Mouselab-MDP: A new paradigm for tracing how people plan

Frederick Callaway [1]
Dept. of Psychology, UC Berkeley
fredcallaway@berkeley.edu

Falk Lieder [1]
Dept. of Psychology, UC Berkeley
falk.lieder@berkeley.edu

Paul M. Krueger [1]
Dept. of Psychology, UC Berkeley
pmk@berkeley.edu

Thomas L. Griffiths
Dept. of Psychology, UC Berkeley
tom_griffiths@berkeley.edu

[1] **These authors contributed equally.**

Abstract

Planning is a latent cognitive process that cannot be observed directly. This makes it difficult to study how people plan. To address this problem, we propose a new paradigm for studying planning that provides experimenters with a timecourse of participant attention to information in the task environment. This paradigm employs the information-acquisition mechanism of the Mouselab paradigm, in which participants click on options to reveal the outcome of choosing those options. However, in contrast to the original Mouselab paradigm, our paradigm is a sequential decision process, in which participants must plan multiple steps ahead to achieve high scores. We release Mouselab-MDP open-source as a plugin for the JsPsych online Psychology experiment library. The plugin displays a Markov decision process as a directed graph, which the participant navigates to maximize reward. To trace the the process of planning, the rewards associated with states or actions are initially occluded; the participant has to click on a transition to reveal its reward. This information gathering behavior makes explicit the states the participant considers. We illustrate the utility of the Mouselab-MDP paradigm with a proof-of-concept experiment in which we trace the temporal dynamics of planning in a simple environment. Our data shed new light on people's approximate planning strategies and on how people prune decision trees. We hope that the release of Mouselab-MDP will facilitate future research on human planning strategies. In particular, we hope that the fine-grained time course data that the paradigm generates will be instrumental in specifying algorithms, tracking learning trajectories, and characterizing individual differences in human planning.

Keywords: planning; process tracing; meta-decision making; pruning; research methods

Acknowledgements

This work was supported by grant number ONR MURI N00014-13-1-0341 and a grant from the Templeton World Charity Foundation.

1 Introduction and Background

Many if not all of us have found ourselves in an unfortunate predicament due to a lapse of forethought. However, despite the ease with which such events come to memory, they are the exception rather than the rule. Humans have an exceptional ability to establish long-term goals and make steady progress towards their completion, often over the course of years or decades. This is only possible through planning. Planing is a broad notion, potentially taking multiple forms (Morris & Ward, 2004); however for present purposes, we define planning as explicitly considering potential future states, the actions one might take in them, the states that might result from those actions, and so on. For example, a student deciding whether or not to attend a class, is engaged in planning when they consider doing homework during class time and skimming the slides on the morning of the following lecture.

Planning is a fundamental aspect of higher-order cognition, and it has accordingly received much attention in cognitive psychology. Planning has been studied with verbal protocol analysis and computer simulation of problem solving (Newell, Simon, et al., 1972), models of the hierarchical structure of human behavior (Miller, Galanter, & Pribram, 1986; Botvinick, Niv, & Barto, 2009), errors and reaction times in sequential decision-making (e.g., Huys et al., 2012, 2015), and neural activity in spatial navigation tasks (Ólafsdóttir, Barry, Saleem, Hassabis, & Spiers, 2015; Simon & Daw, 2011; Balaguer, Spiers, Hassabis, & Summerfield, 2016). Planning has also received considerable attention in artificial intelligence (Russell, Norvig, & Intelligence, 2010; LaValle, 2006), and Markov decision processes (Sutton & Barto, 1998) have emerged as a common mathematical framework for bridging these two literatures.

Research on planning is complicated by the fact that we cannot directly observe the cognitive processes of planning. Thus, researchers must infer this latent cognitive process based on the decisions participants ultimately make, or on neural data which are challenging to collect and interpret. One approach to this problem is to design *process tracing* paradigms that externalize some aspect of the cognitive process. Payne, Bettman, and Johnson (1988) developed one such methodology for studying multi-alternative risky choice: the "Mouselab" paradigm. The paradigm presents a decision problem as a payoff matrix whose entries are initially occluded. Participants choose between multiple gambles whose outcome-dependent payoffs are recorded in the cells of the payoff matrix. Each column of the matrix corresponds to a gamble and each row of the matrix corresponds to one of the possible outcomes. The outcome probabilities are known and the gambles differ only in how much money they assign to each of the possible outcomes. Critically, to find out how much money a gamble pays for each of the outcomes, the participant has to click on the corresponding cell of the payoff matrix. This series of clicks makes experimentally observable the information a participant considers in making a decision. This new form of data led to novel insights into people's decision processes including the discovery of adaptive strategy selection (Payne et al., 1988).

