Supplementary Material

for

Anticipated replanning costs influence resource-rational adaptation of planning depth in probabilistic environments

Sophia-Helen Sass[1], Lorenz Gönner[1], Sascha Frölich[2], Christian Baeuchl[1], Franka Glöckner[2], Michael N. Smolka[1*]

[1]*Department of Psychiatry and Psychotherapy, Technische Universität Dresden, Dresden, Germany*

[2]*Department of Psychology, Technische Universität Dresden, Dresden, Germany*

Probabilistic multi-step planning task: Debriefing questionnaire

Space Adventure Task Debriefing Questionnaire

In order to assess the clarity of the instructions and gain a deeper insight into your decision-making process, we now ask you to fill in the following questionnaire related to the game you just participated in.

1. Which keyboard button did you have to press in order to fly one step in the clockwise direction?

○ M ○ X

2. Which keyboard button did you have to press in order to fly two steps in the clockwise direction?

○ M ○ X

3. Please indicate how many game points could be collected by landing on the respective planets.

○ -20 ○ -10 ○ 0 ○ +10 ○ +20

○ -20 ○ -10 ○ 0 ○ +10 ○ +20

○ -20 ○ -10 ○ 0 ○ +10 ○ +20

○ -20 ○ -10 ○ 0 ○ +10 ○ +20

○ -20 ○ -10 ○ 0 ○ +10 ○ +20

4. Because of the asteroid storm, landing on the target planet became more challenging, increasing the chance of the spaceship not reaching its destination. Please estimate the chance of missing the target planet.

○ 10% ○ 20% ○ 30% ○ 40% ○ 50% ○ 60% ○ 70% ○ 80% ○ 90%

5. How many actions did you typically plan in advance?

○ 1 action ○ 2 actions ○ 3 actions ○ it varied often ○ I don't know.

The following questions are about the strategies you used to collect fuel in the game. Please explain your strategies as precisely as you can.

6. What strategy did you typically use to determine the best flight route in a planet constellation? Please describe your strategy below.

7. Did you include the chance that your spaceship would not reach the planet you were heading for in your planning? If so, please describe how this affected your route choice. If not, please just write "No".

8. Did you change or adapt your strategy within a planet constellation or throughout the game? If so, what circumstances triggered that? If not, please just write "No".

9. How would you describe your action strategy regarding target planets in an asteroid storm?

○ **Risk-taking:** I was rather willing to attempt a landing on a planet in an asteroid storm when it could have been beneficial, even though the outcome was uncertain.

○ **Cautious:** I rather tried to avoid landing on a planet in an asteroid storm, because the outcome was uncertain.

7. Did you include the chance that your spaceship would not reach the planet you were heading for in your planning? If so, please describe how this affected your route choice. If not, please just write "No".

8. Did you change or adapt your strategy within a planet constellation or throughout the game? If so, what circumstances triggered that? If not, please just write "No".

9. How would you describe your action strategy regarding target planets in an asteroid storm?

- **Risk-taking:** I was rather willing to attempt a landing on a planet in an asteroid storm when it could have been beneficial, even though the outcome was uncertain.
- **Cautious:** I rather tried to avoid landing on a planet in an asteroid storm, because the outcome was uncertain.
- **Neither:** I was neither particularly risk-seeking nor cautious; rather, I adapted my behavior to the circumstances of the planetary constellation.
- I don't know.

10. Now you have made it to the end of the space adventure game. Do you have any general feedback or impressions that you would like to share with us? If not, please just write "No".

Submit Form

Figure S 1: Debriefing questionnaire page 1 - 3. Participants filled out the debriefing questionnaire after completing the planning task. It was used to check their understanding of basic task rules and receive qualitative information on their decision strategies.

Psychological measures

Analyses of psychological measures include 53 participants (out of 74) who completed the respective questionnaires.

Impulsiveness

To assess impulsiveness as a covariate for planning task performance and planning depth, we used the 15-item Barratt Impulsiveness Scale (BIS-15)[1]. Participants self-rated statements about their usual behavior on a 4-point Likert scale from 1 (rarely/never) to 4 (almost always). The BIS-15 includes three subscales: non-planning, motor, and attentional impulsiveness. We assessed impulsiveness to account for individual differences in how participants respond to uncertainty induced by probabilistic state transitions in the planning task. We hypothesized that stronger impulsiveness is associated with shallower planning and lower performance in the planning task.

