45$ – for example, $r = .26$ between Operation-Span and Mental Rotation [154], $r = 0.45$ for Rotation-Span versus Mental Rotation[183], and a correlations $r = 0.33$ between Corsi and Rotation [134]. These converging results suggest visuospatial working memory supports the active spatial transformations required for mental rotation. Both WM tasks showed a modest association with Pattern Detection (Corsi: $r = 0.242$; CDT: $r = 0.234$), comparable to reported correlations around $r = 0.38$ in pattern-recognition memory tasks [209]. We observed weak correlations between perseverative errors and both Corsi ($r = 0.156$) and CDT ($r = 0.181$). The literature is mixed – reports range from near-zero [202] to moderate ($r \approx 0.30–0.50$; [230]) – and recent reviews stress large between-study heterogeneity [108]. Mental Rotation correlated moderately with Pattern Detection ($r = 0.325$), echoing correlation finding in Miyake et al. [134] ($r = 0.33$), and factor-analytic evidence that both tasks load on the same factor [30, 86, 101, 126]. Finally, the WCST–Pattern-Detection association was weak ($r = 0.136$), consistent with findings that perseverative errors reflect executive shifting rather than visual pattern detection [108].

Between non-planning tasks and planning tasks In line with previous studies showing that higher fluid intelligence benefits Tower-of-London and strengthens model-based control in the Two-Step task [68, 69, 212, 235], we observed parallel links in our data: Raven's SPM correlated $r = 0.39$ with TOL score and $r = 0.23$ with Two-Step model-based weight.

Designed to parse model-based from model-free decision-making [50, 151], the Two-Step Task model-based weight showed modest yet significant correlations with inhibitory control (WCST; $r = 0.179$). To our knowledge, this relationship has not been reported previously. Clinical evi-

dence converges with the same pattern: alcohol-dependent patients commit more perseverative and non-perseverative errors in WCST [231], and, when assessed after detoxification, they exhibit attenuated model-based control on the two-step task relative to healthy controls [181]. Gillan et al. [70] showed that individuals scoring higher on a compulsivity factor completed fewer categories on the WCST and displayed weaker model-based control on the two-step planning task. Prior research shows mixed results regarding working memory's relationship with model-based planning. Eppinger et al. [56] found that higher working-memory capacity predicted greater model-based control, whereas the relationship was absent in older adults. We found significant but weak correlation with one of our WM task, CDT ($r = 0.153$) but not Corsi task. Additionally, we observed significant but weak correlations between the Two-Step Task and Mental Rotation task($r = 0.184$).

TOL performance correlated modestly with Raven's SPM score($r = 0.394$), closely matching the coefficients reported by Zook et al. [235] and Unterrainer et al. [212] ($r \approx 0.38$–0.40). Its correlation with spatial working memory were modest in our data – Corsi span ($r = 0.22$) and change detection d' ($r = 0.23$). Gilhooly et al. [68] found a modest Corsi–TOL correlation of $r = 0.26$, and Temple, Carney, and Mullarkey [206] reported a stronger link ($r = 0.61$). Joyce and Robbins [102] and Welsh, Satterlee-Cartmell, and Stine [223] reported significant contribution of working memory to TOL performance, whereas two studies found no correlation between the Corsi span and TOL performance[212, 235]. Inhibitory control showed a likewise modest influence on TOL performance($r = 0.272$ with WCST negated perseverative errors), squarely within the 0.22˘0.48 range previously observed for WCST and Stroop interference measuring inhibitory control [122, 206, 223, 235].Finally, TOL performance also showed moderate correlations with mental rotation ($r = 0.393$) and pattern detection ($r = 0.230$). These findings align closely with the mental rotation correlation previously reported by Cheetham et al. [36] ($r = 0.31$) and a somewhat weaker correlation with pattern recognition documented by Robbins et al. [173] ($r = 0.18$). Relatedly, performance on the TOH, a similar task to TOL, has also been shown to correlate moderately

with pattern detection ($r = 0.27$) [134]. Taken together, these results suggests the role of mentally representing and manipulating spatial configurations, as well as maintaining information in working memory in TOL.

The novel FIAR task correlated significantly with all cognitive measures: fluid intelligence (SPM; $r = 0.344$), working memory (Corsi: $r = 0.355$, CDT: $r = 0.321$), inhibitory control (WCST; $r = 0.187$), simple planning tasks (TOL: $r = 0.280$, Two-Step: $r = 0.185$), and spatial abilities (Mental Rotation: $r = 0.213$, Pattern Detection: $r = 0.328$). This broad association profile supports the hypothesis that complex planning tasks require more cognitive resources. [235].

Between planning tasks Correlations among the Three Planning Tasks – Two-Step, TOL, and FIAR – were modest to moderate. The Two-Step Task correlated weakly but significantly with TOL ($r = 0.166$) and FIAR ($r = 0.185$). TOL and FIAR exhibited a somewhat stronger relationship ($r = 0.280$).

Taken together, our correlations replicate and extend previous findings. Despite the potential attenuation of effect sizes in online data collection [111, 171], our findings align with lab-based literature, demonstrating reasonable convergent validity of web-based cognitive assessments. The correlations and theoretical expectations motivated subsequent factor analysis to identify latent cognitive dimensions underlying these intercorrelations.

Factor Analysis

To uncover latent dimensions underlying performance across the nine tasks, we conducted an exploratory factor analysis on the task measures. Sampling adequacy was acceptable ($KMO = 0.78$), and Bartlett's test of sphericity confirmed that the correlation matrix differed significantly from the identity matrix, $\chi^2(36) = 1.5 \times 10^3$, $p = 5.4 \times 10^{-150}$, indicating that the data were

appropriate for factor analysis. We applied principal-axis factoring with Varimax rotation to obtain orthogonal factors; results for Oblimin and Promax were qualitatively consistent.

Rules of thumb such as Kaiser's criterion and parallel analysis become unreliable with fewer than ten variables [46, 139]. Following best-practice recommendations for such circumstances [59, 224], we retained the smallest factor structure that achieved a cumulative explained variance of roughly 50–60%. A three-factor solution met this target, accounting for 57.2% of the total variance; a two-factor alternative explained only 43.9%.

In the three-factor solution(Table 2.2), Factor 1 accounted for 23.7% of the variance. Mental Rotation d′ (0.80), Raven's SPM score (0.76), Tower of London score (0.63), and Pattern Detection d' (0.55) had the highest loadings on this factor. These tasks share high visuospatial processing ability. Factor 2 explained 20.1% additional variance and showed its highest loadings on the Corsi span (0.78), Change Detection d' (0.70), and Four-in-a-Row Elo (0.67). The combination suggests that the factor reflects a reliance on working memory; in Four-in-a-Row, working memory would be needed to maintain simulated sequences of moves. Factor 3 accounted for 13.4% additional variance, showing its highest loadings on the Two-Step model-base weight (0.83) and (negated) WCST perseverative-error (0.62). These loadings suggests that this factor represents inhibition of habitual responses. Two variables exhibited cross-loadings: WCST inhibitory control moderately loaded onto Factor 1 (0.32), suggesting that successful inhibition also demands some visual-relational ability. Pattern Detection performance moderately loaded onto Factor 2 (0.35), suggesting that spotting patterns requires observers to hold recently fixated item locations in working memory while they scan the array and integrate those locations into a configuration.

Overall, the EFA indicates three partially overlapping but distinguishable cognitive factors: (i) visuospatial processing, (ii) working memory, and (iii) inhibitory control, suggesting that there might be no single cognitive mechanism suffices to explain planning performance; instead, distinct but partially overlapping faculties contribute across different planning tasks.

To validate the structure identified through EFA, we tested the three-factor structure in a con-

Table 2.2: Varimax-Rotated Factor Loadings for Three-Factor Solution. Strong loadings (> 0.5) are **bold**, moderate loadings (0.3–0.5) are *italicized*.

Variable	Factor 1	Factor 2	Factor 3
Four-in-a-Row	0.20	**0.67**	0.15
Two-Step Task	0.00	0.18	**0.83**
Tower of London	**0.63**	0.13	0.19
Change Detection	0.21	**0.70**	0.09
Mental Rotation	**0.80**	0.13	0.13
WCST	*0.32*	0.03	**0.62**
Corsi	0.14	**0.78**	0.05
Pattern Detection	**0.55**	*0.35*	-0.10
Raven's SPM	**0.76**	0.26	0.19
Proportion Variance	**23.7%**	**20.1%**	**13.4%**
Cumulative Variance	**23.7%**	**43.9%**	**57.2%**

firmatory factor analysis that specified latent variables for *visuospatial ability and simple planning*, *working memory*, and *inhibitory control*.Model fit was excellent: Comparative Fit Index (CFI) = 0.981, Tucker–Lewis Index (TLI) = 0.972, and Root Mean Square Error of Approximation (RMSEA) = 0.036. The exact-fit test was, as often in large samples, statistically significant, $\chi^2(24) = 39.0$, $p = .027$, indicating a small but detectable residual misfit. Together with EFA, these findings support a differentiated three-factor architecture governing performance across the nine tasks.

Regression Analysis

EFA and CFA tell us which latent abilities covary across tasks, but they do not reveal the cognitive components underpinning each planning task. To pinpoint the predictors for each planning task, we used LASSO regression, controlling for multicollinearity. We performed 5-fold cross-validation with standardized regression coefficients reported for interpretability. Statistical significance was assessed using Bonferroni-adjusted p-values (Table 2.3).

For Four-in-a-Row, cognitive predictors explained 23.8% of the variance. Significant contributors were Corsi (β_{std}=0.199, $p = 1.6\times10^{-5}$), Change Detection (0.147, $p = 1.5\times10^{-3}$), SPM (0.175, $p = 1.2 \times 10^{-3}$), and Pattern Detection (0.186, $p = 4.03 \times 10^{-5}$). For the Tower of London, cog-

nitive predictors explained 21.3% of performance variance, with significant contributions from SPM (0.188, $p = 6.5 \times 10^{-4}$) and Mental Rotation (0.203, $p = 1.1 \times 10^{-4}$).For the Two-Step task, cognitive predictors explained only 7.6% of the variance. SPM (0.134, $p = 2.4 \times 10^{-2}$) and WCST (0.109, $p = 2.09 \times 10^{-2}$) reached significance, but effect sizes were modest.

Regression analysis suggests different cognitive profiles for the three planning tasks. Complex planning tasks such as FIAR may rely heavily on working memory and fluid intelligence, whereas simpler planning tasks such as TOL strongly depends on visuospatial ability and fluid intelligence.

Table 2.3: Lasso regression results for planning tasks

	Outcome measures		
Predictors	4IAR Elo	TOL score	Two-Step
Corsi	0.199^{***}	0.034	0.039
CDT	0.147^{**}	0.071	0.052
SPM	0.175^{**}	0.188^{***}	0.134^{*}
WCST	0.070	0.077	0.109^{*}
Rotation	−0.069	0.203^{***}	0.039
Pattern	0.186^{***}	0.038	0.032

Note. $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$

Unexplained variance

We next asked whether the variance left unexplained by these regressions reflects stable but unmeasured abilities or merely noise. Following Mitko and Fischer [133], we split each task's trials into odd and even halves, refitted the regressions, and correlated residuals across halves. If the unexplained variance represents a separate planning ability not captured by our cognitive measures, we would expect significant reliability in these residuals.

The split-half reliability analysis revealed substantial reliability in the residuals across all planning tasks. Residual correlations were significant (TOL: $r = 0.44$, $p = 6.6 \times 10^{-24}$; FIAR: $r = 0.54$, $p = 2.0 \times 10^{-37}$, indicating the presence of stable unexplained variance beyond what our cogni-

tive measures captured. This suggests there may be additional cognitive components underlying planning performance not fully accounted for by our predictor variables.

However, the cross-task correlations of these residuals were generally low and non-significant. Table 2.4 shows the average cross-half correlation for each pair of tasks. The largest average was between TOL and ELO ($r \approx 0.099$, $p \approx 0.03$), while the TOL–Two-Step and ELO–Two-Step pairs showed near-zero correlations. Overall, cross-task residual correlations were low (mean correlation $r = 0.05$), suggesting that the unexplained variance appears to be task-specific rather than evidence for a single, overarching "planning ability.".

Task Pair	Avg. Correlation	p-value range
TOL–FIAR	0.0986	0.0270–0.0364
TOL–Two-Step	0.0434	0.1891–0.5630
FIAR–Two-Step	0.0039	0.1547–0.2104

Table 2.4: Average cross-task cross-half correlations for each pair of planning tasks. The "p-value range" corresponds to the two cross-half comparisons (e.g., TOL Half1 vs. ELO Half2 and TOL Half2 vs. ELO Half1).

Across complementary analyses, three separable cognitive dimensions emerge as the principal sources of shared variance among the nine tasks. There is no strong evidence of a single planning ability common to different planning tasks. Instead, most of the unexplained variance appears to be task-specific, and each task also contains reliable variance that remains to be explained.

2.4 Discussion

Our study aimed to clarify the cognitive mechanisms underpinning planning and to determine whether planning tasks of varying levels of complexity share common cognitive processes. By assessing participants across three planning tasks—the Two-Step task, Tower of London, and Four-in-a-Row – alongside measures of working memory, fluid intelligence, mental rotation, pat-

tern detection, and inhibitory control, we found evidence of both unity and diversity in cognitive mechanisms related to planning. Factor analysis revealed three cognitive factors across these tasks. The Two-Step Task model-based control and inhibitory control measured from WCST loaded onto the same factor. The Tower of London task performance clustered together with mental rotation, pattern detection and fluid intelligence. Four-in-a-Row performance loaded onto the same factor with working memory and pattern detection task performance. The fact that the planning task metrics loaded onto different cognitive factors challenges simplistic conceptualizations of planning as a unitary construct. Our regression analysis further supported these distinctions, showing that fluid intelligence significantly contributed across all planning tasks but that each task relied on additional cognitive abilities. Notably, working memory and pattern detection predicted performance in the Four-in-a-Row task, mental rotation predicted TOL performance, and inhibitory control was predictive of the Two-Step Task model-based weight. These findings resonate with the proposal by Burgess et al. [25], that planning involves multiple subprocesses – including mental simulation, retrospective and prospective memory, or inhibition – whose relevance varies according to task requirements. In other words, each planning task might draw on a different constellation of cognitive abilities, rather than relying on a single, domain-general planning mechanism.

One possible explanation for this differentiation might be the varying complexity among planning tasks. The Two-Step Task, characterized by minimal look-ahead and small state spaces, likely emphasizes inhibitory control due to the need to suppress impulsive responses toward repeating rewarding actions, rather than extensive mental simulations. This interpretation is consistent with Goel and Grafman [72]'s view of planning impairments in the Tower of Hanoi: they argued that the observed deficits might reflect a generalized response inhibition failure rather than a specialized "planning" dysfunction. In simpler tasks like the Two-Step, a need to suppress habitual responses might become more prominent due to states frequently repeating. In addition, tasks with higher combinatorial complexity, such as FIAR, might impose greater de-

mands on working memory, echoing findings that more complex task (Tower of Hanoi) might increase working memory load [68, 235]. The associations of FIAR with working memory and pattern detection make sense in light of its high combinatorial complexity. The involvement of pattern detection is intuitive, given that Four-in-a-Row requires detecting task-relevant features, echoing findings that feature dropping rate decreases as people gain more experiences in the task[148]. Working memory in FIAR likely operates at multiple timescales: encoding the current board states, maintaining information about potential future moves, and retrieving relevant sequences of moves of the current board, or previously successful strategies across games. Thus, the working memory demands of FIAR may reflect both short-term simulation and the longer-term retention of effective heuristics over multiple games.

However, complexity alone may not fully account for these distinctions; other task characteristics might also play a significant role in differentiating these planning tasks. The relationship between mental rotation and TOL likely arises from the both tasks' inherent spatial component, which requires mental simulation and manipulation of spatial configurations, making spatial abilities critically important.

Our findings have several limitations. We only used three planning paradigms, each differing in complexity. This constrains the generalizability of our conclusions. Systematically controlled tasks–such as parametric variations of the N-in-a-row paradigm (e.g., m-by-n versions)–could help delineate how incremental changes in state-space complexity shape the cognitive mechanisms of planning. Our work also sets the stage for further investigation into the complexity of real-world planning. Real-world planning often goes beyond the scope of typical laboratory tasks – requiring individuals to navigate unclear subgoals, unclear end states and perhaps multiple ways to succeed [6, 25, 72, 83, 84, 132, 137, 178], which may tax cognitive abilities differently than well-defined laboratory tasks we measured. Future research should therefore examine a more diverse array of planning tasks, beyond systematic manipulations of state-space complexity, to better reflect the intricacies of real-world planning scenarios [25, 49, 160].

Our findings inform the ongoing debate surrounding the classification of the Two-Step Task as a measure of planning. Some studies characterize it as a model-based planning task [60, 131], whereas others argue it reflects sophisticated habit formation rather than planning [4, 44]. Feher da Silva and Hare [60] defines model-based behavior as selecting actions based on an internal model of the environment, aligning with definitions of planning behavior. However, engaging in model-based processes may not necessarily involve complex planning. Our results demonstrate that performance in the Two-Step Task loads primarily with inhibitory control, distinguishing it from other planning tasks. Future research would benefit from tasks incorporating larger and more distinctive state spaces, thereby minimizing reliance on habitual responses and providing clearer measures of planning processes.

Our results also contribute to ongoing discussions on mental simulation and planning. At first glance, it seems intuitive to hypothesize that processes involved in mentally simulating future states during planning tasks would overlap significantly with mental simulation involved in manipulating physical objects[195]. Classic research by Shepard on mental rotation demonstrated that larger rotation angles require longer response times, suggesting an internal process akin to "playing out" rotations mentally. Similarly, literature on intuitive physics proposes that intuitive reasoning about physical scenarios also involves mental simulation[133]. Our results, however, present a mixed picture – mental rotation predicted TOL performance, suggesting shared cognitive processes of simulation, but not on the FIAR or Two-Step Task. This suggests that mental simulation processes may differ substantially depending on specific task characteristics – such as reliance on physical spatial configurations versus abstract rules. Future research should further explore these nuanced distinctions between various forms of mental simulation in planning contexts.

In conclusion, our study provides evidence that planning involves a constellation of distinct but interrelated cognitive processes, the precise composition of which depends on task complexity. Recognizing planning as multifaceted and dependent on task complexity enables future

research to more precisely identify the cognitive mechanisms of planning.

Having established that human planning is underpinned by multiple cognitive abilities, the next step is to peer inside the human mind 'black box'. While factor and regression analyses elucidate which cognitive abilities support planning, they do not reveal precisely how plans are constructed moment-to-moment. To capture the real-time unfolding of thought processes underlying planning, we turn in the next chapter to think-aloud protocols, providing a window into participants planing strategies.

3 | Do humans think in trees? Lesson from think-aloud protocal

3.1 Introduction

A challenge for cognitive scientists trying to understand planning is its covert nature: extensive cognitive processing typically precedes observable behavioral choices. The identification of (sub)goals, the roll-out of branches of the decision tree, and the evaluation of future states all occur without external markers. To nail down the processes underlying planning, researchers have resorted to eye movements [26, 54, 61, 73, 148] and neural measures [12, 114, 219]. But how far can we push our understanding of planning through behavioral data alone?

Traditionally, efforts to tackle this question have taken two approaches: (1) increasing the richness of the data and (2) building computational models that can be fitted to human data. At one end of the spectrum, earlier studies often used richer data sources – like think-aloud verbalizations – but paired them with relatively simple, manually specified models [51, 141]. At the other end, newer work relies on simpler behavioral measures (e.g., final choices, response times) but applies advanced modeling techniques to fit parameters [148].

As a data-enrichment method, think-aloud has a long history in cognitive psychology[51, 57, 141]. In think-aloud studies, participants verbalize their ongoing thoughts while performing a task, providing a window into otherwise hidden cognitive processes. Think-aloud studies have

advanced our understanding of cognitive strategies, such as subgoal decomposition and heuristic reasoning. For instance, De Groot [51] showed that chess grandmasters' advantage lies not in simply searching deeper than novices, but in applying more effective heuristics – mental shortcuts or rules of thumb. Likewise, Newell and Simon [141] used think-aloud data to inform their General Problem Solver (GPS) model, a model that approached problem-solving as a heuristic-driven search, decomposing complex problems into subgoals and applying operators to iteratively reduce differences between the current state and a goal state. They argued that participants' step-by-step verbalizations closely matched the solution steps predicted by GPS, establishing a foundational perspective of human problem-solving as heuristic-driven search within a clearly defined problem space consisting of states, actions, and transitions.

Despite these contributions, think-aloud methods have limitations. This method is labor-intensive and reveals only what enters the conscious thought – meaning some unconscious cognitive processes might go unreported [65, 179]. Early think-aloud studies often involved anecdotal or single-digit sample sizes, raising concerns about generalizability. Additionally, early cognitive models (such as GPS) required manual specification of relevant operators, subgoals, and domain knowledge. This manual tuning process risks biasing models towards specific observations [147]. Consequently, the think-aloud method became more niche as cognitive psychology gravitated toward methods perceived as more objective, quantitative, and scalable.[147]

On the other end of the spectrum, researchers have sought to infer covert planning processes by using standard behavioral measures with advanced statistical modeling tools [27, 88, 95, 148]. In the context of planning, these models often assume that people perform heuristic tree search: constructing a mental tree of possible future states, pruning low-value branches, and focusing on promising lines of action. This perspective relates to early AI systems such as the Logic Theorist [194], which proved scores of symbolic-logic mathematical theorems using heuristic-guided search, and on the first full chess programs [17], which carried out a depth-first minimax search. While those early systems were largely hand-coded with heuristics, modern cognitive models

typically fit parameters directly to human data using likelihood-based methods.

Fit alone, however, does not guarantee that the recovered parameters correspond to the mechanisms people actually use. In other words, these models risk becoming "as-if" descriptions that reproduce observed choices without capturing the underlying mechanism. Evidence suggests that individuals rarely conform to the rigid patterns assumed by computational models. For example, they may use progressive deepening [51] – revisiting the same candidate moves multiple times and looking further ahead on each pass, or satisfice by pursuing "good enough" solutions [194]. They may prune large potential losses [96] or reuse partial solutions [95]. All of these strategies reflect cognitive resource constraints [66]. These findings underscore the complexity of human planning – traits that computational models with simplified assumptions might fail to fully capture.

Our study aims to bridge these two lines of work – rich verbalization data and computational models quantitatively fitted to behavioral data – by examining alignment between cognitive insights derived from think-aloud protocols and predictions from computational cognitive models that rely on heuristic tree-search assumptions. Specifically, we ask: How closely do individuals' verbalized cognitive processes correspond to predictions made by heuristic tree search models fitted to human data? The answer is critical for validating the models as faithful representations of internal cognitive mechanisms, rather than simply accurate predictors of final choices.

To address this question, we analyzed think-aloud data from participants playing Four-in-a-Row, a two-player, full-information game with a computational model that explicitly assumes tree-based planning. Our approach involved comparing verbalization-derived metrics against computational model predictions of planning depth. We hypothesized that if people truly perform tree search when playing the game, verbalized planning depth should correlate positively with the model's estimated depth.

The chapter is organized as follows: first, we present descriptive statistics of the think-aloud data, detailing verbalized planning depths and branching behaviors. Second, we examine cor-

relations between verbalization-derived metrics and participants' playing strength. Third, we test alignment between human-generated verbal metrics and computational model derived metrics. Finally, we list qualitative observations of participants' verbal strategies. We concluded by discussing the implications of these findings for future think-aloud research and computational models of human planning.

3.2 Methods

We pre-registered the study on the Open Science Framework (https://doi.org/10.17605/OSF.IO/K7V2P).

Participants

We preregistered a target sample size of 35 participants and continued recruiting until we reached that target. We enrolled 38 individuals through the Sona participant management system and university-wide advertisements. As specified in the preregistration, participants were eligible if they were at least 18 years old and had native or near-native English proficiency. We excluded three individuals who did not meet the language criterion. We also excluded one additional participant (not preregistered for exclusion) because they failed to follow the task instructions during the think-aloud task; they spoke very little and showed minimal engagement during training. The final sample included 34 participants (19 female, 13 male, 2 non-binary; mean age = 22.94 years, range = 18–50 years). Participants received monetary compensation for their participation. We describe the base payment and performance-based bonuses for each stage of the task in the Procedure section. The study was approved by the Institutional Review Board of New York University, and we obtained informed consent from all participants.

Apparatus and Materials

We conducted the experiment in person, in a quiet and dimly lit room to minimize distractions and ensure consistent recording conditions. We implemented the task as a locally hosted webpage coded in JavaScript. Participants used a desktop setup powered by an Intel NUC mini PC (regulatory model: NUC8i5BEH). The system was connected to a 27-inch Acer T272HL touchscreen monitor.

