**Response to reviewer and editor comments**

EDITOR COMMENTS:

In particular, your revision must address the following (as well as all other reviewer comments):  
  
(1) Reviewers 2 and 3 both express doubt over whether your experiments address planning processes. We consider this to be a fundamental point because much of the background and justification of this work is based on earlier work on backward planning in animals (refs 2-5) and forward planning in humans (ref 10). To address this, you should ideally conduct at least one new experiment which properly speaks to backward planning instead of simply providing arguments for your framing of this work as relating to planning.

*We now show in planning problems that require multiple actions (specifically 2-step problems) that findings here hold. Moreover, we emphasize how our earlier studies reflect pre*  
  
(2) All reviewers raise the question about alternative explanations for experiment 4. Please ensure you address their concerns, otherwise it should be removed.

*We have conducted a follow-up experiment to ensure experiment 4 cannot be explained by a base-rate bias when outcome probabilities are very small. Moreover, we have addressed reviewers’ concerns regarding external factors explaining our results. To do so, we have removed all information regarding race, which we agree may bias results given preexisting misconceptions and stereotypes. We show study 4’s findings hold when removing questions involving race/ethnicity, and use only questions based on states as factors leading to professions in a new independent study.*   
  
(3) Please address Reviewer 3’s comment regarding the evidence strength for the backward model.

*We conduct a full model comparison to show that backward prediction model explains data best in divergent tasks, and forward prediction model explains data best in convergent tasks. To do so, we compute the expected values according to each model (PR, SR, Fully Model-Based and Base-Rate Bias), and show that PR best-explains data in divergent tasks, and SR best-explains data in convergent tasks.*   
  
(4) Please improve the reasoning, experiment descriptions, and results reporting towards a more concrete and clearer way.

*Significant revisions were made to the logic of the paper, the explanation of computations, and the clarify of visualizations in figures.*

*To ensure the computation of SR and PR are known, we show now in matrix form a sample computation, with each number in the computation fully visible. Regarding clarifying the reasoning, we refine our explanation of our hypothesis as being about predicting upcoming states. We describe how these state predictions can be used in planning, both in simpler language and in the specific mathematical equations used to for planning. Last, in new studies 5 and 6, we show how PR and SR predictions are used to form plans in tasks that require taking multiple, sequential actions.*

REVIEWER COMMENTS:  
  
Reviewer #1:  
Remarks to the Author:  
In this manuscript, Sharp and Eldar present nice evidence based on a test of planning direction that humans adaptively deploy forward or backward planning depending on the convergence/divergence of the environment. I think this manuscript is really well done, especially considering the pre-registered nature of the study, and the map change experiments (e.g., Fig 1E). The results are conceptually novel, interesting, and rigorous. Overall, I think this manuscript is solid and can even be published as is. I have only a few minor suggestions to improve the manuscript.

1. There is a statement that backward-predictive representations are highly sensitive to the base rate of states. This seems true, but couldn’t one in principle define a base rate subtracted backward representation, which would reflect the backward-predictive representations more cleanly?

*The backward-subtracted predecessor representation, as used in Jeong et al. (2022; called Predecessor Representation Contingency or PRC), which subtracts predecessor base rates from a PR, would result in the present study in negative probabilities, which are difficult to interpret. If instead one divided by base-rates, one would get the SR, which we test. Ultimately, if an agent puts in the extra cost of learning base rates of starting states and PR to compute a PRC, such extra learning should provide a benefit over not learning other strategies that do not require such learning. However, in the present tasks, it is actually optimal to use SR (even if it is less efficient to do so).*

2. Not sure if Fig 4 exclusively requires backward learning. Because the task forces agents to choose one of two options, when the outcome has really low probability (high divergence condition), it is reasonable to choose high baseline probability characteristic. But it doesn’t necessarily mean that they use the backward-predictive representation.

*To address this concern, we ran an independent follow-up study to experiment 4, in which we used low-outcome states from Study 4, and starting states with*

3. Fig 1B & C calculations could benefit from more explanation – wasn’t super clear until I read through the supplementary note about the dot product between reward value and predictive representations. This description seemed a bit unclear in the figure legend since the summing operation of a dot product is not explicitly stated in the legend.

