Revealing human planning strategies with eye tracking

Frederick Callaway (fredcallaway@gmail.com)
Department of Psychology, New York University

Miaomiao Yu (my2689@nyu.edu)
Department of Psychology, New York University

Marcelo G. Mattar (marcelo.mattar@nyu.edu)
Department of Psychology, New York University

Abstract

Most recent research on human planning attempts to adjudicate between different possible models based on their ability to predict choices and perhaps response times. Here, we propose that eye-tracking can provide a crucial additional constraint. Using a simple task that makes gaze highly revealing of internal planning, our results (1) provide a more nuanced perspective on previously proposed reward- and depth-based pruning mechanisms, (2) suggest that people use a planning strategy that incorporates elements of both best-first search and Monte Carlo tree search, and (3) suggest that planning (in our task) does not involve accumulating evidence about reward.

Keywords: planning; eye tracking

We developed a new paradigm that uses eye tracking to directly measure the computations underlying planning. The experimental interface is illustrated in Figure 1A. There are eleven locations (states), each labeled with the number of points one would gain or lose by moving there (rewards). The current state is highlighted in blue and possible actions are indicated by arrows. Participants select actions by clicking on the state they wish to move to, attempting to maximize cumulative reward before reaching a state with no outgoing arrows. Both the rewards and transition structure change on every trial. We use a random circular layout to prevent participants from using spatial heuristics that would not generalize to more naturalistic planning tasks (c.f., Correa, Ho, Callaway, Daw, & Griffiths, 2023; Zhu, Lakshminarasimhan, Arfaei, & Angelaki, 2022). Finally, we impose a time limit of 15 seconds to plan and execute a sequence of moves (in one phase).

We recorded participants' gaze continuously using an EyeLink 1000 with a chin rest. To eliminate any uncertainty about which reward a person is (visually) attending to at each moment, we adopt a gaze-contingent display, such that the reward at a given state is only shown when their gaze is recorded in a region near that state.

Results

We recruited 31 participants from the student pool at NYU. We excluded 3 participant due to poor tracker calibration. We additionally excluded 15 trials on which the participant indicated that the gaze contingency was not working and 200 trials on which the time limit was reached. This left 28 participants and 2585 trials in our final analysis.

Depth- and reward-based pruning Two of the most well-known mechanisms of human planning concern when people decide to "prune" a branch of their decision tree. Figure 2A shows that participants were decreasingly likely to continue searching down a path the deeper they were in the tree (B = -1.076, [-1.388, -0.763], $p < .001$; we only consider cases where a child state is available). However, in contrast to most models of depth-limited search (Keramati, Smittenaar, Dolan, & Dayan, 2016; Krusche, Schulz, Guez, & Speekenbrink, 2018; Snider, Lee, Poizner, & Gepshtein, 2015), we see a smoothly decreasing probability rather than a strict cutoff.

Other work has proposed that people prune paths when they discover a large penalty (Huys et al., 2012). Consistent with this, we found that people were more likely to continue searching down a path when the last-fixated state had a higher reward (Figure 2C; B = 0.019, [0.010, 0.028], $p < .001$). However, the effect was relatively weak, and was roughly linear. This contrasts with the standard pruning model in which large penalties are selectively avoided.

Comparing best-first and Monte Carlo tree search We next sought evidence for more sophisticated search strategies. We focused on the two most widely used search algorithms in AI, both of which have been proposed as models of human planning: best-first search (BFS; van Opheusden et al., 2023; Zhang, Lipovetzky, & Kemp, 2023) and Monte Carlo tree search (MCTS; Éltető & Dayan, 2023).

A key difference between BFS and MCTS is that MCTS

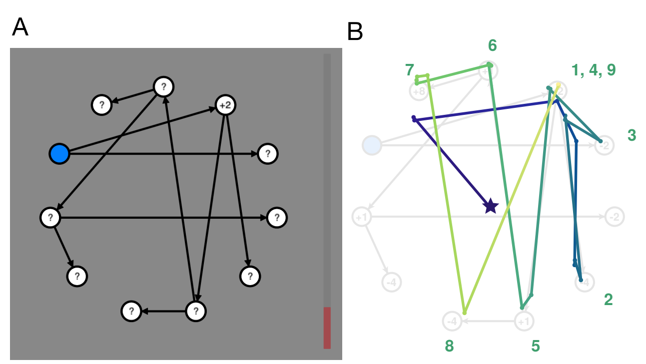

Figure 1: Measuring planning with eye tracking. (A) The task interface as shown to participants. (B) The sequence of fixations before the first action for one trial, starting at the star.