For the Stage 2 experiment, we recorded audio using a Cyber Acoustics CVL-1084 USB cardioid microphone, positioned close to the participant. We captured video of the participant's screen interactions and gestures using an iPhone 14 Pro Max or iPhone 15 Pro, positioned to view the screen area.

We implemented the Four-in-a-Row gameplay as a browser-based JavaScript interface, displayed on the touchscreen monitor. The participant made moves by touching a square on the grid. As in Opheusden et al. [148], the participant played against an AI agent on a 4-by-9 board, with alternating turns to place pieces. The first player to achieve four pieces in a row in any direction (horizontally, vertically, or diagonally) won the game. The visual design, interaction logic, and task instructions matched those used in Opheusden et al. [148].

For stage 2, we constructed a set of 12 Four-in-a-Row puzzles, 2 (K, L) for practice and 10 (A-J) for the official experiment. Each puzzle was based on a specific mid-game board state. It featured a single optimal move that, if selected, allowed the initiating player to force a win within 3, 4, or 5 total moves, assuming perfect play from both players. We balanced turn-taking roles by designing five puzzles in which the participant made a move for black and five in which the participant made a move for white. We designed the puzzles manually and used a custom-built solver (link to solver repository) to verify that each puzzle had exactly one winning move and that the win could be forced within the intended number of moves, regardless of the opponent's responses. Figure 3.1 shows an example puzzle; Figure 3.2 shows the full set of 12 puzzles.

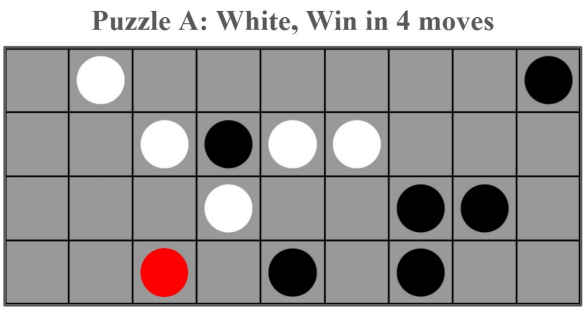


Figure 3.1: Example puzzle from Stage 2 (Puzzle A). It is White's turn. An optimal sequence of moves guarantees a win within 4 moves, assuming perfect play from both players. The red piece indicates the only move that leads to a guaranteed win.

Procedure

The experiment consisted of one session, which consisted of two stages. Before stage 1, the participant completed consent and demographic forms.

Stage 1 (Free play): The participant was instructed on the rules of the Four-in-a-Row game and played two practice games. They then played 40 games of Four-in-a-Row against AI agents. The difficulty of the AI opponents adapted dynamically using a staircasing algorithm, which increased or decreased difficulty by one level after each win or loss, respectively, based on participant performance. Stage 1 was conducted without the presence of an experimenter. We did not impose any time limits.

Participants received a base payment of $10 for completing this stage, with a maximum performance-based bonus of $8. Specifically, they earned $0.20 for each game won, $0.10 for each tie, and no bonus for a loss. Participants were informed that they would not be compensated if they dropped out before completing the session.

Stage 2 (Puzzles): At the beginning of Stage 2, we trained the participant on the think-aloud protocol using two sample problems that were relevant to everyday life but unrelated to the Four-in-a-Row task. Following this, the participant completed two practice puzzles to familiarize themselves with the task. They then proceeded to solve 10 Four-in-a-Row puzzles that were

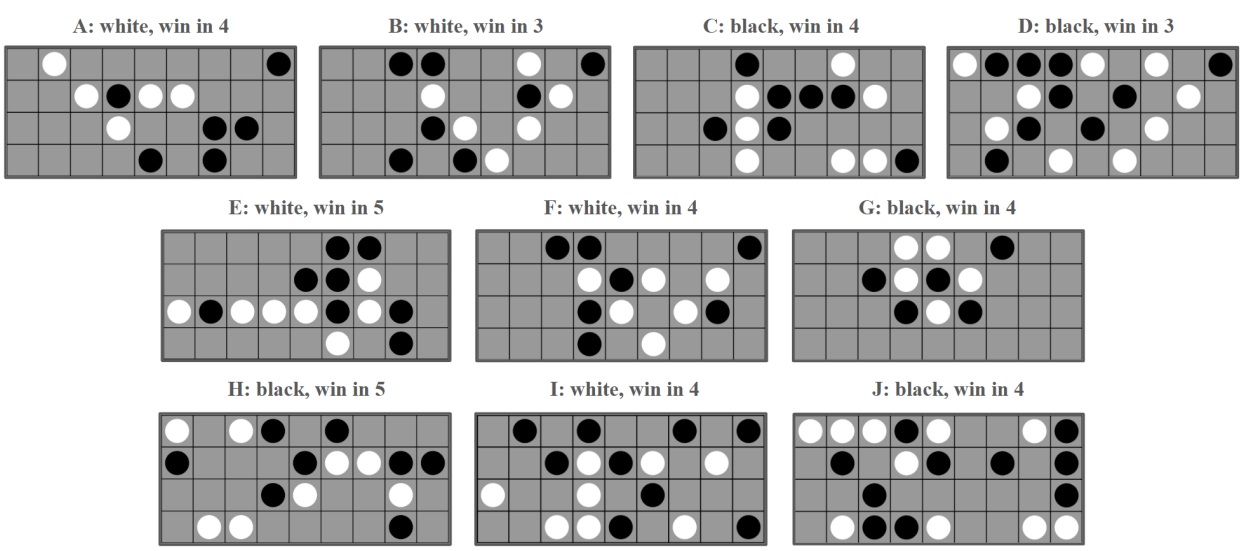


Figure 3.2: Initial board states for all 10 Four-in-a-Row puzzles used in Stage 2. Each puzzle displays its unique ID, starting player color (black or white), and solution depth (3–5 moves) that guarantees a win if played optimally. Puzzle order was randomized during the experiment, and the starting player was counterbalanced across the set.

presented in randomized order. In each puzzle, the participant was instructed to make the best possible move as if continuing a real game from that position. Although each puzzle was designed to have a single winning move, we did not inform participants that a win was guaranteed; instead, we told them only that one move was clearly superior to others. We instructed the participant to verbalize their planning process while indicating board positions through screen touches.

During Stage 2, the experimenter monitored the participant's behavior to ensure adherence to verbalization instructions and to manage task duration. Although there was no strict time limit for each puzzle, the participant had been instructed to aim for a duration between 3 and 8 minutes to balance performance quality with time control. If the participant remained silent for more than 10 seconds, the experimenter presented a visual prompt (a printed cardboard sign reading "Keep Talking.") If the participant's verbalization exceeded 10 minutes, the experimenter instead presented a cardboard reading "Control Time." These prompts were shown silently and served as standardized, nonverbal cues. The participant had previously been instructed to either resume verbalizing or begin wrapping up their reasoning process upon seeing the corresponding

sign. This protocol ensured consistency in engagement and timing across trials while minimizing direct intervention by the experimenter.

After narrating their planning process for each puzzle, the participant tapped an on-screen button to indicate they had done their thinking and were ready to make the move. Right after that, they were required to make a move within 5 seconds; if they did not respond in time, they lost the chance to submit an answer and earn the bonus. After each puzzle, the participant rated how difficult they found the puzzle and how confident they felt about their move using an on-screen survey.

The participant received $10 for completing Stage 2, with an additional performance-based bonus of up to $20. They earned $2.00 for each puzzle in which they selected the best move. Only one best move was eligible for the bonus in each puzzle, and no bonus was awarded otherwise. As with Stage 1, compensation was contingent on completing the study.

Data Coding

Audio and video recordings from Stage 2 were analyzed to extract verbalized planning metrics. Each recording was segmented by participant and puzzle. For each puzzle attempted by the participant, we identified all *depth-1 moves*—defined as the first verbally articulated candidate moves from the initial puzzle state, prior to any internal elaboration or branching.

For every depth-1 move mentioned, we coded a range of behavioral and cognitive metrics to characterize the structure and content of the participant's planning. Table 1 lists all verbalization-derived metrics referenced in this paper, including their definitions and computation methods. A complete list of all coded metrics is available in Supplementary Materials (Section S4).

Two authors (D.L. and S.L.) jointly conducted the data coding using a custom pipeline (See 6). They embedded key codes (including depth-1 moves, board features mentioned during planning (e.g., two-in-a-row), planning depth, and transcripts of verbalization used to determine the maximum stated planning depth per puzzle) directly into the video files as time-aligned labels

to support accuracy and traceability. After completing the initial round of annotations, they reviewed all codes in a second pass to ensure consistency with the protocol. When they disagreed, they discussed their interpretations and resolved all conflicts through consensus. The resulting dataset is publicly available on OSF, and labeled videos are available upon request in accordance with data-sharing guidelines.

Model Fitting and Analysis

We fitted a computational model introduced by Opheusden et al. [148] to participants' Stage 1 free-play data. This model assigns values to board states via a feature-based evaluation function, where each feature (e.g., centrality, connected two-in-a-row, unconnected two-in-a-row, three-in-a-row, and four-in-a-row) is weighted according to its relevance. These weights, alongside other parameters, are estimated from participants' observed choices. The model constructs a decision tree using best-first search. At each iteration, it expands the most promising node based on current value estimates and backpropagates values using a minimax rule. The model includes pruning of low-value branches to approximate cognitive efficiency and incorporates stochastic elements such as Gaussian value noise and feature dropout to simulate variability and attentional lapses. Feature dropout simulates lapses of selective attention by randomly omitting instances of board features (e.g., three-in-a-row patterns at specific locations) from the evaluation function before the tree search begins. These omitted features do not contribute to value computations on that trial, effectively reducing the information available for planning.

The primary analysis (pre-registered) assessed the Pearson correlation between two distinct measures of planning depth: one inferred from the computational model (i.e., the depth of the simulated search tree used to fit participants' freeplay choices), and the other derived from participants' verbalizations during Stage 2 (i.e., the maximum number of sequential moves articulated during thinking aloud). Additional model parameters we used include heuristic quality, computed with the average correlation between game-theoretical values of 5482 test boards used in

Opheusden et al. [148] and the value function output for the participant. Feature dropping rate is one of the fitted parameter of the cognitive model. Additional analyses explored strategy use and its relation to model parameters, as described in the coding materials.

Linear Mixed–Effects Modelling

We fit two families of linear mixed–effects models (LMMs) with `statsmodels`. Models were estimated with restricted maximum likelihood (REML). In all models the continuous predictors were z-scored (mean 0, SD 1) so that fixed–effect coefficients can be interpreted as the expected change in the outcome for a one–SD increase in the predictor.

ELO as a function of verbalization metrics. For each verbalization metric – average planning depth, average branching factor, average tree size, and average number of sentences – we fit an independent model of the form

$$\mathrm{ELO}_i \;=\; \beta_0 \;+\; \beta_1\,\mathrm{Metric}_i \;+\; u_{0i} \;+\; \varepsilon_i,$$

where $u_{0i} \sim \mathcal{N}(0, \sigma^2_{\text{subject}})$ is a random intercept for participant i. This structure accounts for between-participant differences in baseline ELO while testing whether each metric predicts ELO.

Within-puzzle change in planning depth. To examine how planning depth evolves during a single puzzle, we modelled the sequence of depths verbalised by each participant:

$$\mathrm{Depth}_{ijk} \;=\; \beta_0 \;+\; \beta_1\,\mathrm{Step}_{ijk} \;+\; u_{0i} \;+\; v_{0j} \;+\; \varepsilon_{ijk}.$$

Here, i indexes participants, j indexes puzzles, and k indexes step index – the ordinal position $(0, 1, 2, \dots)$ of each depth-1 move the participant verbalized while working through the puzzle. The step index therefore serves as a linear proxy for progression in time within the planning

Table 3.1: Verbalization-derived metrics and their computation

Metric	Definition	How it is computed
Depth-1 move	The first articulated move from the initial puzzle state.	Recorded as the first move explicitly mentioned by the participant during verbal planning (indexed 0–35 in the 4-by-9 board)
Planning depth	The maximum depth of explicitly articulated sequences of moves.	For each puzzle, the longest linear sequence of moves articulated from the root state (initial puzzle state) to the leaf (final state) is recorded.
Average depth	The average maximum depth of explicitly articulated sequences of moves per puzzle.	Averaged the planning depth across the 10 puzzles for each participant
Number of branches	Number of depth-1 moves	The count of depth-1 moves articulated from the initial state is calculated in each puzzle.
Average branch	Average number of depth-1 moves	Average number of branches across puzzles for each participant.
Deeper branch number	The presence of articulated alternative branches in addition to the longest planning sequence	Counted by identifying any additional articulated moves branching off the main linear planning path beyond the initial depth-1 moves.
Average tree size	Average articulated tree size, reflecting both breadth (number of branches) and depth (planning depth).	Computed as the product of *Number of branches* and *Planning depth* for each puzzle, averaged across all puzzles for each participant.
Number of verbalized sentences	Number of spoken sentences during planning	Computed using Google's sentence-level speech parsing API

episode: each time a participant articulates another concrete next move, the index increments by one. Random intercepts are included for both participants ($u_{0i} \sim \mathcal{N}(0, \sigma^2_{\text{subject}})$) and puzzles ($v_{0j} \sim \mathcal{N}(0, \sigma^2_{\text{puzzle}})$) to capture baseline differences in verbalized depth attributable to individual or puzzle characteristics.

3.3 Results

To understand the cognitive processes of planning, we analyzed think-aloud data collected from participants and fitted the cognitive model (See methods) to behavioral data. We begin by summarizing the key characteristics of the think-aloud data. Next, we relate verbal metrics to players' ELO ratings to examine how verbalization patterns correlate with playing strength. Then, we investigate the alignment between verbalized metrics and model metrics. Furthermore, we looked at how planning depth unfolds over the course of problem-solving episodes. Finally, we report qualitative observations that provide insight into the thinking processes.

Summary Statistics

We first examined descriptive statistics of participant's planning behavior derived from their verbalizations in Stage 2 of the experiment (Table 3.2). Each of the 34 participants completed 10 puzzles. Participants spent an average of 168 seconds (SEM = 19) per puzzle, articulating about 24.1 sentences (SEM = 3.0; range: 2.4–68.5 sentences).

Planning depth (See Table 3.1), representing the depth of articulated look-ahead moves, has an average of 3.53 (SEM = 0.27) across participants. Notably, 5 out of 34 participants showed minimal look-ahead behavior, with on average fewer than two moves deep.

We quantified the number of branches by counting the number of verbally articulated depth-1 moves stemming from the initial puzzle state(See Table 3.1). Participants on average explored 5.71 branches per puzzle (SEM = 0.51). Deeper branch number, indicating exploration of alternative branches beyond main planning sequence, was absent in 3 out of 34 participants.

We calculated the average tree size(See Table 3.1), which integrates planning breadth and depth, by multiplying the average number of branches by the average planning depth for each participant. The mean tree size across participants was 23.67 nodes (SEM = 3.54).

Move repetition, defined as the frequency of re-articulating depth-1 moves per puzzle, oc-

curred on average 2.25 times (SEM = 0.32) across participants. This repetition, observed in 22 out of 34 participants, suggests limitations in working memory capacity during planning to hold move sequence.

Additionally, some participants showed limited verbalized planning. approximately 2 out of 34 participants explored on average fewer than two branches per puzzle, and 5 out of 34 subjects never explored alternative moves after the first move (no deeper branching). 5 out of 34 subjects gave up planning before their search reached end state (i.e., they never articulated a complete solution path).

Table 3.2: Descriptive statistics of verbalized planning behaviors (N=34).

Verbalized Metrics	**Mean (SEM)**
Planning depth	3.53 (0.27)
Duration per puzzle (sec)	167.90 (19)
Sentences per puzzle	24.07 (3.0)
Branches per puzzle	5.71 (0.51)
Move repetitions	2.25 (0.32)
Tree size (branches × depth)	23.67 (3.5)
Categorical Observations	**Count**
Zero deeper branch number	3
Average depth < 2	5
Average branches < 2	2
Never planning to puzzle completion	5
Repeating depth-1 moves	22

These statistics provide an initial overview of participants' verbalized planning behaviors.

Subjective Difficulty and Verbalization

As a validation method for planning metrics, we examined whether subjective difficulty ratings correlate with measurable aspects of the planning process. We asked whether puzzles that feel harder are accompanied by (i) more verbalization and (ii) greater search depth (from both verbalized and model-derived depth). Higher subjective difficulty correlated significantly with

increased verbalization metrics, including the number of sentences (r = 0.464, $p < 0.001$) and average planning depth (r = 0.257, $p < 0.001$). However, subjective difficulty did not significantly correlate with computationally derived puzzle depth (r = 0.104, $p = 0.057$), again highlighting a divergence between verbalization-derived metrics and model-derived metrics (Figure 3.3).

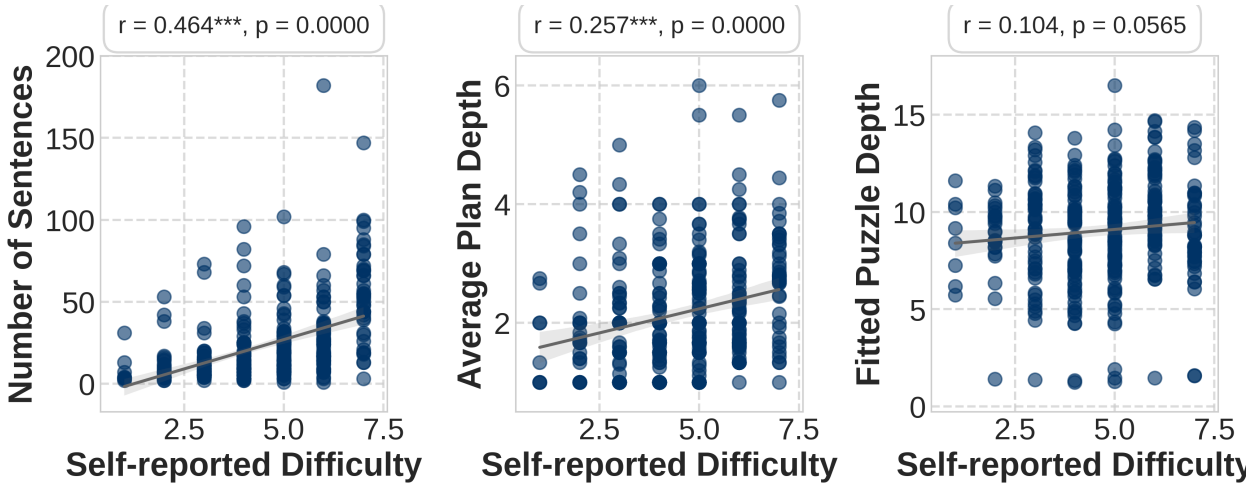


Figure 3.3: Subjective difficulty shows significant positive correlation with verbalization-based metrics (sentences and planning depth), but not with computationally derived puzzle depth

Positive associations with verbal metrics suggest that puzzles judged hard induce more overt cognitive activity. However, without experimental manipulation, we cannot determine whether difficulty prompts additional verbalization or if extensive self-explanation increases perceived difficulty.

Verbalization and Performance

To understand the relationship between verbalized planning metrics and participants' playing strength (measured as Elo scores), we performed linear mixed-effects modeling (LMM) analyses. We included multiple verbalization metrics as predictors, including the average number of sentences, average planning depth, average number of branches explored, and average tree size.

The LMM revealed two significant predictors of playing strength: the average number of sentences verbalized ($\beta = 0.346$, $p = 0.021$) and the average tree size (standardized coefficient = 0.341,

p = 0.024). These results suggest that more extensive verbalization and larger articulated search trees are positively associated with higher performance (Figure 3.4). Other verbalization metrics, such as average planning depth (standardized coefficient = 0.213, p = 0.059) and average number of branches (standardized coefficient = 0.222, p = 0.193), showed positive yet non-significant trends.

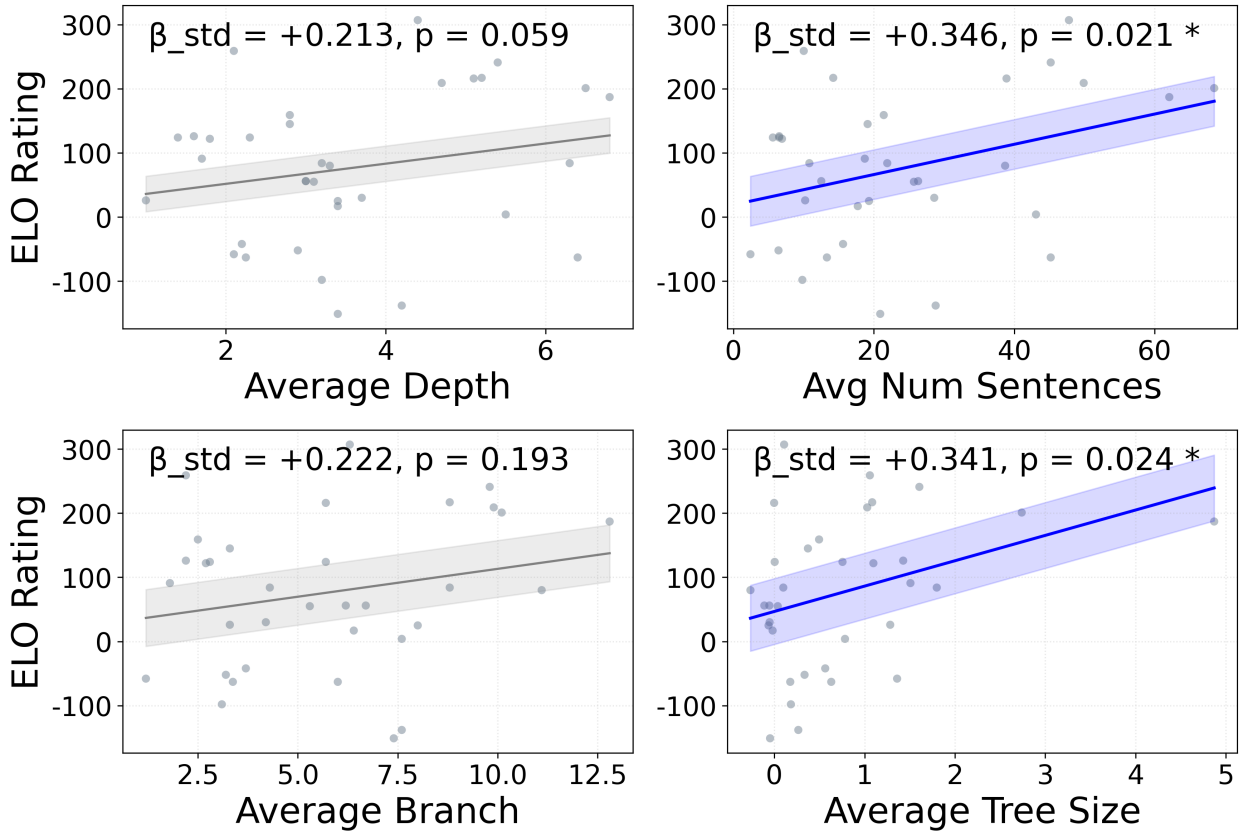


Figure 3.4: Scatterplots show the associations between different verbalization metrics (x-axes) and Elo ratings (y-axes), with regression lines indicating the strength and direction of relationships. Blue regression lines represent statistically significant relationships (p < 0.05). For each metric, the standardized coefficient (β_{std}) (β_{std}) indicates the expected change in Elo rating associated with a one standard deviation increase in the metric

Verbalization metrics vs. Model metrics

To evaluate the alignment between verbalized planning behaviors and computational model-derived metrics, we computed pairwise Pearson correlations between four verbalization measures – average number of sentences, average depth, average branch, and average tree size—and three model-derived indices (fitted search depth, heuristic quality, and feature-dropping rate). All p-values were adjusted for multiple comparisons using the MAX-statistic permutation method (See Table 3.3).

Across participants, no verbal metric was significantly associated with any model index ($|r| < .13$, all $p > 0.56$). For example, average verbalized depth was unrelated to model-estimated depth ($r = -0.06$), heuristic quality ($r = -0.13$), or pruning rate ($r = -0.12$). Similarly, average number did not correlate significantly with model metrics: heuristic quality ($r = -0.03$), fitted depth per puzzle ($r = 0.09$), or feature-dropping rate ($r = -0.09$).

These findings suggests a notable discrepancy between human verbalized metrics and computational model predictions (Table 3.3).

