

## **UC Merced**

### **Proceedings of the Annual Meeting of the Cognitive Science Society**

#### **Title**

Revealing human planning strategies with eye-tracking

#### **Permalink**

<https://escholarship.org/uc/item/8sm0k8z2>

#### **Journal**

Proceedings of the Annual Meeting of the Cognitive Science Society, 46(0)

#### **Authors**

Callaway, Frederick

Yu, Miaomiao

Mattar, Marcelo G

#### **Publication Date**

2024

#### **Copyright Information**

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

# 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 a small set of hypothesized models based on their ability to predict participants' choices, using carefully designed experiments and/or model comparison. Here, we propose an alternative approach. We designed a task in which gaze is highly indicative of participants' planning operations, allowing us to discover properties of human planning from eye-tracking data in a data-driven way. Our results reveal ways that people's planning strategies have both similarities and differences with classical planning algorithms like best-first search and Monte Carlo tree search. They also provide a more nuanced perspective on previously proposed properties of human planning like pruning and depth limits. We conclude that planning research would benefit greatly from an increased use of rich sources of data that provide more direct evidence about the internal processes underlying sequential decision-making.

**Keywords:** planning; eye tracking

## Introduction

If you brought together a room of researchers studying human planning, you could quickly achieve consensus on two (and perhaps only two) points. First, planning is among the most impressive and interesting human cognitive capacities. Second, studying planning is hard. These two facts have a single explanation: planning consists of long sequences of internal cognitive operations operating on rich mental representations. Despite this complexity, substantial progress in the study of human planning has been made using two strategies broadly known to any cognitive scientist. First, one can carefully design experiments such that competing theories make clearly divergent predictions about choice. Alternatively, one can rely on model comparison to identify which theory is statistically most consistent with the data.

The classical approach for studying human planning is well-exemplified by one of the most influential papers in the field. Huys et al. (2012) propose that people “prune” unpromising branches of their decision tree, not fully considering plans that involve paying a large cost early on. They found empirical support for this idea using a task that frequently pits long-term rewards against short-term costs; people failed to act optimally in this specific case, suggesting that large costs prevented them from considering later rewards. But there are other plausible explanations for this observed bias; perhaps people simply discount later rewards, or maybe they overweight large penalties? Here, model comparison comes to the

rescue. Compared to these alternatives, the proposed pruning model assigns the highest likelihood to participant's choices.

While impressive—and ultimately necessary—inferring planning processes from choices alone has a major drawback. Whether these inferences are supported by a carefully engineered behavioral result or formal model comparison, they can only provide relative support for one theory vs. another. Experiments are typically designed to distinguish between two possible models (one of which is often a baseline). And model comparison is literally just that: a *comparison*. If both our experimental designs and our interpretation of behavior depend on the theories the researcher already had in mind, then we will have a hard time discovering anything genuinely new about human planning, and what we do discover will be based on circumstantial evidence.

Is there a better, or at least easier way? What if, instead of *inferring* planning operations, we could simply *measure* them? Interestingly, some of the earliest work on human planning, conducted by Newell and Simon (1972), took this approach. They developed their models based on “think-aloud protocols” in which participants told the experimenter's exactly what they were thinking about; they credit these transcripts with the idea for means-ends analysis, which ultimately became the core component of their theory of human problem solving. Inspired by this approach, recent work has studied planning using more constrained forms of process-tracing, requiring participants to click to reveal rewards at future states (Callaway et al., 2022; Jain et al., 2022). However, such a heavy-handed approach likely elicits different planning strategies than people would use in the real world.

Here, we develop a new paradigm for studying planning using eye tracking. Eye tracking has already had an enormous impact on our understanding of how people make “simple” one-shot choices (e.g., Krajovich, Armel, & Rangel, 2010), and recent work has shown how gaze data can reveal properties of people's planning in sequential tasks (Cristín, Méndez, & Campos, 2022; Kadner, Willkomm, Ibs, & Rothkopf, 2023; Zhu, Lakshminarasimhan, Arfaei, & Angelaki, 2022). Building on this work, we designed a minimalistic paradigm that allows us to determine with a high degree of accuracy exactly which future state people are considering at each moment, while still pushing the limits of their abilities. This allows us to characterize human approximate planning strategies at a level of detail not possible in previous work.