Spearman correlational analyses were performed to assess the association of potential inter-individual differences in trait impulsiveness on relative planning task performance and planning depth across the entire task. The analyses were performed for the BIS-15 score of the three sub-scales and the summarized score separately. There was no significant correlation between any BIS-15 subscale or the total score with relative planning task performance (motor: ρ = -0.026, p = 0.853; non-planning: ρ = 0.122, p = 0.384; attentional: ρ = -0.017, p = 0.901; sum: ρ = 0.055, p = 0.697) or planning depth (motor: ρ = -0.174, p = 0.213; non-planning: ρ = 0.031, p = 0.824; attentional: ρ = -0.030, p = 0.833; sum: ρ = -0.058, p = 0.678).

Barratt impulsiveness scale (BIS, points) [c]	
Motor	9.15 (3.00)
Non-planning	9.94 (2.96)
Attentional	9.57 (2.93)

Table S 1: Descriptive statistics of the BIS-15 Impulsiveness scale. Maximum possible impulsiveness score per subscale is 20.

Risk propensity

To assess risk propensity we used Holt and Laury's lottery choice task (HLL)[2]. Participants made 10 binary choices between two lotteries A and B. In each choice, lottery A featured a lower variance in payoffs, while B had a higher variance. The

payoffs themselves remained constant. However, as participants progressed through the decisions, the payoff probabilities changed, so that the expected value of A decreased as the expected value of B increased. Participants were informed that one of their choices would be played, and they would receive the winnings. We counted the 'number of safe choices' (choosing A) before switching to B. Switching before reaching item 4 indicates risk-seeking behavior, switching at item 4 indicates risk-neutral behavior, and switching after item 4 indicates risk avoidance. Ten participants, who switched back to lottery A after initially switching to B were excluded from the analysis of this measure[3]. The grouping based on the HLL revealed in 26 risk-avoidant and 8 risk-seeking participants.

In our probabilistic planning task of the main experiment, uncertainty in the probabilistic transition can be defined as “expected uncertainty”[4], i.e., it reflects known or learned probabilities of outcomes without knowing which specific outcome will occur. This type of uncertainty is similar to decision-making uncertainty in the HLL. Thus, we expected participants that exhibit risk-seeking behavior in the HLL to be more prone to risk-taking in the planning task.

To account for the effects of risk propensity on planning task behavior we divided the sample into a risk-averse, risk-neutral, and risk-seeking sub-sample based on their risk propensity measured by their HLL responses. The effects of mini-block type regarding replanning costs (within-subjects) and risk-attitude group (between-subjects) on relative performance and planning depth were assessed in respective two-way mixed ANOVAs. Effect sizes are reported as partial η^2.

The effect of risk propensity on planning task performance was not significant ($F = 2.104$, $p = .135$, $\eta^2 = .095$), nor was the interaction with replanning cost condition ($F = 1.009$, $p = .399$, $\eta^2 = .048$). Also for planning depth neither the main effect of group ($F = 0.143$, $p = .867$, $\eta^2 = .007$) nor the interaction of risk propensity group and replanning cost condition ($F = 1.211$, $p = .314$, $\eta^2 = .057$) was significant.

Lottery task (HLL, risk propensity)	
Mean N (SD) of low risk choices	5.37 (2.23)
Proportion risk-avoidant (%)	60.47
Proportion risk-neutral (%)	20.93

Table S 2: HLL risk propensity measure. The maximum number of risky choices is 10. Fewer than 4 indicates risk avoidance, 4 indicates risk neutrality, and more than 4 indicates risk-seeking behavior.

Mathematical model description

We modeled participants' action choices in the probabilistic planning task using reinforcement learning (RL) agents. Each agent could have a planning depth d of one, two, or three steps, respectively, limited by the number of (remaining) available actions. The agents' environmental model includes the available actions $A = \{'one\ step', 'two\ steps'\}$ and states S (planet positions), transition probabilities $p(s_{t+1}|s_t, a_t)$ for reaching a subsequent state s_{t+1} from a given state s_t with action a_t, as well as the immediate reward $r(s_t)$, which is returned upon reaching a state s_t. Agents plan their actions by computing the expected cumulative reward for executing each action (Q-values) with an optimal forward planning algorithm.