Table 3.3: Correlation Matrix of Metrics

	Elo	Fitted Depth	Heuristic Quality	Feature Drop Rate	Average Sent.	Agerage Depth
Fitted Depth	−0.012 p=1.000	—				
Heur. Qual.	0.283 p=0.903	0.172 p=1.000	—			
Feat. Drop Rate	−0.357 p=0.563	−0.120 p=1.000	−0.389 p=0.401	—		
Avg Sent.	0.346 p=0.624	0.088 p=1.000	−0.027 p=1.000	−0.095 p=1.000	—	
Avg Depth	0.213 p=0.998	−0.048 p=1.000	−0.056 p=1.000	−0.132 p=1.000	0.803* p=0.000	—
Tree Size	0.287 p=0.894	−0.021 p=1.000	−0.056 p=1.000	−0.144 p=1.000	0.821* p=0.000	0.915* p=0.000

*significant at p < 0.05 (MAX statistic method).

Dynamics of Verbalization

To understand how planning behavior evolves within a single puzzle-solving episode, we analyzed changes in planning depth over time using LMM models. Each complete puzzle-solving

attempt was treated as a distinct episode, allowing us to track temporal patterns of planning. We found a significant increase in planning depth as participants progressed through a puzzle (standardized coefficient = 0.066, $p < 0.001$). Participants initially engaged in shallower planning, which progressively deepened toward puzzle completion. This within-episode dynamic reveals temporal aspects of planning not captured by computational metrics aggregated per puzzle or participant (Figure 3.5).

These within-episode dynamics highlight aspects of cognitive process that model-derived metrics – limited to one aggregated measure per subject or puzzle – are unable to fully capture.

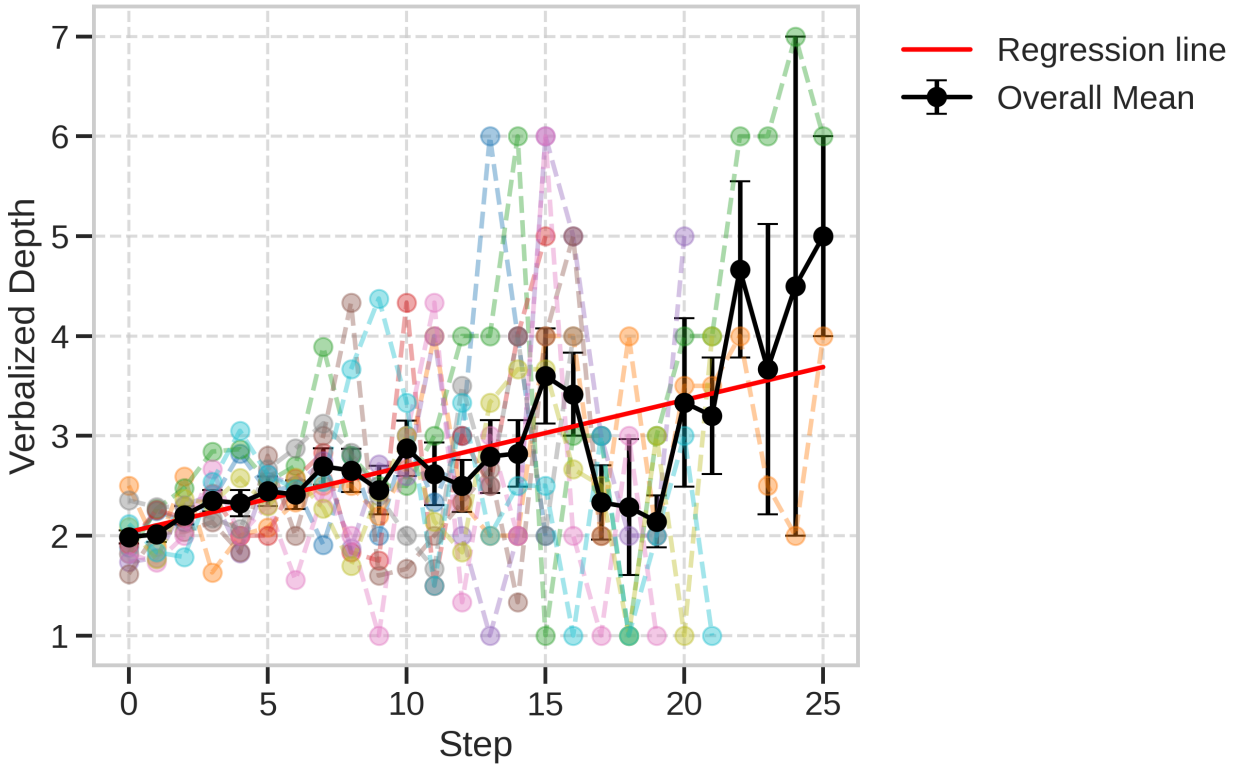


Figure 3.5: Relationship between planning depth and progression through puzzle-solving episodes. Step index represents a verbalized depth-1 move, serving as a temporal proxy within episodes. Colored lines indicate average planning depth across different puzzles. We define *step index* as the sequential position of each articulated depth-1 move within a participant's solution process.

The positive associations with verbal metrics suggest that puzzles judged hard induce more overt cognitive activity, but because we lack experimental manipulation we cannot say whether

difficulty causes extra verbalization or whether extensive self-explanation makes a puzzle feel harder. The absence of correlation to model depth reinforces the earlier theme: verbalization-derived metrics and model-derived metrics are different.

Observations and Examples

To complement the quantitative analyses, we provide representative examples highlighting notable qualitative aspects of verbalized planning behaviors.

Participants frequently demonstrated limited working memory capacity, struggling to retain sequences of moves, thus repeatedly revisiting prior articulations. Example:

> "So if I go here [31], then white goes there [4], then I'm thinking this one [23]. I go here [31], white goes there [4], I go here [23]... I'm just looking at those squares to hold them in working memory... Gosh, this is so many steps... I don't have the visual imagery for this... I might have uh missed something here. Just, very carefully, this square is black [31], this square is white [4]... now they have three... "

Participants were explicitly instructed to verbalize everything that went through their minds, which naturally led to verbalization of random thoughts. Such occurrences reflect the nature of concurrent verbalization protocols. Example:

> "I think it's gotta be something diagonal… It throws people off, my shoulders pop so hard and hurt... I'm thinking about the salad that I'm gonna get after this... Now, okay, I need to focus back on track... Oh crap now I'm thinking about friend group drama... I think my brain has solved it and I need to just listen."

Some participants plan through elimination rather than explicitly plan out the final selected action. Example:

> "So then it has to be this one, this middle one, definitely has to be it, because, well, all the other threats don't work."

This qualitative examples highlights nuanced cognitive processes underpinning planning behaviors that computational modeling through observable behaviors may overlook.

3.4 Discussion

By combining concurrent verbalization with quantitatively-fitted process models of tree search, the present study investigate whether two classic methodological traditions – think-aloud protocols and computational modeling – converge on the same story about how people plan. They do not. Verbalized planning depth bore little relation to the depth estimated from a cognitive model that reliably predicts human moves, even though both measures were obtained in the same participants solving the same Four-in-a-Row puzzles.

Why might depth estimates from verbalizations and model predictions diverge? One plausible explanation is that verbalization itself serves as a cognitive filter. Concurrent reports inherently depend on conscious awareness, potentially excluding automatized or image-based computations crucial to planning but inaccessible to introspection [57, 65, 145]. Alternatively, the divergence could stem from the model's assumption of best-first search guided by a stable value function, which may effectively capture choice outcomes but not reflect the heterogeneity or complexity of actual cognitive processes. Indeed, our think-aloud data revealed substantial variation in strategy use: some participants produced explicit tree-like searches akin to early chess studies [51, 141], others performed minimal explicit look-ahead, and still others relied on heuristics or satisficing strategies outside the model's assumptions [194]. Echoing resource-rational perspectives [119], such varied strategies may reflect adaptive trade-offs between cognitive effort and expected utility, in contrast to the model's assumption of a uniform search strategy across puzzles.

To better capture the complexity of planning process, future models could incorporate flexible parameters such as adaptive depth selection, progressive deepening, or explicit working memory constraints. Think-aloud protocols offer concrete data to inform such modifications. For example,

the frequent re-articulation of the same depth-1 move points to working-memory limits, and verbal indicators of subjective difficulty might be used to dynamically adjust search depth or pruning strategies during particularly challenging puzzles. The current model extrapolates from a participant's typical free-play behavior and therefore misses effortful planning episodes triggered by especially taxing puzzles. Adaptive models whose search parameters depend on difficulty estimates may capture this coupling more faithfully. Embedding these constraints inspired by think-aloud data might yield models that better match cognitive processes during planning.

At the same time, two other verbal indices – the sheer amount of speech and the articulated "tree size" – predicted playing strength, highlighting information that would have been invisible had we relied on behavioral choices alone. The positive link between articulation volume (sentences, tree size) and Elo strength aligns with earlier work showing that deliberate self-explanation often accompanies expertise [57]. Importantly, our design is purely correlational: extensive verbalization might facilitate stronger play by externalizing intermediate representations, or stronger players might simply have more to say. Experimental manipulations that encourage or suppress verbal output will be required to clarify between these explanations.

Our finding that perceived difficulty tracks verbal – but not model-derived depth – suggests the value of measuring subjective experience. Participants talked more and looked marginally deeper when a position felt hard. However, the causal arrow is unclear. Talking aloud may inflate subjective difficulty by spotlighting uncertainties, or genuine difficulty may elicit more overt reasoning. A within-subject design that experimentally manipulates verbalization (e.g. silent vs. prompted explanation) could disentangle these routes.

Several limitations temper the conclusions. First, the sample (N = 34) consisted mainly of novice Four-in-a-Row players; planning in experts may look different. Second, think-aloud protocols are known to miss unconscious processes[65, 179]. Third, verbal data were painstakingly hand-coded; future work should explore automated transcription and coding via large language models to scale up sample sizes. Fourth, all analyses are cross-sectional and correlational; causal

claims about how verbalization metrics and performance interact remain speculative.

These caveats point to several avenues for future research. First, developing joint generative models that simultaneously fit verbal and choice data, treating verbalizations as informative but noisy indicators of internal cognitive states. Second, employing verbal markers of working-memory load and perceived effort could dynamically adjust computational model parameters such as search depth and breadth on a trial-by-trial basis. Finally, multimodal triangulation—combining verbal data with eye-tracking—could identify when unspoken computations occur, overcoming verbalization filtering effects and clarifying the relationship between verbal and hidden cognitive processes.

In sum, our results cast doubt on the common assumption that people always think in trees during planning. Think-aloud data reveal a diversity of strategies and cognitive processes invisible to purely behavioral assessments. Computational models will thus greatly benefit from incorporating verbal protocols not as historical curiosities but as integral constraints guiding the development of process-level accounts of human planning.

While think-aloud data reveal rich strategic differences across individuals, they also highlight gaps in our existing, hand-crafted cognitive models. In the next chapter, we turn to modern machine learning approaches to see if they can better predict human behavior than existing state-of-the-art cognitive model and inspire better cognitive modeling.

Think-aloud data revealed rich, idiosyncratic strategies of human planning. The diversity of strategies might suggest that human planning exhibit long-term strategic biases that current cognitive model fails to capture. To move beyond description we need a flexible function approximator that can absorb those long-range, compelx dependencies. In the next chapter, we introduce modern machine learning approaches –specifically transformer models – to test whether they better predict human behavior than the current state-of-the-art models and inspire more accurate cognitive modeling.

4 | Are humans Markovian planners?

4.1 Introduction

Recently, there has been a growing interest in employing sophisticated computational models to elucidate cognitive mechanisms underlying complex decision-making tasks [43, 93, 128, 129, 156]. To create accurate cognitive models of planning, *games* have shown to be a great testing ground [7]. In particular, games offer an environment that feels intuitive and enjoyable for players, while offering a flexible platform that allows researchers to study complex planning through a well-defined set of rules that encode a task. Critically, many games allow experimenters to scale task complexity far beyond what is feasible in traditional psychological tasks while retaining full experimental control [7].

Four-in-a-Row exemplifies these advantages. The game's state space is large enough to elicit non-trivial planning, but small enough to allow tractable analysis and modeling. Previous cognitive models of Four-in-a-Row have relied on a hand-crafted value function or fully connected neural networks that predict the next human move from the *current* board state alone [112, 148]. Such models inherit a Markov assumption: the decision at time t depends only on state t. Although this assumption simplifies modeling, it ignores the possibility that human players carry forward latent plans, sub-goals, or stylistic biases that unfold over many moves. Are human decisions in fact Markovian, or do they exhibit systematic dependencies on move history?

Transformer architectures [218] offer a natural way to answer this question. Designed for

sequential data, transformers have revolutionized language processing and have recently begun to reshape reinforcement learning and sequential decision-making [31, 39, 98, 120]. Their self-attention mechanism can integrate information over arbitrarily long contexts, making them ideal for detecting long-horizon structure in gameplay. Moreover, transformers have proven adept at *imitating* human behavior in complex domains, opening new avenues for cognitive modeling [182].

While neural network architectures have historically been deployed in games primarily to achieve superhuman performance—most notably through agents such as AlphaZero and AlphaStar [10, 190] – some recent research has shifted focus toward networks designed explicitly to *mimic* human decision-making processes [128, 129, 156]. This shift from optimizing perfect gameplay toward capturing human behavioral fidelity aligns directly with our objective: to elucidate the cognitive mechanisms that underlie human decision-making in Four-in-a-Row.

In this chapter, we introduce **GPT-4IAR**, a transformer network trained to predict human play in Four-in-a-Row. Unlike typical reinforcement learning agents aiming at optimal play, our goal is to replicate human decision-making by conditioning predictions on sequences of moves rather than isolated board states, with the goal of producing a transformer model we can probe to refine cognitive models and theoreis of planning. GPT-4IAR conditions its predictions on longer sequences of previous moves, allowing it to capture and leverage long-term strategic biases evident in human gameplay. In addition to predicting moves, we also explore the network's capacity to predict reaction times, further extending its applicability for modeling complex human behavioral statistics. Our contributions are two:

1. We develop GPT-4IAR, a transformer architecture that ingests tokenized state–action sequences and jointly predicts the next move and an estimate of decision latency.

2. Empirical evidence demonstrating significant predictive advantages of using extended historical contexts, providing strong evidence against the adequacy of Markovian assumptions

for human gameplay in Four-in-a-Row

Collectively, our results establish transformer-based emulators as powerful tools for analyzing human planning behaviors and for benchmarking and refining simpler, more tractable cognitive models.

1. We assemble a large corpus of human Four-in-a-Row games, spanning a wide range of skill levels.

2. We develop GPT-4IAR, a transformer architecture that ingests tokenized state–action sequences and jointly predicts the next move and an estimate of decision latency.

3. We show that longer historical context yields substantial gains in both move and reaction-time prediction, implying that human planning in Four-in-a-Row is better predicted by a non-Markovian process.

These results position transformer-based emulators as a powerful tool for dissecting human planning and for benchmarking simplified cognitive models.

4.2 Methods

In this section, we describe the implementation details of GPT-4IAR, the transformer architecture we developed to tackle human behavioral mimicry in 4IAR. In this work, we are interested in predicting two things: the *action*, which is the square where the player is going to place their next piece, and the *reaction time* (RT), which is the amount of time that a player takes to take an action, measured in seconds.

Data Representation

Inspired by the Trajectory Transformer [98], we decided on a simple tokenization approach to represent the boards, actions, and RTs as tokens, which we can then feed to a standard GPT

architecture with little modification. Figure 4.1 summarizes the process for the tokenization of one *round*, which is composed of one board state and the corresponding action and RT.

We represent each board state **b** as a vector of nine entries, one for each column on the board. Each entry corresponds to a base-3 representation of the respective column, as each square on the board has three possible states (empty, black, white), and each column is given an offset to induce an ordering in the vector representation. Then, each action a is represented by a single scalar. Since the board is composed of $4 \times 9 = 36$ squares, there are 36 possible actions, each of which is represented by a single integer value. Finally, due to the nature of tokenization, each RT t must be discretized. To do this, we took the full empirical distribution of RTs from the entire dataset and binned it into twenty quantiles of equal probability, each having 5% of the total probability mass. We can then use these quantiles as boundaries for each of our bins, which we use to decide which token to assign to a given t. Thus, for GPT-4IAR one round comprises 11 tokens: 9 tokens for the board state, 1 token for the action, and 1 token for the RT.

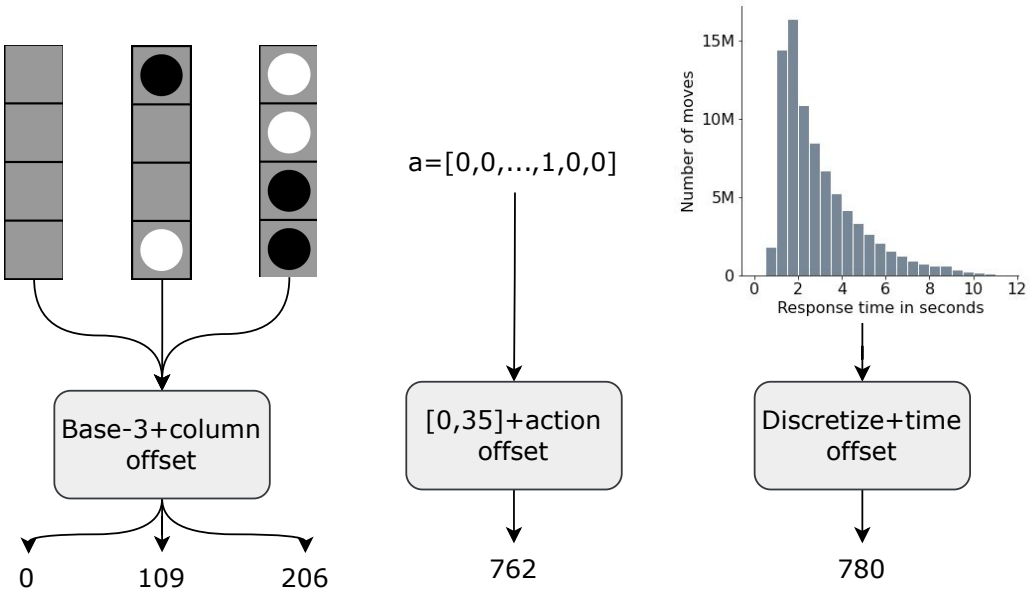


Figure 4.1: Tokenization scheme of board states, actions, and discretized reaction times (RTs).

We define a *trajectory* τ, following the convention used by the Trajectory and Decision Trans-

formers, as a sequence of rounds:

$$\tau = (\mathbf{b}_1, a_1, t_1, \ldots, \mathbf{b}_T, a_T, t_T)\,,$$

where the indices $i \in [1, T]$ indicate the number of the round played, and T is the current or latest board state. Note that in our representation the trajectory only includes the rounds – board $\mathbf{b}_i$ and actions a_i, t_i – of the human user (black pieces). The next board in the sequence, $\mathbf{b}_{i+1}$, already includes the move of the AI opponent (white pieces) as a response to what the user did in the i-th round.

Network Architecture

A general diagram of the architecture of GPT-4IAR is shown in Figure 4.2. Essentially, we follow the architecture of GPT-2 [165], with the sequence of bespoke tokens we have described in the previous section replacing the string-based tokenization used by text-based large language models.

As a training objective, we use a weighted mean cross-entropy loss, assigning a weight of 1 to the action and RT tokens, and $\frac{1}{9}$ to each of the nine board state tokens. Even though we are not interested in predicting board states per se, our preliminary analyses showed that including board states in the learning objective with a small weight achieves overall better predictive performance than having no weight (and also better than unit weight), possibly by helping the network learn an explicit board representation as well as some opponent modeling.

To train the network, we use a 90/5/5 train/validate/test split on our dataset of 10 million games. We use the AdamW optimizer [121] to minimize the target loss with parameters $\alpha = 6 \cdot 10^{-4}$, $\beta_1 = 0.9$, $\beta_2 = 0.95$ and $\lambda = 0.1$, which are the default parameters of the open-source implementation we base GPT-4IAR on.

For model assessment, we go through the whole test set to gather the evaluation metrics.

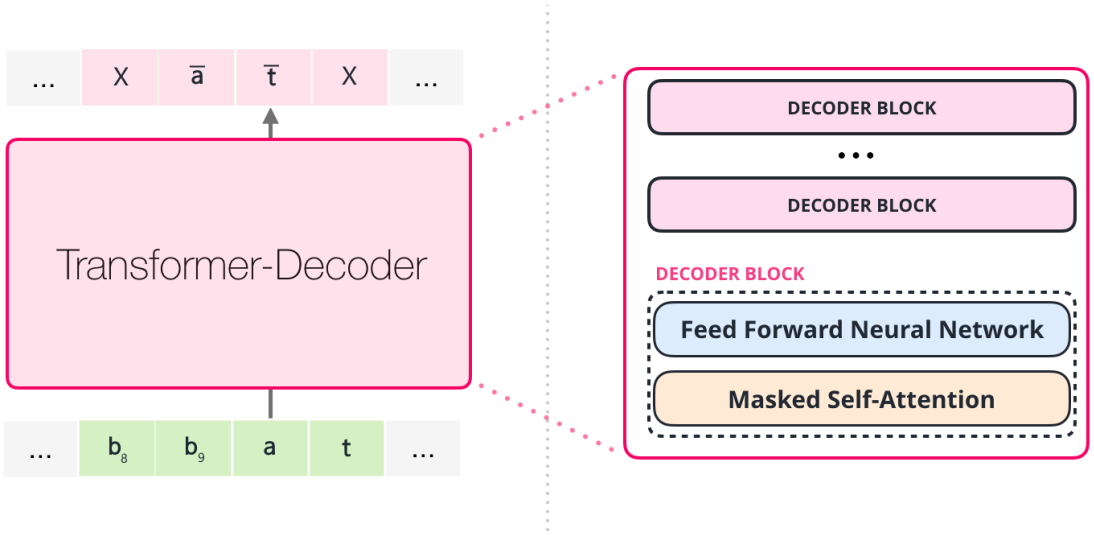


Figure 4.2: GPT-4IAR architecture. The transformer network is trained on predicting the next token in the sequence (board state, actions and RTs). Figure adapted from "The Illustrated GPT-2" [5].

Given a board $\mathbf{b}$, the network outputs a probability distribution over all action tokens which is used to compute the cross-entropy loss. We also pick the most likely token as our action prediction used to compute accuracy. Similarly, we input a board and an action $(\mathbf{b}, a)$ to extract a probability distribution over the RT tokens. To measure accuracy, we pick the most likely RT. Accuracy rate may not be the best metric by which to assess RT prediction, since RT is a (discretized) metric continuum. As such, we also evaluate RT prediction through root-mean-square error (RMSE), i.e., distance between our model prediction and the data. For this, we calculate the expected value of the RT token as our prediction and then compute the RMSE with respect to the true RT for each data point, and we aggregate all errors with an average. Losses are calculated separately for action tokens and RT tokens, which are averaged over the test set to give the final reported values.

For all experiments, unless stated otherwise, the network hyperparameters were fixed to the standard GPT-2 values shown in Table 4.1.

Hyperparameter	Value
Embedding dimensionality	768
Layers	12
Attention Heads	12

Table 4.1: Fixed hyperparameters used for training.

4.3 Results

In this section we report our preliminary results with GPT-4IAR. While thorough statistical testing and additional experiments are needed to draw more definite conclusions, several trends can already be observed in our experiments. With the architecture, tokenization, and training regime in place, we test two predictions: (i) longer context windows should boost move prediction accuracy, (ii) the same contextual information should allow reaction-time prediction.

Training Networks with Different Context Lengths

In order to evaluate the performance of GPT-4IAR at predicting human behavior when more past information is potentially available both during training and at test time, we trained three networks that differed in their *context length* (also known as context window) – the maximum token sequence that the network can process –, all else being equal. We trained three networks with context lengths of 256, 512, and 1024 tokens, respectively. The loss curves from training are shown in Figure 4.3.