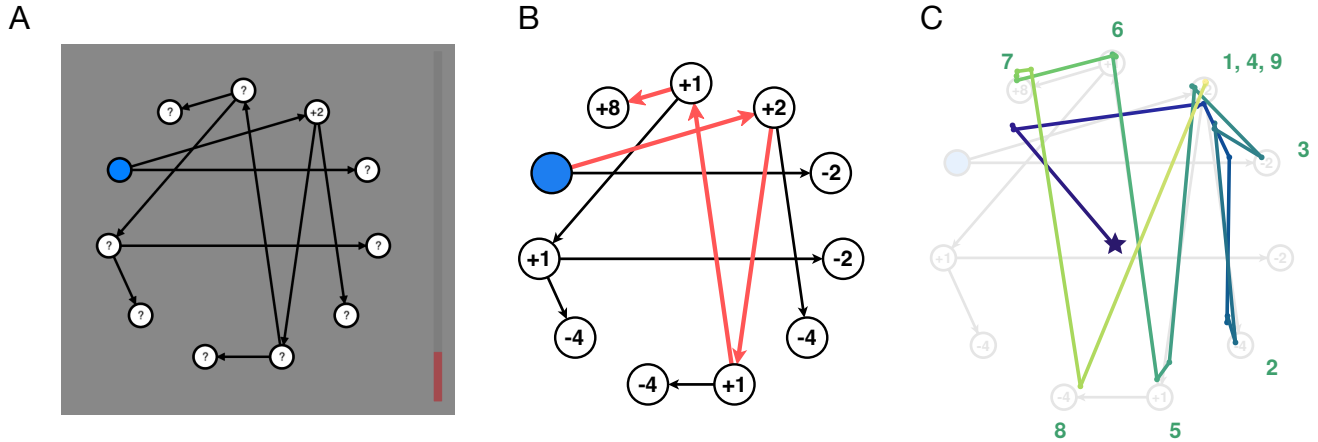


Figure 1: Measuring planning with eye tracking. (A) The task interface as shown to participants. Participants click to navigate between states, following arrows, aiming to collect the most points. Their current state is highlighted in blue. Here, gaze is directed to the state in the top right, so its reward is revealed. (B) The chosen action sequence (red arrows) on an example trial. (C) The sequence of fixations before the first action for the trial in panel B. The sequence begins at the central star (navy) and ends in the top right (yellow) 5.9 seconds later. The order of recorded state fixations (when the gaze-contingent display revealed the reward) are indicated by green numbers. This example had the most visually interpretable gaze data of around fifty random trials we considered. The complete dataset (including videos) and analysis code is available at <https://github.com/fredcallaway/callaway2024revealing>.

### A minimal eye tracking planning task

Our primary contribution is a new paradigm that uses eye tracking to directly measure the computations underlying planning. We designed the paradigm to satisfy five key desiderata: (1) the participant’s gaze clearly indicates which future state they are currently evaluating, (2) decision-time planning (and not across-trial learning) is the only way to make informed decisions, (3) strong performance requires integrating rewards and costs across multiple actions, (4) the task pushes the limits of human working memory capacity and/or processing speed, such that few people can reach near-optimal performance, and finally (5) the task is as simple as possible given the previous constraints.

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. Once the move is made, the decision is final—they cannot reverse the action or revisit previous states. The trial ends when the participant reaches a state with no outgoing arrows. The participant’s goal is to visit a sequence of states that maximizes the total points earned.