Although the traditional Mouselab paradigm is very useful for studying decision making, it cannot be used to study planning because the decision on one gamble does not affect future gambles. Thus, to apply the Mouselab process tracing method to planning, we simply replace the single decision with a Markov Decision Process (MDP), in which a participant must make a sequence of choices, each one affecting the choices that will be available in the future. By releasing the paradigm open-source as a JsPsych plugin (De Leeuw, 2015), we hope to make it possible for experimenters to form detailed analyses of planning processes without extensive programming or expensive equipment.

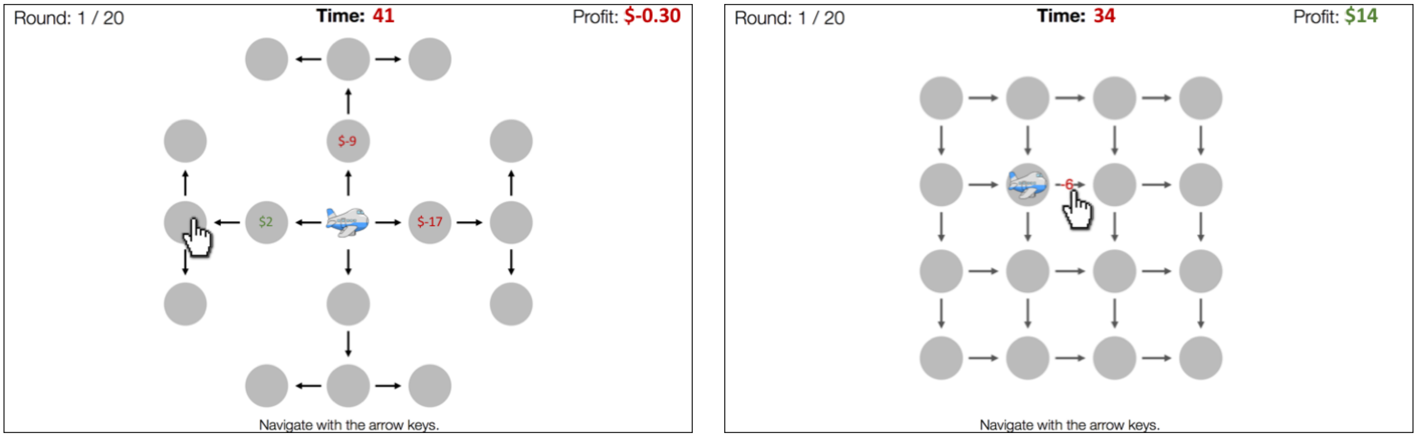
2 Using Mouselab-MDP

Having motivated the paradigm, we briefly describe both the interface through which experimenters specify experiments, and the interface through which participants engage in the task. Two screenshots of the paradigm are shown in Figure 1, and a live demo can be viewed at http://cocosci.dreamhosters.com/webexpt/mouselab-demo/. The code for Mouselab-MDP lives at https://github.com/fredcallaway/Mouselab-MDP.

On each trial an MDP is presented with an intuitive spatial interface: a directed graph with states as nodes and actions as edges. The participant navigates through the graph using the keyboard, attempting to collect the maximal total reward. States or edges are annotated with the reward for reaching the state or taking the action; but crucially, these labels may not be visible when the trial begins. The participant may need to click or hover the mouse over a state/edge to see the associated reward; the timecourse of these information-gathering actions provides fine-grained evidence of the planning process. Furthermore, there may be a price associated with gathering information; this creates a tradeoff between the cost and value of information, which is the focus of work on bounded optimality and resource rationality (Griffiths, Lieder, & Goodman, 2015; Shenhav et al., 2017).

With the Mouselab-MDP plugin you can create a planning experiment by specifying the following critical components:

1. `graph` is a mapping $s \mapsto A$ where A is itself a mapping $a \mapsto (r, s')$. This structure specifies the actions a available in each state, as well as the reward r and resultant state s' associated with each action.
2. `initial` is the state in which the participant begins the trial. Along with `graph`, this defines a deterministic MDP.



a) State values revealed with clicks

b) State values shown while hovering the mouse

Figure 1: Two example paradigms created with the Mouselab-MDP plugin for JsPsych: a) Each state is labeled with the reward for reaching that state; these rewards become visible after they are clicked, with a $0.10 fee per click. b) The reward for making a transition is revealed only while the mouse is hovering over the corresponding arrow.