Planning is constrained by the agent's planning depth. Action selection is modeled for the two available actions (*one-step* and *two-step* action). The choice probabilities are determined by a sigmoid transformation $\sigma(x)$ of the difference between the corresponding state-action Q-values $\Delta Q(s_t, d) = Q(a_t = 'one\ step'|s_t, d) - Q(a_t = 'two\ steps'|s_t, d)$, where the probability of choosing an action increases with its relative Q-value.

$$p(a_t = 'one\ step'|s_t, d) = \sigma(\beta * \Delta Q(s_t, d) + \theta), \quad \text{(S 1)}$$

$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad \text{(S 2)}$$

$$\Delta Q(s_t, d) = Q(a_t = 'one\ step'|s_t, d) - Q(a_t = 'two\ steps'|s_t, d) \quad \text{(S 3)}$$

Choice probabilities for each RL model were additionally modified by a participant-specific inverse decision temperature β and an action bias θ. The inverse decision temperature β controls the sensitivity of the model to differences in Q-values, with β = 0 leading to random behavior. θ denotes an a priori response bias, where positive values imply a bias towards choosing the two-steps action.

For deterministic transitions, the computation of Q-values is identical across models. However, the planning strategies differ in how Q-values are calculated for the probabilistic transition.

In the unbiased full-breadth planning model, Q-values are computed using the planets' true values, weighted with the true transition probabilities of $p(s_{t+1}|s_t, a_t)$ = 0.5

for high-probability transitions and $p(s_{t+1}|s_t, a_t)$ = 0.25 for each low-probability transition. All possible transition outcomes are considered for the computation of Q-values.

In contrast, the low-probability pruning model assumes that low-probability transitions are ignored (pruned) in the planning process to reduce computational demands. This was implemented by setting the belief about the likelihood of a high-probability transition to 1.

In the discounted low-probability pruning model, the action value for the high-probability transition is discounted with a probability discounting factor γ_{prob}.

$$Q(s_t, d) = \gamma_{prob} Q(s_t, d) \quad \text{(S 4)}$$

The discounting factor follows a typical hyperbolic discounting function[6].

$$\gamma_{prob} = \frac{1}{1 + \kappa q_{t+1}} \quad \text{(S 5)}$$

$$q_{t+1} = \frac{1 - \rho_{t+1}}{\rho_{t+1}} \quad \text{(S 6)}$$

The hyperbolic discounting function γ_{prob} is modified by an individual discounting parameter kappa (κ). For κ = 0, the discounted value $\gamma_{prob} Q(a_t, s_t, d)$ for an uncertain planet equals the undiscounted expected value $Q(a_t, s_t, d)$. Larger κ-values indicate stronger discounting of probabilistic outcomes, where actions leading to an uncertain planet have a lower subjective value and become therefore less likely to be chosen in an action sequence, while actions leading to uncertain losses have a higher subjective value and become more likely to be chosen. The κ-values were capped at 30, as values beyond this threshold did not offer additional information.

Planning depth and parameter inference

To infer distributions over the model parameters and planning depth d, we used approximate Bayesian inference with a hierarchical generative model and a hierarchical approximate posterior. This model incorporates participant-specific parameters and mini-block-level information for choices and learning, with planning depth d modeled at the mini-block level and all other parameters at the participant level. We used stochastic variational inference using the Pyro v1.5.2 library[7]. For more details on the inference procedure, see Steffen et al. (2023).

We assume that planning occurs before the first action within each mini-block. Therefore, the primary variable of interest when analyzing planning behavior is the planning depth prior to the first action.