*We have added the explanation of the dot product specific calculation in the legend.*

4. Page 11 line 15 appears incomplete at “efficient”.  
  
  
  
Reviewer #2:  
Remarks to the Author:  
Summary of paper: This paper provides a series of studies (3 lab studies and 1 analysis of real world data) arguing that people can use forward predictive and backward predictive representations to guide their choices in a sensible manner. The first experiment uses a task involving two start states, followed by four possible intermediate states, followed by eight possible final states. Participants were first trained on state-successor state contingencies to learn the predictive representations, then they were given a set of trials in which rewards appeared at states and needed to choose which initial state to begin from; finally, they were given trials in which the initial-state to intermediate-state contingencies changed. The second and third studies made the transition graph "divergent" (out-bound edges > in-bound edges) or "convergent" (in-bound edges > out-bound edges) and showed that behavior was better explained by backwards representations in the former and forwards representations in the latter. Finally, the fourth study collected data on judgments of predicting occupations (a person being a journalist or not) from demographic characteristics (a person being white or asian). Participants were more likely to choose high base-rate characteristics when the occupation was rarer (a proxy for divergence), which is a prediction of the backwards representation model.  
  
Main review: I found this to be a stimulating paper that addresses an interesting question about how people learn forward versus backward representations. The experiments and analyses were well executed, however, I have reservations about the framing of the paper and whether the different components fit together to form a unified picture, which are summarized in the following points:  
  
1. The paper is explicitly framed in terms of backwards versus forwards \*planning\*, however, the main decision-making experiments (exp 1-3) are not planning tasks since they involve selecting a single action (which state to start at) that then initiates a series of events (the intermediate and final state visited). Planning is usually conceptualized as as involving a sequence of interdependent choices (e.g., chess moves) and is computationally distinct from prediction or making a single choice. Additionally, the final survey experiment is not a decision-making task at all, but rather a prediction task where participants provide conditional probability judgments (e.g., people aren't making a choice about a person's demographic characteristics). This ambiguity is also reflected in the inconsistent terminology that is used throughout the paper (backwards planning vs. prediction vs. learning). What unifies these (again, well executed) experiments is not that they involve planning, but that they contrast different kinds of learned predictive models (e.g., learning a successor representation versus a predecessor representation from transitions; neither the SR nor PR is a planning algorithm, at least as used in this paper). The authors should frame these results differently as that is needed to make the scope of the contribution clearer.

*We thank the reviewer for bringing up these points so that we can clarify how we use the terms “planning” and “prediction” and in doing so, improve the framing of the paper as a whole. Firstly, we used the technical definition of planning from a reinforcement learning, whereby task structure (encoded by predictive representations here) is integrated with a reward function (here instructed point values) to enable agents to derive the values of actions on the fly (on a trial-by-trial basis). This stands in contrast to forms of value learning that require individuals to use feedback to update expectations action values. However, given we use only one-step prediction, and most forms of planning involve multi-step prediction, we have changed our language to focus on prediction rather than planning. This is seen in our change in Title as well as throughout the manuscript, changing “forward/backward planning” to “forward/backward prediction”. Second, because we aim to generalize our findings to unequivocal cases of planning, we also conducted two additional preregistered studies wherein participants had to use multi-step predictions in order to plan, and select the best action based off of such multi-step planning.*

2. In study 4, the authors introduce the qualitative prediction around base rates (i.e., only the backward model predicts selection based on base rates). While this seems intuitively true, I don't think the authors make an argument for this or clearly demonstrate why this must be the case. Additionally, it wasn't clear to me whether the first 3 experiments boil down to whether people are more likely to choose based on base-rates.

