A treasure hunt through sequential-valued based decision making

Ursula M. Goldstein

Freie Universität in Berlin

Abstract

In this report we focused on the methods behind the SCHATZ-experiment, the statistics used to analyze the behavioral data and the general conclusion based on the evaluated behavioral data.

This study focused on the concept of sequential-valued based decision making by using the model of POMDP. The POMDP model formalizes the interaction between agents, in this study participants, and the environment, the 5x5 treasure hunt map. The agent choose and action *a* (arrow movement), causing a state transition that generates an observation *z* (light or grey bar)and a reward *r* (two treasure chests). Behavioral data was received and analyzed through matlab. Furthermore the data demonstrated that the tasks was not too difficult nor too easy and that some of the participant did not solve more than 60% of the solvable tasks. Some participants did not even response on some trials and the median reaction time across all decisions was longer compared to the median reaction time for all decision that followed after a single treasure was discovered. Lastly all of the participant solved one of the tasks at least one time on the first attempt.

*Keywords:* POMDP, goal-directed, decision making, uncertain information

A treasure hunt through sequential-valued based decision making

In everyday life we make decision based on how certain we are about the outcome. This is one of the components of sequential-valued based decision making, a subfield of cognitive neuroscience. In this field the fundamental questions are concerned with how humans and other cognitive agents are able to devise multi-step strategies in uncertain environments that are suited to achieve specific goals and how these strategies are conceived, supported, and constrained by the neural architecture. In sequential-valued based decision making, also referred to as "goal-directed decision making", the process of making a decision is based directly on predictions concerning action outcomes and their attendant incentive values. In other words, it requires accesses to stored information about action-outcome contingencies and action-outcome knowledge must be integrated with incentive knowledge (Solway and Botvinick, 2012). Furthermore, besides the definition of the concept, when we combine conflicting sensory information or cues, more uncertain information deserves less weight in such decision making (Bach and Dolan, 2012). Therefore if one presents a stimuli with probabilistic information the decision making and the reaction time might be influenced. Lastly, concerning the neural substrates, model-based mechanism such as goal-directed decision making are associated with activation of the medial prefrontal cortex and the dorsomedial striatum (Dayan and Daw, 2008).

In the following study participants performed runs of an interactive visual treasure hunt task. The task required the participants to take actions, causing a state transition that generated a cue or observation. This interchange repeats after the agent chooses another action until they reached the second reward (treasure chest). These are parts of the process of a model called partially observable Markov decision process (POMDP). The POMDP model formalizes the interaction between agents, in this study participants, and the environment, the 5x5 treasure hunt map. Generally, the agent chooses action *a*, causing a state transition that generates an observation *z* and a reward *r*, these finite set have to be specified and their function detailing their relationships. Also the agent does not simply know its state, but they must take actions to gain state-related information, this is a hallmark of the POMDP model (Littman, 2009). In this study the action for the participants was movement in a specific direction (up,down,left,right) and the observation was a cue of their distance to the reward, represented by visual cues such as light or dark grey bars. The reward included two treasure quests and was only submitted as received if both treasure chest were found under specific condition that are further explained in the methods section.

In this report we focused on the methods behind the experiment, the statistics used to analyze the behavioral data and the general conclusion based on the evaluated behavioral data. The method section will include a further explanation of the actual experimental paradigm and how the behavioral data was received and analyzed. The results will evaluate and discuss the experimental and behavioral indices. Lastly the discussion will give a summary and general conclusion about the participant's behavior on the treasure hunt task based on the evaluated behavioral data.

**Method**

**Participants**

The experimental data set was acquired in a sequential value-based decision making paradigm hereinafter referred to as “treasure hunt”. Behavioral and FMRI data was recorded from a group of 20 participants (11 female, 19 right-handed, 1 male participant both-handed with left hand preference, mean age 26.95 years, standard deviation age 2.93 years) with no reports of neurological or psychiatric disorders after providing written informed consent. The study was in line with the human subject guidelines of the Declaration of Helsinki and was approved by the ethics committee of the German Psychological Society (Deutsche Gesellschaft für Psychologie).

