Reviewer #2: (Please distinguish between essential and non-essential revisions, as indicated in instructions at the top of this page, and please do not indicate whether or not paper should be published.)  
  
The effects of time horizon and guided choices on explore-exploit decisions in rodents  
Wang / Gerken / Wieland / Wilson / Fellous  
  
In this manuscript, the authors translate an elegant exploration/exploitation task originally developed for humans to rats. They find that rats show different effects of time horizon than humans do. While I applaud the goal of this project and both versions of the task (human and rodent) are elegant, there are too many differences between the tasks and the subjects to draw the conclusions they want to.  
  
Overview: I think this could be an absolutely excellent and groundbreaking study. But doing so requires additional experimentation and additional analyses. In the manuscript's current form the data is inadequate to draw conclusions from.  
  
Comments:  
  
- [ESSENTIAL REVISION] Sex of the subjects. The experiment was done with 6 (very small n!) of only male rats. The human studies were done with 45 psychology undergraduates (14m, 31f). Given that we know there are sex differences in both human and non-animals on exploration/exploitation choices (for example, the cited study by Chen…Grissom), it is inadequate to compare 6 male rats to a human distribution that is 70% female. (Note that I don't have a problem with using undergraduates as the human model species. It's probably an appropriate comparison to the laboratory rat.) [However, isn't it NIH policy now that all non-human animal experiments \*must\* include both male and female animals unless a scientific reason can be given for limiting it? Nevertheless, whatever the NIH policy is, the rat cohort is simply too small and limited for the conclusions.]

We thank the reviewer for raising the question about gender differences.

1. We agree that the issue of sex differences is important, but we do not propose or formulate any hypothesis about sex differences in this task. Recent published work using similar task designs to ours reported no sex differences in either directed or random exploration measures (Smith et al 2021). Also, in our own human data for Experiment 4, a Two-way ANOVA (horizon by gender) analysis showed a significant main effect for horizon (p < 0.001) and a non-significant effect of sex (p = 0.59).
2. To properly study how female rats might behave differently using our design would require an assessment of their estrous cycles, and possibly the inclusion of ovariectomized animals. This work, which is not funded by our current grants, would add 2 years of additional work. Again, acknowledging the need for more work on female animals, we leave such work for future studies.
3. The rodent portion of our study is not a NIH funded study, and is not subject to any requirement regarding sex-balance. Our emphasis is on developing and characterizing a new rodent model of the explore-exploit task compatible with a human task. Study of how populations characteristics (e.g. sex, age), while certainly important, are left for future work.

We thank the reviewer for raising the concern about our sample size.

1. Despite having small number of rats (6), we have a very large number of games from each rat. Excluding pretraining, we collected a total of 530 sessions, 67781 trials for our rat experiments. Our human study has a total of 70400 trials which is similar and comparable to that of our rats. The number of trials are matched. Through hierarchical Bayesian modeling (which pools all trials from all subjects in model fitting), we believe our human/rat comparison was carried out in a fair way.
2. For Experiment 1 and 2, we have over 30000 trials in each experiment. This number is comparable to the trial sizes used in the literature. Here we list of the number of animals/trials used in previous explore-exploit rodent studies.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Maze or Box | Number of rodents | Number of sessions | Total running time | Number of trials |
| Our study | Maze | 6 | 530  (200+ per experiment) | ~530 hours  (1-2 hours/session) | 67781 total (30000+ per experiment) |
| ﻿Beeler et al., 2010 | Box | 10  (2 groups) | 130 | Unknown | 68096 |
| ﻿Laskowski et al.,  2016 | Maze | 22  (2 groups) | 352 | ~120 hours  (20 min/session) | ~29300 |
| ﻿Chen et al., 2021 | Box | 32  (2 groups) | 256 | 512 hours  (up to 2 hours/session) | 70656 |
| ﻿Cinotti et al., 2019 | Box | 23 | 184 | Unknown | 52992 |
| ﻿Verharen et al., 2020 | Box | 60 | 196 | Unknown | ~49000 |

Although our number of subjects is 6, our number of trials is comparable to that of other rodent explore-exploit studies (especially the Laskowski paper that also uses a spatial maze design). Also note that we reported another 70000 trials in humans in our study. Since our design compares H = 1 and H = 6 within subjects, we only needed half as many rats as in a between-subject design study.

1. We are minimizing the use of animals as per IACUC rules. We did not report claims in our results based on a trend, all our main results are statistically significant.
2. We acknowledge that our Experiment 3 should be considered preliminary. However, Experiment 3 is not essential to the conclusions of this study.

