The neurocomputational mechanisms of human sequential decision making under uncertainty in a spatial search task

Dirk Ostwald[1,2] (dirk.ostwald@fu-berlin.de), Lilla Horvath[1]
[1]Freie Universität Berlin, Habelschwerdter Allee 45, 14195 Berlin, Germany
[2]Max Planck Institute for Human Development, Lentzeallee 94, 14195 Berlin, Germany

Abstract

Sequential decision making under uncertainty is a ubiquitous aspect of human behaviour. For example, when trying to locate the source of radioactive decontamination, one will typically rely on a gamma counter, which signals the direction of the highest radiation with some associated observation noise. Similarly, humans often have to navigate abstract representational spaces, such as career paths, financial investment schemes, or maintaining a healthy lifestyle. Across all these domains, in order to achieve a set goal, humans can only rely on cues that are inflicted with uncertainty and which provide only partial information about how to reach the goal in an efficient way. In the current project, we used a simulated spatial search task to study the cognitive-behavioural and neural underpinnings of such sequential decisions under uncertainty. By combining artificial agent models rooted in the theory of partially-observable Markov decision processes with behavioural experimentation and functional magnetic resonance imaging, we provide evidence for human decision strategies that share similarities with real-time dynamic programming and rely on the neural representation of task-specific belief states.

Keywords: Sequential decisions; uncertainty; behaviour; fMRI

Overview and outline

For the current project we acquired behavioural and fMRI data from a group of 20 human participants (11 female, effective sample size n = 19) performing a sequential decision making task. The behavioural and fMRI data were then analysed using a model-based approach. In the following, we first discuss the experimental task ("Behavioural methods and results: Spatial search task") and then detail our behavioural modelling approach ("Behavioural methods and results: Agent models and behavioural data analysis"). After presenting the thus obtained behavioural results, we address the project's neuroimaging component ("fMRI methods and results").

Behavioural methods and results

Spatial search task. The participants' task was to uncover two treasures in a 5-by-5 cell grid-world (Figure 1A). On each task attempt, participants were initially positioned in the upper-left grid cell and had a limited number of steps at their disposal. If the participant failed to visit both treasure locations within the available step limit, the task treasure configuration was considered unresolved and the participant relocated to the start position on the next task attempt. Participants had

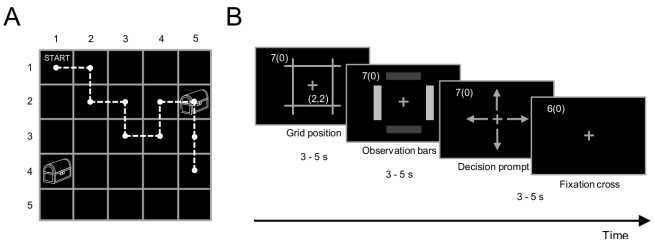

Figure 1: Spatial search task. (A). Bird's eye perspective of the grid-world. Participants were aware of the overall grid-world layout, but not of the locations of the treasures, and on each task trial were presented only with the information available to them at their current location. The dashed white line depicts an exemplary path on which the participant discovers one treasure before running out of available steps two steps later. (B). Trial layout. On each trial of the task, participants were first presented with the grid cell they currently occupied, including its row and column indices. Next, probabilistic sensor readings on the location of the treasures were presented in the form of light and dark observation bars. Finally, participants were prompted to choose one of the location-dependent possible directions. Upon their choice, the next trial commenced with the presentation of the resulting grid cell.

a maximum of three task attempts available for a given task configuration (treasure locations) and completed 16 randomly assigned task configurations in total. Crucially, participants were sequentially presented with the information available to them from a first-person grid cell perspective (Figure 1B): on each trial of the task, participants were first presented with the grid cell they currently occupied, including its row and column indices. Next, participants were presented with a collection of light and dark observation bars towards the adjoining grid cells. These observation bars conveyed uncertain information about the treasure locations and could be interpreted as noisy signals of a "treasure-sensor". Specifically, the sensor always returned a dark bar for directions leading away from the treasures, while it displayed either a light or a dark bar in the direction of the treasures. The sensor's accuracy of correctly returning a light bar towards the treasure locations depended on the participants' current L1-distance from the treasures: its readout was completely unreliable at the most distant grid cell position and parametrically increased in accuracy as the participant moved towards the treasures. Following the presentation of the observation bars, participants were asked to decide to move into one of the available directions, i.e., to any

of the neighbouring grid cells. Diagonal steps or steps off the grid were not allowed. Upon the participants' decision, a post-decision fixation cross was presented for a few seconds, after which the next trial commenced with the presentation of the resulting grid cell position. Participants were informed about the number of remaining steps and the number of treasures visited throughout a task attempt.