The curves show an asymptotic reduction of both training and validation loss as a function of context length (lowest loss achieved by the different colored lines), suggesting that the network is able to extract more information from observing further into the past to improve predictions. Moreover, it seems likely that further extending the context length would still improve performance. However, in practice there are well-known computational limitations to implementations with longer context windows due to the quadratic scaling of the standard attention mechanism

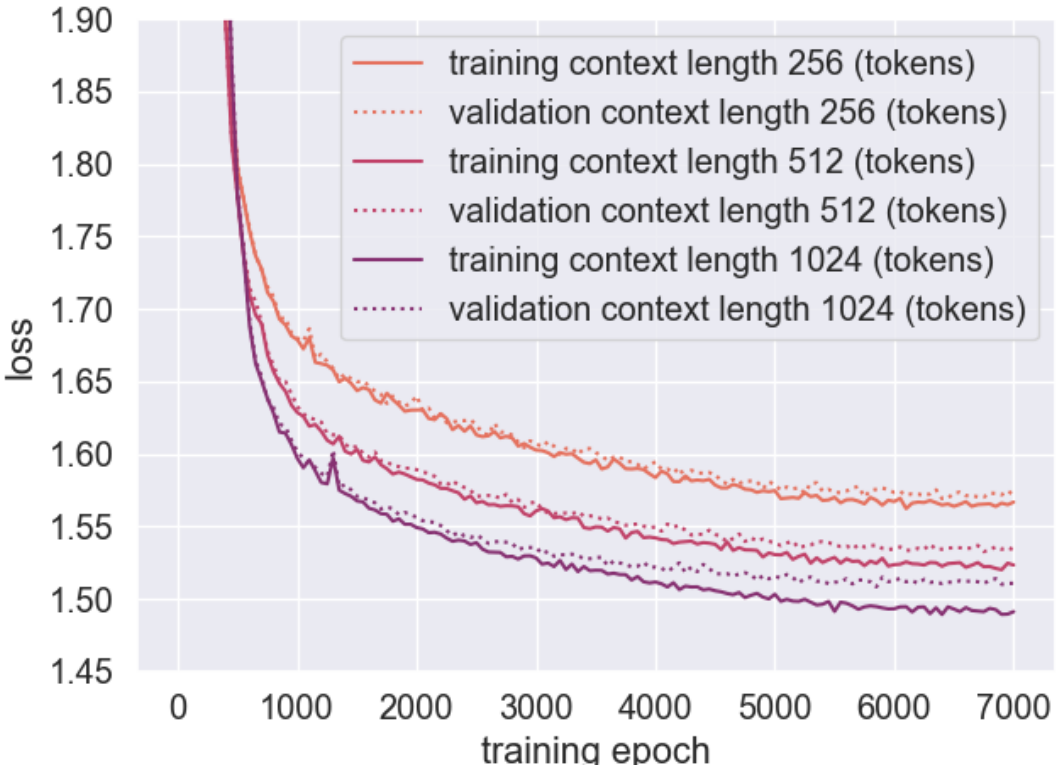


Figure 4.3: Training and validation loss for GPT-4IAR networks with different context lengths.

as a function of context length [218].

Performance Under Different Context Lengths

Now that we have trained networks with different context lengths, we can evaluate the role of *sequence length* on performance at inference time, by systematically varying the length of the sequence of past rounds the network can use to make predictions. Clearly, the maximum number of rounds each network can process is limited by its context length. Remember that a round is 11 tokens long, so our networks can store in context from up to 23 rounds for our smallest network to up to 93 rounds for our largest network. Considering that a game is on average 7.3 rounds in our dataset, even our smallest network can store in context a few past games.

Moreover, we provide comparisons with the previous state-of-the-art prediction results [112],

particularly in terms of accuracy on next-move-prediction on the test set, and some qualitative assessments on the output of the network on single boards.

Action Prediction.

The accuracy of action prediction as a function of the size of a trajectory of past game states, actions and RTs received as context is shown in Figure 4.4. We observe a substantial, positive effect on prediction accuracy of increasing the sequence length. In particular, we observe an improvement of up to around $6-7\%$ (from about 41% to 48% accuracy) when we include more past rounds into the context compared to only having one round (the current board). These results ostensibly indicate that the transformer is able to better predict behavior based on long-term dependencies in decisions. Notably, performance does not seem to be plateauing, suggesting that networks could be able to exploit even longer temporal correlations.

Finally, while all GPT-4IAR models exhibit similar performance when evaluated on sequences of the same length, there is perhaps a small advantage in using larger networks trained on longer contexts. This result, if confirmed, would suggest that networks trained on longer contexts are better even when limited to short sequences. More analyses are needed to assess statistical significance of this finding.

We compare our best performing GPT-4IAR model against the previous state-of-the-art fully connected network model by Kuperwajs, Schütt, and Ma [112] in Table 4.2.

Prediction	Metric	Fully Connected	GPT-4IAR
Actions	Accuracy	41.71 %	48.08 %
	Loss	1.866	1.504
RTs	Accuracy	—	14.69 %
	Loss	—	1.508
	RMSE	—	5.16 bins

Table 4.2: Comparison between the fully connected model and GPT-4IAR at predicting actions and RTs, using different metrics: accuracy, cross-entropy loss, and root mean squared error (RMSE). Only GPT-4IAR predicts RTs.

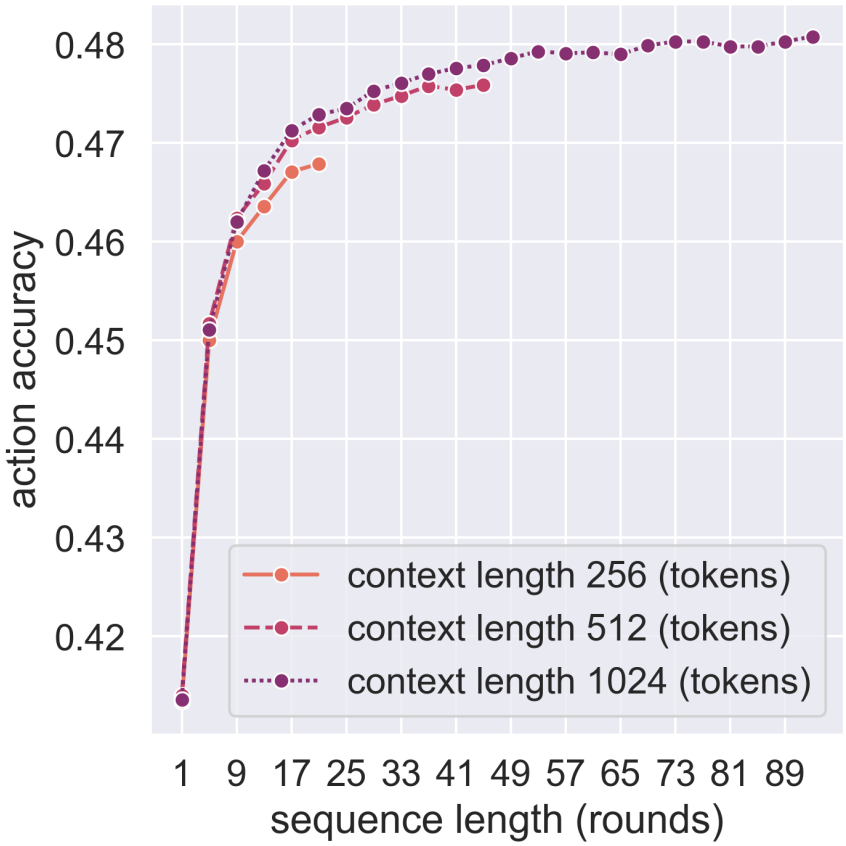


Figure 4.4: Action prediction accuracy of GPT-4IAR models with different context length as a function of provided sequence length (one round = 11 tokens).

Some examples of the actions predicted by GPT-4IAR are shown in Figure 4.5. Qualitatively, observing Figure 4.5(a), we can see that the network is able to capture lapses in human gameplay. The optimal move would be to block white from connecting four on a diagonal, but we can see a low, but non-zero probability of making moves that would try to develop a win condition for black. In the board shown in Figure 4.5(b) we can also see that the network is able to capture uncertainty on "harder" boards too, with diffuse probabilities across the board, e.g. on the second row, seventh and ninth column, to make a decision.

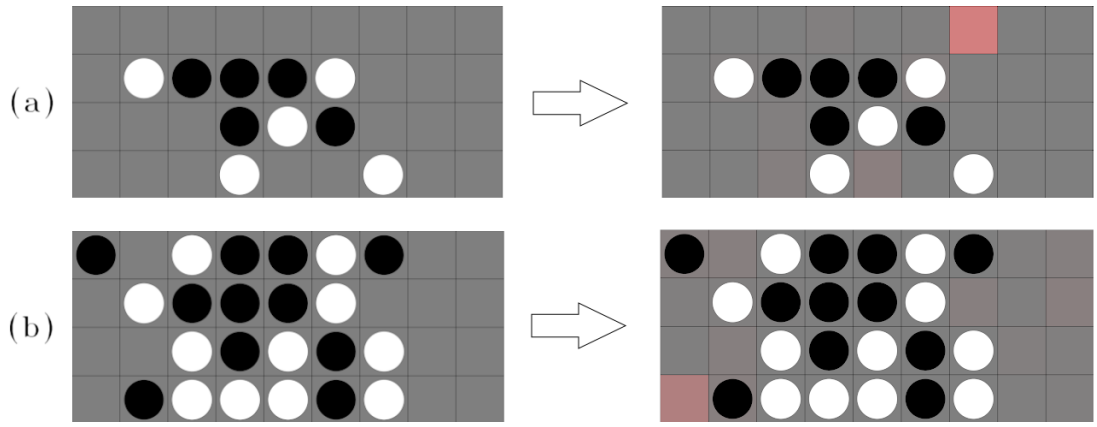


Figure 4.5: Example distributions of predicted moves (right) given two distinct input boards (left). Intensity of red in each square indicates the action probability assigned by the GPT-4IAR network.

Reaction Time Prediction.

We also evaluate the performance of the network at predicting the reaction time of a player. First, we measure prediction accuracy by choosing the most likely RT token given a trajectory of past boards, moves and RTs. RT prediction accuracy improves with the length of the provided sequence, as shown in Figure 4.6, reaching a maximum of approximately 14.69%.

Since RT is a continuous variable, accuracy may not necessarily be the best measure of performance for prediction, e.g. it may still be acceptable if we predict one bin up or down from the most likely value. Hence, we also study the RMSE of RT prediction, measured in terms of bin distance, shown in Figure 4.7. Similarly to the accuracy results, RMSE improves as a function of sequence length. For reference, the RMSE of the RT data with respect to a constant prediction is 6.66 bins.

4.4 Discussion

This chapter introduced GPT-4IAR, a transformer that learns to predict human play in Four-in-a-Row by conditioning on a rich history of prior moves. Three main insights emerge.

First, move context matters. If decisions depended only on the current state, additional move

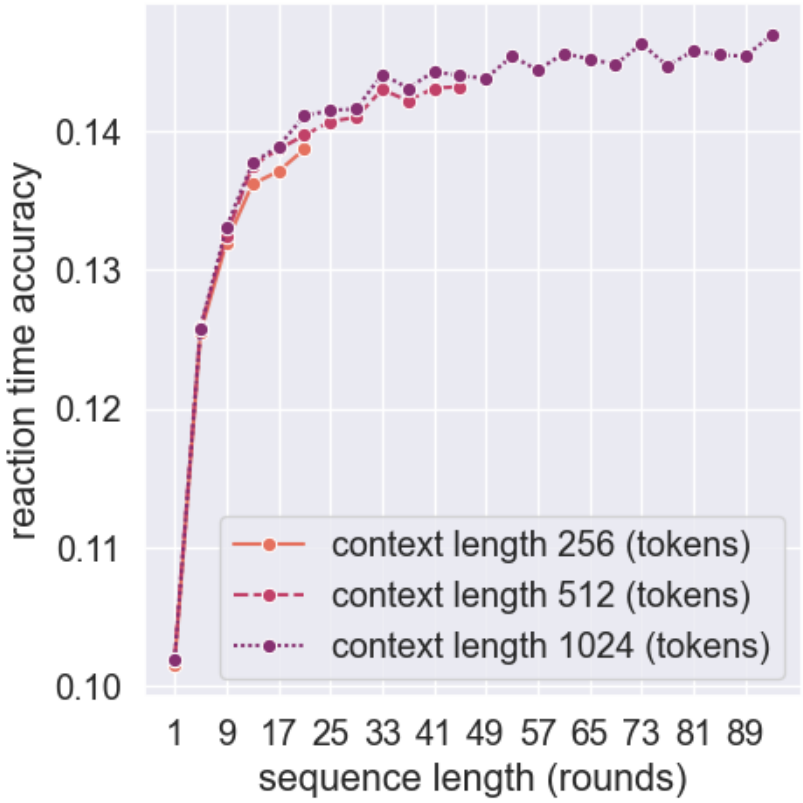


Figure 4.6: Reaction time prediction accuracy of GPT-4IAR models with different context length as a function of provided sequence length (one round = 11 tokens).

history would add no predictive power. Yet GPT-4IAR's action accuracy climbs steadily with extra context. Our findings indicate a clear advantage of considering long-term move sequences, reinforcing the hypothesis that human players employ intricate strategic biases and long-horizon planning.

Second, GPT-4IAR predicts not only which move a player will choose but also how long that choice will take. The network's ability to predict reaction times further enriches its utility, highlighting possible interactions between decision complexity, player expertise, and cognitive processing speed that researchers might be able to probe.

Third, an essential implication of our work is the potential of transformer-based models to serve as emulator of human behavior. By providing more accurate human-like gameplay predic-

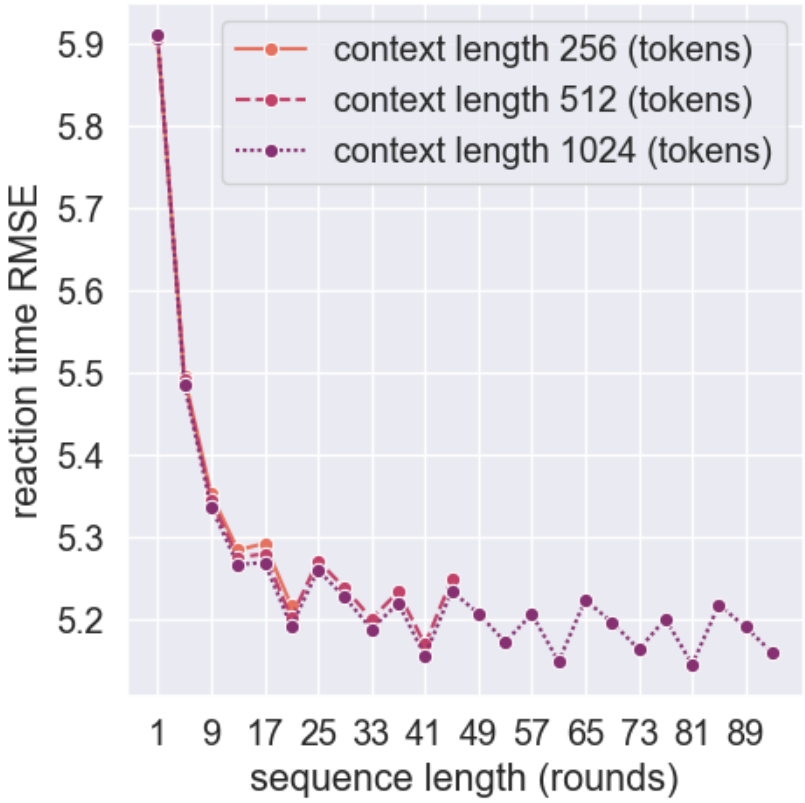


Figure 4.7: Reaction time root mean squared error (RMSE) as a function of provided sequence length, measured in bins.

tions, GPT-4IAR enables detailed comparative analyses against interpretable cognitive models. The model opened a two-way street between black-box deep nets and interpretable cognitive theories. By comparing GPT-4IAR's predictions against classic process models, we can locate mismatches that inspire principled refinements to those cognitive models.

This study has three immediate limitations. First, hyper-parameter scope: apart from context length, we left learning rates, embedding sizes, and label-smoothing untouched; second, move encoding: we currently flatten each board into a 1-D token string, but graph encodings or 2-D convolutional patches passed to the transformer might capture spatial regularities more efficiently. Third, behavioral prediction breadth: the model now predicts moves and latencies only; extending our framework to predict other human behavioral metrics, such as skill ratings (e.g.,

Elo scores), could further augment our understanding of human planning.

Finally, the insights and methodologies presented here hold broader applicability across various combinatorial games and cognitive tasks. Integrating transformer architectures with tree-search methods, as explored in parallel research on “tree-of-thought" for LLM reasoning [229], could further enrich computational models of human planning and decision-making processes.

Having seen how machine learning can enhance our understanding of human planning with extensive human data, we now invert the perspective. Instead of asking whether an AI tool can mimic human behavior, we ask what we can learn from the learning dynamics of AlphaZero, a self-optimizing agent.

5 | What machines can learn from humans? Lessons from AlphaZero

5.1 Introduction

For decades, research on artificial and human intelligence has advanced in parallel, each field offering models, theories, or data that help the other [113, 142, 211]. Deep learning has intensified this exchange: convolutional and recurrent networks reproduce hallmarks of human vision [227] and reward learning [220], while neuroscientists adopt network-level analyses inspired by AI to probe biological brains [172]. Yet the experimental substrates on which the two disciplines test their ideas often sit at opposite ends of a complexity spectrum. While complex board games such as Chess and Go serve as grand challenges for AI research [190, 192], they remain intractable for detailed human behavior modeling. Conversely, traditional cognitive planning studies favor simpler tasks (e.g., the two-step task [50]) that, while tractable, may be too simple to capture the strategic complexities observed in humans.

A fruitful middle ground emerges in tasks of *intermediate complexity*: challenging enough to demand non-trivial planning, yet still amenable to computational analysis. Again, we used Four-in-a-Row. The AlphaZero family learns two coupled functions from self-play—*policy* and *value*—and fuses them with Monte-Carlo Tree Search (MCTS) to select moves [190, 192]. Although its success on Go, Chess, and Shogi is well documented, much less is known about how

its internal planning signals evolve in a domain where human cognition can be measured at comparable resolution. Recent evidence in Chess suggests blind spots: AlphaZero struggles on specialized puzzles requiring short, forced sequences of moves that are easy for human players to reason through in Chess[200].

In this chapter, we asked two questions: What exactly does it learn through self-play, and where does its strategy break down? To answer the first question, we track the trajectory of policy quality, value accuracy, and search depth across training checkpoints, and we use targeted "swap" experiments (e.g. replacing a strong value network with a weak one) to isolate each component's causal impact on playing strength. We show that all these components improve with training in a manner partially reminiscent of human learners. In addition, we dissect the feature representations in AlphaZero's hidden layers to investigate its learned features.

To answer the second question, we curate a battery of forced-win puzzle positions that require short, forced sequences. Despite near-perfect performance in standard games, trained agents fail on 93% of these puzzles. Augmenting the search with human-inspired heuristics rescues 15% of the errors, underscoring gaps in the learned evaluation function.

We close with a Discussion that places these results in the broader context of human planning, considering plausible differences and similarities in how humans and machines allocate resources to searching deeper or searching "smarter".

5.2 Method

Our overarching goal is to examine AlphaZero's learning dynamics in Four-in-a-row and compare them to known human data. We also investigate puzzle-like scenarios to test AlphaZero's ability to plan in short forced-move puzzles.

AlphaZero

The deep RL agents trained to play Four-in-a-row are slightly modified versions of AlphaGo Zero [192] and AlphaZero [190]. The modifications we made should not affect our main conclusions (details in later sections). The agent consists of two components, a DNN, for evaluating boards and providing a policy/prior over actions; and MCTS, for simulating future outcomes and making a final decision. We recapitulate only the important aspects here. For full detail, see [192].

Neural Network Architecture

The DNN takes a board and optionally color as inputs, passes through three or nine residual blocks and then two fully-connected layers before outputting a policy / action prior and a value (Figure 5.1. Notice that the original AlphaZero does include player color as an input feature, but that information is not essential to solving the game. We therefore omit it among some agents. The value can be interpreted as the chance of winning/losing for the current player, and the policy provides a prior probability of choosing each move before the tree search.

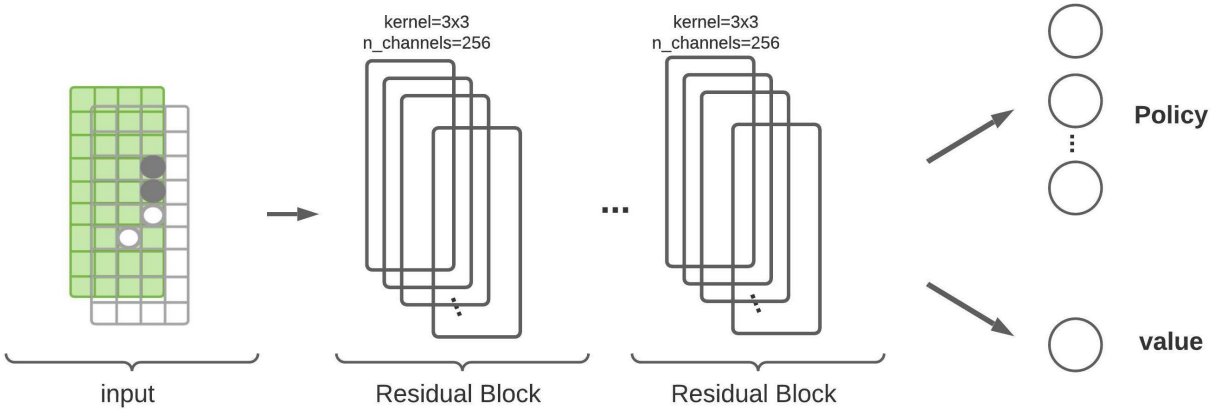


Figure 5.1: AlphaZero Network Architecture

Monte Carlo Tree Search (MCTS)

To make a move given a board, the agent builds a tree starting from the current board as the root. Nodes in the tree are board positions s and edges are legal actions ($a \in \mathcal{A}(s)$). Each edge stores the visit count $N(s,a)$ and the mean action value $Q(s,a)$, which gets updated after each search. Each simulation traverses down the tree until reaching a leaf node, by selecting actions that maximize $(Q(s,a) + c_{\text{PUCT}} P(s,a) \frac{\sqrt{\sum_b N(s,b)}}{1+N(s,a)}$, where $P(s,a)$ is the action prior given by the DNN and c_{PUCT} is a hyperparameter controlling the exploration-exploitation trade-off. This selection criterion favors moves with a high prior, low visit count, and high mean action value. Once the leaf node is reached, the algorithm expands the tree by selecting the action with the highest prior. It then evaluates the immediate value of the result of this action using the DNN and backpropagates this value to update the Qs and Ns of the ancestor edges. After a fixed number of searches, set as a hyperparameter N_{MCTS} (by default 100), the agent selects a move a in the root position s_0, with probability $\pi(s_0,a)$ proportional to its exponentiated visit count: $\pi(a|s_0) = \frac{N(s_0,a)^{1/\tau}}{\sum_b N(s_0,b)^{1/\tau}}$, where τ is a temperature parameter. The smaller it is, the more deterministic the move selection becomes. MCTS can be understood as improving the prior probability given by the DNN into a more sophisticated action probability π through searching into the future [208].

Training

For each training iteration, we use the current best agent to play 100 games against itself to generate the training examples. During the first 15 steps of each self-play game, the temperature is set to 1 to induce variability in the data. In the AlphaZero paper, the temperature is later set to 0. Here we include one Network whose temperature does not switch. For some Networks, during self-play, Dirichlet noise with a hyperparameter Dir_α is added to the current root of the MCTS tree to encourage exploration. We also include Networks without Dirichlet noise, which deviates from AlphaZero. The motivation for different hyperparameters is exploratory and not

systematic.

Each training example contains a tuple of (board positions s, MCTS output $\pi(a|s)$, game outcome r). The DNN value output v is trained to match the game result r under a mean-squared error loss and the policy output $\mathbf{p}$ is trained to match the action probabilities π under a cross-entropy loss, with $L2$ weight regularization. The DNN parameters are optimized by the Adam Optimizer, using the training examples from the last 20 training iterations, in mini-batches of 64 examples. During each training iteration, the DNN is trained for 10 epochs. The updated network will play 30 games against the current best. If the updated network can win more games than it loses, it will be accepted and become the current best network for data generation and network comparison. We use this looser selection criterion compared to AlphaGo Zero to encourage easier network update. Because if the updated network is not accepted, we revert it back to the current best network, forgoing the parameter update (different from AlphaGo Zero). If the “continuous training" hyperparameter is true, the updated network continues training in the next iteration, similar to AlphaGo Zero.

Measuring playing strength

We hold a tournament in which each agent plays against every other agent once as both colors. There are 789 agents in total, including all accepted iterations from the thirteen Networks, as well as the agents whose N_{MCTS} have been modified, and those whose value or quality functions have been swapped. The temperature is fixed at 0.1. Playing strength is quantified by Elo ratings, computed by the BayesElo program [48]. The Elo ratings are computed such that the difference between the Elo ratings of two players maps monotonically to the probability that one player will defeat the other.

Probe boards and game-theoretic values

The probe boards are all positions (5482 positions) which occurred in human-vs-human experiments conducted by Opheusden et al. [148]. The game-theoretic values of these boards are defined as game outcomes in which both sides play perfectly. Opheusden et al. [148] approximated the game-theoretic values by searching each board for 200,000 iterations using the cognitive model. The result for most boards converges to a game-theoretic value, while the undetermined ones are assigned a 0 value, indicating a draw. We computed the correlation between AlphaZero's value outputs and pre-computed game-theoretic values to obtain value quality. After computing the game-theoretic values for each child board of all probe boards (used in value quality and depth calculation), we applied softmax to the values of those children boards to get an "optimal" policy for each probe board. We then concatenate these optimal policies across all probe boards and correlate the resulting vector with the concatenated policy vector returned by a DNN.

