**Background**

We designed the following study to build on our previous preregistered study (https://osf.io/kcaqh) wherein we showed humans display evidence of using a predecessor representation (PR) over a successor representation for planning in a diverging environment. Specifically, we found evidence that an agent uses PR for backward planning in a one-step planning problem. However, that study left open the question of whether the same behavior indicative of backward planning would be observed in a multi-step planning problem, where individuals must plan out multiple actions.

In the present study we thus tackle whether humans show evidence of using a predecessor representation (PR) as opposed to a successor representation (SR) to plan in a two-step planning problem in a diverging environment.

**Main hypothesis**

Individuals will leverage PR-based planning in a two-step planning problem in a diverging environment.

**Task**

To build on our results in Study 2 (<https://osf.io/9yzex>), wherein we showed humans utilize a predecessor representation for planning, we designed the following task to test the hypothesis that, in a diverging state space that involves planning multiple actions, humans predominantly use a predecessor representation (PR) for planning, as opposed to a successor representation (SR), for planning.

Participants first complete 520 trials of a forced choice learning phase to ensure they know the state space and transitions well, and that each participant sample the same number of times from each state transition in the state space. Unlike Study 2, the present study involved subjects also taking an action at the second stage to reach the third stage. Every 30 or so trials during this forced choice learning phase, participants are given a memory quiz where they must choose one of the two first-stage states that has the best chance of transitioning them to a 3rd stage state.

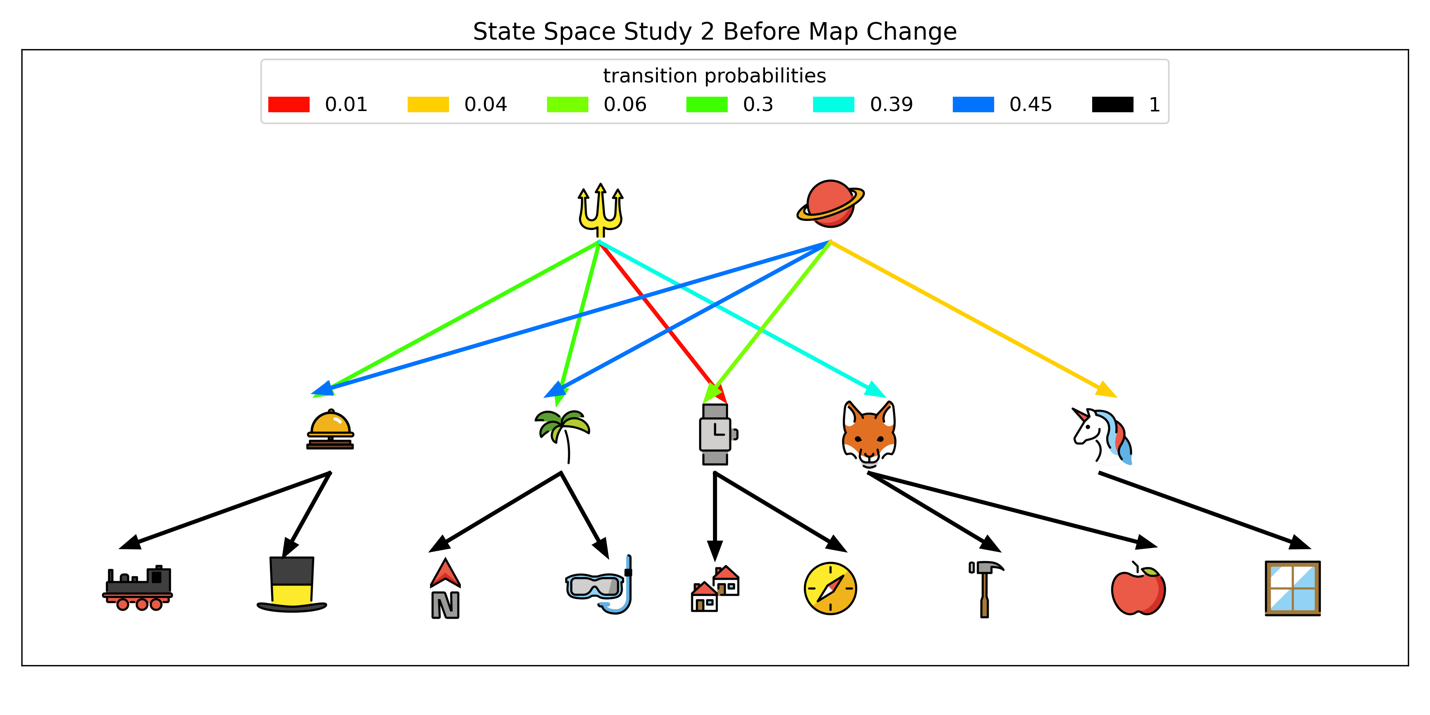
After this planning phase, participants face planning queries that ask them to use what they learned about the probability of transition to certain states in the preceding training phase to win the most money.