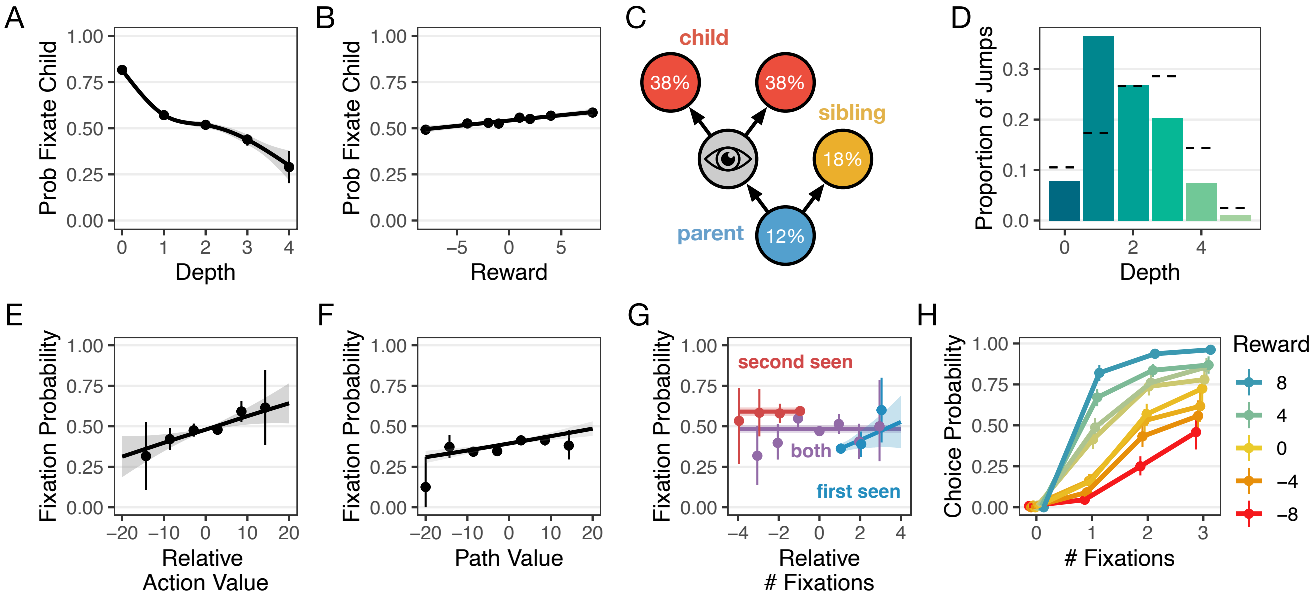

Figure 2: Results. (A) The probability that a "child" (see panel C) of the previously fixated state is fixated as a function of the depth of the previously fixated state. (B) Like A but for the reward of the previously fixated state. (C) The three types of saccades that occur above chance level. The eye indicates the previously fixated state. States are labeled with the total proportion of fixations of the given type (38% is the total probability across both child states). (D) The proportion of non-child saccades, or "jumps", to states of each depth. Chance is shown by dashed lines. (E) Conditioning on one of the two children of the last fixated state being fixated next, the probability of fixating an arbitrarily chosen one as a function of the difference in total attainable reward from visiting the two states ($Q(\cdot, s_1) - Q(\cdot, s_2)$). Unseen rewards are set to zero when computing action value. (F) Conditioning on a frontier state being fixated next, the probability of fixating an arbitrarily chosen one as a function of the cumulative reward up to (but not including) the given state. (G) Like E but as a function of the relative number of previous fixations to each state, split by whether each state had been fixated at least once. (H) The probability of visiting a state next (a binary choice) as a function of the number of fixations to that state and its reward. Legend excludes rewards of ± 1 and ± 2.

is constrained to simulate states in temporal order (rollouts), whereas BFS can consider states in arbitrary order. As shown in Figure 2C, people indeed most often fixated a *child* of the previously fixated state, consistent with the rollouts of MCTS. However, they also often fixated a *sibling*. This corresponds to considering an alternative action from the previous (parent) state, something BFS does frequently.

In cases where people don't fixate a child, MCTS predicts that they will "jump" back to the initial state. As shown in Figure 2D, this was uncommon in the data. However, this could be because people do not need to fully fixate the initial state to see where the arrows point. In this case, a new rollout would appear to begin in a depth-1 state, and indeed such states accounted for the greatest share of jump saccades (37%). However, 56% of jump saccades went to states at depth 2 and greater, inconsistent with MCTS.