**Materials and Procedure**

Participants performed four consecutive runs of an interactive visual treasure hunt task in the MR scanner after having been familiarized with the task outside the MR environment. In brief, the participants’ task was to find two “treasures” in a 5 x 5 cell grid-world (Figure 1) based on a limited number of available steps. On each trial of the task (Figure 2), participants were visually presented with the grid-cell that corresponded to their current position in the grid-world. Upon being informed about their position, participants were presented with (uncertain) information about the potential location of the two treasures.

Notably, the bars could be interpreted as “noisy sensors” for the location of the treasures. Importantly, the participants were informed that they would always observe a dark grey bar truly indicating the wrong direction and *probabilistic information* for the direction leading towards the treasures (that is, they could either observe a dark grey bar (falsely marking the right direction) or a light grey bar (truly marking the right direction)). The probability of the bars for the directions leading towards the treasures to assume the light grey state was modeled by a decreasing function of distance to the treasures. However, the participants were not informed about the functional relationship between the distance to the treasures and the probability of obtaining dark or light grey bars. The bars were presented for a duration uniformly sampled from an interval of 3 to 5 seconds. Finally, on each trial of the task, participants were visually presented with a set of arrows that indicated the possible directions of the next step and prompted the participant’s decision on the current trial. Notably, diagonal steps and steps off the 5 x 5 grid-world were not possible and thus no arrows presented for these directions. Participants indicated their preferred direction by means of pressing one of four buttons with the index to little finger of their right-hand. If participants did not respond within a given response time window, they maintained their current position on the next trial. Additionally, while being presented with the arrows participants were also visually informed about the number of remaining steps to find the two treasures and how many treasures they had already discovered. The arrows and step number was presented until the participants’ indicated their decision by the button press, but no longer than for a duration uniformly sampled from an interval of 3 to 5 seconds, upon which the response was defined as missed and the participant’s location remained unchanged.

The entire experiment comprised a collection of experimental “tasks” that corresponded to a specific location setting of the two treasures. For each task, participants had a maximum of three “attempts” or "blocks” to solve the task. A task was defined to be solved if and only if both treasures were discovered during a single attempt. If in a previous attempt participants discovered no or only a single treasure before the number of available steps was exhausted, they were “reset” to the left upper grid cell, and if they found a treasure in the former block, to solve the task they had to revisit the location of the previously discovered treasure (remained unchanged) and in addition, discover the previously undiscovered treasure. The available number of steps was randomly assigned for each attempt and sometimes sufficed and sometimes did not suffice to discover the treasures in a single attempt. Specifically, the number of optimal steps for a given treasure location was determined using Dijkstra’s algorithm (Dijkstra, 1959) and served as the expectation of a discrete distribution over ±2 step sizes (probabilities f 𝑝𝑝(±0 𝑠𝑠𝑠𝑠𝑠𝑠𝑠𝑠𝑠𝑠)=0.4,𝑝𝑝(±1 𝑠𝑠𝑠𝑠𝑒𝑒𝑒𝑒)=0.2,𝑝𝑝(±2 𝑠𝑠𝑠𝑠𝑠𝑠𝑠𝑠𝑠𝑠)= 0.1). Given the 5 x 5 grid-world layout the maximal number of optimal steps was 12, which resulted in a maximum number of 14 trials per attempt, in which each attempt could last for a maximum of 210 seconds.

Each attempt ended with the presentation of the grid-cell location corresponding to the chosen action from the previous trial. If participants, on a given attempt, solved a task, they were subsequently informed that both targets were found during a single attempt and a new task was being created. If participants did not solve the task on a given attempt because the available number of steps was exhausted, but one or two more attempts for the current task remained available, participants were informed that their current step limit was reached, and their position was being reset. Finally, if participant did not solve the task upon completion of third attempt, they were informed that both their step and attempt limit were reached (presented for a duration uniformly sampled from an interval of 3 to 5 seconds ) and a new task was being created (a novel location for the two treasures). For each experimental FMRI run, participants were presented with four tasks and thus a minimal number of four attempts, if all tasks were solved in the first attempt, and a maximal number of 12 attempts, if all task were only solved on the last attempt or not solved at all. Given the presentation durations and attempt scheme, the maximal task duration was 630 seconds. Finally, the maximal duration for an FMRI run was 2440 seconds. Additionally, to allow for a “baseline” MR signal measurement after every fourth attempt only the stimulus fixation cross was presented for a duration uniformly sampled from an interval of 6 to 8 seconds. All stimulus presentation durations were chosen for optimized GLM contrast sensitivity.