For each planning query, a single reward location is instructed on the screen, and the participant is told to use this reward information to choose two actions that will maximize their reward. In order to do so, they should use what they have learned about the state space in the preceding training phase, which determines the likelihood of transitioning to successive states given an initial action. Participants know that if they reach a location, they will receive the reward found there (i.e., the reward function is deterministic).

To isolate the effect of the type of probabilistic representation used in planning (PR or SR), we instructed the same reward magnitude (100 points) for each instructed reward location across all query types. However, participants are never shown whether they receive the reward; they are merely told the computer is simulating whether they in fact reached the reward, and this will affect the bonus money they receive at the experiment’s conclusion. This is intended to keep learning unaffected by reward receipt, removing the possibility that model-free learning impacts subsequent decisions. Participants are given unlimited time to deliberate before taking an action.

To dissociate PR- and SR-based planning, we designed the state space to include initial actions that differ in base-rates, which only PR-based planning is sensitive to. Thus, the low base-rate starting state was experienced 160 times whereas the high base-rate starting state was experienced 360 times. The degree to which the base-rates differed was determined together with the state-transition probabilities so as to allow for planning queries by which SR preferred low base-rate actions and PR-based planning preferred high base-rate actions.

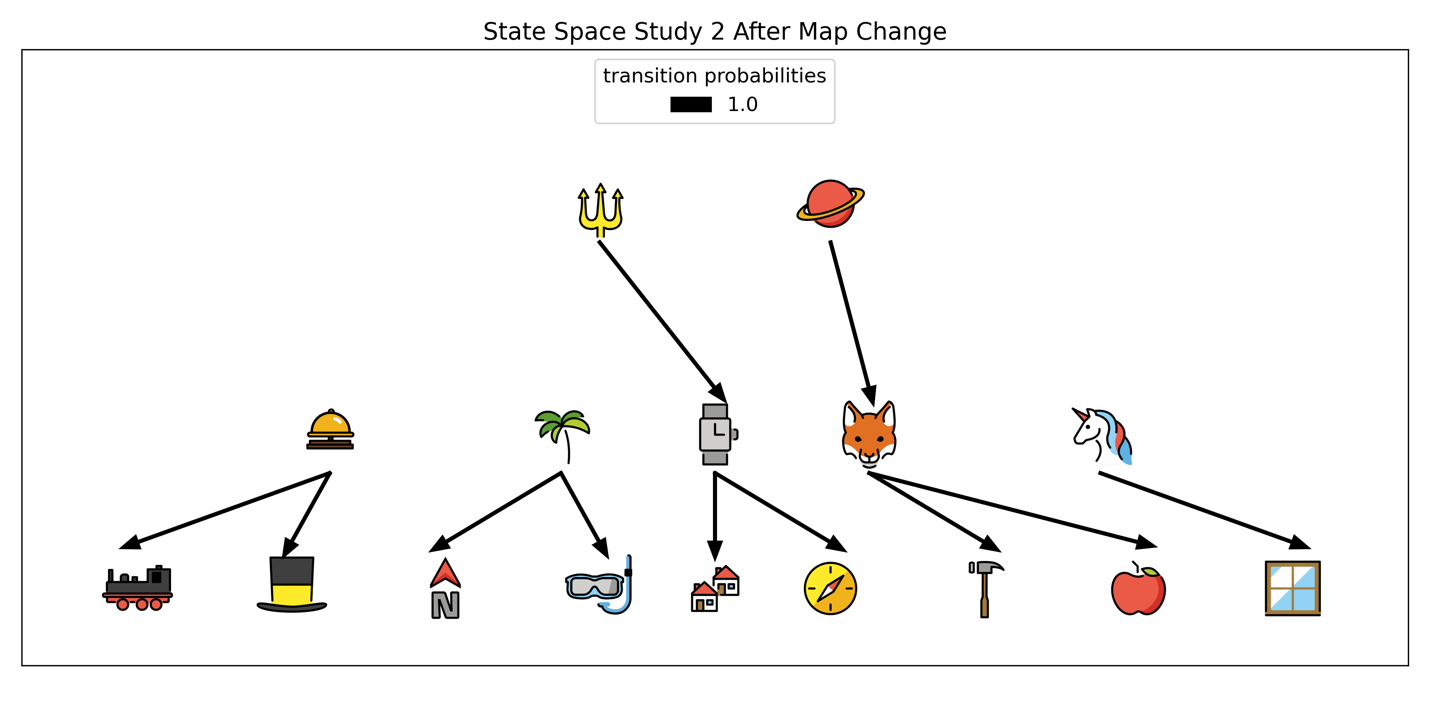
To test our main hypothesis, we designed 4 single-goal planning queries wherein PR-based and SR-based planning arrive at different conclusions regarding which action is best to take, but holding constant across strategies the ratio of best:worst action value identical to that of Study 2 (<https://osf.io/kcaqh>). In this Study, unlike our previous Study 2, participants were required to plan out their first and second actions by entering them one after the other in a single trial after observing which distal state contains reward.