Agent models and behavioural data analysis. To formally describe a variety of sequential decision-making strategies on our spatial search task, and as a basis for behavioural and fMRI data analyses, we designed a set of nine agent models. This agent model set varies along two dimensions (Figure 2). First, the agents differ in their internal representation of the environment (belief state-free vs. belief state-based): the belief state-free agents do not encode a probabilistic representation of the latent treasure locations and for their decisions only rely on the information immediately available to them. In contrast, the belief state-based agents entertain a belief state in the form of a probability distribution over the latent treasure locations, which is dynamically updated on every trial in a normative Bayesian fashion. They subsequently use this belief state when selecting an action. Second, the agents of our model set differ with respect to their optimization goal (purely exploitative vs. explorative and exploitative): the purely exploitative agents use their current knowledge about the environment merely to collect the treasures, whereas the explorative and exploitative agents base their decisions also on the the goal of reducing their uncertainty about the latent treasure locations.

In more detail, the agent models implement the following sequential decision-making strategies: the belief state-free agent A1 chooses its actions (i.e., step directions) uniformly at random, while the belief state-free agent A2 relies on the observation bars and always chooses a step direction marked with a light bar. Agents A3a to A7 encode belief states about the location of the latent treasures. Based on its trial-by-trial belief state representation, agent A3a identifies the grid cell it believes to contain a treasure with the highest probability and moves towards this cell in a minimum L1-distance sense. Formally, this agent implements a heuristic real-time dynamic programming approach originally proposed by (Korf, 1990) and later elaborated on by (Geffner and Bonet, 1998). Agent A3b employs the same strategy as A3a with the difference that, until there is only one treasure left, the agent identifies the most probable location of both treasures and moves towards the cell which is closer to its position. Agent A4a does not identify single grid cells possibly containing a treasure. Instead, agent A4a takes into account its belief state distribution over all grid cells and, by using the same heuristic real-time dynamic programming approach as agents A3a and A3b, moves towards that part of the grid where it expects to find a treasure with the fewest steps. Agent A4b employs the A4a strategy until it encounters an ambiguous decision situation, i.e., a trial on which there appears to be more than one best action. In such trials, agent A4b switches to the A3b strategy and re-evaluates the

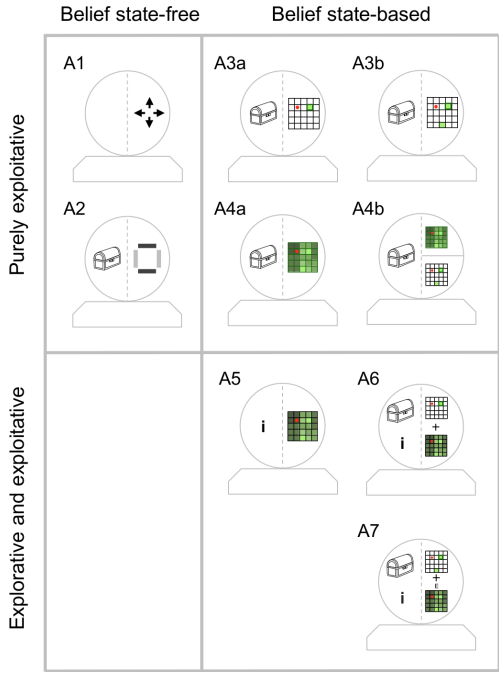

Figure 2: Agent model space.

available actions accordingly. Agents A5, A6 and A7 use their belief state representation not only to reach the treasures as fast as possible, but also to identify choice options that minimize their uncertainty about the treasure locations. Specifically, agent A5 chooses directions that promise the largest information gain in the sense of a maximized expected Bayesian surprise (i.e., the expected KL-divergence between the current and hypothetical belief states). Finally, the hybrid agents A6 and A7 combine both explorative and exploitative strategies: agent A6 combines the strategies of agents A3a and A5, whereas agent A7 combines the strategies of A3b and A5.