Both the rewards and transition structure change on every trial; this ensures that decision-time planning is the only way to make informed choices. It also encourages participants to look at states as they consider them, to discover or remind themselves of the reward there. We additionally 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. This resolves any uncertainty about which reward a person is

(visually) attending to at each moment.

A key feature of the task is the random, circular layout (c.f., Correa, Ho, Callaway, Daw, & Griffiths, 2023; Zhu et al., 2022), which prevents participants from using physical proximity as a cue to connectivity. This layout has two important consequences. First, it prevents one from quickly scanning to identify regions of high reward. Second, it requires participants to internally represent and track the sequence of actions they would take to get to a current state, as the layout does not make this information immediately apparent.

Finally, we impose a time pressure of 15 seconds to complete each trial, with the time remaining indicated by an animated bar on the right side of the display. If the time runs out, the bar flashes red and random actions are forced until a terminal state is reached (one with no outgoing connections). Participants perform all their planning and actions within this single 15-second period; we do not impose separate planning and execution phases.

Before we continue, we must address an obvious critique of this paradigm, that eye movements reflect information-seeking and not *true* planning. While we agree that studying planning that occurs entirely in the head is worthwhile, there are at least two reasons to believe that our task engages and reveals planning processes. First and foremost, much of the planning that people do in the real world involves concurrent information seeking. We use maps to plan routes, calendars to schedule events, and we look at the board when playing board games (indeed, gaze correlates with model-predicted planning operations in this context; van Opheusden et al., 2023). Second, most work employing eye tracking in non-sequential choices makes the “eye-mind assumption” (Just & Carpenter, 1976) that gaze indicates internal processing.

## Methods

**Stimuli** On each trial, the points on each state were drawn from the set  $\{-8, -4, -2, -1, +1, +2, +4, +8\}$  with equal probability. The connections were sampled from the set of unbalanced binary trees with 11 nodes. That is, each state besides the initial state has one *parent* (incoming arrow, a state that could be visited immediately before) and either two or zero *children* (outgoing arrows, states that could be visited next; see Figure 2A for an illustration). This yielded trees with maximal depth between three and five.

**Eye tracking** Participants' gaze was recorded monocularly at 500 Hz using an EyeLink 1000 Plus with a chin rest. Each trial began with a drift check, and we recalibrated the eye tracker whenever the drift check failed repeatedly. As discussed above, we employed a gaze-contingent display; each reward was only visible when the participants' gaze was recorded to be within a circular region centered on each state. The size of this region was calibrated separately for each participant such that they could quickly and reliably trigger the display when instructed to fixate on each state.

**Procedure** The experiment began with an interactive instruction phase that explained the rules of the task. They then completed three practice trials, on which they had to earn the maximum number of points before moving on. After three failures, the experimenter intervened and ensured the participants understood the instructions before they could move on. Participants then completed the eye tracking calibration. Finally, they completed 100 trials of the main task.

**Participants and exclusions** We recruited 31 participants from the student research participation pool at New York University. We excluded 3 participants due to poor eye 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.

## Results

We begin by analyzing participants' behavioral performance. We then analyze their eye movements. Finally we consider the relationship between eye movements and choice.

### Behavioral performance

We began by analyzing the participant's selection of paths. Overall, participants selected an optimal path on 63% of trials, earning 75% of the possible reward. Note that chance performance is 0 reward, since the expected reward at each state, and thus also the total reward under random behavior, are both zero. At the level of individual actions, participants selected a correct successor state (one that is on an optimal path from their current state) 84% of the time (chance: 53%). Although participant performance was only slightly better than that attained by myopically maximizing immediate reward (73% of possible reward), their choices were sensitive to future rewards (the optimal state-value function,  $V(s)$ ;

$B = 0.192 [0.150, 0.235]$ ,  $p < .001$ ) as well as immediate reward ( $B = 0.336 [0.306, 0.367]$ ,  $p < .001$ ).<sup>1</sup> Note that  $B$  (vs.  $\beta$ ) indicates a non-standardized coefficient; thus, dividing the coefficients, we see that future rewards were weighed only 57% as heavily as immediate rewards, indicating steep temporal discounting.