3. `layout` is a mapping $s \mapsto (x, y)$ that specifies the location of each state on the screen.

Specifying only these settings will result in a graph with rewards shown on the edges between nodes and no labels on the states. This corresponds to a standard MDP with a known transition and reward function. To take advantage of the Mouselab features, the user must specify at least one of the following optional properties:

1. `stateLabels` is a mapping $s \mapsto \ell$ that specifies the labels to be shown on each state.
2. `stateDisplay` ∈ { 'never', 'hover', 'click', 'always' } specifies when state labels are displayed. When set to 'click', clicking on the state causes the label to appear and remains on the state for the duration of the trial. The optional parameter `stateClickCost` specifies the cost (a negative number) for clicking on a single state. When set to 'hover', the label appears only while the mouse is hovering over the associated edge. There is no cost for this option due to the likelihood of accidentally passing the mouse over an edge.
3. `edgeLabels` is analagous to `stateLabels`, except that it defaults to the rewards associated with each edge.
4. `edgeDisplay` is analagous to `stateDisplay`. `edgeClickCost` specifies the cost.

This novel paradigm offers three advantages over previous behavioral paradigms for the study of planning:

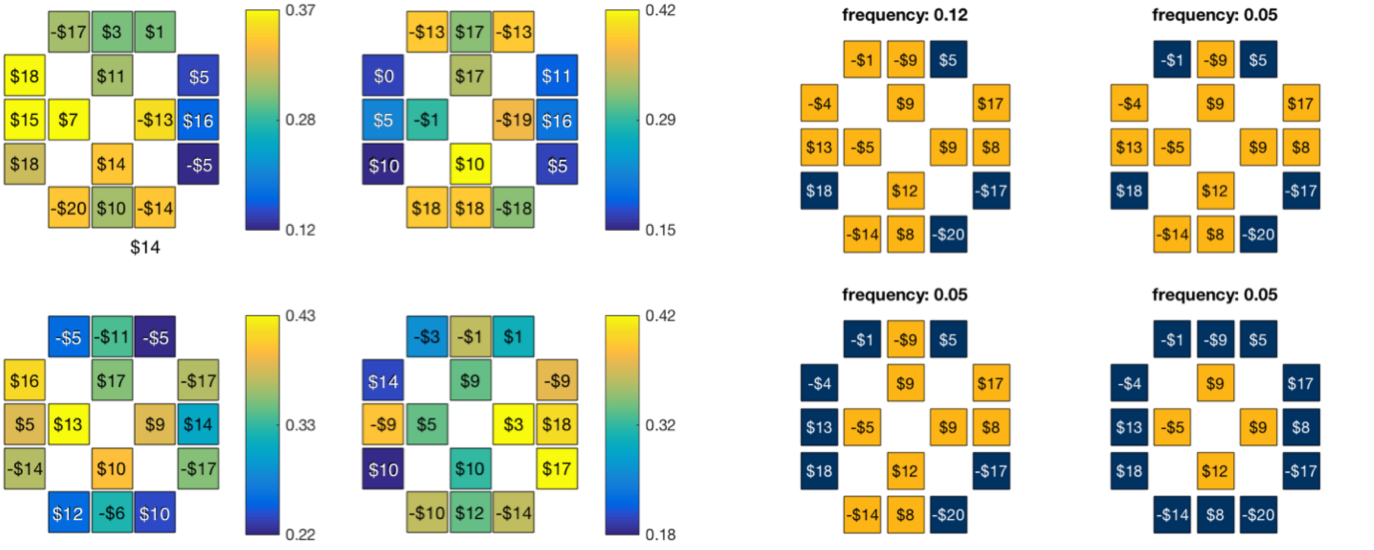
1. The latent cognitive process of planning is made observable by recording the participant's clicks. This allows the experimenter to test hypotheses that would otherwise be difficult to confirm or refute based on overt behavior.
2. A wide variety of state-transition and reward structures can be displayed automatically. This allows an experimenter to create a large number of highly variable stimuli (potentially automatically with a computer program) in relatively little time (compared to using an image-editing software).
3. The paradigm is packaged as a JsPsych plugin with a concise, yet flexible stimuli specification. This allows experimenters with only basic knowledge of Javascript to use our plugin to create a wide range of qualitatively novel experiments that can be run online with crowd-sourcing services such as Amazon Mechanical Turk.