Puzzles

To evaluate AlphaZero's problem-solving ability, we designed a set of puzzles derived from Four-in-a-row game states. Each puzzle presents a scenario where there is a forced win for the current player within five moves. Solving these puzzles requires constructing sequential threats and anticipating the opponent's responses, thus testing the agent's strategic planning and sequential reasoning.

5.3 Results

We first report on how AlphaZero's performance, as well as its internal planning metrics, evolve across training in Four-in-a-row self-play. We then turn to the puzzle-solving experiments

and show how AlphaZero's limitations can be partially mitigated by introducing human-inspired features.

AlphaZero's learning dynamics in Four-in-a-Row

Playing Strength and Elo Rating

AlphaZero's playing strength increases over training across all Networks (Figure 5.2A; left). To obtain a human benchmark, we have the strongest human player we could find play 4 games each against 8 selected agents, with Elo ratings ranging from 140 to 242 (the best). Agents at middle training iterations already surpass the human bench mark, with later agents lying above the 95 % confidence interval. Human learning curve from the previous study is recreated here (5.2A; right). Our question is what aspects of the agents' capacity have improved to enable such an improvement in playing strength.

Evidence for smarter trees

Prior human modeling studies in 4-in-a-row used planning metrics to explain human learning and playing strength. Value function quality measures how closely people's heuristic evaluation of a board aligns with the game-theoretic value of a board (see Methods). Planning depth reflects how many steps into the future can one look ahead. Feature dropping rate reflects how often people ignore features on the board [148]. The study showed a learning effect on value function quality when the initial quality is not too high, on planning depth, and on feature dropping rate. Since AlphaZero agents don't have human-like attentional lapses, nor are their values directly computed from explicit features (like controlling a 3-in-a-row on the board), feature dropping rate is not included in our comparisons.

For each AlphaZero agent, we compute the value function quality by calculating the Pearson correlation between the DNN-returned immediate values of the probe boards and their game-

theoretic values obtained in previous work (see Methods). We measure planning depth by having each agent make a move at probe boards, and average across the probes the length of the deepest branch of each resulting MCTS tree. Both value function quality and planning depth increase as training progresses (Figure 5.2B and C; left). We plot previous human results here to aid comparison (Figure 5.2B and C; right) [148]. Compared to human learning, the increases in planning metrics are more drastic in AlphaZero, which is expected given that human is not a blank slate to begin with.

Entropy of action prior mediates the increase in planning depth

When the total search budget is fixed, one possible mechanism for the increase in planning depth could be a more targeted and less scattered search process. In AlphaZero the targetedness of the search is largely modulated by the policy. The policy starts out uniform and evolves to match the post-MCTS action probabilities. Since the search process makes the action probabilities be less uniform, the policy should become less uniform and thus have a lower entropy over training, defined as $H(s) = -\sum_a p(a|s) \log p(a|s)$. A decrease in entropy over training is confirmed in Figure 5.3B. (Similar to planning depth, the entropy here is also averaged across probe boards.)

Policy quality improves

A more concentrated prior does not necessarily imply “smarter" searches. A bad prior can lead a deep but misguided search. We therefore develop a metric, policy quality, to quantify how good AlphaZero’s policies are. Policy quality reflects the correlation between AlphaZero’s policies and the optimal policies, derived from the game-theoretic values (see Methods). The policy quality improves over training for all Networks (Figure 5.3A). So not only are the priors more concentrated, but they also align better with optimal policies, and thus lead the search in more promising directions.

The increase in playing strength is mediated by planning metrics

Playing strength of AlphaZero agents increases with training (Fig 5.2A; left), and we hypothesize that the effect of training on playing strength is mediated by planning metrics. Mediation analysis shows that policy quality, value function quality and planning depth all act as significant mediators of Elo ratings (Figure 5.4). (We test the significance using bootstrapping procedures.)

Policy quality matters the most

Mediation analysis on each planning metric shows that learning-induced Elo changes are mediated by planning metrics. However we cannot reliably conclude the relative contribution of each metric, since the metrics are correlated with each other (value-policy:0.94, value-depth:0.80, depth-policy:0.85). We first demonstrate the dominant contribution of policy quality to performance through observational data and then in the later sections use causal manipulations to dissect the role of value function quality and planning depth.

Policy quality, planning depth and value function quality together explain 0.95 of the variance in Elo in a linear regression, with weights $\beta_{policy} = 0.92$ ($p < 10^{-20}$), $\beta_{depth} = 0.09$ ($p < 10^{-5}$), and $\beta_{value} = -0.03$ ($p = 0.334$), respectively. We also test the dominance of policy quality by first regressing out all the other confounders (training iterations, value quality and depth) from Elo. Policy quality explains the residuals significantly well ($F = 29.11$, $p < 10^{-6}$). By contrast, neither a similar residual regression for value function quality ($F = 0.18$, $p = 0.668$) nor one for depth ($F = 2.69$,$p = 0.101$) is significant.

Causal manipulations reveal contributions from all planning metrics

The planning depth and value function quality do not show significant contribution to Elo once all the confounding factors are regressed out. But this fact by itself does not rule out the possibility of their contribution. To arbitrate the role of planning depth in playing strength,

we causally manipulate planning depth for each iteration in the best Network, while holding everything else about an agent constant. To do this, we replicate an agent and then set its number of MCTS searches (N_{MCTS}) to four different levels. The resulting agents are then included in the tournament. For the same iteration (dots connected by the same line in Figure 5.5A.), a higher N_{MCTS} induces a high planning depth, which correlates positively with Elo in all iterations. As training progresses, the positive effect of planning depth on Elo diminishes, as seen from the decreasing of the slopes of the lines in Figure 5.5A. We perform a linear regression (Elo ~ depth) for each iteration within the Network to obtain its "depth efficiency" (Figure 5.5B). The depth efficiency decreases as training progresses. One possible explanation for why the benefit of depth diminishes is that the good action priors of well-trained models are sufficient to guide actions. Adding more depth in the search might not advise major changes to the preferences provided by the prior.

For the value and policy manipulation, we select eleven models from a Network that spans early, middle and late epochs of training (this Network has the highest policy quality and is different from the one in the N_{MCTS} manipulation). We swap either the value or the policy function of a model with the value/policy function of low, middle or the best quality. These swap targets come from the initial, a middle (iter 22) or the final iteration of the model within the Network, respectively. Models from one training epoch do not have swaps with models from the same training epoch (e.g. the policy/value functions of models from early iterations (iter 1-20) will only be replaced by those from the middle and final models).

The result shows that both value and policy contribute to performance, as equipping early models with a well-trained policy or value function improves their Elo (Figure 5.6). The gain is larger with the policy swap initially, but the value swap catches up during the middle epoch (20-35 iters), suggesting value and policy quality can complement each other in this intermediate range, i.e. a good performance does not require both value and policy to be really high, but only one to be high and the other intermediate. Swapping the policy function of a well-trained model to a

naive policy is unambiguously more disruptive than swapping the value function, again echoing the previous section in terms of the overall dominance of policy. Swapping the components of the late models to those of the middle ones produces qualitatively similar but quantitatively less drastic reduction, compared to swapping to early ones. Surprisingly, replacing the value functions of the early models with those of the middle ones provides a larger improvement than swapping the policy functions. This phenomenon awaits further investigations.

Probing for Human-Used Features

To understand how AlphaZero became proficient at winning games, we used a feature probing techniques – concept activation vectors [105]. This approach allowed us to detect features used by human players, such as "3-in-a-row" and "2-in-a-row" configurations, identified by Opheusden et al. [148]. We trained classifiers using activations from specific layers of the neural network during training to predict the presence of these human-used features.

Our analysis revealed that the network acquired the crucial "3-in-a-row" feature in both the value head and intermediate layers, even without exposure to human-generated data (Figure 5.7). By contrast, the "2-in-a-row" feature was not prominently represented in the network. This suggests potential limitations in AlphaZero's ability to learn the full spectrum of strategic features used by humans.

Unsupervised Feature Representation

To further explore what AlphaZero learned through self-play without predefined concepts, we applied a well-established method, Nonnegative Matrix Factorization (NMF), to extract and visualize latent features from hidden layers [115, 127]. We concatenated activations from 14,907 random game states into a matrix and approximated it as the product of weight and feature matrices, minimizing reconstruction error. The resulting factors provided insights into the network's understanding of the game by highlighting important activation patterns.

NMF analysis revealed interpretable factors in the network's intermediate layers, even though AlphaZero was never exposed to human data. (Figure 5.8). These factors captured diagonal, vertical, and horizontal patterns, suggesting AlphaZero's ability to represent various game-relevant features that are interpretable to humans.

Puzzle Testing

Despite its strong playing strength, AlphaZero showed a surprising 93% failure rate in finding the best move to solve the puzzles. These puzzles required constructing a logical sequence of moves that forces a win within a limited number of turns. In some instances, the agent displayed overly defensive play, neglecting opportunities to build offensive threats (Figure 5.9(a)). This observation suggests a gap between AlphaZero's learned strategies and the specific reasoning path used by humans in planning ahead.

Incorporating Human-Inspired Features

We hypothesized that incorporating a cognitive value function, as described by [148], with a linear combination of human-used features, could enhance AlphaZero's puzzle-solving performance. Specifically, we added this cognitive value function output to both the policy and value outputs of AlphaZero network, leveraging features not typically captured in AlphaZero's self-learned heuristics and strategy, such as the "2-in-a-row" and "unconnected-2-in-a-row" configurations. This augmentation of the value and policy predictions led to a 15% improvement in puzzle-solving accuracy (Figure 5.9(b)). We observed that Cognitive Models demonstrated the highest performance with an accuracy of 0.28 ± 0.03. In contrast, AlphaZero Agents exhibited substantially lower accuracy, achieving 0.08 ± 0.01. When the cognitive function was incorporated into the Hybrid Agents, their performance improved to 0.21 ± 0.03. This finding highlights the potential of incorporating human cognitive insights to augment AI performance in tasks re-

quiring specific strategic reasoning path, such as solving puzzles optimally.

5.4 Discussion

Our findings provide a detailed look into the blackbox: how AlphaZero learns and where it can fail in Four-in-a-row, a game of intermediate complexity that permits direct comparisons with human data. Our results demonstrate that AlphaZero's learning in Four-in-a-row unfolds along multiple dimensions: the value function becomes more accurate, the policy becomes more selective, and the tree search depth grows. In particular, the policy proves crucial to overall playing strength, showing that a strong prior for which moves to explore can, in many cases, be more important than searching deeply. This resonates with previous work on MuZero [180] and other variants, highlighting the multifaceted nature of planning: Value provides a context-independent evaluation; policy provides a context-dependent prior over actions; The policy stands out as the primary engine of performance: once it reaches a sufficiently high quality, deep searches bring diminishing returns.

Interestingly, this contrasts with cognitive modeling results in human Four-in-a-row, where increased experience corresponds to expansions in the number of searched moves [148]. In AlphaZero, however, a more concentrated policy yields deeper exploration of each promising branch with fewer total expansions. One might speculate that humans, too, engage in "smarter" search once they develop heuristics, which resonates with a classic finding by Groot [76], who investigated the thinking processes of Chess experts versus novices. De Groot observed that while experts did not necessarily search deeper or consider a dramatically larger number of moves, they were considerably more efficient in selecting and evaluating promising lines of play. By contrast, existing cognitive models of Four-in-a-row Opheusden et al. [148] have primarily explained increases in search depth through an expanding quantity of moves considered, rather than an increasingly sophisticated or selective search mechanism. In addition, the models do

not include a component analogous to the *policy* network in AlphaZero, which guides search toward promising branches by adjusting their visit probabilities. Although these Four-in-a-row models capture many nuances of human planning, they do not yet allow for the possibility that individuals might learn to prioritize moves more effectively over time or from prior experiences.

Inspired by AlphaZero, one promising direction for future cognitive research is to incorporate a "policy-like" component in human planning models. Such a component could learn to rank or weight different branches according to heuristics gleaned from experience or from prior experience. An interesting divergence arises in late-stage skill. At high levels of expertise, AlphaZero's performance plateaus in terms of how beneficial additional search is. Future human experiments on tasks like Four-in-a-row could test whether experts also exhibit a decreased marginal benefit from deeper searching once they have robust heuristics.

Furthermore, our analysis revealed a duality: while AlphaZero successfully learned certain human-interpretable features, such as 3-in-a-row patterns, it struggled to fully capture the breadth of features employed by humans. Despite having very strong heuristics for winning games, AlphaZero's feature learning appears incomplete when compared to that of humans, as evidenced by the absence of features like 2-in-a-row in its learned representations. Despite achieving superhuman playing strength, AlphaZero struggled with puzzles requiring a logical sequence of reasoning [200]. These results point to a fundamental limitation in AlphaZero's self-play training regime: it excels at winning games but falls short in tasks requiring strategic, human-like planning. Notably, AlphaZero is not optimized for winning games in the shortest way possible. As long as it secures a victory, the efficiency of the path taken is of little consequence. Therefore, it is perhaps less surprising that AlphaZero's performance declines in scenarios where the shortest, most logical sequence is crucial. By incorporating human-inspired features into AlphaZero's policy and value estimations, we observed an improvement in puzzle-solving accuracy, demonstrating that human cognitive insights can be leveraged to enhance AI performance in tasks requiring sophisticated strategic reasoning. This improvement underscores the value of blending human intuition with

AI learning models, suggesting that hybrid approaches could help address some of the gaps in AI planning capabilities.

However, there are limitations to our approach that warrant further investigation. The hyperparameter variations are not systematic. While introducing human-inspired value function led to measurable gains in performance, it raises questions about the generalizability of these improvements. AlphaZero's enhanced performance may be specific to the puzzles tested, and those puzzles represent a narrow set of scenarios.

Taken together, our results show that AlphaZero learns both a stronger value function and a deeper, more selective search policy over the course of training in Four-in-a-row. Yet the dominant contributor to final playing strength is the policy itself: as the network becomes confident in the best moves, the value function and deep search provide diminishing but still notable improvements. In puzzle-like scenarios demanding short, forced sequences, AlphaZero's self-play training can fail to learn essential intermediate features that humans leverage routinely, though augmenting the agent with human-inspired value heuristics helps address these failures.

From a cognitive science standpoint, these findings generate new hypotheses for human studies: Do people improve search depth primarily by searching more moves, or do they also become more selective in branching, as AlphaZero does? Could humans similarly benefit from a hybrid approach that unifies an objective "value" representation with direct "policy" learning in complex tasks? Moreover, the fact that high-level strategic success does not always translate to puzzle-solving mastery highlights the importance of evaluating planning beyond raw win rates.

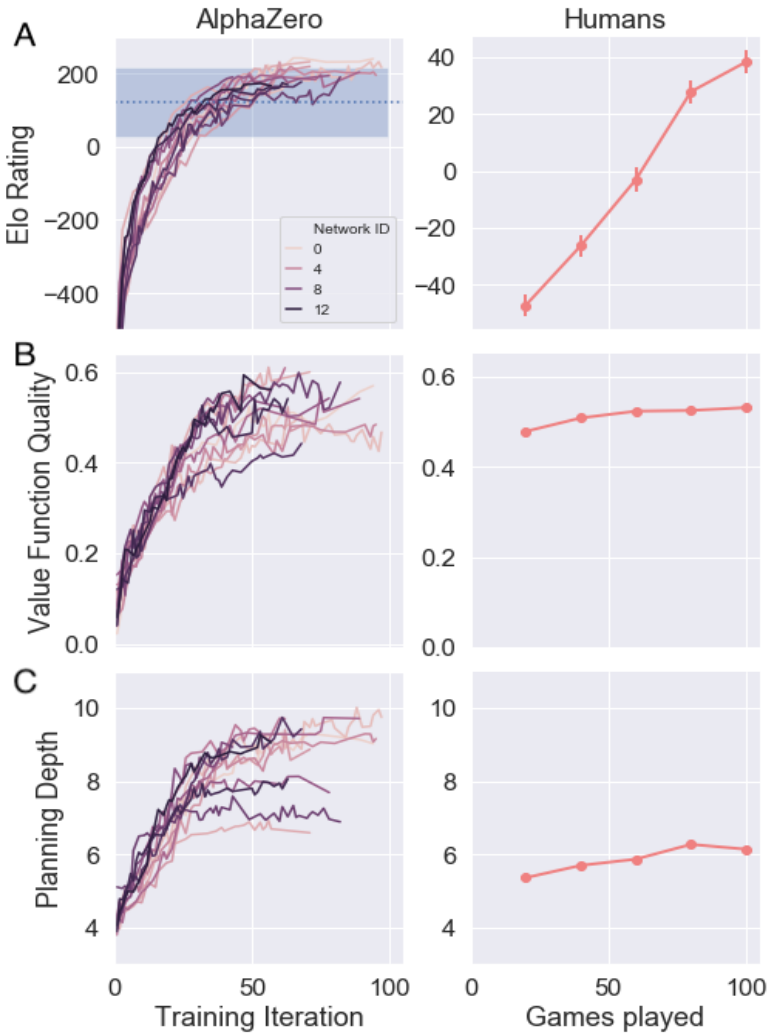


Figure 5.2: Elo and planning metric comparison between AlphaZero and human. Playing strength (Elo rating, A), value function quality (B) and planning depth (C) of both AlphaZero and human increase with training. Solid lines represent Networks. Dotted line in (A) represents the Elo of a strong human player, and the shade reflects the 95-confidence interval of this Elo estimate. Human results are reproduced from data in Opheusden et al. [148]. The scale of Elo ratings are different between AlphaZero (left) and human (right) and the numbers are not directly comparable because there is no tournament between human players from prior study with our AlphaZero agents.

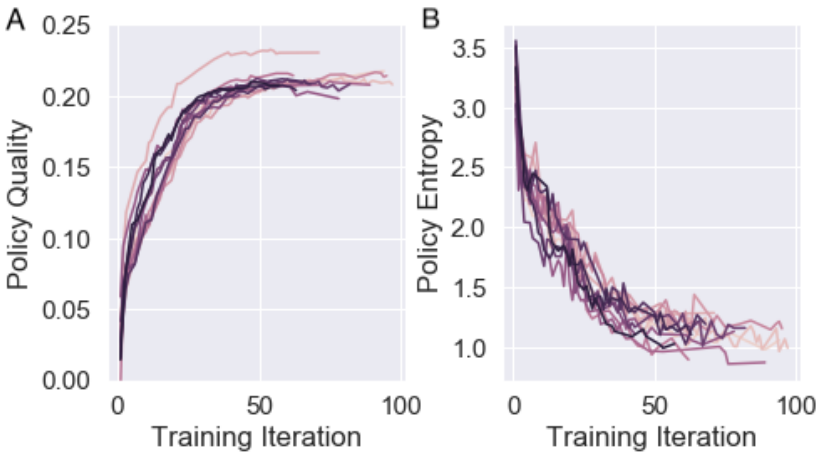


Figure 5.3: A) Policy quality of AlphaZero agents increases with training. B) Policy entropy of AlphaZero agents decreases with training

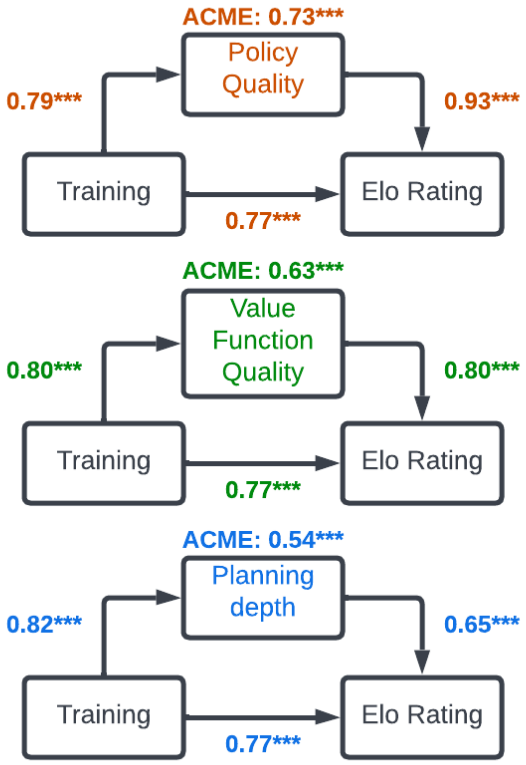


Figure 5.4: Mediation analysis: illustration and results. The effect of training on Elo ratings is mediated via three planning metrics: policy quality, value quality and planning depth. ACME is the average causal mediation effect. Numbers next to arrows represent the regression coefficients between variables. Asterisks denote statistical significance.

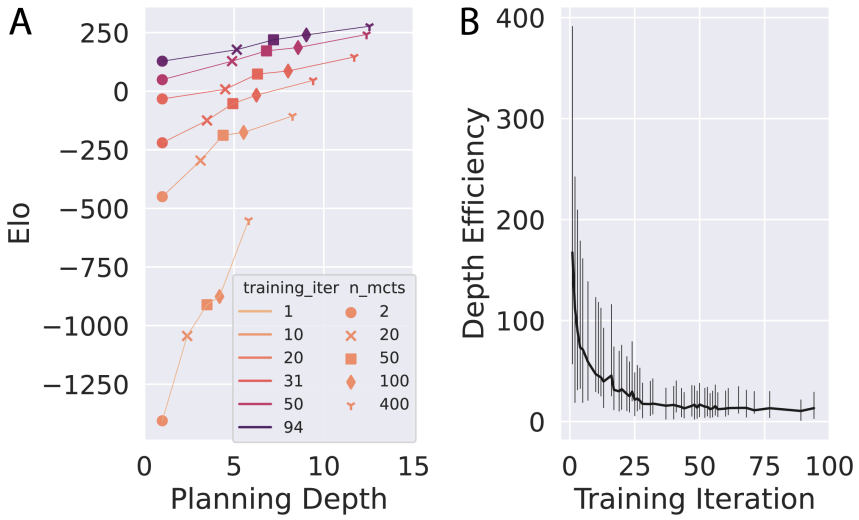


Figure 5.5: The effect of N_{MCTS} manipulation on depth and Elo. A) Elo vs planning depth for selected iterations and all N_{MCTS} manipulation of the chosen Network. Color indicates training iteration, and marker style indicates the number of MCTS searches. Agents with the same training iteration are connected by a line. B) Depth efficiency vs Training iteration for all agents. Depth efficiency is defined as the slope of each line in A (as well as the lines for other iterations not shown in A), which represents the efficiency of depth increase in increasing Elo. Error bars reflects the 95% confidence intervals.

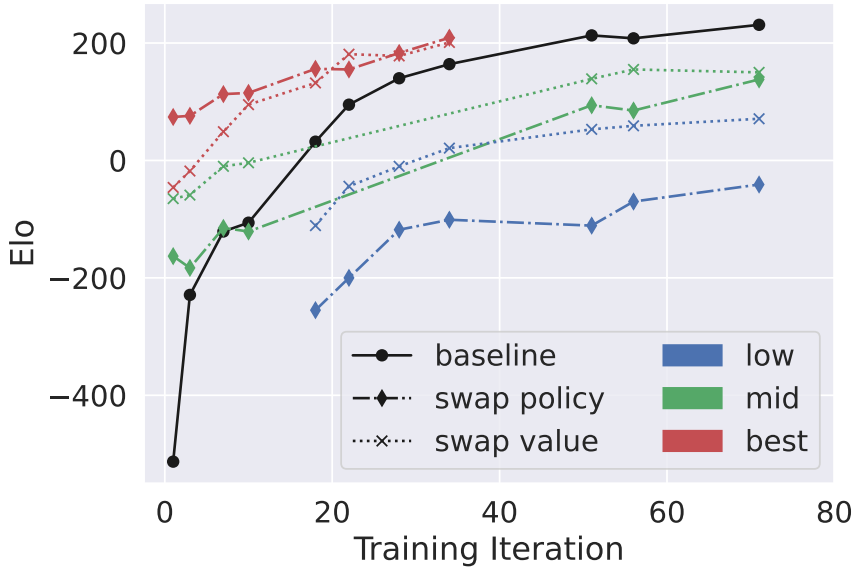


Figure 5.6: Elo ratings as a result of the value or policy function manipulation. Black markers (solid line) are the original agents across training. Colored diamond markers (half solid line) are agents with policy functions swapped. Colored cross markers (dotted lines) are agents with value functions swapped. Color reflects the quality of the target of the swap: low-blue, middle-green, best-red.