**Results**

A Matlab .m-file function that evaluates basic aspects of the participants’ behavior in the experimental paradigm was created in the program Matlab. Presentation script in a cell array variables are called “runlog” and were received in .mat files SCHATZ0XX\_Run\_Y.mat, where the XX denoted the participant number (03, 04,…, 22) and the Y denoted the run number (1,2,3,4). For each participant, there exists a folder called “SCHATZ0XX” in the project folder “Data” which comprises up to four of these .mat files containing the runlog variables. There were a total of 20 Participants labeled 1 through 20 in Figures 3 to 12. Figure 3 depicts the average number of solvable tasks, tasks for which the number of given steps allowed for the discovery of both tasks. The mean for solvable tasks was 96.15 with a standard deviation (std) of 4.34 tasks, therefore most of the tasks were solvable. The next Figures, Figure 4a and b, visualize the average task and run duration in number of decisions. The average task duration in number of decision overall participants was 13 decisions with a std of 1.74 decisions, compared to run duration with 52 decisions and a std of 7.07 decisions. To follow are Figures 5a and b which depict the average task and run duration in seconds. A task lasted on average 182 seconds with a std of 23.05 seconds, compared to a run with an average duration of 727 seconds with a std of 94.21 seconds. Figure 6 evaluates the average proportion of solved tasks across all solvable tasks. Out of all the solvable tasks, 75% tasks were solved with a std of 17.19 %. Out of all the solvable tasks, Figure 7 a, b and c show the tasks that were solved with optimal steps, optimal steps plus 1 step and optimal steps plus 2 steps. On average 68.24 % percent were solved with optimal steps (std = 15.63), 4.83 % with optimal steps plus 1 step (std = 5.47) and 2.2 % were solved with optimal steps plus 2 (std = 3.80). For reaction time, Figure 8 shows the median reaction time in milliseconds (ms) across all decisions. The median reaction time on average across all participants was 717.5 ms with a std of 165.55 ms. Furthermore Figure 9 depicts the median reaction on attempts that followed and attempt on which a single treasure was discovered. The reaction time on average there was 698.30 ms with a std of 162.18 ms. Figure 10 illustrates the mean reaction time across all decisions, which was on average 846.33 ms with a std of

183.06 ms. The average number of solved tasks on the first attempt is depicted in Figure 11. On average 20% of the tasks were solved on the first attempts (std = 3.27). Finally, Figure 12 shows

the average number of how often a participant did not respond. Overall participants on average did not respond 2.1 times (std = 3.23).

**Discussion**

As seen in the results most of the tasks were solvable, therefore each participant had multiple chances to solve a task, and most of the time participants solved at least 50% of the solvable tasks. Rarely anybody solved the tasks on the first attempt, this shows that the tasks were also not too easy. Participant 8 is the only one that did not solve more than 50 % that were solvable and also did not respond all the time. If one was to analyze more data, participant 8 maybe not be included since 50 % is at a chance rate. Interestingly most of the tasks were solved with optimal steps rather than optimal steps plus 1 or 2 steps. One would think the uncertain information would lead to more steps than optimal. Median reaction time is shorter on trials where a single treasure was discovered. The uncertain information is more certain if the participant repeats the same maze with the treasure in the same location, therefore the median reaction should be quicker. Mean reaction time compared to median reaction time across all the decision, is larger and has a larger standard deviation. Median reaction time represents the reaction time for the task better, since mean reaction time might not be robust enough against higher values (Whelan, 2008). Figure 11 demonstrates that all of the participants at least one time solved the tasks on the first attempt. Again this shows that the task was not too difficult and some learning might have occurred over the blocks. Lastly Figure 12 depicts if the participants did not respond. Interestingly participant 16 did not respond multiple times but this solved 80% of the solvable tasks.