3 Revealing Planning Strategies

Here, we present a proof-of-concept case study to illustrate the utility of the Mouselab-MDP paradigm. The data from this experiment sheds new light on the well-studied notion of pruning (Huys et al., 2012).

3.1 Methods

We used the Mouselab-MDP plugin for JsPsych. In each trial participants route an airplane from the center of the screen to one of eight final destinations via two intermediate locations (see Figure 1a). We specified `stateLabels` as the rewards associated with the edge leading to each state. We set `stateDisplay` to 'click' and 'stateClickCost' to 0.10; thus participants could click on a state to reveal the reward for traveling to that state, at the price of $0.10. Participants



a) Frequency of sampling by location, across four MDPs.

b) The four most common sets of clicks for one MDP.

Figure 2: Patterns of mouse clicking in humans. a) Each subplot shows the location and monetary value of each state for a different MDP. The colors indicate how often each state was inspected before the first move. b) Frequencies indicate the portion of participants who selected each set, for one particular MDP. Inspected states are gold. The monetary values indicate the reward for flying to that state.

were required to spend at least 45 seconds on every trial to prevent time cost from discouraging participants from clicking and planning.

We recruited 31 participants on Amazon Mechanical Turk. The task took 31 minutes on average and participants were paid $\$1.50$ plus a bonus equal to 5% of their earnings on one randomly selected trial (average bonus: $\$1.69$). Participants completed a total of 20 trials in randomized order. The trials differed only in rewards. All rewards were integers between $-\$20$ and $+\$18$ (mean: 4.2, SD: 10.3). The reward functions were designed so that 3-step planning with full information led to significantly higher returns than 2-step planning, 1-step planning, or random choice ($\$38.35$ vs. $\$16.06$–$\$18.30$).

3.2 Results

Figure 2a shows the frequency with which participants inspected each of the sixteen states of the task environment (Figure 1a) prior to selecting the first flight, in four of the 20 MDPs. Each MDP corresponds to a decision tree with four branches each of which comprises a stem and three outer nodes. Examining Figure 2a (and similar plots for all the MDPs), the frequency with which participants inspected the outer nodes of a branch appears to increase with the reward at the stem of the branch. Conversely, participants were less likely to inspect the outer nodes of a branch when its inner node was a large loss. This is indicative of pruning. To quantify this intuitive result, we define pruning as inspecting none of the subsequent nodes of a branch after inspecting the stem. Figure 3 shows that the probability with which people prune a branch of their decision tree increases gradually with the magnitude of the loss at its stem. A logistic regression analysis accounting for individual differences in the base rate of pruning confirmed that the increase in pruning with the decrease of the reward at the stem of the branch was statistically significant ($t(1276) = -2.01, p < 0.05$). To our knowledge, this is the first analysis to find that pruning changes gradually with the magnitude of the loss or gain.

Our results are consistent with those of Huys et al. (2012) who showed that a model with a "specific pruning parameter" (i.e. a larger discount γ applied to states after a large loss) fit participant choices better than a model without such a parameter. However, their data and model are also consistent with an alternative explanation: people may consider all branches of the tree but discount distant rewards while overweighting immediate losses (Green & Myerson, 2004). Our data, in contrast, provide direct evidence that participants did not consider states after costly transitions. This indicates that the continuous discounting parameter of Huys et al. (2012) may have captured the average effect of many discrete pruning decisions.

Figure 2b shows the four most common sets of clicks (before the first flight) for a one MDP. The most frequent strategy inspected exactly one complete path on each of the four branches, and the second most common strategy pruned one or more branches whose reward at the stem was lowest.