### Fixations

We next aimed to characterize participants' fixation behavior. Note that throughout the paper, we use the term "fixation" to refer to a period of time in which gaze was recorded near a given state (and thus, the reward was displayed in the gaze-contingent design). Sequential fixations to the same state are merged. Thus, each of our "fixations" typically includes many true eye fixations.

In the analyses below, we exclude fixations that immediately precede a move to the fixated state (15% of all fixations). These fixations could have been initiated purely to guide the cursor to the chosen state, thus reflecting an already-made choice rather than planning. This could lead to making spurious conclusions about planning (for example, that people preferentially consider rewarding states) based on fixations that actually reflect choice (people move to rewarding states).

**Search focuses on proximal states** If people are performing any sort of non-exhaustive forward search, they would show an overall attentional bias towards states closer to their current state. Figure 2A thus shows the average number of fixations that states of different depths receive before the first action is taken on each trial. Indeed, we see that states that can be reached in one step from the initial state (depth 1) receive on average 1.55 fixations before the first action, while states at all other depths receive less than one on average, falling to only 0.26 for the deepest states, five steps into the tree. This reduced attention to more distal states is consistent with the earlier result that future rewards are down-weighted relative to future rewards. We return to this point below.

**Saccades reflect local search** We next asked whether participants' fixations depended on the available transitions between states. To that end, we labeled each pair of sequential fixations, or "saccade", with the graphical relationship between the two fixated states. In the tree-structured graphs we use, this relationship can be defined by the number of steps backwards (against the arrows) and then forwards (with the arrows) one would take to get from one state to the other. Figure 2 illustrates the three types of saccade that occurred more often than chance (all  $p < .001$  in proportion tests): fixations to a *child* of the last-fixated state (38% of all saccades), a *sibling* (17%), or a *parent* (12%). That is, people were most likely to consider a state they could visit immediately *after* the last-considered state, or one they could visit

<sup>1</sup> All reported regressions include random (uncorrelated) slopes and intercepts (except in the case where they can be specified *a priori*), implemented with the lmer R package. Brackets indicate 95% CI. We always run a single joint regression for each dependent variable.

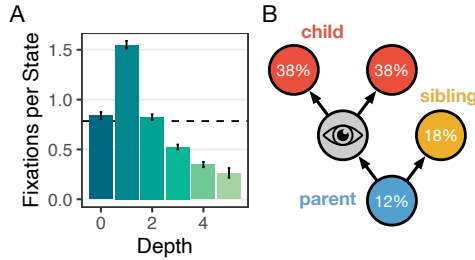


Figure 2: Fixations reflect transition structure. (A) The average number of fixations per state at each depth in the tree. This plot only includes fixations made before the first move. The dashed line indicates chance (average number of fixations divided by the number of states). (B) An illustration of 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).

instead of the last-considered state, occasionally one they would visit *before*. This pattern of results is broadly consistent with best-first search, which tends to consider children and siblings of the last-considered state (and, in versions with explicit backtracking, parents). Although the high frequency of child saccades is suggestive of rollouts (purely forward simulations), rollout-based algorithms like Monte Carlo tree search (MCTS) never predict fixations to a sibling or parent.

In the next three sections, we aimed to characterize in greater detail the factors driving people’s search. Given the frequency of “child” saccades identified above, we broke down the decision of which state to fixate next into three components: (1) If the last-fixated state has children, do you fixate one of them next? (2) If so, which child do you fixate? (3) If not, which state do you fixate instead?

**Continue or switch?** As shown above, participants frequently fixated children of the previously-fixated state, which we can interpret as further considering a given path. However, they also frequently violated this general tendency. What determines when people abandon the path they were considering and switch to an alternative?

