Reviewers' comments:

Reviewer #1: This manuscript by Wang et al. describes a novel rodent task for exploring explore-exploit decision making across varying "time horizons". Their approach is based on a similar task characterized humans in an earlier paper, and the current study includes further examination of human explore-exploit decisions on this task. The influence of relevant variables on task performance are described (though inferential statistics are scant), and this is supplemented with Bayesian model fitting to characterize parameters associated with two established heuristics for exploration: directed exploration (exploration threshold parameter) and random exploration (decision noise parameter). Like humans, rats performed the task well and were able to guide their choices by comparing recent information from guided trials with prior knowledge of possible outcomes. A major species difference was that humans increased their initial exploration of the unguided (unknown) option with time horizon (i.e., number of future opportunities to exploit feedback from guided and early choice trials), which represents a rational, directed exploration heuristic. Rats, instead, showed the reverse relationship, and additional data indicated that their exploratory behavior may have been driven in part by volatility in reward contingencies. Rats also appeared to show improved performance and less initial exploration when allowed to self-guided vs. being forced to sample one of the two response options. This is a really interesting study and the authors do a good job of highlighting the potential advantages and novel elements of this task versus other rodent tasks for measuring explore-exploit decision making. Despite some current issues (listed below), I believe the manuscript would be of interest to many in the field.  
  
1) The conceptual framework for the current study is not very well developed, particularly in the Introduction, which assumes quite a bit of readers. The main question at hand - i.e., how time horizon relates to explore/exploit decisions is not really discussed until late in the paper (in Discussion), which would make it difficult for readers to understand the purpose of experimental design parameters, as well as predictions about results. Also, much of the introduction lists limitations of existing reversal learning tasks but it is not very clear how the current approach will improve on this. For instance, without providing a clearer idea of the structure of the current task and how it relates to relevant decision making variables, it is not obvious why the current task does not have the same limitations. The current task also involves a 2-option choice with differential reward, so why is it better than reversal learning?

* We will add a paragraph on the introduction of time horizon in the introduction
* The current design is better than probabilistic reversal learning (most commonly used in rodents) in several ways:
  + In exploration, both good and bad outcomes should occur. However, in reversal learning, after the reversal point, “exploring” the previously suboptimal option will always lead to a better outcome compared to the currently bad option. But in our design, exploring the unguided option can lead to both better or worse outcomes.
  + In order to study directed exploration, there needs to be a difference in uncertainty in the two options. For reversal learning, this uncertainty difference is implicit in that the less chosen option has more uncertainty, since the less chosen option in reversal learning also has a lower estimated value, the reward value and level of uncertainty are not independent. However in our design, the uncertainty is created by the design that the rat has visited one feeder but not the other, and that this uncertainty bias is independent of the value of the exploit option. Because of this dissociation between uncertainty and reward, we are able to quantify directed exploration in our design (but not easily in reversal learning).
  + In most reversal learning paradigms, the rewards are probabilistic. Value estimation of the exploratory option can look like exploiting in that they both involve the animal picking the same option several times in a row. Exploration happens in both stay and switch trials in reversal learning. However, since our rewards are deterministic, exploration almost exclusively happens in the switch trials. (this point may be problematic, I don’t know if we have a better way of saying this, since random exploration can make both stay/switch trials to be exploratory).

2) There are very few statistical results provided. Instead, the Results consists of general descriptions of group level performance without evidence of the significance of findings. This is generally problematic but especially for subtle effects (e.g., constant vs. random reward differences in Fig 10; horizon effects on model parameter estimates in Figure 8; reference to horizon effect on switch in Fig 4D on p. 14).

Our main results have significance of findings:

1. Rats and humans both use prior information in guiding exploration (p < 0.001)
2. Rats and humans switched more when the guided reward is low (p < 0.001)
3. Directed exploration increases with horizon in humans (significant + previous literature support). Directed exploration decreases with horizon in rats (overall significant main effect, p = 0.007. Significant in the Bayesian way only when nGuided = 0, when nGuided = 1/3, the Bayesian results are in the right direction but non-significant)
4. Free vs guided (significant in the Bayesian way… I can compute the traditional statistics, I am pretty sure this is significant)

For constant vs random reward difference, it’s not one of the main results, and is significant.

I will compute maybe a Bayes factor to show significance of the posterior fits.

3) Some additional methodological details should be provided or clarified. What volume were the sugar water drops? Were the rats food/water restricted? How many trials/games per session? Were rats trained through different phases at different rates based on performance? Were rewards ever symmetrical across options within games and if so how was this dealt with for analysis?

1. @Jean-Marc Do we have this info? If not, @Julia, can you measure the volume of sugar water per drop? I think I did this very early in the training.
2. The rats were food restricted but not water restricted.
3. I can compute this. Roughly 100 trials per session, that’s about 15 games
4. Rats were trained at different rates based on performance.
5. Yes, the number of symmetrical trials were too low and were ignored in the analysis.

4) For Experiment 2, it was unclear exactly how H conditions were organized within session, though it seems to be the case that they were strictly alternated when confounded with home base. This should be in Methods. Also for this experiment, how were nG 0, 1, and 3 conditions organized (e.g., blocks of sessions, randomly across sessions).

1. Horizons were alternated (each horizon is tied to a home base within session). But the two home bases can be linked to different horizons across sessions.
2. nG = 0,1,3 were done in blocks of sessions.

5) For the human task, does the schematic in Figure 3 represent actual task stimuli and procedures? For example, were subjects given a tally of past reward histories for all trials within a game? This should be indicated in Methods.