Behavioral simulations

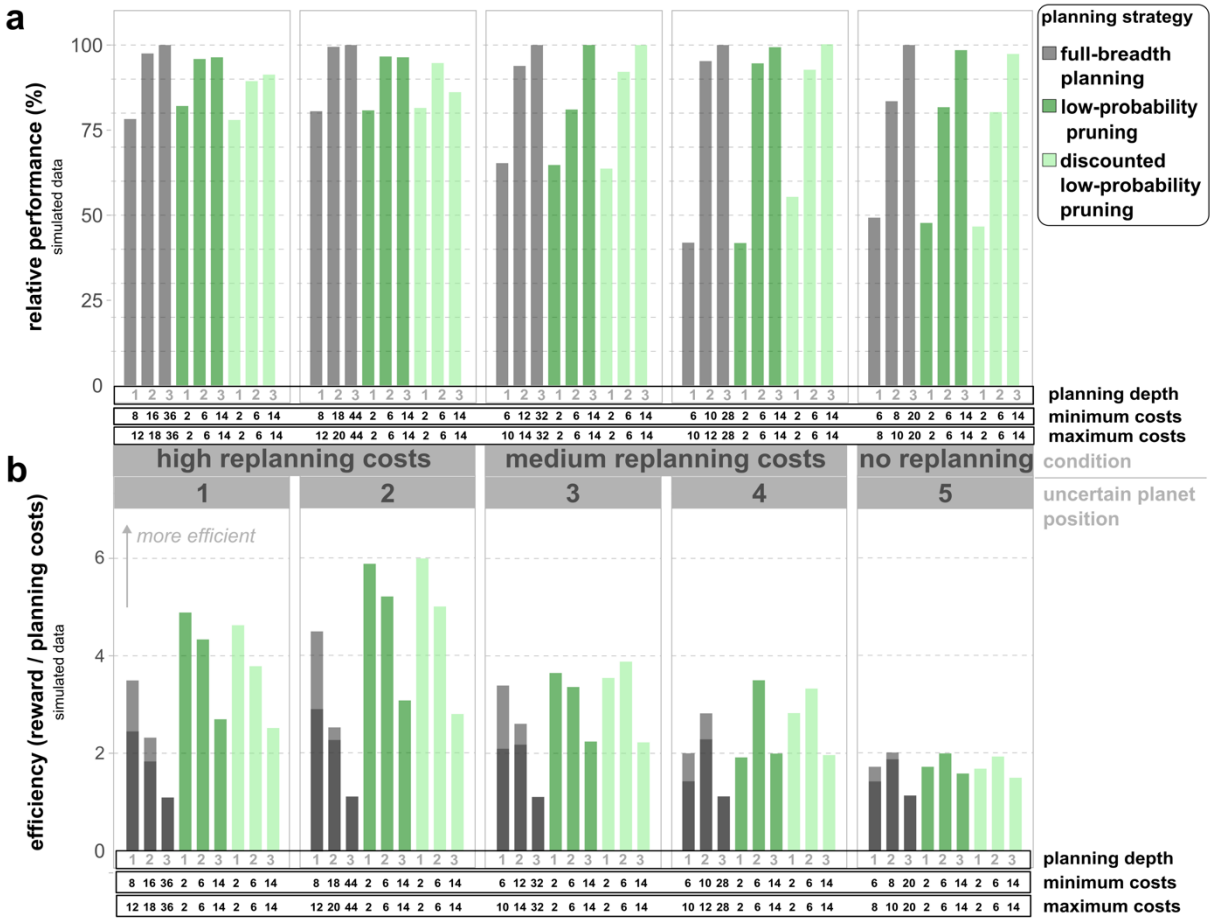

Figure S 2: Comparison of points earned and planning costs in the probabilistic planning task for various strategies (simulated data). The numbers 1 to 5 for each subplot indicate the position of the uncertain planet within the planet constellation of a mini-block, relative to the planner's starting position. These categories represent different mini-block types, each associated with varying (re-)planning costs. **a** Relative performance for each mini-block type was scaled between optimal performance (full-breadth planning strategy, planning depth three) and random performance as null reference. Relative performance for each strategy is informed by the sum of the average gain of points per mini-block of 1,000 agents. The discounting parameter for the discounted low-probability pruning model was set to $\kappa = 5$ (S 5). **b** Efficiency of each strategy as the relative performance divided by the relative initial planning costs as approximated by the number of decision tree nodes considered.

Goodness of model fit and model comparison

We compared the fits of three RL models averaged across the planning task. As a raw measure of model fit, we first computed the negative log-likelihood $l(\phi)$ (NLL, Equation (S 7) for each model. The NLL denotes the log-likelihood of participants' action choices $\{a_t\}$ given the inferred set of parameters summarized as ϕ.

$$l(\phi) = -\log P_\phi[\{a_t\}_{t=1}^N] = -\sum_{t=1}^{N} \log p_t(a_t/\phi) \quad \text{(S 7)}$$

Choice probabilities $p_t(a_t|\phi)$ denote the average of individual choice probabilities, weighted by the probability inferred for each planning depth per mini-block. Based on the NLL we then computed pseudo Rho-squared (ρ^2)[8] for each model as a standardized measure of model fit as a likelihood ratio index (Equation (S 8).