One influential theory is that people “prune” paths when they discover a large penalty (Huys et al., 2012). Figure 3A thus plots the continuation probability as a function of reward of the last fixated state. We see a significant, but fairly weak effect ( $B = 0.019$  [0.010, 0.028],  $p < .001$ ). In contrast to the standard pruning model, which posits a specific aversion to large negative rewards, we see a roughly linear effect, more in line with best-first search. However, in contrast with best-first search, we did not see an effect of the rewards leading up to the last-fixated state ( $B = 0.000$  [-0.019, 0.019],  $p = .975$ ). Nor did we see a (positive) effect of the rewards following it ( $B = -0.043$  [-0.062, -0.024],  $p < .001$ ), as predicted by an earlier pruning model (Dayan & Huys, 2008).

Another commonly proposed idea is that people limit the

depth of their search. Indeed, Figure 3B shows that people were decreasingly likely to continue search the deeper they got into the tree ( $B = -1.076$  [-1.388, -0.763],  $p < .001$ ; note that we only consider cases where a child-state is available). However, in contrast to extant models of depth limits in human planning (Keramati, Smittenaar, Dolan, & Dayan, 2016; Krusche, Schulz, Guez, & Speekenbrink, 2018; Snider, Lee, Poizner, & Gepshtein, 2015), this tendency expresses itself in a continuous way rather than as a strict cutoff, even at the individual level.

**Continuation policy** When continuing to search down a path, what determines which direction one searches? That is, how do people decide which of the two children of the last-fixated state to fixate next? One intuitive idea is that people would seek to balance *exploitation* of states they already know lead to high rewards with *exploration* of states they haven’t considered much. This is the strategy adopted by the standard version of Monte Carlo tree search, specifically UCT (Kocsis & Szepesvári, 2006).

Starting with exploitation (reward-seeking), Figure 4A shows that participants were more likely to fixate the child state with higher *action value* ( $Q$  value), that is, the sum of immediate and future rewards (assuming optimal choice). Considering the two components separately, people were sensitive to both immediate reward ( $B = 0.049$  [0.035, 0.062],  $p < .001$ ) and future reward ( $B = 0.058$  [0.038, 0.078],  $p < .001$ ), weighed roughly equally. Note that for this analysis, we set unseen rewards to zero (since the expected reward at each state is zero). Thus, people were more likely to consider states that they had previously found to yield better outcomes.

Turning to exploration (uncertainty-seeking), Figure 4B shows that people were more likely to fixate the child that had

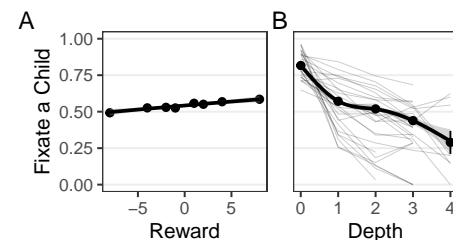


Figure 3: Continue or switch? Conditioning on an additional fixation being made, each panel shows the probability that a child of the last-fixated state is fixated (when one is available) as a function of the (A) reward and (B) depth of the last fixated state. Depth is relative to the current state. In panel A, cases where the last-fixated state is the current state are excluded (as the reward has already been consumed). In panel B, faint lines show individuals. Here, and in all future plots, smooth lines show fits and standard error of a generalized additive model. Points show (possibly binned) means and 95% bootstrapped confidence intervals. We exclude bins containing fewer than ten data points.

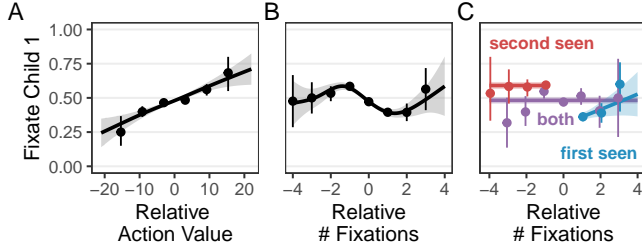


Figure 4: Continuation policy. Conditioning on one of the two children of the last fixated state being fixated next, each panel shows the probability of fixating an arbitrarily chosen “child 1” as a function of the difference in (A) total attainable reward from visiting the two states ( $Q(s_1) - Q(s_2)$ ), (B) the number of fixations to the two states, and (C) number of fixations, split by whether each state had been fixated at least once. In panels A and B, unseen rewards are set to zero.