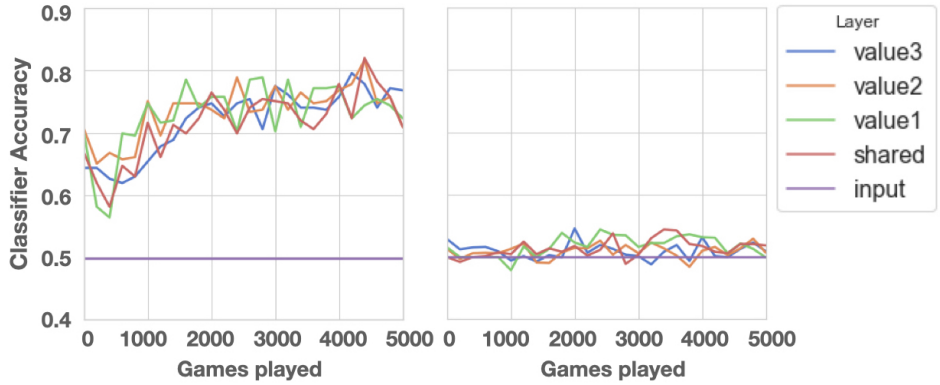


Figure 5.7: Feature Probing Analysis. Detection of "3-in-a-row" (left) versus "2-in-a-row" (right). Activations from the value head and an intermediate layer show learning of the "3-in-a-row" feature. Control inputs are included for reference.

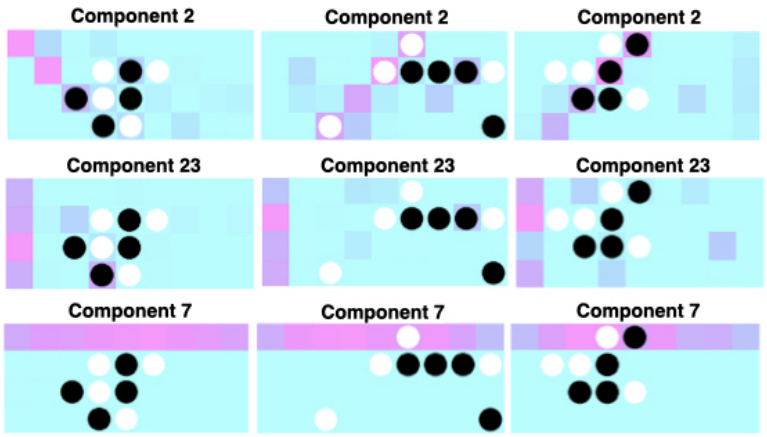
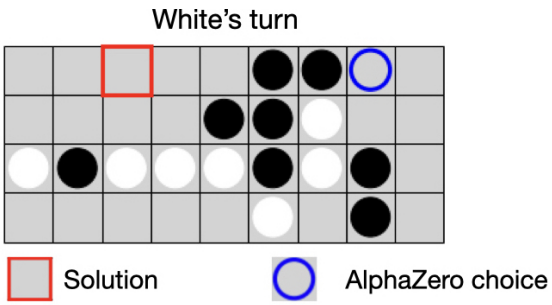
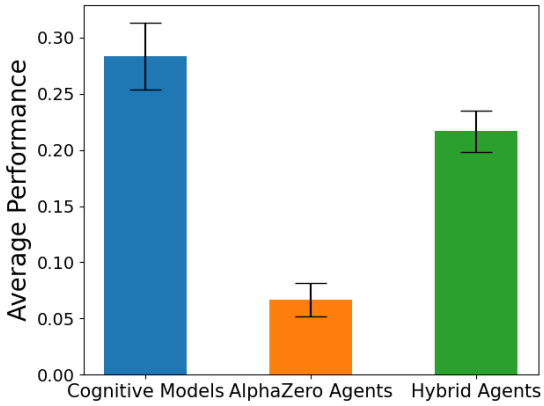


Figure 5.8: Visualization of NMF for selected factors. Panels show features captured by different residual blocks: diagonals, verticals, and horizontals.



(a) An example of AlphaZero's failure. The solution is to build a 3-in-a-row thread (highlighted in blue).



(b) Puzzle-solving accuracy comparison.

Figure 5.9: (a) An example of AlphaZero's failure and (b) Puzzle-solving accuracy comparison.

6 | Conclusion

Planning is the mind's way of living in the future: Before any overt action, we imagine several futures, reject the unpromising, and commit to the sequence most likely to close the gap between the current state and a distant goal. Yet exactly how such prospective control is realized in both brains and algorithms has remained only partly understood. This dissertation used the game *Four-in-a-Row* as a laboratory: rich enough to demand non-trivial look-ahead, yet tractable enough for formal modeling. Across four complementary chapters we asked (i) *what* cognitive components underpin planning, (ii) *how* are plans constructed, (iii) how well can powerful sequence models like transformers predict human planning behavior, and (iv) what the learning trajectory and blind spots of a state-of-the-art artificial planner reveal about planning.

Drawing the studies together yields four headline messages.

1. **Planning is an orchestration of different cognitive abilities that re-weight with task demands.** Individual differences study uncovered the component processes of planning tasks. As state spaces expand, working-memory becomes critical, reflecting the heightened demands of maintaining many future states.

2. **Human planning is richly contextual and non-Markovian.** Think-aloud data and transformer modeling uncover strategies and long-range dependencies that cognitive models with Markovian assumptions miss, demonstrating that priors such as past threats, stylistic preferences, and working-memory constraints shape upcoming moves. Those insights generate concrete hypotheses for improving existing models of planning.

3. **Expertise hinges on smarter search** AlphaZero improves playing strength primarily by learning a strong prior policy. Yet its blind spots on tactical puzzles show that neural-network-based value function still benefit from human-inspired heuristics for logical-sequence reasoning.

4. **Cognitive science and AI advance fastest in dialogue.** Behavioral probes suggest new algorithmic parameters (e.g., working-memory limits, learned priors), while computational models expose hidden theoretical assumptions, creating a virtuous feedback loop that sharpens both disciplines.

What Is Planning Made Of? Chapter 1 tackled the question of whether "planning ability" is unitary or coalition of basic cognitive abilities. Crossing three planning tasks that span fifteen orders of magnitude in state-space size with six basic cognitive tasks in 476 adults produced a stable, three-factor architecture: (i) visuo-spatial processing, (ii) working-memory capacity, and (iii) inhibitory control. Two-Step task depended mainly on suppressing habitual responses; Tower-of-London performance relied mainly on spatial manipulation; Four-in-a-Row drew heavily on working memory. These results corroborate Burgess et al. [25]'s multi-component view of planning.

How Are Plans Built? Factor analysis speaks to latent factors underlying planning, not to the orchestration of those abilities in planning moment by moment. Chapter 2 therefore collected think-aloud data while participants solved Four-in-a-Row puzzles. The transcripts revealed striking heterogeneity. Some players verbalized explicit tree searches, echoing De Groot [51] and Opheusden et al. [148]; others relied on shallow heuristics, such good enough moves (satisficing) or fast elimination rules that bypass search. Verbal depth increased within an episode – evidence of progressive deepening – but did not correlate with the depth inferred from a best cognitive model that fits final choices. Two interpretations are plausible. Either concurrent reports omit

unconscious or imagistic computations [57, 145], or the model's tree search template fails to capture strategic diversity. Taken together, think-aloud data suggests that modeling efforts might benefit from fitting to verbalization data to better capture the intricacy of planning.

Predicting Human Play with Transformers In Chapter 3, we demonstrated that transformer-based models could significantly surpass traditional cognitive models by leveraging move history information. GPT-4IAR, a transformer-based model trained on ten million human games, improved next-move prediction by seven percentage points and log-likelihood by 3.6 percentage points when given up to ninety prior moves. These gains caution any assumption that humans plan Markov-wise from the current board alone, an assumption implicit in many cognitive and RL models. Our results suggest that long-horizon context such as stylistic preferences and memory of earlier threats clearly shapes upcoming moves. Therefore, GPT-4IAR sets a new behavioral ceiling: cognitive models that ignore cross-game history now have a measurable gap to close.

What AlphaZero Learns and Misses Chapter 4 dissected the learning trajectory of an AlphaZero-style agent in *Four-in-a-Row*. Self-play improved three metrics – value quality, policy quality, and search depth – but causal manipulation experiments showed that policy quality drives most of the Elo gains. Once the network learns which branches matter, searching deeper offers diminishing returns – a machine echo of de Groot's claim that grand-masters recognize better patterns and recall more "chunks" without necessarily examining more moves [51]. AlphaZero's learning curve also mirrors the pattern reported by Opheusden et al. [148]: expert human players and AlphaZero both reduce the number of sibling branches explored at each play (lower feature-drop rate and higher heuristic quality in human cognitive model). Interestingly, the same superhuman agent consistently failed at logical sequence puzzles solvable by human experts. Injecting human-inspired features mitigated these failures. This finding exposes a blind spot of pure self-play curricula: winning games does not guarantee coverage of all tactically relevant micro-patterns, and human expertise still benefit the state-of-the-art agents.

Conceptual Integration

Taken together, these results sketch a unified picture of planning as a negotiation between expensive look-ahead and cheap heuristics. In humans, demands on working memory shift as state-space complexity grows, producing the factor structure uncovered in Chapter 1. Expert behavior, whether biological or artificial, therefore rests less on uniformly deeper search than on the ability to channel search effort towards high-value branches. GPT-4IAR's success with long contexts and AlphaZero's reliance on a powerful policy network both exemplify that principle: when the prior is informative, the system can reap most of the benefits of foresight without excessive enumeration.

At the same time, both humans and machines reveal blind spots. Think-aloud data showed that people repeatedly revisit shallow moves due to working-memory constraints, whereas AlphaZero overlooked short tactical sequences because its self-generated training distribution rarely rewards minimal-length victories. These complementary limitations highlight the utility of a comparative cognitive-AI approach, exposing assumptions and driving iterative improvements in both theories and algorithms. Our studies also caution against one-size-fits-all models. The weak alignment between verbalized and model-predicted depth shows that a single tree-search template might be insufficient to explain everyone's planning strategy. Adaptive models that allow depth, breadth or pruning thresholds to vary with board states – or that embed working memory constraints and strategy priors directly – may better capture human planning.

Limitations and Future Directions

Several caveats deserve emphasis. First, our tasks represent a limited subset of planning tasks and omit many complexities intrinsic to real-world planning scenarios, such as multiple simultaneous goals and stochastic contingencies. Second, Think-aloud data may miss unconscious pro-

cesses and required labor-intensive coding. Third, GPT-4IAR's interpretability remains limited: causal probes such as representation dissection or in-silico lesion require further investigation. Finally, our current puzzle set emphasizes short, forcing lines. Constructing a larger, systematically varied puzzle bank – including defensive problems,and deceptive traps – would enable finer-grained diagnostics of human heuristics and machine blind spots

Addressing these limitations suggests clear future research avenues. A natural next step is to construct a parametric "ladder" of $m \times n$ board variants that smoothly scales state-space complexity, allowing researchers to systematically examine how cognitive components change as state space scales. Process-level alignment could be advanced by fitting joint generative models that treat choices *and* verbal utterances as noisy emissions from an underlying search. On the AI side, the puzzle shortfall invites new curriculum design that interleave self-play with adversarial, human-crafted tactical tests, forcing agents to attend to the micro-features their current policies ignore.

Closing Reflection

In the opening pages of this thesis, I argued that planning is the engine behind achievements as mundane as getting to work on time and as challenging as writing a dissertation. The work that followed shows that this engine is assembled from specialized parts – some shared with AI, others uniquely human. The transformer teaches us that long-term memory matters; AlphaZero teaches us that a sharp prior can eclipse sheer depth; think-aloud protocols remind us that models fitted only to choices capture only part of the picture. The future might be trending toward a world where humans and machines plan side by side, each compensating for the other's blind spots. Building planners that combine these strengths, and diagnosing where each falls short, will require a science of planning that is both computationally and psychologically grounded. The work presented here offers a small, concrete step toward that goal.

Supplementary Materials

S1. Correlation Analysis

Table 1 presents correlations among task metrics along with p values, including cognitive tasks and questionnaires.

Table 1: Correlation table of task metrics with p values

	Matrix	Corsi	Rot	WCST	CDT	Pattern	TOL	Two-Step	4IAR	FOS	BIS
Matrix		0.325***	0.591***	0.295***	0.323***	0.409***	0.394***	0.233***	0.344***	0.072	0.009
		(0.00e+00)	(0.00e+00)	(0.00e+00)	(0.00e+00)	(0.00e+00)	(0.00e+00)	(0.00e+00)	(0.00e+00)	(9.81e-01)	(1.00e+00)
Corsi			0.269***	0.156*	0.416***	0.242***	0.215***	0.141	0.355***	-0.047	0.039
			(0.00e+00)	(2.00e-02)	(0.00e+00)	(0.00e+00)	(2.00e-04)	(7.18e-02)	(0.00e+00)	(1.00e+00)	(1.00e+00)
Rot				0.252***	0.324***	0.325***	0.392***	0.184**	0.213***	0.050	-0.017
				(0.00e+00)	(0.00e+00)	(0.00e+00)	(0.00e+00)	(2.10e-03)	(2.00e-04)	(1.00e+00)	(1.00e+00)
WCST					0.181**	0.136	0.221***	0.179**	0.187**	0.028	0.066
					(2.40e-03)	(9.93e-02)	(1.00e-04)	(3.10e-03)	(1.60e-03)	(1.00e+00)	(1.00e+00)
CDT						0.234***	0.249***	0.153*	0.321***	0.006	-0.040
						(0.00e+00)	(0.00e+00)	(2.65e-02)	(0.00e+00)	(1.00e+00)	(1.00e+00)
Pattern							0.230***	0.136	0.328***	0.043	-0.062
							(0.00e+00)	(9.56e-02)	(0.00e+00)	(1.00e+00)	(1.00e+00)
TOL								0.166*	0.280***	0.077	0.059
								(9.40e-03)	(0.00e+00)	(1.00e+00)	(1.00e+00)
Two-Step									0.185**	-0.049	-0.046
									(2.00e-03)	(1.00e+00)	(1.00e+00)
4IAR										0.107	-0.057
										(9.98e-01)	(1.00e+00)
FOS											0.000
											(1.00e+00)
BIS											

Significance levels:

* $p \leqslant 0.05$ ($|r| \geqslant 0.152$) ** $p \leqslant 0.01$ ($|r| \geqslant 0.171$) *** $p \leqslant 0.001$ ($|r| \geqslant 0.197$)

S2. Reliability Analysis

Reliability estimates (split-half) for each task are summarized in Table 2. Comparisons with established literature values are provided to contextualize the psychometric strength of each task. Excellent reliability is defined as > 0.9, good as $0.7 - 0.9$, fair as $0.4 - 0.7$, and poor as < 0.4 based on established psychometric criteria [40]. Our findings align generally with prior literature, demonstrating variability across tasks consistent with psychometric expectations.

Table 2: Comparison of Split-half Reliability Estimates for Each Task

Task	Reliability (Current Study)	Reliability (Literature)	References
Raven's SPM	0.92	0.81–0.89	[3, 18]
Mental Rotation	0.93	0.79–0.92	[87, 116]
Pattern Recognition	0.77	N/A	N/A
Four-in-a-row	0.77	N/A	N/A
Change Detection	0.72	0.79	[153]
WCST (pers. errors)	0.78	0.83–0.95	[109, 201, 233]
Tower of London	0.69	0.60–0.72	[104, 110, 210, 214]
Corsi Block-Tapping	0.55	0.70–0.79	[149, 155]
Two-Step Task	0.46	0.00–0.55	[22]

Note: Excellent reliability is defined as > 0.9, good as $0.7 - 0.9$, fair as $0.4 - 0.7$, and poor as < 0.4 based on established psychometric criteria [40].

S3. Factor Analysis

Table 3 presents the results of exploratory factor analysis with varimax rotation. We report both three-factor and two-factor solutions for comparative purposes, highlighting factor loadings greater than 0.30. The two-factor solution explained 47.0% of total variance, whereas the three-factor solution explained 57.2%

Three-factor vs. Two-Factor Solution

For comparison, we also examined a two-factor solution (Table 3), which explained **47.0%** of the total variance.

Table 3: Varimax-Rotated Factor Loadings for Three-Factor and Two-Factor Solutions. Bolded loadings exceed 0.30.

Three-Factor Solution				Two-Factor Solution		
Variable	**F1**	**F2**	**F3**	**Variable**	**F1**	**F2**
Four-in-a-row	0.20	**0.67**	0.15	Four-in-a-row	0.18	**0.67**
Two-step	0.00	0.18	**0.83**	Two-step	**0.49**	0.06
TOL	**0.63**	0.13	0.19	TOL	**0.60**	0.24
Change Detection	0.21	**0.70**	0.09	Change Detection	0.16	**0.71**
Mental Rotation	**0.80**	0.13	0.13	Mental Rotation	**0.70**	0.28
WCST Errors	**0.32**	0.03	**0.62**	WCST Errors	**0.62**	0.01
Corsi	0.14	**0.78**	0.05	Corsi	0.06	**0.78**
Pattern Recognition	**0.55**	**0.35**	-0.10	Pattern Recognition	**0.35**	**0.47**
SPM	**0.76**	0.26	0.19	SPM	**0.70**	**0.39**
Var. Expl.	**23.7%**	**20.1%**	**13.4%**	**Var. Expl.**	**23.8%**	**23.1%**
Cumulative	**23.7%**	**43.9%**	**57.2%**	**Cumulative**	**23.8%**	**47.0%**

S4. Methods for Think-Aloud Video Analysis

This section outlines the procedure for analyzing think-aloud video recordings from participants solving Four-in-a-Row puzzles. Each participant completed 10 different puzzles (A-J), with recordings coded into spreadsheets identified by subject ID (01-48).

Each spreadsheet contains three sets of rows for each puzzle: Articulated Tree rows, Feature rows, and Qualitative Description rows. Articulated Tree rows document each first move considered by the participant in the initial puzzle state, presented in order of articulation. Feature rows document board features or groups of features that the participant articulates or indicates through pointing. The Qualitative Description provides overall impressions of the participant's verbal and non-verbal behavior.

S4.1 Recording Criteria for Articulated Tree

Each Articulated Tree column corresponds to one first move. The following five criteria were used to determine when to add a column (all criteria must be satisfied):

	subject ID	
Puzzle identity	Puzzle#_	
	PuzzleID	
Articulated tree (defined by the mention of a first move)	First_move_location	Articulated Tree rows: information about a specific first move and its subsequent moves
	First_move_time	
	First_move_side	
	First_move_goal	
	Feature	
	Plan_depth	
	Plan_depth_max	
	Plan_until_end	
	Side_count	
	Sequence_strategy	
	planning_sequence	
Feature	Articulated_feature	Feature rows: information about
	Articulated_feature_time	
	Articulated_feature_plan	
Qualitative Description	Original_speech	Qualitative Description rows: overall
	strategy	

Figure 1: Example spreadsheet for think-aloud protocol analysis showing Articulated Tree rows (yellow), Feature rows (green), and Qualitative Description rows (red).

1. **Intent to make a move**, satisfied by either:

   (a) Articulation of intent (e.g., "I am going to make a move now," "If I go here...")

   (b) Context implying intent to move (e.g., evaluating a move and proceeding to indicate subsequent moves)

2. **Indication of a specific square**, satisfied by one of:

   (a) Touching or pointing at one square

   (b) Verbal indication through articulating a goal or features formed by the move

   (c) Touching multiple squares with subsequent moves only making sense for one

   (d) Touching multiple squares with one being touched closer in time to the expressed intent

3. **Indication of player** (self or opponent), satisfied by:

(a) Articulating the player (e.g., "if I/they..." or "if black/white...")

(b) Indicating the player through articulated goals or features

4. **Confirmation as a first move rather than subsequent move**, satisfied by:

   (a) Time discontinuity with previously articulated moves

   (b) Clear transitions between previous and current first moves (e.g., using "or")

   (c) Logical inconsistency if counted as a subsequent move

5. **Not a repetition**, satisfied by at least one:

   (a) Different square from the previous column

   (b) Different player side from the previous column

   (c) Time discontinuity with the previous column

S4.2 Coded Data for Articulated Tree

For each first move, the following data were coded:

1. **First_move_square**: Square position (0-35) on the board
2. **First_move_time**: Time period of articulation (format: Minute' Second')
3. **First_move_side**: Player side ("s" for self, "o" for opponent)
4. **First_move_goal**: Goal of the move ("a" for attack, "d" for defend, "b" for both, "u" for unclear)
5. **Feature**: Patterns of pieces/empty squares articulated for the move
6. **Plan_depth**: Maximum depth of the planning tree
7. **Plan_depth_max**: Whether this subtree is the deepest of all side branches ("T" or "F")

8. **Plan_until_end**: Whether planning leads to a win/loss ("T" or "F")
9. **Side_count**: Count of nodes in side branches
10. **Sequence_strategy**: Description of special strategies used
11. **Planning_sequence**: Sequence of nodes visited with maximum planning depth

S4.3 Feature Notation System

Features were coded using the following notation:

- **Player prefix**: 'o-' for opponent features; no prefix for self features
- **Empty squares**: 'e-v-' (vertical), 'e-h-' (horizontal), 'e-d-' (diagonal)
- **1-in-a-row**: '1-v-', '1-h-', '1-d-' for vertical, horizontal, diagonal respectively
- **2-in-a-row**: '2-v-', '2-h-', '2-d-'
- **Unconnected 2-in-a-row**: 'un-2-v-', 'un-2-h-', 'un-2-d-'
- **3-in-a-row**: '3-v-', '3-h-', '3-d-'
- **Unconnected 3-in-a-row**: 'un-3-v-', 'un-3-h-', 'un-3-d-'
- **Other notations**: 'b' (blocked), 'tri' (triangle)
- **Connectors**: '-' connects aspects of pattern; '+' connects multiple features

S4.4 Recording Criteria for Feature

Each Feature column corresponds to a feature or group of features articulated by the participant. A feature was considered articulated when the participant touched or pointed at multiple squares belonging to a group of pieces or empty squares.

For each articulated feature, the following data were coded:

1. **Articulated_feature**: Notation of the articulated feature
2. **Articulated_feature_time**: Time when the feature was articulated
3. **Articulated_feature_plan**: Whether the feature was used for planning ("T" or "F")

If multiple features were articulated simultaneously with shared squares, or if multiple features were used for planning the same first move, they were connected with a "+" and placed in the same column.

Bibliography

[1] Luigi Acerbi and Wei Ji Ma. "Practical Bayesian Optimization for Model Fitting with Bayesian Adaptive Direct Search". In: *Advances in Neural Information Processing Systems 30*. 2017, pp. 1834–1844.

[2] P. L. Ackerman, M. E. Beier, and M. O. Boyle. "Working memory and intelligence: The same or different constructs?" In: *Psychological Bulletin* 131.1 (2005), pp. 30–60. DOI: 10.1037/0033-2909.131.1.30.

[3] Riaz Ahmad, Sarwat J. Khanam, and Zaeema Riaz. "The Standard Progressive Matrices in Pakistan". In: *Uses and Abuses of Intelligence: Studies Advancing Spearman and Raven's Quest for Non-Arbitrary Metrics*. Ed. by John Raven. Royal Fireworks Press, 2006, pp. 404–412.

[4] Thomas Akam, Rui Costa, and Peter Dayan. "Simple plans or sophisticated habits? State, transition and learning interactions in the two-step task". In: *PLoS Computational Biology* 11.12 (2015), e1004648. DOI: 10.1371/journal.pcbi.1004648.

[5] Jay Alammar. *The Illustrated GPT-2 (Visualizing Transformer Language Models)*. 2019.

[6] Nick Alderman et al. "Ecological validity of a simplified version of the multiple errands shopping test". In: *Journal of the International Neuropsychological Society* 9.1 (2003), pp. 31–44. DOI: 10.1017/S1355617703910046.

[7] Kelsey R Allen et al. *Using Games to Understand the Mind*. 2023. DOI: 10.31234/osf.io/hbsvj.

[8] L. Victor Allis. "Searching for Solutions in Games and Artificial Intelligence". PhD thesis. Maastricht, The Netherlands: University of Limburg, 1994.

[9] Yuichiro Anzai and Herbert A. Simon. "The Theory of Learning by Doing". In: *Psychological Review* 86.2 (1979), pp. 124–140. DOI: 10.1037/0033-295X.86.2.124.

[10] Kai Arulkumaran, Antoine Cully, and Julian Togelius. "AlphaStar: An Evolutionary Computation Perspective". In: *Proceedings of the Genetic and Evolutionary Computation Conference Companion*. GECCO '19. Prague, Czech Republic: Association for Computing Machinery, 2019, pp. 314–315. ISBN: 9781450367486. DOI: 10.1145/3319619.3321894.

[11] Alan D. Baddeley. *Working Memory*. Oxford, UK: Oxford University Press, 1986. ISBN: 9780198521167.

[12] Jan Balaguer et al. "Neural Mechanisms of Hierarchical Planning in a Virtual Subway Network". In: *Neuron* 90.4 (2016), pp. 893–903. DOI: 10.1016/j.neuron.2016.03.037.

