Comments:   
In this paper the authors seek to show that at least in some circumstances humans select "goals" based on "model-free" valuations of the goals, even though information is available for model-based assessment of options.   
  
The authors attempt to explore this question using a task that is modification of the two-step task that has be used to study model-free versus model-based choice. In the most basic version in state (stage) 1 of the task subjects are presented two out of a universe of 4 actions, and must choose one. Two of the 4 actions lead to a "blue" reward distribution 80% of the time, and two lead to the "red" distribution 80% of the time. All choices lead to the green distribution 20% of the time. The various distributions were truncated slowly varying truncated Gaussians. Note that the identical actions were never paired (e.g. if 1 had a .8 probability to transition to blue and so did 3, the they weren't paired). Subjects are told the transition probabilities, and have practice trials. There are then rewarded trails, and a subset of critical trials. The critical trials are preceded by a set-up trial where green is rewarded (with a boost to hit the boundary, depending on the sign of the  
reward first drawn randomly); in the critical the complementary pair of choices is presented (if 1,3 are "blue" and 2,4 are red, the if (1,2) is the set-up choice (3,4) is the critical choice.   
  
Basic Claim: if (for example) in a set-up trial the subject chooses 1 and green is heavily rewarded, then in the critical trial, the subject's propensity to choose 3 will be increased. Moreover the authors conceptually merge choices (for the example here) 1,3 into a latent goal "blue" and base their analysis on this idea - that subjects use model-free (cached -based) learning based on this representation of the state-action space to make choices.   
  
Major Issues.   
  
In general the use of the word "goal" here seems too strong. First, the "goal" of a subject is (presumably) to do as well as possible, so to assert that your goal is to get to the blue by choosing 1 (using the example in the paper) assumes too much - you value 1 more because of the probabilities and values, so you tend to choose it. At any rate, it seems like it is simpler to just note that it seems like subjects understand (and they should since they were told the probabilities) that (again using the example above) that 1 = 3 - they are identical, so that really the state-action space is collapsed into (1,3) and (2,4). Once this is noted, then there really is no mystery and the result is that here people do use "model-free" learning on these reduced state-action pairs.   
  
Hence the statement at the end of the discussion of the first experiment is a perplexing to me, as is the need for the second experiment. In the second experiment the state-action space is (color, 1) (color,2) etc. (shape,1), (shape,2) ... and then, as before various pairs, are statistically equivalent. As for the third task the addition of the additional letter which is identified with the number does not seem to add anything, despite the author's assertion at the end of the experiment three discussion. In other words once one posits a reduced state-action space (blue, red), again since they are identical statistically, and one asserts that the associations (A,B,C,D) -> (1,2,3,4) are deterministic, the result is not surprising.   
  
A more interesting question would be to understand in a learning process how state-action representations change.   
  
Minor issues:   
  
1. In general it was difficult to understand exactly how the task(s) went. In particular did each participant do all three tasks? Did they do 75 practice trials and 175 rewarded trials on each?   
  
2. Tables should be provided for each regression.   
  
3. In the methods section it is not at all clear what models there go with the models in the results section. This is very frustrating and needs to be clarified. Ideally equations should be provided for each regression clearly stating what it is for and what the variables are.   
  
4. The use of the full generative model to simulate data and show that the stylized facts of the real data were replicated is interesting, but why not fit the real choices with the generative model?   
  
5. In the MB aspect of the model why learn the transition probabilities when they were told (extensively)? Were these probabilities estimated over practice? Over just the rewarded trials - same question for values of the states too.