In conclusion the data depicted needs further analyzing for significance but overall it shows that there is some behavioral difference between participants and that some learning might have occurred over blocks or runs.

References

Bach, D.R., & Dolan, R.J. (2012). Knowing how much you don't know: a neural organization of uncertainty estimates. *Nature Reviews, 13,* 572-586. doi:10.1038/nrn3289

Dayan, P., & Daw, N.D. (2008). Decision theory, reinforcement learning, and the brain.  *Cognitive, Affective, & Behavioral Neuroscience, 8,* 429-453.

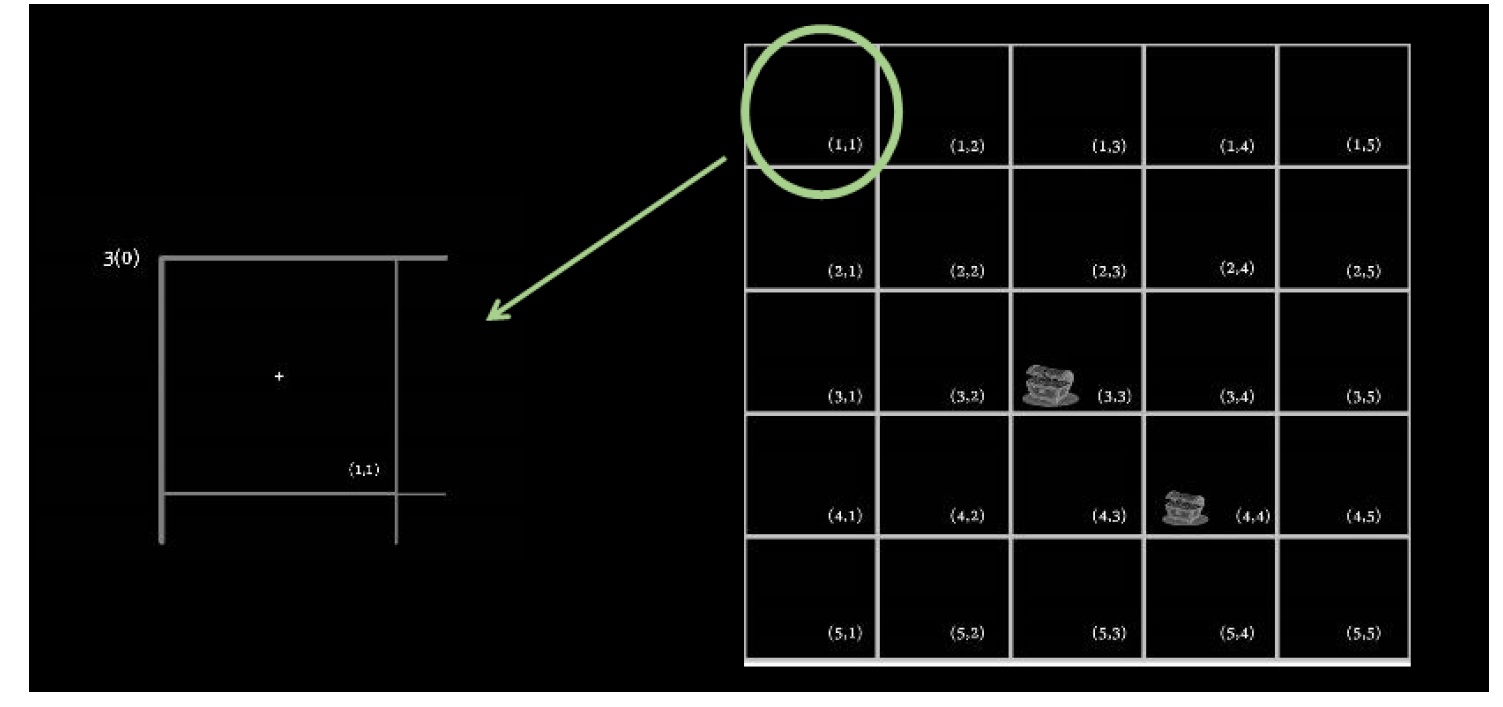
doi:10.3758/CABN.8.4.429

Dijkstra, E. W. (1959). A note on two problems in connection with graphs. *Numerische Mathematik*, *1*, 269-271. doi:10.1007/BF01386390