*We now clarify the computations involved in PR versus SR to demonstrate why the former is sensitive to base rates while the latter is not. Specifically, we now say:*

*“SR probabilities by definition conditionalizes on the initial states (s). By contrast, PR probabilities are modulated as a function of base rates (s). Thus, it is*

*By definition, base-rate sensitivity must be the case for PR-based planning and base-rate neglect must be the case for SR-based planning.*

*We first show how environmental divergence leads to more efficient representations in PR and less in SR.*

*Then, we show that the only behavioral difference in PR and SR, base-rate sensitivity, differs according to environmental divergence.*

*Thus, we have a model, which according to a novel principle, accounts for base-rate sensitivity in the context of multi-step prediction.*

3. More generally, I found it difficult to understand the import of the experimental results since they were not reported in terms of qualitative differences in behavior that would be predicted by the different accounts. Rather, experimental results were mainly described in terms of theoretical constructs (e.g., saying that participants' choices indicated backward planning because they were consistent with choices that backward planning assigns higher value). As someone who is familiar with this family of modeling techniques, I found it hard to figure out what was meant concretely, and I imagine that non-computational readers will find it especially opaque.

*We thank the reviewer for pointing out places where confusion arose. We have revised our description which specifies how our models do indeed generate different qualitative differences in behavior. We do this first by specifying the computations in Fig 1, and showing how SR generates no clear policy, whereas PR does. In studies 2 and 3 we show how the same computations lead to opposite action predictions.*

*STATE COMPUTATION*

4. I found figures for experiments 1-3 difficult to interpret since it wasn't clear when things represented rewards versus expected values and when they represented one-step probabilistic transitions versus successor/predecessor representations.

*We now include in Fig. 1 a full description of the computation used.*

5. Could the results of exp 4 be explained instead by a representativeness heuristic? Namely, for rarer occupations (more divergent ones), people fall back on heuristics such as the most common predecessor type.

*If the rareness of occupations was the best predictor of base-rate sensitivity, then our control analysis for rareness of occupation would not have panned out.*

6. The authors state that backwards representations are more "efficient" in diverging environments. I don't share this intuition and this seemed to be asserted without much evidence.

*We now describe specifically why PR is more efficient in divergent environments in the introduction.*

*When there are more successors then predecessors, a PR representation that is used to guide choice towards a future goal has less representations than an SR. This is due to a simple principle: in a diverging environment, SR will make non-zero predictions about many possible goal states that may not be desired. By contrast, because PR conditions on goals, it only needs to include the desired goal and possible predecessor states, which are far fewer in diverging environments than successor states.*   
  
7. I found the description of the procedure for experiment 4 inconsistent - in the main text it suggests that the DV was choosing a characteristic (demographic feature) but in the methods it says "all queries involved the conditional likelihood of engaging in an occupation..." suggesting its a numerical response.

*We now clarify in methods that “all queries* ***are based on conditional likelihoods, whose objective values were not stated in the queries, of engage***

8. Table S1 has an extra caption

***Removed.***  
  
  
  
Reviewer #3:  
Remarks to the Author:  
Sharp and Eldar conduct four studies to examine whether humans use backward-predictive representations in planning. The first three experiments involve a learning phase where participants observe sequences of transitions, and their understanding is occasionally tested. A subsequent planning phase asks participants to select a starting state that would maximize rewards based on a given set of rewards. This is followed by a transition revaluation phase where they adapt to changes in the transition structure.  
In Experiment 1 included a diverging state space and showed that backward planning was evident in about 58% of choices. Experiment 2, similar to the first but with a single reward and pitting forward and backward choices more directly against each other, showed a preference for backward planning in approximately 61% of choices. Experiment 3, using a converging state space, demonstrated a preference for forward planning. Experiment 4 probed backward planning in a real-world diverging state space with different base rates, revealing a preference for options with a higher real-world environmental divergence rate.  
  
The study addresses an interesting topic, experiments are well designed and the paper is generally well written. I do have several concerns:  
  
MAJOR  
  
1. The authors should clarify the theoretical framing and novelty of their work. Although the authors refer to a planing process throughout their paper, they don’t present clear evidence to that claim in my opinion. The proposed PR model learns cached associations between each state and the states that precede it, akin more to a backwards associative process than a true planing process that involves a step-by-step procedure.