[13] Ernest S. Barratt. "Anxiety and impulsiveness related to psychomotor efficiency". In: *Perceptual and Motor Skills* 9.3 (1959), pp. 191–198. DOI: 10.2466/pms.1959.9.3.191.

[14] Daniel B. Berch, Robert Krikorian, and Eileen M. Huha. "The Corsi block-tapping task: Methodological and theoretical considerations". In: *Brain and Cognition* 38.3 (1998), pp. 317–338. DOI: 10.1006/brcg.1998.1039.

[15] Esta A. Berg. "A simple objective technique for measuring flexibility in thinking". In: *Journal of General Psychology* 39.1 (1948), pp. 15–22. DOI: 10.1080/00221309.1948.9918159.

[16] W. K. Berg and D. L. Byrd. "The Tower of London spatial problem-solving task: Enhancing clinical and research implementation". In: *Journal of Clinical and Experimental Neuropsychology* 24.5 (2002), pp. 586–604. DOI: 10.1076/jcen.24.5.586.1006.

[17] Alex Bernstein and Michael de V Roberts. "A chess playing program for the IBM 704". In: *Proceedings of the Western Joint Computer Conference* (1958), pp. 157–159.

[18] Haifa Al-Bokaia and Ali A. Al-Subaihib. "Standard Progressive Matrices (SPM): validity and reliability". In: *International Journal of Innovation, Creativity and Change* 15.4 (2021), pp. 276–293.

[19] Matthew Botvinick et al. "Reinforcement Learning, Fast and Slow". In: *Trends in Cognitive Sciences* 23.5 (2019), pp. 408–422.

[20] Matthew M. Botvinick and Marc Toussaint. "Planning as Inference". In: *Trends in Cognitive Sciences* 16.10 (2012), pp. 485–488. DOI: 10.1016/j.tics.2012.08.006.

[21] Timothy F. Brady and Joshua B. Tenenbaum. "A probabilistic model of visual working memory: Incorporating higher order regularities into working memory capacity estimates". In: *Psychological Review* 120.1 (2013), pp. 85–109. DOI: 10.1037/a0030779.

[22] Victoria M. Brown et al. "Improving the Reliability of Computational Analyses: Model-Based Planning and Its Relationship With Compulsivity". In: *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging* 5.6 (2020), pp. 601–609. DOI: 10.1016/j.bpsc.2019.12.019.

[23] William Brown. "Some Experimental Results in the Correlation of Mental Abilities". In: *British Journal of Psychology* 3.3 (1910), pp. 296–322. DOI: 10.1111/j.2044-8295.1910.tb00207.x.

[24] Victoria M. Bryan and John D. Mayer. "A Meta-Analysis of the Correlations Among Broad Intelligences: Understanding Their Relations". In: *Intelligence* 83 (2020), p. 101469. DOI: 10.1016/j.intell.2020.101469.

[25] P. Burgess et al. "The search for specific planning processes". In: *The cognitive psychology of planning*. London, UK: Psychology Press, 2005, pp. 199–227.

[26] Frederick Callaway, Miaomiao Yu, and Marcelo G. Mattar. "Revealing Human Planning Strategies with Eye-Tracking". In: *Proceedings of the 46th Annual Meeting of the Cognitive Science Society*. Cognitive Science Society, 2024, pp. 219–225.

[27] Frederick Callaway et al. "Mouselab-MDP: A New Paradigm for Tracing How People Plan". In: *Proceedings of the 44th Annual Conference of the Cognitive Science Society*. Cognitive Science Society, 2022, pp. 381–387.

[28] Murray Campbell, Jr. A. Joseph Hoane, and Feng-hsiung Hsu. "Deep Blue". In: *Artificial Intelligence* 134.1–2 (2002), pp. 57–83.

[29] Murray Campbell, A. Joseph Hoane Jr., and Feng-hsiung Hsu. "Deep Blue". In: *Artificial Intelligence* 134.1–2 (2002), pp. 57–83.

[30] John B. Carroll. *Human Cognitive Abilities: A Survey of Factor-Analytic Studies*. Cambridge, UK: Cambridge University Press, 1993. ISBN: 9780521387125.

[31] Micah Carroll et al. "Uni[MASK]: Unified Inference in Sequential Decision Problems". In: *Advances in Neural Information Processing Systems*. Ed. by Alice H. Oh et al. 2022.

[32] B. J. Casey. "Beyond Simple Models of Self-control to Circuit-based Accounts of Adolescent Behavior". In: *Annual Review of Psychology* 66 (2015), pp. 295–319. DOI: 10.1146/annurev-psych-010814-015156.

[33] Raymond B. Cattell. "Theory of Fluid and Crystallized Intelligence: A Critical Experiment". In: *Journal of Educational Psychology* 54.1 (Jan. 1963), pp. 1–22. DOI: 10.1037/h0046743.

[34] William G. Chase and Herbert A. Simon. "Perception in Chess". In: *Cognitive Psychology* 4.1 (1973), pp. 55–81.

[35] William G. Chase and Herbert A. Simon. "Perception in Chess". In: *Cognitive Psychology* 4.1 (1973), pp. 55–81. DOI: 10.1016/0010-0285(73)90004-2.

[36] J. Marcus Cheetham et al. “Visuospatial over verbal demands in predicting Tower of London planning tasks”. In: *British Journal of Psychology* 103.1 (2012), pp. 98–116. DOI: 10.1111/j.2044-8295.2011.02049.x.

[37] Li Chen et al. “Planning-oriented Autonomous Driving”. In: *arXiv preprint arXiv:2212.10156*. 2023.

[38] Lili Chen et al. “Decision Transformer: Reinforcement Learning via Sequence Modeling”. In: *arXiv preprint arXiv:2106.01345* (2021).

[39] Lili Chen et al. “Decision Transformer: Reinforcement Learning via Sequence Modeling”. In: *Advances in Neural Information Processing Systems*. Ed. by M. Ranzato et al. Vol. 34. Curran Associates, Inc., 2021, pp. 15084–15097.

[40] Domenic V. Cicchetti. “Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology”. In: *Psychological Assessment* 6.4 (1994), pp. 284–290. DOI: 10.1037/1040-3590.6.4.284.

[41] Janet Cockburn. “Performance on the Tower of London test after severe head injury”. In: *Journal of the International Neuropsychological Society* 1.6 (1995), pp. 537–544. DOI: 10.1017/S1355617700000667.

[42] Ronald L. Cohen and Tor Sandberg. “Intelligence and Short-Term Memory: A Clandestine Relationship”. In: *Intelligence* 4 (1980), pp. 319–331. DOI: 10.1016/0160-2896(80)90026-4.

[43] Anne G. E. Collins and Amitai Shenhav. “Advances in modeling learning and decision-making in neuroscience”. In: *Neuropsychopharmacology* 47.1 (Jan. 2022), pp. 104–118. ISSN: 1740-634X. DOI: 10.1038/s41386-021-01126-y.

[44] Anne G.E. Collins and Jeffrey Cockburn. “Beyond dichotomies in reinforcement learning”. In: *Nature Reviews Neuroscience* 21 (2020), pp. 576–586. DOI: 10.1038/s41583-020-0355-6.

[45] Philip M. Corsi. “Human memory and the medial temporal region of the brain”. Unpublished doctoral dissertation. Doctoral dissertation. Montreal, Canada: McGill University, 1972.

[46] A. B. Costello and J. W. Osborne. “Best Practices in Exploratory Factor Analysis: Four Recommendations for Getting the Most from Your Analysis”. In: *Practical Assessment, Research, and Evaluation* 10.7 (2005). DOI: 10.7275/jyj1-4868.

[47] Jean-Baptiste Couëtoux et al. “Continuous Upper Confidence Trees with Progressive Widening”. In: *Proceedings of the 23rd IEEE International Conference on Tools with Artificial Intelligence (ICTAI)*. 2011, pp. 120–127.

[48] Rémi Coulom. “Whole-history rating: A Bayesian rating system for players of time-varying strength”. In: *International Conference on Computers and Games*. Springer. 2008, pp. 113–124.

[49] S. P. Davies. “Planning and problem solving in well-defined domains”. In: *The Cognitive Psychology of Planning*. Ed. by R. Morris and G. Ward. Vol. 35. Psychology Press, 2005.

[50] Nathaniel D. Daw et al. “Model-Based Influences on Humans’ Choices and Striatal Prediction Errors”. In: *Neuron* 69.6 (Mar. 2011), pp. 1204–1215. DOI: 10.1016/j.neuron.2011.02.027.

[51] Adriaan D. De Groot. *Thought and Choice in Chess*. The Hague: Mouton, 1965.

[52] Johannes H. Decker et al. “From Creatures of Habit to Goal-Directed Learners: Tracking the Developmental Emergence of Model-Based Reinforcement Learning”. In: *Psychological Science* 27.6 (2016), pp. 848–858. DOI: 10.1177/0956797616639301.

[53] Adele Diamond. “Executive Functions”. In: *Annual Review of Psychology* 64 (2013), pp. 135–168. ISSN: 0066-4308. DOI: 10.1146/annurev-psych-113011-143750.

[54] Jonathan S. Diamond, Daniel M. Wolpert, and J. Randall Flanagan. "Rapid Target Foraging with Reach or Gaze: The Hand Looks Further Ahead than the Eye". In: *PLOS Computational Biology* 13.7 (2017), e1005504. DOI: 10.1371/journal.pcbi.1005504.

[55] Arpad E. Elo. *The Rating of Chessplayers, Past and Present*. New York: Arco Publishing, 1978.

[56] B. Eppinger et al. "Of goals and habits: Age-related and individual differences in goal-directed decision-making". In: *Frontiers in Neuroscience* 7 (2013), p. 253. DOI: 10.3389/fnins.2013.00253.

[57] K. Anders Ericsson and Herbert A. Simon. *Protocol Analysis: Verbal Reports as Data*. Revised. Cambridge, MA: MIT Press, 1993.

[58] Michael D. Ernst. "Permutation methods: A basis for exact inference". In: *Statistical Science* 19.4 (2004), pp. 676–685. DOI: 10.1214/088342304000000396.

[59] Leandre R. Fabrigar et al. "Evaluating the use of exploratory factor analysis in psychological research". In: *Psychological Methods* 4.3 (1999), pp. 272–299. DOI: 10.1037/1082-989X.4.3.272.

[60] Carolina Feher da Silva and Todd A. Hare. "Humans Primarily Use Model-Based Inference in the Two-Stage Decision Task". In: *Nature Human Behaviour* 4.10 (2020), pp. 1053–1066. DOI: 10.1038/s41562-020-0905-y.

[61] Susann Fiedler and Andreas Glöckner. "The Dynamics of Decision Making in Risky Choice: An Eye-Tracking Analysis". In: *Frontiers in Psychology* 3 (2012), p. 335. DOI: 10.3389/fpsyg.2012.00335.

[62] Richard E. Fikes and Nils J. Nilsson. "STRIPS: A New Approach to the Application of Theorem Proving to Problem Solving". In: *Proceedings of the 2nd International Joint Conference on Artificial Intelligence (IJCAI)*. 1971, pp. 608–620.

[63] Richard E. Fikes and Nils J. Nilsson. "STRIPS: A New Approach to the Application of Theorem Proving to Problem Solving". In: *Artificial Intelligence* 2.3–4 (1971), pp. 189–208.

[64] L. Flower et al. "Planning in writing: The cognition of a constructive process". In: *A rhetoric of doing: Essays on written discourse in honor of James L. Kinneavy*. Ed. by S. P. Witte, N. Nakadate, and R. D. Cherry. Southern Illinois University Press, 1992, pp. 181–243.

[65] Mark C. Fox, K. Anders Ericsson, and Russell Best. "Do Procedures for Verbal Reporting of Thinking Have to Be Reactive? A Meta-Analysis and Recommendations for Best Reporting Methods". In: *Psychological Bulletin* 137.2 (2011), pp. 316–344. DOI: 10.1037/a0021663.

[66] Xavier Gabaix et al. "Costly Information Acquisition: Experimental Analysis of a Boundedly Rational Model". In: *The American Economic Review* 96.4 (2006), pp. 1043–1068.

[67] Giorgio Ganis and Rogier A. Kievit. "A New Set of Three-dimensional Shapes for Investigating Mental Rotation Processes: Validation Data and Stimulus Set". In: *Journal of Open Psychology Data* 3.1 (2015), e3. DOI: 10.5334/jopd.ai.

[68] K. J. Gilhooly et al. "Visuo-spatial and verbal working memory in the five-disc Tower of London task: An individual differences approach". In: *Thinking & Reasoning* 8.3 (2002), pp. 165–178. DOI: 10.1080/13546780244000006.

[69] Claire M. Gillan et al. "Characterizing a psychiatric symptom dimension related to deficits in goal-directed control". In: *eLife* 5 (2016), e11305. DOI: 10.7554/eLife.11305.

[70] Claire M. Gillan et al. "Comparison of the Association Between Goal-Directed Planning and Self-reported Compulsivity vs Obsessive-Compulsive Disorder Diagnosis". In: *JAMA Psychiatry* 77.1 (2020), pp. 77–85. DOI: 10.1001/jamapsychiatry.2019.2998.

[71] Fernand Gobet and Herbert A Simon. "Templates in chess memory: A mechanism for recalling several boards". In: *Cognitive Psychology* 31.1 (1996), pp. 1–40.

[72] Vinod Goel and Jordan Grafman. “Are the frontal lobes implicated in "planning" functions? Interpreting data from the Tower of Hanoi”. In: *Neuropsychologia* 33.5 (1995), pp. 623–642. DOI: 10.1016/0028-3932(95)90866-p.

[73] Jeremy R. Gordon, Jason Chuang, and Giovanni Pezzulo. *Gaze Dynamics Prior to Navigation Support Hierarchical Planning*. bioRxiv preprint. 2025. DOI: 10.1101/2025.01.16.633460.

[74] David M. Green and John A. Swets. *Signal Detection Theory and Psychophysics*. New York: John Wiley & Sons, 1966.

[75] James G. Greeno. “Natures of Problem-Solving Abilities”. In: *Handbook of Learning and Cognitive Processes. Vol. 5: Human Information Processing*. Ed. by William K. Estes. Hillsdale, NJ: Erlbaum, 1978, pp. 239–270.

[76] Adriaan D. de Groot. *Thought and Choice in Chess*. The Hague, Mouton, 1965.

[77] Joseph F. Hair et al. *Multivariate Data Analysis*. 4th ed. New Jersey: Prentice Hall, 1995.

[78] Jessica B. Hamrick. “Analogies Between AlphaZero and the Brain”. In: *Trends in Cognitive Sciences* 23.7 (2019), pp. 569–572.

[79] Shiyang Hao et al. “Reasoning with Language Model Is Planning with World Model”. In: *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (2023), pp. 8154–8173.

[80] Peter E. Hart, Nils J. Nilsson, and Bertram Raphael. “A Formal Basis for the Heuristic Determination of Minimum Cost Paths”. In: *IEEE Transactions on Systems Science and Cybernetics* 4.2 (1968), pp. 100–107. DOI: 10.1109/TSSC.1968.300136.

[81] J. R. Hayes and J. G. Nash. “On the nature of planning in writing”. In: *The science of writing*. Chicago: Routledge, 2013, pp. 29–55.

[82] John R. Hayes and Herbert A. Simon. “Thinking and the Development of Problem-Solving Skills”. In: *British Journal of Psychology* 68.1 (1977), pp. 1–16.

[83] Barbara Hayes-Roth and Frederick Hayes-Roth. “A Cognitive Model of Planning”. In: *Cognitive Science* 3.4 (1979), pp. 275–310. DOI: 10.1207/s15516709cog0304_1.

[84] Barbara Hayes-Roth. *Flexibility in Executive Strategies*. Tech. rep. R-1170. Santa Monica, CA: RAND Corporation, 1980.

[85] Robert K. Heaton. *Wisconsin Card Sorting Test manual*. Odessa, FL: Psychological Assessment Resources, 1981.

[86] Mary Hegarty and David Waller. “A Dissociation between Mental Rotation and Perspective-Taking Spatial Abilities”. In: *Intelligence* 32.2 (2004), pp. 175–191. DOI: 10.1016/j.intell.2003.12.001.

[87] Gerrit Hirschfeld, Meinald T. Thielsch, and Boris Zernikow. “Reliabilities of Mental Rotation Tasks: Limits to the Assessment of Individual Differences”. In: *BioMed Research International* 2013 (2013), p. 340568. DOI: 10.1155/2013/340568.

[88] Mark K. Ho et al. “People Construct Simplified Mental Representations to Plan”. In: *Nature* 606 (2022), pp. 129–136. DOI: 10.1038/s41586-022-04743-9.

[89] Robert R. Hoffman, Beth Crandall, and Nigel Shadbolt. “Use of the Critical Decision Method to Elicit Expert Knowledge: A Case Study in the Methodology of Cognitive Task Analysis”. In: *Human Factors* 40.2 (1998), pp. 254–276. DOI: 10.1518/001872098779480442.

[90] Eleanor Holton et al. “Disentangling the Component Processes in Complex Planning Impairments Following Ventromedial Prefrontal Lesions”. In: *Journal of Neuroscience* 45.12 (2025), e1814242025. DOI: 10.1523/JNEUROSCI.1814-24.2025.

[91] John L. Horn. “A Rationale and Test for the Number of Factors in Factor Analysis”. In: *Psychometrika* 30.2 (1965), pp. 179–185. DOI: 10.1007/BF02289447.

[92] Jiawen Huang et al. “Schema-based predictive eye movements support sequential memory encoding”. In: *eLife* 12 (2023), e82599. DOI: 10.7554/eLife.82599.

[93] L. T. Hunt et al. “Formalizing planning and information search in naturalistic decision-making”. In: *Nature Neuroscience* 24.8 (2021), pp. 1051–1064. ISSN: 1546-1726. DOI: 10.1038/s41593-021-00866-w.

[94] David R. Hunter. “MM Algorithms for Generalized Bradley–Terry Models”. In: *Annals of Statistics* 32.1 (2004), pp. 384–406.

[95] Quentin J. M. Huys et al. “Interplay of Approximate Planning Strategies”. In: *Proceedings of the National Academy of Sciences of the United States of America* 112.10 (2015), pp. 3098–3103. DOI: 10.1073/pnas.1414219112.

[96] Quentin JM Huys et al. “Bonsai trees in your head: how the Pavlovian system sculpts goal-directed choices by pruning decision trees”. In: *PLoS Computational Biology* 8.3 (2012), e1002410.

[97] Quentin JM Huys et al. “Interplay of approximate planning strategies”. In: *Proceedings of the National Academy of Sciences* 112.10 (2015), pp. 3098–3103.

[98] Michael Janner, Qiyang Li, and Sergey Levine. “Offline Reinforcement Learning as One Big Sequence Modeling Problem”. In: *Advances in Neural Information Processing Systems*. 2021.

[99] Petra Jansen and Martin Heil. “The Relation Between Motor Development and Mental Rotation Ability in 5- to 6-Year-Old Children”. In: *European Journal of Developmental Science* 4.1 (2010), pp. 66–74. DOI: 10.3233/DEV-2010-4105.

[100] Arthur R. Jensen and Richard A. Figueroa. “Forward and Backward Digit Span Interaction with Race and IQ: Predictions from Jensen’s Theory”. In: *Journal of Educational Psychology* 67.6 (1975), pp. 882–893. DOI: 10.1037/0022-0663.67.6.882.

[101] W. Johnson and Jr. Bouchard T. J. "The structure of human intelligence: It is verbal, perceptual, and image rotation (VPR), not fluid and crystallized". In: *Intelligence* 33.4 (2005), pp. 393–416. DOI: 10.1016/j.intell.2004.12.002.

[102] Eileen M. Joyce and Trevor W. Robbins. "Frontal lobe function in Korsakoff and non-Korsakoff alcoholics: planning and spatial working memory". In: *Neuropsychologia* 29.8 (1991), pp. 709–723. DOI: 10.1016/0028-3932(91)90067-I.

[103] Henry F. Kaiser. "An Index of Factorial Simplicity". In: *Psychometrika* 39.1 (1974), pp. 31–36. DOI: 10.1007/BF02291575.

[104] Christoph P. Kaller, Josef M. Unterrainer, and Christoph Stahl. "Assessing Planning Ability with the Tower of London Task: Psychometric Properties of a Structurally Balanced Problem Set". In: *Psychological Assessment* 24.1 (2012), pp. 46–53. DOI: 10.1037/a0025174.

[105] Been Kim et al. "Interpretability beyond feature attribution: Quantitative testing with concept activation vectors (tcav)". In: *International conference on machine learning*. PMLR. 2018, pp. 2668–2677.

[106] Donald E. Knuth and Ronald W. Moore. "An Analysis of Alpha-Beta Pruning". In: *Artificial Intelligence* 6.4 (1975), pp. 293–326.

[107] Levente Kocsis and Csaba Szepesvári. "Bandit Based Monte-Carlo Planning". In: *Proceedings of the 17th European Conference on Machine Learning*. Vol. 4212. Lecture Notes in Artificial Intelligence. Springer, 2006, pp. 282–293. DOI: 10.1007/11871842_29.

[108] B. Kopp et al. "A meta-analysis of relationships between measures of Wisconsin Card Sorting and intelligence". In: *Brain Sciences* 9.12 (2019), p. 349. DOI: 10.3390/brainsci9120349.

[109] Bruno Kopp, Florian Lange, and Alexander Steinke. "The Reliability of the Wisconsin Card Sorting Test in Clinical Practice". In: *Assessment* 28.1 (2021), pp. 248–263. DOI: 10.1177/1073191119866257.

[110] Lena Köstering et al. “Assessment of Planning Performance in Clinical Samples: Reliability and Validity of the Tower of London Task (TOL–F)”. In: *Neuropsychologia* 75 (2015), pp. 646–655. DOI: 10.1016/j.neuropsychologia.2015.07.017.

[111] John H Krantz and Reeshad Dalal. “Validity of Web-based psychological research”. In: *Psychological experiments on the Internet*. Ed. by Michael H Birnbaum. San Diego, CA: Academic Press, 2000, pp. 35–60.

[112] Ionatan Kuperwajs, Heiko H. Schütt, and Wei Ji Ma. “Using deep neural networks as a guide for modeling human planning”. In: *Scientific Reports* 13.1 (Nov. 2023), p. 20269. ISSN: 2045-2322. DOI: 10.1038/s41598-023-46850-1.

[113] Brenden M Lake et al. “Building machines that learn and think like people”. In: *Behavioral and brain sciences* 40 (2017).

[114] Níall Lally et al. “The Neural Basis of Aversive Pavlovian Guidance during Planning”. In: *Journal of Neuroscience* 37.42 (2017), pp. 10215–10229. DOI: 10.1523/JNEUROSCI.0085-17.2017.

[115] Seung Lee. “Algorithms for Non-negative Matrix Factorization”. In: *NIPS* (2000).

[116] Jennifer Lehmann, Claudia Quaiser-Pohl, and Petra Jansen. “Correlation of Motor Skill, Mental Rotation, and Working Memory in 3- to 6-Year-Old Children”. In: *European Journal of Developmental Psychology* 11.5 (2014), pp. 560–573. DOI: 10.1080/17405629.2014.888995.

[117] Harvey S. Levin et al. “Dimensions of cognition measured by the Tower of London and other cognitive tasks in head-injured children and adolescents”. In: *Developmental Neuropsychology* 12.1 (1996), pp. 17–34.

[118] Falk Lieder and Thomas L. Griffiths. “Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources”. In: *Behavioral and Brain Sciences* 43 (2020), e1–e60.

[119] Friedrich Lieder and Thomas L. Griffiths. "Resource-Rational Analysis: Understanding Human Cognition as the Optimal Use of Limited Computational Resources". In: *Behavioral and Brain Sciences* 43 (2020), e1.

[120] Tianyang Lin et al. "A survey of transformers". In: *AI Open* 3 (2022), pp. 111–132. ISSN: 2666-6510. DOI: https://doi.org/10.1016/j.aiopen.2022.10.001.

[121] Ilya Loshchilov and Frank Hutter. "Decoupled Weight Decay Regularization". In: *International Conference on Learning Representations*. 2019.

[122] M. Luciana et al. "Tower of London performance in healthy adolescents: The development of planning skills and associations with self-reported inattention and impulsivity". In: *Developmental Neuropsychology* 34.4 (2009), pp. 461–475. DOI: 10.1080/87565640902964540.

[123] Robert C MacCallum et al. "Sample size in factor analysis". In: *Psychological Methods* 4.1 (1999), pp. 84–99. DOI: 10.1037/1082-989X.4.1.84.

[124] Neil A. Macmillan and C. Douglas Creelman. *Detection Theory: A User's Guide*. 2nd ed. Mahwah, NJ: Lawrence Erlbaum Associates, 2005.