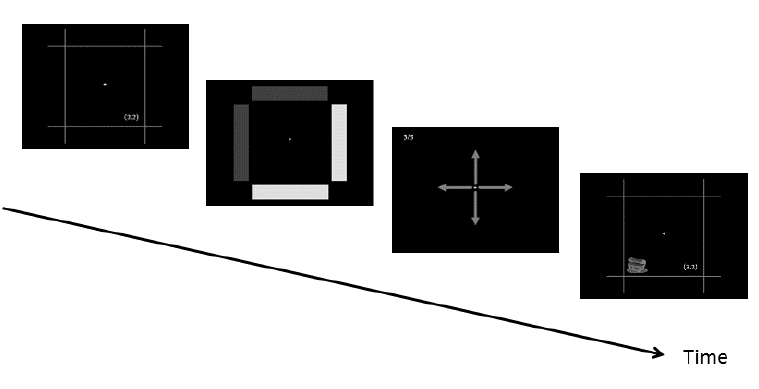
Littman, M.L. (2009). A tutorial on partially observed Markov decision processes. *Journal of Mathematical Psychology, 53,* 119-125. doi:10.1016/j.jmp.2009.01.005

Solway, A., & Botvinick, M.M. (2012). Goal-directed decision making as probabilistic inference: a computational framework and potential neural correlates. *American Psychological Association, 119,* 120-154. doi: 10.1037/a0026435

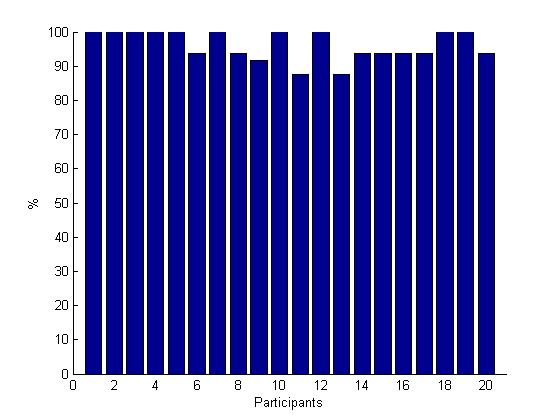
Whelan, Robert (2008) "Effective analysis of reaction time data," The Psychological Record, *58*, 475-482.



*Figure 1.* 5 x 5 cell grid-world with two “treasures” at cells (3,3) and (4,4), respectively. Participants viewed the grid-world from the perspective shown on the left, i.e., only saw a single cell from above at a given.



*Figure 2.* An example of an experimental trial.



*Figure 3.* Percentage of tasks that were solvable for each participant



*Figure 4a.* The average task duration in number of decisions for each participant



*Figure 4b.* The average run duration in number of decisions for each participant



*Figure 5a.* The average task duration in seconds for each participant



*Figure 5b.* The average task duration in seconds for each participant



*Figure 6.* The average proportion of solved task for each participant



*Figure 7a.* The average proportion of solved task with optimal steps for each participant



*Figure 7b.* The average proportion of solved task with optimal steps plus 1 step for each participant



*Figure 7c.* The average proportion of solved task with optimal steps plus 2 steps for each participant



*Figure 8.* The median reaction time across all decision for each participant



*Figure 9.* The median reaction time on attempts that followed an attempt on which a single treasure was discovered for each participant



*Figure 10.* The mean reaction time across all decision for each participant



*Figure 11.* The average proportion a task was solved on the first attempt



*Figure 12.* The number of times a participant did not respond