*Planning involves using knowledge of an environment in order to determine which actions are best to take. This could involve just a single action, as someone for instance imagines which of many actions will transition them to their goal. As such, single-step planning differs from other forms of retrieving cached predictions about which action has most value, which involves no kind of However, typically, this involves taking action* ***sequences*** *to solve a problem that requires taking multiple actions. The motivation for the successor representation was to solve multi-action planning problems by avoiding step-by-step operations algorithmically. Instead, one takes cached multi-step predictions (e.g., action 1 predicts state 1 and state 2 by probability X1 and X2), and uses a simple linear operation with a vector of where reward is to solve this problem.*

*While studies 1 and 2 could use this kind of planning in a 1-step problem, we agree it blurs the line between planning and prediction. For this reason, we shift the focus to forward and backward prediction. However, to demonstrate that humans use these predictions to solve putative multi-action planning problems, we carried out two new preregistered experiments that require individuals to take multiple actions in order to reach reward. In doing this, we show that our key principle holds. Backward multi-action prediction is used to solve a planning problem in a diverging state space, and forward multi-action prediction is used to solve a planning problem in a converging state space.*

Figure 1 describes the computation as if participants integrate reward information about every step along a path. Yet, the SI shows that subjects seem to simply focus on the highest reward state, using cached information about which state has the strongest association to it.

*We largely agree with the reviewer that Studies 1-3 focus on forward and backward prediction as opposed to planning, and have thus change our text accordingly to focus on prediction (both in the title, in the abstract and throughout the paper).*

*While it is true that our SI analysis shows that participant may not utilize all available reward information, we regard those findings as preliminary with respect to whether individuals truly can integrate more than a single reward in their planning, as other evidence suggests they can. Thus, we opted for displaying in figure 1 the classical form of successor representation (Dayan, 1994), and its counterpart in predecessor representation, that has ample evidence. Moreover, because we use single-goal studies in subsequent experiments, we regard the question about how many reward individuals use in different prediction or planning problems as outside the scope of the present manuscript.*

*WE also now have done two additional preregistered experiments to justify our statement that SR and PR can be used for planning. In these studies, participants had to plan out more than one action.*

In Experiment 2, “planning” only involves 1 reward, and hence further blurs the line between associative retrieval and true planning. Experiment 4 only involves a 1-step association. I therefore think the authors should be more clear about the process under investigation, and incorporate the literature on backwards associative learning, which goes as far back as Ebbinghaus.

*We agree with the reviewer, and other reviewer comments, and have thus changed the paper to focus instead on multistep prediction and not on planning. We thank the reviewer for pointing us to Ebbinghaus. We see his ideas as critical fundaments upon which we extend backward prediction to a successor representation framing.*

2. The indirect support for the backward model through null effects of the MB model and the base rate influence is not sufficient. A more robust approach would be to directly compare predictions from forward, backward, simple base rate, choice bias, and MB models, utilizing all available data points.

*We could get log likelihoods for each model, assuming that if the participant chooses in-line with the model’s predicted action, it is ~1 and ~0 if not. This is like taking assuming a beta\*EV high enough to*

*Or we can take the approach of regressing EVs onto choice as we did for study S1.*  
3. The paper should report more detail, particularly regarding the models central to the results. Known names should be used (e.g., successor representation model for the forward "planning" model) and models described in the main text.

Information about frequencies and transition probabilities during the learning phase, and how participants' knowledge developed over time, would be useful for assessing the efficiency benefit of backward planning

*We have fixed this both in-text and in Fig.1 by including proper names for the algorithms we focus on (successor representation – SR, predecessor representation – PR). Additionally, we now display how each model operates, including a sample EV calculation for SR and PR in Fig. 1. We include the frequencies and transition probabilities in main text as well.*

*We do not have access to how participants knowledge developed over time. We simply have their performance on memory probes during learning, which we use to exclude participants. We agree that future work should try to more directly probe how participant knowledge of state spaces evolves over time.*  
4. Experiment 4 has several issues that need addressing, such as

the 1-step transition structure's compatibility with planning,

*We change this to focus on prediction instead of planning in line with changes made to studies 1-3*

clarity on participants' knowledge of base rates,

*We make the assumption that participants are influenced by the true base-rates of starting states. We now make clear this is a potential limitation of the work, as peoples’ knowledge of base-rates could deviate widely from true base-rates*

and potential external factors influencing participants' data.