- [ESSENTIAL REVISION] The rat experiment starts with a subset of cued responses. It is not clear whether the rats treated the cued responses and the uncued as part of the same "condition". I don't know how to fix this, but it needs to be considered. The authors do acknowledge that the rats may have been making some decisions based on overall volatility rather than the actual time-horizon experiment desired, but they do not disprove this hypothesis. I think it is necessary for the authors to directly test these alternative theories (using Bayesian model fits?) and show that these alternate explanations are not as good explanations as their intended experimental logic.

We do not believe the “cued” and “uncued” conditions affect the validity of our results for the following reasons.

1. There are no “uncued” conditions in our design. In the guided trials, light at one of the two feeders is on and the rat is guided to visit that feeder. In the unguided trials, both lights are on at the two feeders. Rats are pretrained to understand that they are allowed to travel to either feeder locations with lights on to obtain rewards. Blinking the light is a necessary step to draw the attention of the rat to those two feeders that are active (there are 6 other inactive feeders at the periphery). We have revised and expanded the method section to clarify this point.
2. The fact that rats adapted the percentage of switching in the first “uncued” choice based on the rewards gained in the “cued” responses, shows that rats seemingly transferred the learned value from the “cued” phase to the “uncued” phase. This shows that rats did not treat them as different conditions.

We thank the reviewer for raising about the issue of the effect of volatility on exploratory behavior.

1. Our results actually suggested that the horizon difference can not be accounted for by volatility alone. In Experiment 3, our results showed that increase volatility in the task (the random condition) increased both directed and random exploration. Since only directed exploration and not random exploration selectively changed between horizons in rats, volatility does not account for the horizon difference we observed in Experiment 2.
2. We have modified both the results and the discussion section of the paper to emphasize the above point.

- [ESSENTIAL REVISION] The rat experiments used the same 6 rats for all experiments. This is a major problem as the rats are almost certainly going to come to the subsequent experiments with learned expectations from the previous ones. The authors should, instead, identify the number of rats needed for each cohort and do a separate experiment with each group of rats naïve to the experimental paradigm, so that they all have the same training experience.

We thank the reviewer for raising the question about the carry-over effect across experiments.

1. From a training perspective, in order to teach a rat to learn Experiment 2, it is actually necessary that they perform Experiment 1 first. For all our rats, they first learned how to perform the task in a fixed horizon condition (Experiment 1), then we introduced both horizon conditions within each session (Experiment 2). Naïve rats are in our hands incapable of doing Experiment 2 without going through Experiment 1 first.
2. Any carry-over effects from Experiment 1 to Experiment 2 should affect both horizon conditions equally in Experiment 2. The horizon difference we observed in Experiment 2 cannot arise from previous Experiment 1 exposures, because we randomly assigned the 2 horizon conditions to the 2 homebases each day. This has now been indicated in the method section.
3. Experiment 3 may have some carry-over effects from Experiment 1 and 2. We do not think it is necessary to run Experiment 3 in naïve rats, however. First, Experiment 3 is not essential to the main conclusions of the paper. We acknowledged that exposure to Experiment 1 and 2 may contribute to the high threshold (more switching) in Experiment 3, but it is unlikely that the high decision noise parameter arises from any carry-over effects from Experiment 1 and 2. Since increasing volatility increases decision noise, it still holds from Experiment 3 that volatility itself does not account for the horizon differences we observed in rats. We made this clearer in the text.

- There remain some real differences between the tasks. In particular, the human rewards are not consumed in the present, and thus form an amortized goal that can only be used (can they? were the points used for anything? That was unclear) in total. In contrast, the rats are receiving a reward with direct intrinsic value (sugar water). This means that the rewards are both immediate and biologically necessary for the rats, but neither for the humans. It would be better to try to match these if possible. That being said, I am not as concerned about the physical differences between tasks. I agree that rats treat space and levers differently, as do humans. The real question is whether the authors can show somehow that the computational algorithms \*necessary\* to solve these tasks are equivalent or whether alternative computational algorithms are possible. If there are alternative computational algorithms possible, then the authors need to show that they do not match with the observed behavior.

We thank the reviewer for the question regarding the differences across tasks.