To assess the agent models' behaviour on the spatial search task, we ran a series of simulations using the same task configurations as for the human participants (e.g., treasure locations, number of available steps). To evaluate the agent models in light of the experimentally acquired human behavioural data, we further used a combination of maximum-likelihood model estimation and Bayesian-information criterion (BIC)-based model evaluation. Specifically, for each agent model and participant we first maximized the log probability of the participants' choices, then computed the participant- and model-specific BIC scores, and finally evaluated the model-specific group BIC scores using a random-effects Bayesian model selection procedure (Rigoux et al., 2014).

Behavioral results for participants and agents. We first evaluated the performance of the human participants. On average, the participants solved 11.26 (SEM $\pm$ 0.67) of 14.95 solvable tasks (Figure 3A, upper panel). Due to the randomly

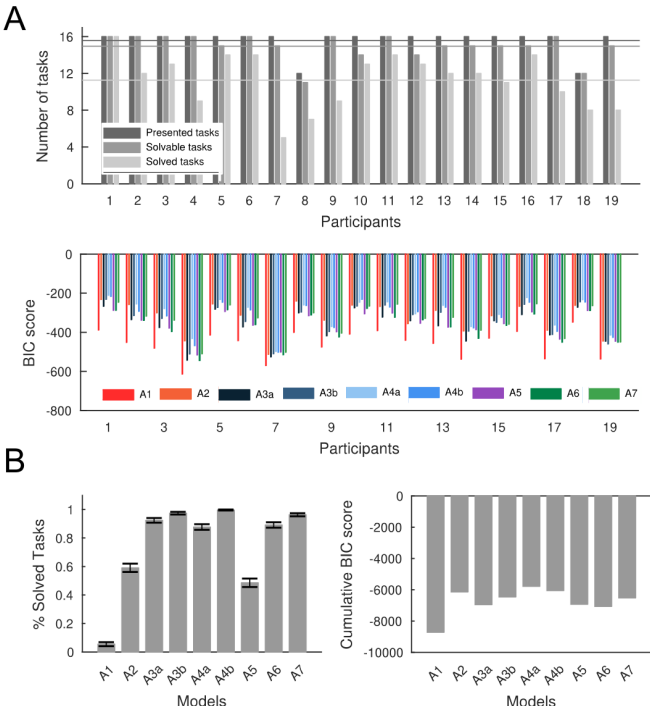

Figure 3: Behavioural results. (A). Participant performance evaluated by the number of solved tasks (upper panel) and model log evidence for each participant and agent model (lower panel). (B). Simulated task performance of the agent models (left panel) and cumulative BIC scores over participants (right panel).

allocated available steps a minority of the presented tasks was not solvable. The participants' performance was stable throughout the experiment with no differences between runs ($F(3,70) = 0.72$, $p = 0.55$). Moreover, we found that 59.12 % (SEM $\pm$ 3.54) of the tasks were solved within one or two attempts and significantly less within three attempts (16.09% SEM $\pm$ 2.19%). This indicates that participants were able to solve most tasks using only two attempts and thus validates our choice of limiting the available attempts to three for a task. Next, to compare the performance of human participants to that of the agents, we performed a series of agent simulations (Figure 3B, left panel). Here, the performance of the belief state-free random choice agent model A1 was the lowest followed by a substantially higher performance of the belief state-based directed information-seeking A5 and the belief state-free reward-driven A2 agent models. Agents A3a, A4a, A6 and A7 solved approximately 11 of the tasks with participant configurations, which corresponded to the average participant performance. The best performing agents were agents A3b and A4b, with A4b solving almost all tasks. Finally, we evaluated the agent models in light of the participants' choice data to identify the agent model that best describes the participants' behaviour. As visualized in the lower panel of Figure 3A, for 15 of the 19 participants, the BIC score was maximal under agent model A4a. This resulted in the highest cumulative BIC score for this model (Figure 3D). Moreover, the protected model exceedance probability of the group-level random-effects Bayesian model analyses for model A4a was $p \geq 0.99$. This strongly supports the conclusion that agent A4a was the most frequently applied strategy within the group of participants. Based on the pseudo-r^2 statistic (McFadden, 1974) we found that on average, A4a explained 35.65 % (SEM $\pm$ 2.06%) of the participants choice variance. In summary, our behavioural modelling initiative indicates that for making sequential decisions on the current spatial search paradigm, humans use a probabilistic representation of the task environment and primarily make decisions with the aim of uncovering the treasures, rather than exploring their environment.