*We agree that external factors could have significant influence on decision making for Study 4 queries. However, we chose facts about the US and only chose individuals in the US to try to reduce this problem, in addition to sampling from a large population (n=1000) to reduce these between-subject differences. However, we now note this as a major limitation – future studies need to assess subjective probabilities of base-rates to validate our assumption that on average, these match the objective base-rate probabilities*

I am therefore not sure if questions regarding race are the most valid to investigate planning directionality.

The use of the term divergence in this experiment seems inconsistent with its use in other studies.

*We used an extended definition of divergence in Study 4 to gain greater regularity on the influence of state space structure on backward and forward prediction. In addition to the ratio of starting : end states, which is how we defined divergence in Studies 1:3, the probabilities of specific transitions can also modulate divergence as a function of how much probability density is in a specific transition. Consider for example an environment with 10 starting states and 5 end states. This is converging according to the definition in Studies 1-3. But now imagine that one of the 10 starting states has a very high base rate, 99%. This starting state will be experienced basically each time, and be spread across the 5 end states. Thus, in this case, the transition appears to take on the structure of a diverging state space, because the other starting state are essentially negligible. From this example, we can glean that in addition to the number of states in the state space, the specific ratio of probabilities can also shape environmental divergence and convergence.*

MINOR  
5. More information about participants reaction times would be useful (see Mommenejad et al., 2017 for arguments linking RTs to SR processes). I could see forward learning and backward planning to be associated with additional cost for inversion. Particularly interesting would be comparing errors vs correct trials and post map rearrangement.

*The setting here is different than Momennejad et al., 2017, in that here we give participants unlimited time with the reward information, and then let them choose when they want to enter their action in a separate phase. Thus, because participants could take more time for various reasons (including taking breaks from the task) we don’t believe it is useful. Moreover, we don’t have any clear hypotheses regarding how RT would differentiate between PR and SR.*

*I don’t know what an inversion cost is and how that bears on patterns of correct responding.*

6. In Figure 1, panels are out of order (panel A appears between panels B and C).

*Fixed*

7. Evidence for forward planning (Study 3) appears stronger than evidence for backwards planning (Study 1/2). Why?

*The modal values more different than mean values. However, this could be the case because SR is a prior over PR and thus is easier to leverage. Secondly, when PR is used in the diverging study 2, it is actually not optimal, because base-rate sensitivity conflicts with optimal decision-making. We did this so that SR and PR could be differentiated. But in many real-world settings, base-rate sensitivity does not conflict with optimal decision-making (i.e., it is* ***not*** *the case in many real-world settings that high base-rate starting states are less likely to transition to desired goals).*

8. Could you clarify whether there was a control condition with equal base rates of both starting states where backward planning would not have an efficiency benefit?

*This is a necessary condition to test in further work that can look at neural correlates of backward prediction/planning. IN the current study, because behavior between PR and SR is only differentiable as a function of base-rate sensitivity, we wouldn’t be able to distinguish PR from SR in the proposed control condition. However, in future work, if we can detect backward prediction neurally, then this control condition could be tested.*   
  
9. Does "forced choice starting state" mean participants were exposed to only one state?

*Participants were shown which action to take for the starting state underneath the state.*   
  
10. Did participants learn equally well for both starting states?

*In the diverging design, yes, there’s no evidence to suggest participants learned more for the high base-rate starting state, as evidenced by the null effect in the base-rate queries.*

11. For the predictive representations check: was there choice consistency among participants, and did more backward-planning subjects drop in accuracy in those trials?

*Will check if PR-conforming subjects did worse in post map-change phase.*   
  
12. Some proofreading is needed. For instance, the Discussion section (line 365 citation; Line 354 efficient [planning]) and Methods section (Line 392 detailed [in]; Line 530 same) contain incomplete sentences. The SI: Table S1 caption also requires attention.

*Will proofread*