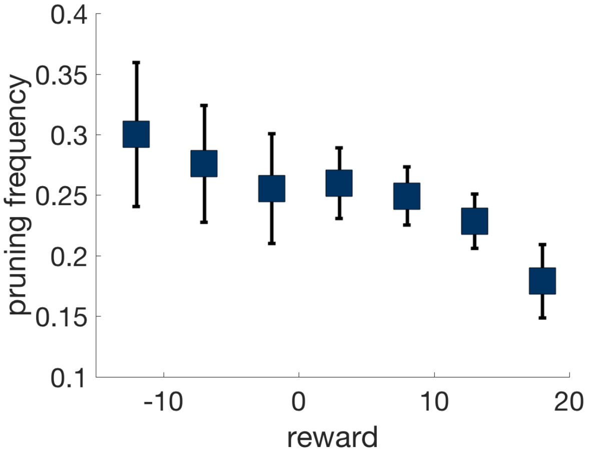


Figure 3: Frequency of pruning ±1SEM (vertical axis) given the reward at the stem of the branch (horizontal axis). Pruning was defined as inspecting none of a branch's outer states (before the second move) after having inspected the reward at its stem (before the first move). Each point shows the mean frequency of pruning.

4 Conclusion

We have developed a process-tracing paradigm for the study of planning and demonstrated that it can reveal the processes of planning at a higher level of resolution than standard behavioral paradigms. We showed that the paradigm provides a new look at one aspect of planning: pruning. However, many questions remain. Do people plan in a breadth-first or depth-first fashion? Do they plan and then act all at once, or do they interleave planning and acting? How do they balance the cost of planning with the potential for increased reward? Do these aspect of planning vary based on the problem structure? If so, how do people meta-reason about what planning strategies to employ? How do people learn and refine their planning strategies? How and why do people differ in their planning abilities? No single paradigm can give an answer to all these questions. However, Mouselab-MDP can produce data that speaks to all of them.

Future work will extend the Mouselab-MDP plugin to stochastic environments and enable tracing additional aspects of the planning process, including representation of a transition model and heuristic evaluation of state values.

References

Balaguer, J., Spiers, H., Hassabis, D., & Summerfield, C. (2016). Neural mechanisms of hierarchical planning in a virtual subway network. *Neuron*, *90*(4), 893–903.

Botvinick, M. M., Niv, Y., & Barto, A. C. (2009). Hierarchically organized behavior and its neural foundations: A reinforcement learning perspective. *Cognition*, *113*(3), 262–280.

De Leeuw, J. R. (2015). jspsych: A javascript library for creating behavioral experiments in a web browser. *Behavior Research Methods*, *47*(1), 1–12.

Green, L., & Myerson, J. (2004). A discounting framework for choice with delayed and probabilistic rewards. *Psychological bulletin*, *130*(5), 769.

Griffiths, T. L., Lieder, F., & Goodman, N. D. (2015). Rational use of cognitive resources: Levels of analysis between the computational and the algorithmic. *Topics in cognitive science*, *7*(2), 217–229.

Huys, Q. J., 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 Comput Biol*, *8*(3), e1002410.

Huys, Q. J., Lally, N., Faulkner, P., Eshel, N., Seifritz, E., Gershman, S. J., ... Roiser, J. P. (2015). Interplay of approximate planning strategies. *Proceedings of the National Academy of Sciences*, *112*(10), 3098–3103.

LaValle, S. M. (2006). *Planning algorithms*. Cambridge university press.

Miller, G. A., Galanter, E., & Pribram, K. H. (1986). *Plans and the structure of behavior.* Adams Bannister Cox.

Morris, R., & Ward, G. (2004). *The cognitive psychology of planning*. Psychology Press.

Newell, A., Simon, H. A., et al. (1972). *Human problem solving* (Vol. 104) (No. 9). Prentice-Hall Englewood Cliffs, NJ.

Ólafsdóttir, H. F., Barry, C., Saleem, A. B., Hassabis, D., & Spiers, H. J. (2015). Hippocampal place cells construct reward related sequences through unexplored space. *Elife*, *4*, e06063.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1988). Adaptive strategy selection in decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(3), 534.

Russell, S., Norvig, P., & Intelligence, A. (2010). *Artificial intelligence: A modern approach* (3rd ed., Vol. 25). Prentice-Hall, Englewood Cliffs.

Shenhav, A., Musslick, S., Lieder, F., Kool, W., Griffiths, T. L., Cohen, J. D., & Botvinick, M. M. (2017). Toward a rational and mechanistic account of mental effort.

Simon, D. A., & Daw, N. D. (2011). Neural correlates of forward planning in a spatial decision task in humans. *Journal of Neuroscience*, *31*(14), 5526–5539.

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction* (Vol. 1) (No. 1). Cambridge, MA: MIT press.