**Choice 1**: Trident or Planet? After this choice, transition to next state as a function of transition probabilities

**Choice 2**: Choose left or right to always go to where arrow points (e.g., left at bell always transitions to train)

**Figure 1.** State space before transition revaluation. The first choice involves participants selecting which state they can start from: the trident or the planet. The second choice involves choosing left or right at the second-stage state (5 possible). These actions are deterministically related to the final states.



**Choice 2**: Choose left or right to always go to where arrow points (e.g., left at bell always transitions to train)

**Choice 1**: Trident or Planet? After this choice, transition to next state via instructed deterministic transitions

**Figure 2.** State space *after* transition revaluation. 1st stage states now deterministically lead to a single successor state. The instructions are as follows: “The picture world you learned has now CHANGED! Now, the TRIDENT leads only to the WATCH, and the PLANET leads only to the FOX. Look below! (We display the planet pointing to only the bell, and the planet pointing only to the bell).

**Statistical Analysis**

**Main hypotheses**

**Evidence of PR:** We model the four 1st-step choices (i.e., choosing trident or planet in Figure 1) participants made during planning queries for which PR- and SR-based planning prefer different actions using a hierarchical beta-binomial model, that estimates the degree to which participants chose in line with PR-based planning predictions. Note we do not model 2nd-step choices because PR and SR converge on the preferred action for these transitions. A beta distribution at the group level defines the tendency of the group to choose in line with PR-based planning. This group-level tendency is defined by its mode, and the spread of possible parameter values around the mode is captured by the distribution’s concentration. This parameterization is recommended for beta distributions (Kruschke, 2014). At the subject-level, a binomial parameter defining the tendency to choose in line with PR-based planning is drawn from the group-level beta distribution. Thus, a binomial likelihood function serves to account for the four 1st-step choices each participant made.