* + - 1. We agree with the reviewer that there are differences between the rat and human version of the task (e.g. juice for rats & points for humans). Despite the physical differences, we want to highlight that the underlying structure of the two tasks are identical. So computationally, any algorithm that can solve the human version of the task, can identically solve the rodent version of the task. Many researchers have worked on comparing various computational models in various explore-exploit tasks (for example, see Gershman et. al, 2018). In the human version of our task, the computational model that separates directed vs random exploration has been well validated (Wilson et al., 2014). The interest of our paper is not to compare different explore-exploit algorithms, instead, we applied the established model in Wilson et al. 2014 to estimate directed and random exploration in both humans and rats and compare how directed vs random exploration parameters differ across horizon conditions and between species. We do find the suggestion of the reviewer that different algorithms might be worthy of study. In this work, we do not have any specific hypotheses on alternative models however, so we did not explore those possibilities. They will be left for further work. We make a mention of the reviewer’s suggestion in the text.

References:

Gershman SJ. Deconstructing the human algorithms for exploration. Cognition. 2018 Apr;173:34-42. doi: 10.1016/j.cognition.2017.12.014. Epub 2017 Dec 29. PMID: 29289795; PMCID: PMC5801139.

Wilson RC, Geana A, White JM, Ludvig EA, Cohen JD. Humans use directed and random exploration to solve the explore-exploit dilemma. J Exp Psychol Gen. 2014 Dec;143(6):2074-81. doi: 10.1037/a0038199. Epub 2014 Oct 27. PMID: 25347535; PMCID: PMC5635655.

* + - 1. We agree that there is room to make the two tasks more identical in future studies. We edited the discussion to raise this issue. Since our research interest is in the change of exploration strategy with time horizon, within each species, the differences in physical implementation of the task should affect both horizon condition equally and hence should have minimal impact on the differences in horizon adaptive exploration.
      2. Human participants, receive research credits after completing the task. However, the points in the task do not convert to monetary or other forms of physical reward.

While the authors do acknowledge many of these caveats (not all) in the discussion, I think that the experiment itself is (in its current form) too damaged by these caveats for us to take their conclusions.

We have substantially revised the interpretations of the results and added new figures. Our main contributions are:

* 1. We developed a rodent task that addressed the limitation of reversal learning paradigms (See introduction of the manuscript) and allows for a separation between directed and random exploration.
  2. We showed that rats were able to use prior information to guide exploration (Figure 4C trial #1, Figure 7C).
  3. We assessed how rats explored under different time horizon. We found that unlike humans, rats decrease their decision threshold (which quantifies directed exploration) in longer horizon condition (Figure 9).
  4. We reported difference in decision thresholds between self-guided vs cue-guided exploration (Figure 11, 12).

We hope that the new figures, and re-write of the results clarify the conclusions, and make the caveats and limitations of the study more explicit.  
  
- The description of the tasks is very poorly written. I did not understand the task as described in the experimental methods on my initial readthru and was only able to make sense of it as I read through the paper and kept coming to "wait, that doesn't work unless they did…" "oh, I see, yes, OK, they did." All of the factors do seem to be in the methods, but it was very hard to understand on a first read through.

We apologize for the confusions, which likely to have subtended some of the comments above. We have added a paragraph on a high-level description of the task before diving into the details in the methods section. We also have edited multiple sections of the manuscript to clarify. we hope this has improved the readability of our methods section.

- The Bayesian analysis of the tasks is very elegant, as is the separation of directed and random exploration.

Thank you.  
  
- Figure 4 seems to show that there is no effect of horizon on the choices. I'm confused how this supports their conclusions. Figure 5 suggests that the choice is made less based on exploration than on whether the guided cue is near the boundary. (As the authors acknowledge, if the cue is near the boundary, then expected value of the other option is known to be higher [if guided is low] or lower [if guided is high].) Figure 5 seems to show that the entire effect is due to these boundary effects. Figure 7 argues that the humans are showing a subtle effect, but the rats are not [the decrease isn't significant, is it?]. Figures 8 and 9 seem to argue that there is an effect, particularly in the directed exploration parameter. How do all of these results fit together? How much of the difference is due to n(rats)=6 and n(humans)=45?