In contrast, BFS predicts that fixations will be directed to states on the *search frontier*, those which have not been previously fixated but whose parent state has been fixated. Consistent with this, most fixations were on the search frontier (52%) and people were significantly more likely to fixate such states (B = 1.098, [0.976, 1.219], $p < .001$). However, there were also a substantial number of refixations to previously fixated states (41%), inconsistent with BFS.

A more subtle difference between BFS and MCTS lies in the influence of reward. In MCTS, search is directed towards states that have been found to lead to high *future* rewards. In contrast, BFS focuses search on states with high *past* reward. People were sensitive to both types of value. Focusing on cases where a child state was fixated (consistent with MCTS), Figure 2E shows that people were more likely to fixate the state that led to higher rewards deeper in the tree (B = 0.050, [0.036, 0.064], $p < .001$). Focusing on the cases where a frontier state was fixated (consistent with BFS), Figure 2F shows that people were more likely to fixate states at the end of paths with high cumulative reward (B = 0.026, [0.006, 0.047], p = .013). However, this effect was fairly weak.

MCTS and BFS also make different predictions regarding the tendency to seek out new information. MCTS predicts that people will preferentially fixate states that have received fewer fixations, whereas BFS predicts that people will specifically seek out states that have never been fixated. Focusing on the case where a child state was fixated, Figure 2G shows that people were indeed more likely to fixate the state that had not been seen yet (B = -0.489, [-0.627, -0.351], $p < .001$), but they were not strongly sensitive to the number of additional fixations (B = -0.024, [-0.114, 0.065], p = .593).

No evidence for evidence accumulation Finally, we consider how fixations relate to participants' ultimate choices. We predicted that repeated fixations would accumulate evidence for/against visiting states with positive/negative reward. Evidence accumulation underlies standard models of fixations in non-sequential choice (e.g., Krajbich, Armel, & Rangel, 2010) as well as most planning algorithms that revisit states, including MCTS and any approach based on "backups" (see also Solway & Botvinick, 2015). Contrary to our prediction, Figure 2H shows that additional fixations to a state strictly increased the probability that it was visited, regardless of reward. Considering only cases where both possible next states have been fixated at least once, the interaction between reward and fixations was *negative* (B = -0.055, [-0.071, -0.040], $p < .001$). In ongoing research, we are attempting to better understand the functional role of refixations in our task.

References

Correa, C. G., Ho, M. K., Callaway, F., Daw, N. D., & Griffiths, T. L. (2023, June). Humans decompose tasks by trading off utility and computational cost. *PLOS Computational Biology*, *19*(6), e1011087.

Éltető, N., & Dayan, P. (2023, June). *Habits of Mind: Reusing Action Sequences for Efficient Planning* (No. arXiv:2306.05298). arXiv.

Huys, Q. J. M., Eshel, N., O'Nions, E., Sheridan, L., Dayan, P., & Roiser, J. P. (2012). Bonsai trees in your head: How the Pavlovian system sculpts goal-directed choices by pruning decision trees. *PLOS Computational Biology*, *8*(3), e1002410.

Keramati, M., Smittenaar, P., Dolan, R. J., & Dayan, P. (2016, November). Adaptive integration of habits into depth-limited planning defines a habitual-goal–directed spectrum. *Proceedings of the National Academy of Sciences*, *113*(45), 12868–12873.

Krajbich, I., Armel, C., & Rangel, A. (2010). Visual fixations and the computation and comparison of value in simple choice. *Nature Neuroscience*, *13*(10), 1292–1298.

Krusche, M. J. F., Schulz, E., Guez, A., & Speekenbrink, M. (2018). Adaptive planning in human search. In *Proceedings of the Annual Meeting of the Cognitive Science Society.*

Snider, J., Lee, D., Poizner, H., & Gepshtein, S. (2015, September). Prospective Optimization with Limited Resources. *PLOS Computational Biology*, *11*(9), e1004501.

Solway, A., & Botvinick, M. M. (2015). Evidence integration in model-based tree search. *Proceedings of the National Academy of Sciences*, *112*(37), 11708–11713.

van Opheusden, B., Kuperwajs, I., Galbiati, G., Bnaya, Z., Li, Y., & Ma, W. J. (2023, May). Expertise increases planning depth in human gameplay. *Nature*, 1–6.

Zhang, C., Lipovetzky, N., & Kemp, C. (2023). Comparing AI Planning Algorithms with Humans on the Tower of London Task. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *45*(45).

Zhu, S., Lakshminarasimhan, K. J., Arfaei, N., & Angelaki, D. E. (2022, May). Eye movements reveal spatiotemporal dynamics of visually-informed planning in navigation. *eLife*, *11*, e73097.