We test our hypothesis by extracting the mode of the posterior group-level beta distribution, which defines the population’s tendency to choose in line with PR-based planning, and comparing the mode to a range of null values. A value of 1 for the mode indicates subjects chose perfectly in line with PR-based planning, and 0.5 or below indicates SR-based planning. We hypothesize that the mode of the posterior will be significantly greater than 0.5, indicating evidence in favor of PR-consistent choice. To do so, we define a region of practical equivalence (ROPE) around the null value of 0.5 following Kruschke’s (2014) guidelines. Specifically, we take the standard deviation of the percentage of times subjects chose in line with PR-based planning and multiplied this value by 0.1 to define effects that are too small to be considered significant (here, +/-0.01). Statistical significance is defined by the highest-density interval (HDI) being entirely non-overlapping with the ROPE.

We also test for transition revaluation the same as in Study 1 to ensure participants use predictive representations to plan and not a full transition map as used by model-based planning. We use a simple Bayesian model to estimate the mean of the difference between the cost of transition revaluation minus the cost of the reward revaluation, where this mean difference is drawn from a normal distribution with prior mean and variance on the scale of the data.

**Manipulation Checks**

To test whether the main result did not merely reflect a general, reward-independent bias to select high-base-rate states, we used another set of single-reward queries, wherein backward and forward planning favored the high-base-rate starting state in half of the queries and the low-base-rate starting state in the other half. To estimate whether this occurred, we will fit the same beta-binomial hierarchical model described to model evidence for our main hypothesis on these queries. The test is passed if the HDI does not contain extreme low (0.3) or high (0.7) values.

To ensure individuals could enact 2-step planning, which is necessary to verify multi-step planning was possible in our task, we test whether the percentage of correct answers on 2-step planning problems is significantly greater than the chance level of 0.25 (50% chance of being correct on each step). To do so, we fitted the same beta-binomial model described for the main hypothesis on the 4 planning queries for which PR and SR enact the same policy. We compare the estimate of the posterior mode to a value of 0.25, signifying chance-level correctness.

**Data exclusion**

Subjects that fail on at least 5 learning trials during learning are automatically excluded.

Subjects that score worse than 75% on their memory probes will be eliminated. According to pilot data, we expect this to be around 10% of the collected data, and have incorporated this expected data loss into our sample-size calculations.

**Preliminary Results**

Existing data are from pilot data (n=9). None of these data will be included in the preregistered study analysis. These data comprise the same OSF repository attached to the present preregistration (see here: <https://osf.io/s286z/>).

Participants: Collected online through Prolific and PsychoPy (Pavlovia). Only requirements are that subjects speak English fluently.

Main hypothesis:

Posterior mode percent to choose the first-step in-line with PR- as opposed to SR-based planning = 0.72, HDI=[0.513,0.947].

Manipulation checks:

Posterior mode percent to choose high base-rate starting state = 0.51, HDI=[0.30,0.76]. No base-rate bias value = 0.50

Posterior mode percent correct on two-step planning problems = 0.66, HDI=[0.42,0.92]. Chance-level correctness = 0.25.

**Power Calculation**

Subjects = 75. (expected 10% exclusion due to poor memory of state transitions).

We followed Kruschke’s (2014) procedure for Bayesian Power Analysis. Using pilot data (n=9), we fitted the hierarchical Bayesian Beta-Binomial model described below to test our main hypothesis, and extracted the group-level posterior distributions for our power analysis. For 100 iterations, we randomly drew from these posteriors (now our priors for our preregistered analysis), and generated 100 simulated datasets, each with our set sample size of n=75. On each of these simulated datasets, we re-fit our hierarchical model, and tallied the number of times highest density interval (HDI) of the group mode (the key parameter defining the tendency in the group to engage PR-based planning) did not contain the null values. This resulted in an estimate that n=75 yields 96% power to detect a significant effect for our main hypothesis.