received fewer fixations. However, this effect appears to be driven by the center of the scale (-1 to 1). Notably, this range captures the specific case where one state had been fixated and the other had not. Figure 4C thus shows the effect of fixation count separately for the case where only the first, only the second, or both child states had been previously fixated. We see that people were more likely to fixate unseen states ( $B = -0.489 [-0.627, -0.351]$ ,  $p < .001$ ); but controlling for this effect, they were not (strongly) sensitive to the relative fixation count ( $B = -0.024 [-0.114, 0.065]$ ,  $p = .593$ ). This is inconsistent with MCTS, which predicts a negative effect of each additional fixation.

**Switching policy** Finally, when one chooses not to fixate a child of the last-fixated state—or there is no child available—where does one fixate instead? Echoing Figure 4A, participants tended to switch to states with higher previously-seen reward (Figure 5A;  $B = 0.032 [0.015, 0.050]$ ,  $p < .001$ ). However, in contrast to the continuation policy, we did not see an effect of future rewards on switching decisions ( $B = 0.010 [-0.009, 0.029]$ ,  $p = .309$ ). Nor did we see an effect of the rewards leading up to the state, as predicted by best-first search ( $B = 0.012 [-0.006, 0.030]$ ,  $p = .196$ ).

Rollout-based algorithms like MCTS make a strong prediction that non-child saccades (switches) should exclusively be directed to the current state, to begin a new rollout. Interestingly, we instead found that participants showed a strong tendency to switch to *children* of their current state (Figure 5B;  $B = 0.851$ , 95% CI  $[0.679, 1.023]$ ,  $p < .001$ ), with an otherwise smoothly decreasing probability of switching to deeper states ( $B = -0.274 [-0.353, -0.194]$ ,  $p < .001$ ). Although this appears to contradict MCTS, it’s possible that people implicitly begin the rollout at the current state, but can either recall or see where its children are located without fixating it.

Echoing Figure 4C, Figure 5B also shows that participants were more likely to switch to states that had not previously been fixated. However, this was only true for depth-1 states

( $B = 0.675 [0.423, 0.927]$ ,  $p < .001$ ), not for other states ( $B = -0.104 [-0.321, 0.114]$ ,  $p = .350$ ). Controlling for the seen/unseen effect, we again found no effect of number of previous fixations on the switching policy ( $B = -0.007 [-0.098, 0.084]$ ,  $p = .885$ )—in contrast to the prediction of MCTS if these fixations reflected the second step of a rollout implicitly begun at the current state.

## Fixations and choice

At long last, we now turn to the relationship between participant’s fixations and their choices.

**Inattention (mostly) explains discounting** Recall that we found a considerable degree of myopia in people’s choices; future rewards were weighed about 57% as heavily as immediate rewards. However, we also found that people directed their fixations towards proximal states (Figure 2A). Does reduced attention to distal rewards explain the apparent discounting of those rewards? Repeating the regression from “Behavioral performance”, but computing immediate reward and future value only using rewards that were fixated at least once, we find that immediate and future rewards exert nearly equal influence on choice (immediate:  $B = 0.462 [0.412, 0.512]$ ,  $p < .001$ ; future:  $B = 0.435 [0.386, 0.484]$ ,  $p < .001$ ), roughly corresponding to a discount rate of 94%.<sup>2</sup> This suggests that, in our task, the behavioral discounting of distal rewards is almost entirely explained by reduced attention to those rewards.