[125] Marcelo G. Mattar and Máté Lengyel. "Planning in the brain". In: *Neuron* 110.6 (2022), pp. 914–934. DOI: 10.1016/j.neuron.2021.12.018.

[126] M. G. McGee. "Human Spatial Abilities: Psychometric Studies and Environmental, Genetic, Hormonal, and Neurological Influences". In: *Psychological Bulletin* 86.5 (1979), pp. 889–918. DOI: 10.1037/0033-2909.86.5.889.

[127] Thomas McGrath et al. "Acquisition of chess knowledge in alphazero". In: *Proceedings of the National Academy of Sciences* 119.47 (2022), e2206625119.

[128] Reid McIlroy-Young et al. "Aligning Superhuman AI with Human Behavior: Chess as a Model System". In: *26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*. July 2020.

[129] Reid McIlroy-Young et al. *Learning Personalized Models of Human Behavior in Chess*. Aug. 2020.

[130] George A. Miller, Eugene Galanter, and Karl H. Pribram. *Plans and the Structure of Behavior*. New York: Henry Holt and Company, 1960. DOI: 10.1037/10039-000.

[131] Kevin J. Miller, Matthew M. Botvinick, and Carlos D. Brody. "Dorsal hippocampus contributes to model-based planning". In: *Nature Neuroscience* 20.9 (2017), pp. 1269–1276. DOI: 10.1038/nn.4613.

[132] Eliane C. Miotto and Robin G. Morris. "Virtual Planning in Patients with Frontal Lobe Lesions". In: *Cortex* 34.5 (1998), pp. 639–657. DOI: 10.1016/S0010-9452(08)70770-6.

[133] Ariana Mitko and Jason Fischer. "When it all falls down: The relationship between intuitive physics and spatial cognition". In: *Cognitive Research: Principles and Implications* 5 (2020), Article 24. DOI: 10.1186/s41235-020-00224-7.

[134] A. Miyake et al. "How are visuospatial working memory, executive functioning, and spatial abilities related? A latent-variable analysis". In: *Journal of Experimental Psychology: General* 130.4 (2001), pp. 621–640. DOI: 10.1037/0096-3445.130.4.621.

[135] A. Miyake et al. "The unity and diversity of executive functions: A latent variable analysis". In: *Cognitive Psychology* 41.1 (2000), pp. 49–100. DOI: 10.1006/cogp.1999.0734.

[136] Volodymyr Mnih et al. "Human-Level Control Through Deep Reinforcement Learning". In: *Nature* 518.7540 (2015), pp. 529–533.

[137] Robin Morris, M. Kotitsa, and Jessica Bramham. "Planning in Patients with Focal Brain Damage: From Simple to Complex Task Performance". In: *The Cognitive Psychology of Planning*. Ed. by Robin Morris and Geoff Ward. Psychology Press, 2005, pp. 213–240.

[138] Nicole J. Mulcahy and Josep Call. "Apes Save Tools for Future Use". In: *Science* 312.5776 (2006), pp. 1038–1040. DOI: 10.1126/science.1125456.

[139] D. J. Mundfrom, D. G. Shaw, and T. L. Ke. “Minimum Sample Size Recommendations for Conducting Factor Analyses”. In: *International Journal of Testing* 5.2 (2005), pp. 159–168. DOI: 10.1207/s15327574ijt0502_4.

[140] Allen Newell, J. C. Shaw, and Herbert A. Simon. “Report on a general problem-solving program”. In: *Proceedings of the International Conference on Information Processing*. 1959, pp. 256–264.

[141] Allen Newell and Herbert A. Simon. *Human Problem Solving*. Prentice-Hall, 1972.

[142] Allen Newell and Herbert A. Simon. *Human Problem Solving*. Englewood Cliffs, NJ: Prentice-Hall, 1972. ISBN: 9780134454030.

[143] Thomas E Nichols and Andrew P Holmes. “Nonparametric permutation tests for functional neuroimaging: A primer with examples”. In: *Human Brain Mapping* 15.1 (2002), pp. 1–25. DOI: 10.1002/hbm.1058.

[144] Thomas E. Nichols and Andrew P. Holmes. “Nonparametric permutation tests for functional neuroimaging: A primer with examples”. In: *Human Brain Mapping* 15.1 (2002), pp. 1–25. DOI: 10.1002/hbm.1058.

[145] Richard E. Nisbett and Timothy D. Wilson. “Telling More Than We Can Know: Verbal Reports on Mental Processes”. In: *Psychological Review* 84.3 (1977), pp. 231–259. DOI: 10.1037/0033-295X.84.3.231.

[146] Kate Nussenbaum et al. “Moving Developmental Research Online: Comparing In-Lab and Web-Based Studies of Model-Based Reinforcement Learning”. In: *Developmental Science* 23.4 (2020), e12968. DOI: 10.1525/collabra.17213.

[147] Stellan Ohlsson. “The Problems with Problem Solving: Reflections on the Rise, Current Status, and Possible Future of a Cognitive Research Paradigm”. In: *The Journal of Problem Solving* 5.1 (2012), pp. 101–128. DOI: 10.7771/1932-6246.1144.

[148] Bas van Opheusden et al. “Expertise increases planning depth in human gameplay”. In: *Nature* 618.7967 (2023), pp. 1000–1005. ISSN: 1476-4687. DOI: 10.1038/s41586-023-06124-2.

[149] Arturo Orsini. “Corsi’s block-tapping test: Standardization and concurrent validity with WISC–R for children aged 11 to 16”. In: *Perceptual and Motor Skills* 79.3_suppl (1994), pp. 1547–1554.

[150] A. R. Otto et al. “The curse of planning: Dissecting multiple reinforcement-learning systems by taxing the central executive”. In: *Psychological Science* 24.5 (2013), pp. 751–761. DOI: 10.1177/0956797612463080.

[151] A. R. Otto et al. “Working-memory capacity protects model-based learning from stress”. In: *Proceedings of the National Academy of Sciences* 110.52 (2013), pp. 20941–20946. DOI: 10.1073/pnas.1312011110.

[152] Adrian M. Owen. “Cognitive planning in humans: neuropsychological, neuroanatomical and neuropharmacological perspectives”. In: *Progress in Neurobiology* 53 (1997), pp. 431–450.

[153] Hrag Pailian and Justin Halberda. “The Reliability and Internal Consistency of One-Shot and Flicker Change Detection for Measuring Visual Working Memory Capacity”. In: *Memory & Cognition* 43 (2015), pp. 397–420.

[154] J. L. Pardo-Vázquez and J. Fernández-Rey. “Working memory capacity and mental rotation: Evidence for a domain-general view”. In: *The Spanish Journal of Psychology* 15.3 (2012), pp. 881–890. DOI: 10.5209/rev_SJOP.2012.v15.n3.39381.

[155] J. J. de Paula, L. F. Malloy-Diniz, and M. A. Romano-Silva. “Reliability of working memory assessment in neurocognitive disorders: a study of the Digit Span and Corsi Block-Tapping tasks”. In: *Brazilian Journal of Psychiatry* 38 (2016), pp. 262–263. DOI: 10.1590/1516-4446-2015-1879.

[156] Tim Pearce et al. “Imitating Human Behaviour with Diffusion Models”. In: *The Eleventh International Conference on Learning Representations*. 2023.

[157] Marjorie A. Pett, Nancy R. Lackey, and John J. Sullivan. *Making Sense of Factor Analysis: The Use of Factor Analysis for Instrument Development in Health Care Research*. California: Sage Publications, 2003.

[158] L. H. Phillips et al. “Mental planning and the Tower of London task”. In: *Quarterly Journal of Experimental Psychology Section A* 54.2 (2001), pp. 579–597. DOI: 10.1080/713755977.

[159] L. H. Phillips et al. “The role of memory in the Tower of London task”. In: *Memory* 7.2 (1999), pp. 209–231. DOI: 10.1080/741944066.

[160] Louise H. Phillips, Margaret S. MacLeod, and Matthias Kliegel. “Adult aging and cognitive planning”. In: *The Cognitive Psychology of Planning*. Ed. by R. Morris and G. Ward. Psychology Press, 2005, pp. 111–134.

[161] Tracey C. S. Potter, Nessa V. Bryce, and Catherine A. Hartley. “Cognitive components underpinning the development of model-based learning”. In: *Developmental Cognitive Neuroscience* 25 (June 2017), pp. 272–280. DOI: 10.1016/j.dcn.2016.10.005.

[162] Thomas Pouncy, Pedro Tsividis, and Samuel J. Gershman. “What Is the Model in Model-Based Planning?” In: *Cognitive Science* 45.1 (2021), e12928. DOI: 10.1111/cogs.12928.

[163] Kristopher J. Preacher and Robert C. MacCallum. “Repairing Tom Swift’s electric factor analysis machine”. In: *Understanding Statistics* 2.1 (2003), pp. 13–43. DOI: 10.1207/S15328031US0201_02.

[164] Clare R. Raby et al. “Planning for the Future by Western Scrub-Jays”. In: *Nature* 445 (2007), pp. 919–921. DOI: 10.1038/nature05575.

[165] Alec Radford et al. *Language Models are Unsupervised Multitask Learners*. 2019.

[166] John Raven, John C. Raven, and John Hugh Court. *Manual for Raven's Progressive Matrices and Vocabulary Scales*. San Antonio, TX: Harcourt Assessment, 2003.

[167] John C. Raven. *Guide to the Standard Progressive Matrices*. London: H. K. Lewis, 1960.

[168] Eyal M Reingold et al. "Visual span in expert chess players: Evidence from eye movements". In: *Psychological Science* 12.1 (2001), pp. 48–55.

[169] Eyal M. Reingold and Neil Charness. "Perception in Chess: Evidence from Eye Movements". In: *Cognitive Processes in Eye Guidance*. Ed. by Geoffrey Underwood. Oxford: Oxford University Press, 2005, pp. 325–354.

[170] Eyal M. Reingold et al. "Visual Span in Expert Chess Players: Evidence from Eye Movements". In: *Psychological Science* 12.1 (2001), pp. 48–55. DOI: 10.1111/1467-9280.00309.

[171] Ulf-Dietrich Reips. "Web-based research in psychology: A review". In: *Zeitschrift für Psychologie* 230.4 (2022), pp. 250–258. DOI: 10.1027/2151-2604/a000475.

[172] Samuel Ritter et al. "Cognitive psychology for deep neural networks: A shape bias case study". In: *International conference on machine learning*. PMLR. 2017, pp. 2940–2949.

[173] Trevor W. Robbins et al. "A study of performance on tests from the CANTAB battery sensitive to frontal lobe dysfunction in a large sample of normal volunteers". In: *Journal of the International Neuropsychological Society* 4.5 (1998), pp. 474–490. DOI: 10.1017/S1355617798455073.

[174] S. Ian Robertson. *Problem Solving: Perspectives from Cognition and Neuroscience*. Hove, UK: Psychology Press, 2001. ISBN: 9780415203005.

[175] John Ruscio and Brendan Roche. "Determining the number of factors to retain in an exploratory factor analysis using comparison data of known factorial structure". In: *Psychological Assessment* 24.2 (2012), pp. 282–292. DOI: 10.1037/a0025697.

[176] Stuart J. Russell and Peter Norvig. *Artificial Intelligence: A Modern Approach*. 3rd. Upper Saddle River, NJ: Prentice Hall, 2010. ISBN: 9780136042594.

[177] Stuart J. Russell and Peter Norvig. *Artificial Intelligence: A Modern Approach*. 3rd. Upper Saddle River, NJ: Prentice Hall, 2010.

[178] Maureen Schmitter-Edgecombe, Caitlin McAlister, and Alyssa Weakley. "Naturalistic assessment of everyday functioning in individuals with mild cognitive impairment: The day-out task". In: *Neuropsychology* 26.5 (2012), pp. 631–641. DOI: 10.1037/a0029352.

[179] Jonathan W. Schooler. "Introspecting in the spirit of William James: comment on Fox, Ericsson, and Best (2011)". In: *Psychological Bulletin* 137.2 (Mar. 2011), pp. 345–350. DOI: 10.1037/a0022390.

[180] Julian Schrittwieser et al. "Mastering atari, go, chess and shogi by planning with a learned model". In: *Nature* 588.7839 (2020), pp. 604–609.

[181] Miriam Sebold et al. "Model-based and model-free decisions in alcohol dependence". In: *Neuropsychobiology* 70.2 (2014), pp. 122–131. DOI: 10.1159/000362840.

[182] Nur Muhammad Mahi Shafiullah et al. "Behavior Transformers: Cloning k modes with one stone". In: *Thirty-Sixth Conference on Neural Information Processing Systems*. 2022.

[183] Priti Shah and Akira Miyake. "The Separability of Working Memory Resources for Spatial Thinking: An Individual-differences Approach". In: *Journal of Experimental Psychology: General* 125.1 (1996), pp. 4–27. DOI: 10.1037/0096-3445.125.1.4.

[184] Tim Shallice. "Specific Impairments of Planning". In: *Philosophical Transactions of the Royal Society of London. B, Biological Sciences* 298.1089 (1982), pp. 199–209.

[185] Tim Shallice. "Specific impairments of planning". In: *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 298.1089 (1982), pp. 199–209. DOI: 10.1098/rstb.1982.0082.

[186] Tim Shallice and Paul W. Burgess. “Deficits in strategy application following frontal lobe damage in man”. In: *Brain* 114.Pt 2 (1991), pp. 727–741. DOI: 10.1093/brain/114.2.727.

[187] Claude E. Shannon. “Programming a Computer for Playing Chess”. In: *Philosophical Magazine* 41.314 (1950), pp. 256–275.

[188] Roger N. Shepard and Jacqueline Metzler. “Mental rotation of three-dimensional objects”. In: *Science* 171.3972 (1971), pp. 701–703. DOI: 10.1126/science.171.3972.701.

[189] Heather Sheridan and Eyal M Reingold. “Chess players’ eye movements reveal rapid recognition of complex visual patterns: Evidence from a chess-related visual search task”. In: *Journal of Vision* 17.3 (2017), p. 4.

[190] David Silver et al. “A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play”. In: *Science* 362.6419 (2018), pp. 1140–1144. DOI: 10.1126/science.aar6404.

[191] David Silver et al. “Mastering the Game of Go Without Human Knowledge”. In: *Nature* 550.7676 (2017), pp. 354–359.

[192] David Silver et al. “Mastering the game of go without human knowledge”. In: *nature* 550.7676 (2017), pp. 354–359.

[193] Herbert A Simon. “The structure of ill structured problems”. In: *Artificial Intelligence* 4.3-4 (1973), pp. 181–201.

[194] Herbert A. Simon. “Rational Choice and the Structure of the Environment”. In: *Psychological Review* 63.2 (1956), pp. 129–138. DOI: 10.1037/h0042769.

[195] Kevin Smith et al. “To what extent does the brain simulate the external world?” In: *Proceedings of the Conference on Cognitive Computational Neuroscience (CCN)*. 1330. Proposal for a Grouped Author Contribution (GAC) at CCN 2022. 2022, pp. 1–6.

[196] Andrew Solway, Carlos Diuk, and Natalia et al. Córdova. "Optimal Behavioral Hierarchy". In: *PLoS Computational Biology* 10.3 (2014), e1003779. DOI: 10.1371/journal.pcbi.1003779.

[197] Charles Spearman. "Correlation Calculated from Faulty Data". In: *British Journal of Psychology* 3.3 (1910), pp. 271–295. DOI: 10.1111/j.2044-8295.1910.tb00206.x.

[198] Harold Stanislaw and Natasha Todorov. "Calculation of Signal Detection Theory Measures". In: *Behavior Research Methods, Instruments, & Computers* 31.1 (1999), pp. 137–149. DOI: 10.3758/BF03207704.

[199] Laurence Steinberg et al. "Age differences in future orientation and delay discounting". In: *Child Development* 80.1 (2009), pp. 28–44. DOI: 10.1111/j.1467-8624.2008.01244.x.

[200] Hedinn Steingrimsson. "Chess fortresses, a causal test for state of the art Symbolic [Neuro] architectures". In: *2021 IEEE Conference on Games (CoG)*. IEEE. 2021, pp. 1–8.

[201] Alexander Steinke, Bruno Kopp, and Florian Lange. "The Wisconsin Card Sorting Test: Split-Half Reliability Estimates for a Self-Administered Computerized Variant". In: *Brain Sciences* 11.5 (2021), p. 529. DOI: 10.3390/brainsci11050529.

[202] P. Stratta et al. "Is Wisconsin Card Sorting Test performance related to 'working memory' capacity?" In: *Schizophrenia Research* 27.1 (1997), pp. 11–19. DOI: 10.1016/S0920-9964(97)00090-X.

[203] Thomas Suddendorf and Michael C. Corballis. "The Evolution of Foresight: What Is Mental Time Travel, and Is It Unique to Humans?" In: *Behavioral and Brain Sciences* 30.3 (2007), pp. 299–313. DOI: 10.1017/S0140525X07001975.

[204] Richard S. Sutton. "Dyna, an Integrated Architecture for Learning, Planning, and Reacting". In: *SIGART Bulletin* 2.4 (1991), pp. 160–163. DOI: 10.1145/122344.122377.

[205] Ryo Tachibana, Yukihiko Namba, and Yasuki Noguchi. "Two Speed Factors of Visual Recognition Independently Correlated with Fluid Intelligence". In: *PLOS ONE* 9.5 (2014), e97429. DOI: 10.1371/journal.pone.0097429.

[206] Christine M. Temple, Rebecca A. Carney, and Sean Mullarkey. "Frontal lobe function and executive skills in children with Turner's syndrome". In: *Developmental Neuropsychology* 12.3 (1996), pp. 343–363. DOI: 10.1080/87565649609540657.

[207] Siyu Teng et al. "Motion Planning for Autonomous Driving: The State of the Art and Future Perspectives". In: *arXiv preprint arXiv:2303.09824* (2023).

[208] Gerald Tesauro and Gregory Galperin. "On-line policy improvement using Monte-Carlo search". In: *Advances in Neural Information Processing Systems* 9 (1996), pp. 1068–1074.

[209] A. Toornstra et al. "Measuring Visual, Spatial, and Visual Spatial Short-Term Memory in Schoolchildren: Studying the Influence of Demographic Factors and Regression-Based Normative Data". In: *Journal of Pediatric Neuropsychology* 5 (2019), pp. 119–131. DOI: 10.1007/s40817-019-00070-6.

[210] Jennifer Tunstall. "Improving the utility of the Tower of London, a neuropsychological test of planning". PhD thesis. University of Stirling, 1999.

[211] Alan M Turing. "Computing machinery and intelligence". In: *Parsing the turing test*. Springer, 2009, pp. 23–65.

[212] J. M. Unterrainer et al. "Planning abilities and the Tower of London: Is this task measuring a discrete cognitive function?" In: *Journal of Clinical and Experimental Neuropsychology* 26.6 (2004), pp. 846–856. DOI: 10.1080/13803390490509574.

[213] Josef M. Unterrainer and Adrian M. Owen. "Planning and problem solving: From neuropsychology to functional neuroimaging". In: *Journal of Physiology-Paris* 99.4-6 (2006), pp. 308–317. DOI: 10.1016/j.jphysparis.2006.03.014.

[214] Josef M. Unterrainer et al. “Assessing Planning Ability Across the Adult Life Span in a Large Population-Representative Sample: Reliability Estimates and Normative Data for the Tower of London (TOL-F) Task”. In: *Journal of the International Neuropsychological Society* 25.5 (2019), pp. 520–529. DOI: 10.1017/S1355617718001248.

[215] Karthik Valmeekam et al. “Large Language Models Still Can’t Plan”. In: *arXiv preprint arXiv:2306.06224*. 2023.

[216] Bas van Opheusden, Luigi Acerbi, and Wei Ji Ma. “Unbiased and Efficient Log-Likelihood Estimation with Inverse Binomial Sampling”. In: *PLOS Computational Biology* 16.12 (2020), e1008483. DOI: 10.1371/journal.pcbi.1008483.

[217] Steven G. Vandenberg and Alan R. Kuse. “Mental rotations, a group test of three-dimensional spatial visualization”. In: *Perceptual and Motor Skills* 47.2 (1978), pp. 599–604. DOI: 10.2466/pms.1978.47.2.599.

[218] Ashish Vaswani et al. “Attention is All you Need”. In: *Advances in Neural Information Processing Systems*. Ed. by I. Guyon et al. Vol. 30. Curran Associates, Inc., 2017.

[219] Oliver Vikbladh, Evan Russek, and Neil Burgess. *Consolidation of Sequential Planning*. bioRxiv preprint. 2024. DOI: 10.1101/2024.11.01.621475.

[220] Jane X Wang et al. “Prefrontal cortex as a meta-reinforcement learning system”. In: *Nature neuroscience* 21.6 (2018), pp. 860–868.

[221] Xinyun Wang et al. “Self-Consistency Improves Chain of Thought Reasoning in Language Models”. In: *International Conference on Learning Representations (ICLR)* (2023).

[222] Jason Wei et al. “Chain-of-Thought Prompting Elicits Reasoning in Large Language Models”. In: *Advances in Neural Information Processing Systems (NeurIPS)* 35 (2022).

[223] Marilyn C. Welsh, Trey Satterlee-Cartmell, and Michelle Stine. “Towers of Hanoi and London: contribution of working memory and inhibition to performance”. In: *Brain and Cognition* 41.2 (1999), pp. 231–242. DOI: 10.1006/brcg.1999.1123.

[224] Brett Williams, Andrys Onsman, and Ted Brown. “Exploratory factor analysis: A five-step guide for novices”. In: *Australasian Journal of Paramedicine* 8.3 (2010). DOI: 10.33151/ajp.8.3.93.

[225] Anderson M. Winkler et al. “Permutation inference for the general linear model”. In: *NeuroImage* 92 (2014), pp. 381–397. DOI: 10.1016/j.neuroimage.2014.01.060.

[226] Joost C. F. de Winter, Dimitra Dodou, and Peter A. Wieringa. “Exploratory factor analysis with small sample sizes”. In: *Multivariate Behavioral Research* 44.2 (2009), pp. 147–181. DOI: 10.1080/00273170902794206.

[227] Daniel LK Yamins and James J DiCarlo. “Using goal-driven deep learning models to understand sensory cortex”. In: *Nature neuroscience* 19.3 (2016), pp. 356–365.

[228] Shunyu Yao et al. “ReAct: Synergizing Reasoning and Acting in Language Models”. In: *arXiv preprint arXiv:2210.03629* (2022).

[229] Shunyu Yao et al. “Tree of thoughts: Deliberate problem solving with large language models”. In: *Advances in Neural Information Processing Systems* 36 (2023), pp. 11809–11822.

[230] Denise Balem Yates et al. “WCST and NEUPSILIN: Relationships among Executive Functions, Attention, Memory and Language”. In: *Psicologia: Reflexão e Crítica* 26.3 (2013), pp. 506–515. DOI: 10.1590/S0102-79722013000300010.

[231] Che-Hung Yen et al. “Reduced Dopamine Transporter Availability and Neurocognitive Deficits in Male Patients with Alcohol Dependence”. In: *PLOS ONE* 10.6 (2015), e0131017. DOI: 10.1371/journal.pone.0131017.

[232] Ernst Zermelo. “Die Berechnung der Turnier-Ergebnisse als ein Maximumproblem der Wahrscheinlichkeitsrechnung”. In: *Mathematische Zeitschrift* 29.1 (1929), pp. 436–460. DOI: 10.1007/BF01180541.

[233] Zhengkang Zhang et al. “Split-Half Reliability of an Online Card Sorting Task in Young and Older Adults”. In: *Behavior Research Methods* 56.2 (2024), pp. 1039–1051. DOI: 10.3758/s13428-023-02104-6.

[234] Zirui Zhao, Wee Sun Lee, and David Hsu. “Large Language Models as Commonsense Knowledge for Large-Scale Task Planning”. In: *Advances in Neural Information Processing Systems (NeurIPS)*. Vol. 36. 37th Conference on Neural Information Processing Systems. 2023, pp. 1–15.

[235] N. A. Zook et al. “Working memory, inhibition, and fluid intelligence as predictors of performance on Tower of Hanoi and London tasks”. In: *Brain and Cognition* 56.3 (2004), pp. 286–292. DOI: 10.1016/j.bandc.2004.07.003.

[236] Zhaoyu Zuo et al. “Working Memory Guides Action Valuation in Model-based Decision-making Strategy”. In: *Journal of Cognitive Neuroscience* 37.1 (Jan. 2025), pp. 86–96. DOI: 10.1162/jocn_a_02237.

[237] William R. Zwick and Wayne F. Velicer. “Comparison of five rules for determining the number of components to retain”. In: *Psychological Bulletin* 99.3 (1986), pp. 432–442. DOI: 10.1037/0033-2909.99.3.432.