$$\rho^2 = 1 - \frac{l(\phi)}{l_{random}} \quad \text{(S 8)}$$

The compared models have different numbers of free parameters: Three in the full-breadth planning model (β, θ, d), three in the low-probability pruning model (β, θ, d), and four in the discounted low-probability pruning model (β, θ, κ, d). Since models with more parameters tend to have a better fit than models with fewer parameters, we additionally computed the Bayesian Information Criterion (BIC)[9] to determine the quality of model fit, adjusted for the number of free model parameters m with the number of observations n (S 9).

$$BIC = 2l(\hat{\phi}) + m \log n \quad \text{(S 9)}$$

To quantify and interpret the strength of evidence for each model according to the BIC, we calculated the difference ΔBIC (S 10) for each pair of models where $k1$ and $k2$ represent two out of the three models being compared[10]. Model evidence interpretation is based on Neath & Cavanaugh (2012) with "bare mention" for 0 ≤ ΔBIC ≤ 2, "positive" for 2 < ΔBIC ≤ 6, "strong" for 6 < ΔBIC ≤ 10, and "very strong" for 10 < ΔBIC.

$$\Delta BIC = BIC(k1) - BIC(k2) \quad \text{(S 10)}$$

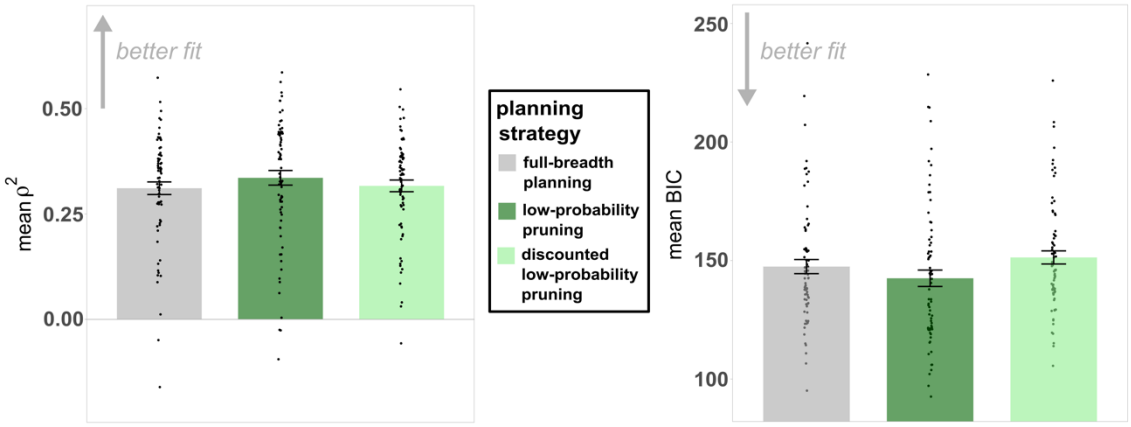

Figure S 3: Results of model comparison. Model fit comparisons averaged across mini-block types for each participant. Error bars indicate standard error of the mean. Higher ρ^2 values and lower BIC values indicate better model fit.

Averaged across participants, the low-probability pruning model fits the behavioral data best compared to the model alternatives with positive evidence for the low-probability pruning model compared to the full-breadth planning model (ΔBIC = 4.93) and strong evidence compared to the discounted low-probability pruning model (ΔBIC = 8.79). Three participants were excluded from all following analyses as the low-probability pruning model explained their data below chance-level, indicated by $\rho^2 \leq 0$, hence parameters could not be inferred reliably. Excluding them did not change the results of the model comparison (full-breadth planning vs. low-probability pruning: ΔBIC = 5.86; discounted low-probability pruning vs. low-probability pruning: ΔBIC = 11.01). Note that only the cleaned values are reported in the main text.

Model cross-fitting

In our study, we employed the following model cross-fitting procedure to evaulate the accuracy of our model selection. We generated a set of simulated data consisting of 100 samples using our candidate models: full-breadth planning, low-probability pruning, and discounted low-probability pruning. The respective parameters for each model were set as follows: β = 3, θ = 0, κ = 10. Subsequently, we fitted the simulated data to each of the three alternative models, performing 30 runs with 500 iterations for each fit. To determine the best-fitting model among these alternatives, we employed the BIC for comparison. We selected the model exhibiting the lowest BIC as the best-fitting model and generated confusion matrices and inversion matrices. These matrices quantified the probability of correctly identifying the true underlying planning model from the set of alternative models (confusion matrix) and the probability of the identified best-fitting model being the true underlying planning model (inversion matrix). This approach enabled us to assess our approach's feasibility to differentiate and reliably identify planning strategies within the human data. A summary of the results can be found in Figure S 4.

