Cognitive Psychology 101 (2018) 29–49

Contents lists available at ScienceDirect

Cognitive Psychology

journal homepage: www.elsevier.com/locate/cogpsych

# The detour problem in a stochastic environment: Tolman revisited

Pegah Fakhari*, Arash Khodadadi, Jerome R. Busemeyer

*Indiana University, Department of Psychological and Brain Sciences, Bloomington, IN, United States*

A B S T R A C T

We designed a grid world task to study human planning and re-planning behavior in an unknown stochastic environment. In our grid world, participants were asked to travel from a random starting point to a random goal position while maximizing their reward. Because they were not familiar with the environment, they needed to learn its characteristics from experience to plan optimally. Later in the task, we randomly blocked the optimal path to investigate whether and how people adjust their original plans to find a detour. To this end, we developed and compared 12 different models. These models were different on how they learned and represented the environment and how they planned to catch the goal. The majority of our participants were able to plan optimally. We also showed that people were capable of revising their plans when an unexpected event occurred. The result from the model comparison showed that the model-based reinforcement learning approach provided the best account for the data and outperformed heuristics in explaining the behavioral data in the re-planning trials.

## 1. Introduction

Humans deal with planning problems in their everyday situations. One of the very familiar situations is to navigate from one place to another in a neighborhood or city. In this scenario, usually there is more than one path to choose, and, depending on the goal, one might select the shortest path, the city roads, a bypass/highway outside the traffic area or a path with the minimal traffic lights. Once he chooses the highway, he still needs to decide whether to take the toll line and/or where to exit. On the other hand, if he chooses the city roads, he would need to decide which intersection to go, to use the main street or to use shortcuts, etc. In other words, after selecting a general path (plan), there are still small paths decisions.

This is an example of a more general problem in which one needs to optimally plan a sequence of interdependent choices to accomplish a goal. In some cases, the shortest path is the optimal path and in others his goal might be to avoid the traffic at all costs.

In multistage decision making, unlike isolated choices, the focus is on how people analyze the interrelated choices to make an *optimal* sequence of choices, Hotaling, Fakhari, and Busemeyer (2015) and Gonzalez, Fakhari, and Busemeyer (2017). Usually, sequential decisions are represented in a decision tree in which the result of an action at one stage (e.g. a decision node) will be fed into the next stage, which might be another decision node or possibly an output node. Consider a decision tree with two decision nodes (yellow circles) and five possible (green) paths that is represented in the upper panel of Fig. 1. Given the starting position, **S** and the goal, **G**, the paths are Path B, Path A1A2, Path A1C2, Path C1C2 and Path C1A2. The black dashed line in the grid separates path B and path A1A2.

In order to make optimal choices, one should know the actual output of each decision node and the uncertainty of each transition. For instance, although the number of steps (or actions) between the starting position and the next decision node (or the goal) is not depicted in Fig. 1, based on the expected losses (*El*), we know that path B is the best path to go to the goal position. In an experience-

* Corresponding author at: Indiana University, Department of Psychological and Brain Sciences, 1101 E. 10th Street, 47405-7007 Bloomington, IN, United States.
*E-mail address:* pegah.fakhari@gmail.com (P. Fakhari).