fMRI methods and results

fMRI data acquisition and preprocessing Simultaneously with the behavioural data, fMRI data was collected on a 3T Siemens Magnetom TrioTim syngo scanner (Siemens, Erlangen) with a 12-channel head coil. 36 interleaved axial slices (flip angle: 80, slice thickness: 3 mm, voxel size: 3 $\times$ 3 $\times$ 3 mm, distance factor: 20 %) of echo-planar T2*-weighted images (field of view 216 mm) were acquired with a TR of 2 seconds. fMRI data were analysed using SPM12. fMRI data preprocessing included motion-correction, spatial normalization to the MNI-EPI reference template, resampling to 2 mm isotropic voxel size, and spatial smoothing using an 8 mm FWHM isotropic Gaussian kernel.

Model-based fMRI GLM analysis Given the results of the behavioural modelling initiative, we were primarily interested in identifying the neural correlates of the putatively encoded dynamic belief state representation and the ensuing action selection process. To this end, the preprocessed fMRI data were analysed using a model-based mass-univariate general linear model (GLM) approach with model-based regressors based on the group-favoured agent model A4a (Cohen et al., 2017). Participant-level GLM design matrices comprised five regressors of interest: the first regressor modelled trial events in a boxcar fashion with onsets corresponding to the time of the grid position presentation and with participant response time-dependent durations. The second regressor constituted a parametric modulation of the first regressor by the trial-by-trial Bayesian surprise of the agent model A4a's belief state representation. The third regressor constituted a parametric modulation of the first regressor by the trial-by-trial entropy of the action choice distribution of agent A4a, which we used as a proxy for the participants' action selection process. The fourth regressor modelled additional task information provided to the participants during the recording session and the fifth regressor modelled task breaks after every fourth task. All regressors of interest were convolved with a canonical hemodynamic response function. Additionally, all participant-level design matrix included a constant run offset and six spatial realignment parameters as nuisance regressors of no interest. Participant-level voxel-wise GLMs were estimated using

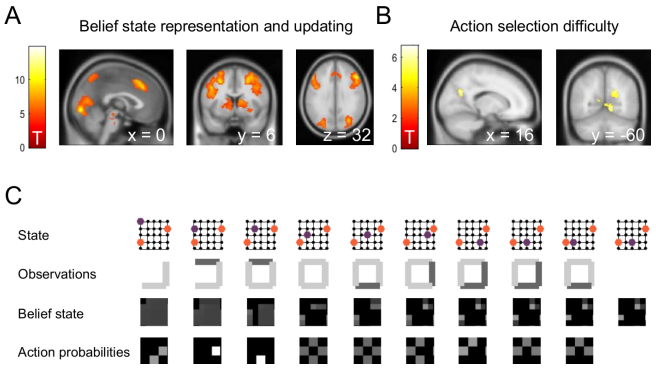

Figure 4: fMRI results. (A). The model-based fMRI data analyses suggest that a large network cortical-subcortical newtork including parts of occipital, parietal cortex, and frontal, as insula and striatum supports belief state representation and update. (B). Action selection difficulty appears to be encoded in posterior cingulate and cuneal cortices. (C). The model-based GLM results of (A) and (B) are based on the dynamic behaviour of the group-favoured agent model A4a, which is shown here for a single task attempt. The first row visualizes the problem state comprising the evoling location of the agent (blue dot) and the two treasures (orange dots). The second row visualizes the observation bars. The third row visualizes the agent's belief state, from which the Bayesian surprise regressor used for the evaluation of (A) was derived. The last row visualizes the choice probabilities for each accessible action (north, east, south, west), the entropies of which formed the basis for the action selection difficulty regressor used for the evaluation of (B).