**No evidence for evidence accumulation** We next considered the effect of repeated fixations on the same state. We expected to see a robust interaction between the number of fixations to a state and its reward in predicting choice. Specifically, repeated fixations to a positive reward should increase choice probability, while repeated fixations to a negative re-

<sup>2</sup>Note that we apply a single discount factor to all future rewards, whereas true discounting compounds for rewards that are multiple steps away. The true discounting rate is thus closer to 1 than what we report.

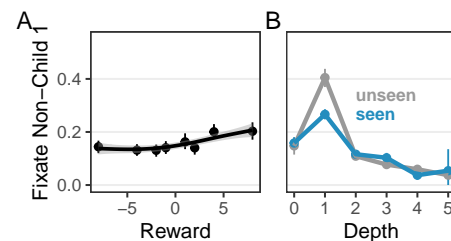


Figure 5: Switching policy. Conditioning on a *non-child* being fixated, each panel shows the probability that an arbitrarily chosen non-child state is fixated next as a function of the (A) reward and (B) depth, of the given state. In panel A, we exclude states which have not been previously fixated (reward unknown). In panel B, we separate the cases where the state had been previously fixated or not.

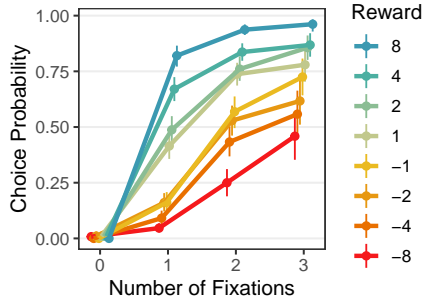


Figure 6: Fixations and choice. The probability of visiting a state next (a binary choice) as a function of the number of fixations to that state and its reward.

ward should decrease choice probability. Such an interaction would reflect evidence accumulation dynamics, in which repeated fixations to a state collect additional evidence about the state’s reward, thus increasing its influence on decisions. This kind of attention-weighted evidence accumulation is the core mechanism in standard models of fixations in non-sequential choice, such as the attentional drift diffusion model (Krajibich et al., 2010). It is also predicted by almost all planning algorithms that revisit states, including MCTS and any approach based on “backups”, such as Dyna (Sutton, 1990).

Figure 6 shows the probability of visiting a state (conditional on being in its parent) as a function of the number of previous fixations to that state and its reward. Unsurprisingly, states with higher reward are more likely to be visited, as long as they have been fixated at least once (the reward was displayed). Slightly more surprisingly, states that have never been fixated are almost never visited, but this makes sense given that it would be unnatural to click a state without looking at it. Most surprising, however, we see that additional fixations to a state strictly increase the probability that it is visited, regardless of reward. In fact, considering only cases where both possible next states have been fixated at least once, the interaction between reward and fixations is *negative* (interaction:  $B = -0.055 [-0.071, -0.040]$ ,  $p < .001$ ; reward:  $B = 0.425 [0.384, 0.467]$ ,  $p < .001$ ; fixations:  $B = 0.178 [-0.040, 0.397]$ ,  $p = .109$ ).<sup>3</sup> Note that, despite the seemingly robust main effect of repeated fixations on choice shown in Figure 6, there is substantial uncertainty in the size of the effect in the mixed-effects regression.

## Discussion

Here, we have presented a new experimental paradigm for studying human planning with eye tracking. Taken together, our results support previous theories of human planning in many ways. However, we also find several departures from previous models, ranging from subtle to downright puzzling.

Consistent with depth limits, we found that people largely

focused their search on proximal states, often not considering states more than one or two steps from their current state before making a choice (Figure 2A). However, contrary to most concrete implementations of this idea, we found that this tendency was graded rather than following a strict cutoff (Figure 3B). We also found that this attentional bias explained the majority of behavioral discounting of future rewards.

Consistent with pruning, we found a modest effect of reward on the decision to continue searching down a path (Figure 3A). However, the effect was relatively weak, and was roughly linear, in contrast to the standard model in which large penalties are selectively avoided. One plausible explanation for this difference is that pruning depends on a stable reward structure, which allows a habitual avoidance response to be learned. Note that cost magnitude cannot explain the difference, as the -8 in our task is actually slightly more extreme than the -140 in Huys et al. (2012), relative to the other possible rewards.