https://doi.org/10.1016/j.cogpsych.2017.12.002
Accepted 23 December 2017

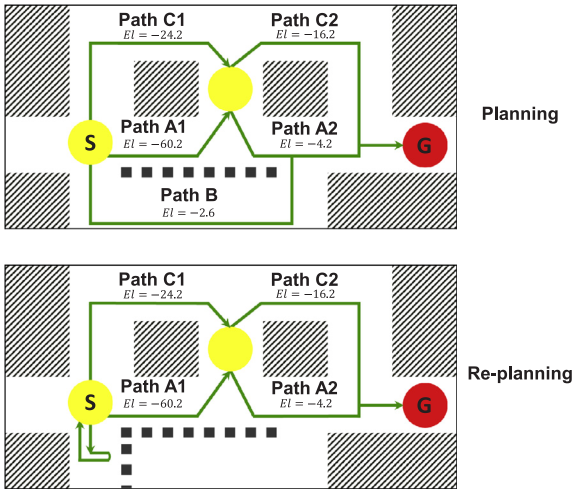


**Fig. 1.** General grid world used in our experiments. The decision nodes are represented by yellow circles and the goal is depicted by red circle. The green paths are available to participants. Note that participants cannot see the obstacles depicted by shadow areas. The black dashed line in the grid separates path B and path A1A2. Top: In the planning trials, for this current start and goal positions, 5 paths are available to participants and they need to find the optimal path. Down: In the re-planning trials, the optimal path, path **B** is blocked and participants need to find the detour. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

based decision tree, this knowledge is established from (an individual's) experience, Daw, Gershman, Seymour, Dayan, and Dolan (2011), Momennejad et al. (2016), Huys et al. (2012), Dezfouli and Balleine (2012) and Keramati, Smittenaar, Dolan, and Dayan (2016) while in a description-based version, it is provided by the experimenter (available during the task), Hotaling and Busemeyer (2012), Hey and Knoll (2011), Drner and Schaub (1994), Johnson and Busemeyer (2001) and Johnson and Busemeyer (2005). However, having this knowledge cannot guarantee the optimal behavior, Hey (2005), Sims, Neth, Jacobs, and Gray (2013), Yechiam, Erev, Yehene, and Gopher (2003), Hotaling and Busemeyer (2012), Huys et al. (2015), Momennejad et al. (2016), Huys et al. (2015), Botvinick, Niv, and Barto (2009) and Keramati, Dezfouli, and Piray (2011).

### *1.1. Examining the optimality of re-planning in sequential decision making tasks*

In many of the previous studies on planning in decision trees, the environment is not dynamically changing, Nassar, Wilson, Heasly, and Gold (2010). Once the participant learns about the risky sequential decision environment, she can *optimally* plan her actions and does not need to update her knowledge later. But in real life, our environment is always changing and unexpected events happen. Usually, we have two strategies to deal with these situations: reevaluate our plans with the new information (which is also known as re-planning) or ignore the new information and stick to our original plan.

In this article, we extend previous planning experimental designs to situations in which the participants experience random changes in the environment and need to modify their original plans to get to the goal position. We test learning to plan and re-planning in one unique framework: a 4 by 7 grid world with stochastic losses. In the learning phase of our experiments, we look into the planning behavior and how participants can learn to find the optimal sequence of choices. Then, in the test phase, we block the optimal path randomly and ask our participants to find a detour path (re-planning behavior) based on what they have learned during the learning phase as illustrated in the bottom panel of Fig. 1. On 30% of the trials, path B (the *optimal* path) is randomly blocked and becomes unavailable [1]. The vertical black dashed line that is added to the grid shows the wall that makes path B unavailable. In this situation, path C1A2 is the best path to choose (optimal re-planning behavior).

It is important to emphasize that the starting and goal positions are not fixed in our design. We randomize these pairs for three reasons: (1) to make sure that our participants have a fair exposure to different aspects of the grid world environment; (2) a random starting point and goal makes it more challenging to discriminate different models and their predictions; (3) to examine human planning behavior from different arbitrary decision nodes located in different layers of a decision tree.

In Section 4.1, we discuss how we fit different models and compare the results in detail. However, it is important to highlight that our design includes a generalization test that fits model parameters to the planning phase (with no blocks in the optimal path), and then subsequently uses these same parameters to predict re-planning in a generalization test when blocks are introduced, Busemeyer and Wang (2000). This provides a very strong test of the competing models that vary in number of parameter and model complexity. In other words, our model comparison is not restricted to how different models can learn the model of the environment and whether or not they can predict planning (which they have been trained for), but involves a more rigorous test on how they can perform in an unexperienced environment.

[1] Participants do not experience the blockage unless they check the optimal path.