We thank the reviewer for bringing up these questions. We answer them in order below:

1. Figure 4 shows that both humans and rats have learned the task and performed well above chance, and that both humans and rats switched more on the 1st free choice compared to later trials. Figure 4 does not show measures of exploration. Despite having similar accuracy, exploration strategies can be very different. In the literature for example, Chen et al. showed that male and female rats have different learning rates and decision noises despite achieving a similar level of accuracy (Chen et al., 2021). In our task, exploration strategies were quantified using two parameters: “decision threshold” and “decision noise” using hierarchical Bayesian analysis. The differences in exploration parameters were shown in Figure 8 and 9.
2. We have clarified Figure 5. The good vs bad option in Figure 5 is relative, for example, if rats were guided to 1 drop and the unguided option offered 0 drop, then we consider the 1 drop option to be the good option, similarly if the guided option has 4 drops but the unguided option has 5 drops, then the 4 drops option is considered a bad option. So, Figure 5 does not indicate boundary effects. We have modified the manuscript accordingly to avoid the confusion.

In fact, if an agent (rat or human) acts completely based on boundary effects, then we would predict that the agent will choose identically in both horizon conditions, since a boundary effect does not change with horizon condition. The fact that we observed significant changes in decision thresholds across horizon conditions indicated that these choices cannot arise entirely from boundary effects.

Figure 7C shows how our rats use a more general form of such “boundary effects” in solving this task (It is referred to as the ‘win-stay lose-shift strategy’ in our manuscript). Rats switch more when the guided reward is low (so do humans). But this does not contradict exploration for the reasons stated in the previous paragraph. We also would like to point out that at 5 drops, rats still switch more than 20% of the times, this also can not arise from pure boundary effect.

1. We found that humans increased their decision threshold with horizon (consistent with the existing human published literature), and that rats decreased their decision threshold with horizon (novel results from our paper).
   * + - 1. In Experiment 1 (Figure 7 and 8), there is a significant main effect of horizon on the threshold parameter for humans (p < 0.001), but there is no significant effect of horizon on the threshold parameter for rats (p > 0.05). In other words, when different horizon conditions were run between sessions (there is only one horizon condition for each session), despite the numerical trend, we didn’t find a statistically significant horizon effect on directed exploration in rats. One critical difference between Experiment 4 and Experiment 1 was that horizon conditions were run within-session for humans and between sessions for rats. Third variables like training effects make it difficult to detect horizon dependent changes in exploration in Experiment 1. As a result, we conducted Experiment 2 (within design) in which rats also have both horizon conditions in the same session.
         2. In Experiment 2 (Figure 9), we found a significant main effect of horizon on the threshold parameter for rats (p < 0.001). There was also a significant main effect of horizon on the model-free p(explore) measure (p = 0.003). Bayesian analysis also shows clear separation of exploration thresholds between horizons (but not decision noise). Together, we showed a significant horizon effect on directed exploration in rats.
         3. The lack of significance in Experiment 1 could occur for many reasons. Firstly, in Experiment 1, different horizons were run in different blocks of days and sessions. Consequently, hidden variables such as the amount of exposure to training before each horizon condition (depending on the order of blocks), time of the day, weight of the rat, etc. could not be controlled and therefore could influence the horizon difference. The problem of latent variables as mentioned above was completely resolved in Experiment 2, since both horizons occurred in the same session. Secondly, experiencing both horizon conditions at once, magnifies the difference between horizon conditions. (Pilot human data in the lab also suggests that the horizon effect is weaker if different horizon conditions were blocked rather than interleaved)
2. The difference in the number of subjects between rats and humans should not affect our conclusions because we matched the number of trials and because the number of games are commensurate with that published by others (see above). Furthermore:
   * + - 1. Our main result was that rats decrease decision threshold with horizon, while humans increase decision threshold with horizon. The horizon difference was assessed separately within each species, using the same quantitative analyses.
         2. Given the difference in the nature of data acquisition for human participants and rats, we unavoidably have to deal with the fact that we have more subjects and fewer trials per person in humans, and fewer subjects but more trials per subject in rats. By using Hierarchical Bayesian analysis we consider both variances across trials (within each subject) and variances between subjects.

- I also notice that the rats are showing much more random exploration (broader sigma density 8G/H) than humans (8C/D). Given that exploration is explained by a combination of directed [theta] and random [sigma] exploration, what is the effect of these sigma differences on the theta distributions?

This is an important control, thank you. We answer the reviewer question using parameter recovery. In Supplementary Figure 1, we simulated behavior using 50 combinations of independent samples of all the parameters used in the model, for each set of simulated behavior, we fit the simulated behavior and compare the recovered model parameters with the ground truth. Our model does a near perfect job in recovering both threshold and noise parameters in the experiment. This result suggests that the variation of sigma does not affect the model estimates of theta (threshold) in our model.