Comparing the two most common search algorithms from AI, best-first search and Monte Carlo tree search (MCTS), our participants appeared to employ something like a hybrid of the two. Like best-first search (and unlike MCTS), they often made saccades between sibling states, evaluating two alternative actions from a given state in sequence (Figure 2B). Furthermore, they selectively explored never-before-seen states (4C and 5B) but did not differentially explore states that had been fixated many vs. few times. However, like MCTS (and unlike best-first search), they frequently revisited states, especially those already found to lead to large rewards (Figure 4A), and they often jumped back to an early depth-1 state to begin a new search trajectory (Figure 5B).

A hybrid best-first-MCTS model might also explain our most puzzling result, that repeated fixations to a state did not increase the influence of its reward on choice (Figure 6). We tentatively hypothesize that these repeated fixations did not serve to estimate rewards, but instead functioned solely as a way to navigate the transition structure. That is, like best-first search, people may (roughly) faithfully track all the rewards they have seen and attempt to direct search towards promising and underexplored regions of the graph. But they cannot store the full search frontier in working memory, and so are forced to navigate through the graph, perhaps even “searching” for the states that they want to consider next.

Researchers studying planning are faced with an especially challenging version of a problem facing all cognitive scientists: inferring complex cognitive processes underlying comparably sparse behavior. Here, we have developed a task that makes gaze maximally informative about internal planning operations, and shown how this approach allows us to reveal both intuitive and surprising properties of human planning using simple exploratory data analyses. We hope that this work will inspire other planning researchers to draw on rich sources of data beyond decisions and response times when conceiving and evaluating theories of human planning.

<sup>3</sup>This regression predicts choice between the two possible next states based on the relative reward, the relative number of fixations, and the interaction. The interaction term is  $\text{nfix}_1 r_1 - \text{nfix}_2 r_2$ .

## References

- Callaway, F., van Opheusden, B., Gul, S., Das, P., Krueger, P. M., Griffiths, T. L., & Lieder, F. (2022, August). Rational use of cognitive resources in human planning. *Nature Human Behaviour*, 6(8), 1112–1125.
- 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.
- Cristín, J., Méndez, V., & Campos, D. (2022, December). Informational Entropy Threshold as a Physical Mechanism for Explaining Tree-like Decision Making in Humans. *Entropy*, 24(12), 1819.
- Dayan, P., & Huys, Q. J. M. (2008, February). Serotonin, Inhibition, and Negative Mood. *PLOS Computational Biology*, 4(2), e4.
- 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.
- Jain, Y. R., Callaway, F., Griffiths, T. L., Dayan, P., He, R., Krueger, P. M., & Lieder, F. (2022, July). A computational process-tracing method for measuring people’s planning strategies and how they change over time. *Behavior Research Methods*.
- Just, M. A., & Carpenter, P. A. (1976). Eye fixations and cognitive processes. *Cognitive Psychology*.
- Kadner, F., Willkomm, H., Ibs, I., & Rothkopf, C. (2023). Finding your Way Out: Planning Strategies in Human Maze-Solving Behavior. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 45(45).
- 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.
- Kocsis, L., & Szepesvári, C. (2006). Bandit Based Monte-Carlo Planning. , 282–293.
- 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*.
- Newell, A., & Simon, H. A. (1972). *Human problem solving* (Vol. 104). Prentice-hall Englewood Cliffs, NJ.
- Snider, J., Lee, D., Poizner, H., & Gepshtein, S. (2015, September). Prospective Optimization with Limited Resources. *PLOS Computational Biology*, 11(9), e1004501.
- Sutton, R. S. (1990). Integrated architectures for learning, planning, and reacting based on approximating dynamic programming. In *Proceedings of the seventh international conference on machine learning* (pp. 216–224).
- 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.
- 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.