The results show that agent behavior coming from different underlying models can be differentiated and identified in our model comparison based on BIC, indicated by the high probabilities along the diagonal of the matrices.

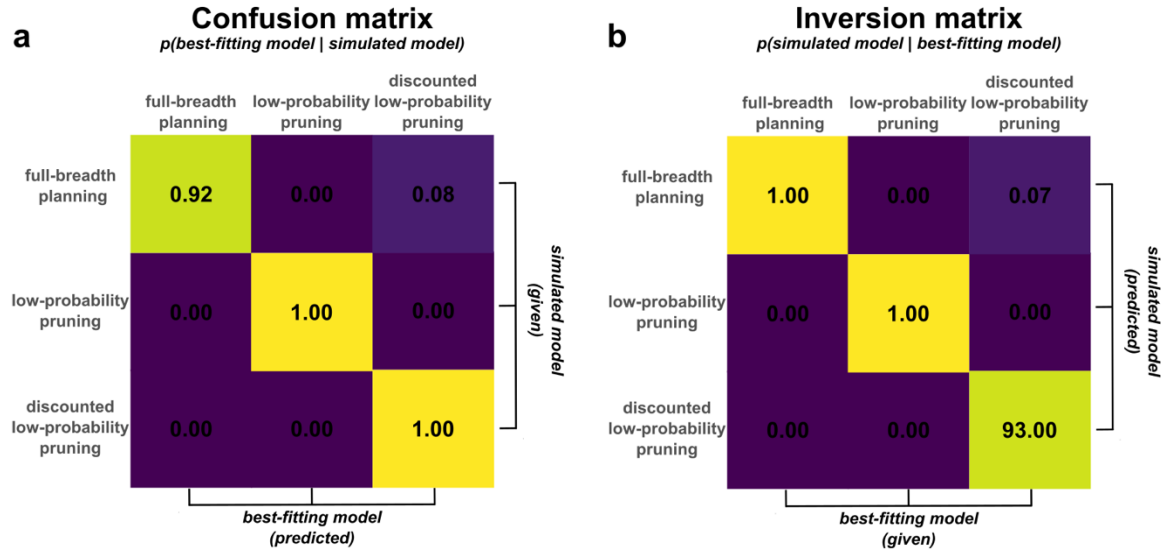

Figure S 4: Cross-fitting for model selection based on BIC. To assess the accuracy of our model selection, confusion and inversion matrices were generated. Diagonals from top left to bottom right indicate the true positives. **a** Confusion matrix for model selection as the probability of a model being identified as best-fitting based on the BIC given the simulated data from the respective model. **b** Inversion matrix for model selection as the probability of the simulated data being based on a respective model, given that it was identified as the best-fitting model based on the BIC.

References

1. Spinella, M. NORMATIVE DATA AND A SHORT FORM OF THE BARRATT IMPULSIVENESS SCALE. *Int. J. Neurosci.* **117**, 359–368 (2007).
2. Holt, C. A. & Laury, S. K. Risk Aversion and Incentive Effects. **92**, (2024).
3. Hirschauer, N., Musshoff, O., Maart-Noelck, S. C. & Gruener, S. Eliciting risk attitudes – how to avoid mean and variance bias in Holt-and-Laury lotteries. *Appl. Econ. Lett.* **21**, 35–38 (2014).
4. Bland, A. R. & Schaefer, A. Different Varieties of Uncertainty in Human Decision-Making. *Front. Neurosci.* **6**, (2012).
5. Sutton, R. S. & Barto, A. G. *Reinforcement Learning: An Introduction*. (The MIT Press, Cambridge, Massachusetts, 2018).
6. Green, L. & Myerson, J. A Discounting Framework for Choice With Delayed and Probabilistic Rewards. *Psychol. Bull.* **130**, 769–792 (2004).
7. Bingham, E. *et al.* Pyro: Deep Universal Probabilistic Programming. (2018) doi:10.48550/ARXIV.1810.09538.
8. Simon, D. A. & Daw, N. D. Neural Correlates of Forward Planning in a Spatial Decision Task in Humans. *J. Neurosci.* **31**, 5526–5539 (2011).
9. Schwarz, G. Estimating the Dimension of a Model. *Ann. Stat.* **6**, (1978).
10. Neath, A. A. & Cavanaugh, J. E. The Bayesian information criterion: background, derivation, and applications. *WIREs Comput. Stat.* **4**, 199–203 (2012).