SPM12's restricted maximum likelihood scheme. Parameter contrasts of interest were obtained using unit vector contrast weights and analysed at the group level using one-sample t-tests. The resulting t statistic maps were thresholded at a cluster-forming threshold corresponding to $p < 10^{-4}$ (uncorrected) and the cluster significance assessed using their spatial extent (Friston et al., 1994).

Model-based fMRI results We found a positive relationship between trial-by-trial belief state updating of the group-favoured agent model A4a and the group fMRI data in a number of areas that are considered to be part of the task positive network (Spreng, 2012; Shulman et al., 1997) (Figure 4A). These regions include, bilaterally, the occipital cortex, the inferior parietal cortex (IPC), the superior parietal cortex (SPC) (bilateral: cluster size: 19678, peak voxel coordinates: x = -26, y = -88, z = 18, peak voxel t-value: 14.74), the lateral frontal cortex (LFC), the medial frontal cortex (MFC), the anterior cingulate cortex (ACC) and the insula (left: cluster size: 2474, peak voxel coordinates x = -28, y = 4, z = 48, peak voxel t-value: 11.54; right: cluster size: 4982, peak voxel coordinates: x = 48, y = 28, z = 28, peak voxel t-value: 11.27). Moreover, we observed increased neural activity with increased Bayesian surprise in left and the right striatum (left: cluster size: 553, peak voxel coordinates: x = -32, y = 20, z = 0, peak voxel t-value: 8.36; right: cluster size: 977, peak voxel coordinates: x = 32, y = 22, z = -2, peak voxel t-value: 10.07). Second, we found a positive relationship between choice difficulty as assessed by the entropy of agent A4a's choice probabilities and the group fMRI data in medial-posterior regions (Figure 4B). Specifically, significantly active clusters were identified in the cuneus (bilateral: cluster size: 235, peak voxel coordinates: x = -2, y = -90, z = 8, peak voxel t-value: 6.76), precuneus (left: cluster size: 55, peak voxel coordinates: x = -6, y = -70, z = 18, peak voxel t-value: 5.34; right: cluster size: 138, peak voxel coordinates: x = 16, y = -62, z = 26, peak voxel t-value: 6.22) and the right posterior cingulate cortex (PCC; right: cluster size: 148, peak voxel coordinates: x = 12, y = -52, z = -2, peak voxel t-value: 6.17).

Conclusion

In summary, we found behavioural support for human sequential decision making strategies that mimic look-ahead strategies reminiscent of real-time dynamic programming algorithms and which are based on cognitive representations of the uncertain task environment (belief states). Representing and updating such belief states appears to involve a large network of cortical and subcortical areas, and utilising them in difficult choice situations appears to involve the posterior cingulate cortex. This work may form the basis for the developing neural population models that implement the suggested algorithmic computations and testing these models directly on human behavioural and neuroimaging data.

References

Cohen, J. D., Daw, N., Engelhardt, B., Hasson, U., Li, K., Niv, Y., Norman, K. A., Pillow, J., Ramadge, P. J., Turk-Browne, N. B., et al. (2017). Computational approaches to fmri analysis. *Nature neuroscience*, 20(3):304.

Friston, K. J., Worsley, K. J., Frackowiak, R. S., Mazziotta, J. C., and Evans, A. C. (1994). Assessing the significance of focal activations using their spatial extent. *Human brain mapping*, 1(3):210–220.

Geffner, H. and Bonet, B. (1998). Solving large pomdps using real time dynamic programming, 1998. In *Working notes. Fall AAAI symp. on POMDPs*.

Korf, R. E. (1990). Real-time heuristic search. *Artificial intelligence*, 42(2-3):189–211.

Rigoux, L., Stephan, K. E., Friston, K. J., and Daunizeau, J. (2014). Bayesian model selection for group studies revisited. *Neuroimage*, 84:971–985.

Shulman, G. L., Corbetta, M., Buckner, R. L., Fiez, J. A., Miezin, F. M., Raichle, M. E., and Petersen, S. E. (1997). Common blood flow changes across visual tasks. *Journal of cognitive neuroscience*, 9(5):624–647.

Spreng, R. N. (2012). The fallacy of a task-negative network. *Frontiers in psychology*, 3:145.