The schematic in Figure 3 represents the actual task stimuli. Subjects were given a tally of past reward histories for all trials.

6) Do the data presented in figures represent all the data from all relevant sessions or were they restricted to sessions after rats had time to learn about the new task contingencies? For example, rats presumably took some time to learn about the change in time horizon across blocks of sessions in Experiment 1. And the same goes for when they switched to the within-session analysis of horizon in Experiment 2, and the random task in Experiment 3. As noted by the authors, performance in the random task shows some clear carryover from the earlier phases of testing. This should be specified as good practice but also raises questions about species differences.

The data presented represent all the data from all relevant sessions. I will repeat our analysis removing these “learning” blocks and show the results here.

7) Humans seem to have little trouble deploying a directed exploration strategy based on time horizon and guided choice feedback (reward size). They seem to explore the unguided option during the first free choice if there is any question about what the best option is, particularly when there is a long horizon to exploit that information. The authors state (p. 10) while showing similar early exploration, "it took longer for rats to switch back," referring to their persistent switching behavior in Fig 4 and 5. But these data are really able to get at this question precisely because they don't describe whether the switch is moving away or toward the guided choice. Given the rats' generally poorer performance and persistent tendency to switch within games, even with long horizons, suggests that they were switching back and forth from the best option. This is later discussed in the context of the random reward task, but the authors should avoid giving the wrong impression when discussion Fig 4 and 5.

I can repeat the switch analysis for moving away vs toward the guided choice.

8) The cross species comparisons are a bit strained despite the general similarities across tasks. For instance, humans appear to receive a continuous tally of past reward within each game, which explains their lack of later exploration once they understand the basic task. But rats could forget this information and decide to re-explore the options, particularly in games with long horizons though reward volatility on short horizon games may provide a separate reason to explore.

It’s true that there are differences in this task. Since the task is so easy for the humans, we doubt that by removing the history would change human performance.

The reviewer is right that rats could re-explore because of memory issues. In the nGuided = 3 condition, the rat does visit the same location 3 times before making a free choice. This repeated visit of the guided option helps with the forgetting issue.

Nevertheless, one of the main findings is that rats can adapt their exploration strategies to the horizon context, but does so in an opposite dependence to humans. This horizon comparison is done within species (the difference in task does not contribute to this), and we showed a qualitative difference of this comparison between species.

9) The authors suggest that rats may be performing the task to satisfice instead of optimize, like their human counterpart. This is related (p. 17) to the higher tendency for humans to explore than rats. This conclusion is partly based on the idea that rats must exert more effort and may be less willing to explore (a point also made earlier in the manuscript). But of course this specific data refers to first trial exploration. Rats were not less willing to explore on long horizon games. As indicated above, they tended to switch back and forth throughout the game (i.e., even when they had sampled both reward contingencies).

We should probably emphasize on the utility of 4 and 5 drops. To satisfice, the rat has a utility function that 3 and 5 drops are similar, whereas humans have a utility function that 5 is much better than 3. As a result, humans have higher threshold for exploiting (correspond to a higher tendency to explore on trial 1). The later switching in rats is explained by high decision noise instead of this threshold in trial 1. A higher decision noise can also arise from “to satisfice” as opposed to “to optimize”.

I can check: switch as a function of last reward in later trials… It may be flat.

10) The greater preference for high reward option and lower level of switching in the free choice vs. guided choice is interesting but one account not considered is that rats some feeder/spatial preferences that bias them on these free choice first trials that continuous to bias their performance to the same degree within each game. This make sense generally and also explains why the effect of free vs. guided choice is so stable across variables like guided reward size and horizon. This could be a long-term feeder bias or perhaps something more dynamic (e.g., like a preference for whichever feeder has been paying off better in recent sessions).

Interesting point.

I can check our data on this. It’s complicated to check the influence of feeder preferences from previous sessions, before having to go there, I can first check if the rat consistently prefer one feeder within a session on the 1st trial in the nGuided = 0 sessions (if not, then this tells us the repeating behavior is not driven by long-term general preference for that feeder).

11) On a related note, a spatial bias parameter was computed in the model fitting but was not discussed.

Actually, I ended up fitting the model without the bias term. The bias will show up in the noise term. I forgot to edit the formula in the paper.   
Minor issues:  
12) P. 10 - statement about boredom or motor error seems to refer to residual responding at last trial but worded as if about the reason for the decrease in switching.

We can re-phrase this.  
13) P. 11 - it says "3 or 4 drops for humans"

???  
  
  
  
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.]

For sample size,

1. We are minimizing the use of animals as per IACUC rules. We do not need more rats for our conclusions (i.e. we do not claim a result based on a trend… all our results are statistically significant…). We could think of some statistical test, showing that the results (or a subset of the results) does not depend on the games of a single rat (delete x% of games from rat 1, no difference. Repeat for each rat à no single rat drives the results).

I will work on this part, basically the same response to reviewer comment 2 can go here.

1. We have a very large quantity of games from each rat. 500 sessions total, with 65000 trials total for rat experiments (plus 70000 trials for humans). For Experiment 1 and 2, we have 30000+ trials in each rat experiment, this is comparable to our number of trials in human experiments 4 and 5.

Here is a list of the number of animals/trials used in previous explore-exploit rodent studies. Ours is at a similar magnitude of trials.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Box/maze | Number of rat/mouse | sessions | time | Number of trials |
| ﻿Beeler, Daw, Frazier, & Zhuang, 2010 | In cage lever | 10? | 10+3 |  | ? |
| ﻿Laskowski et al.,  2016 | Maze/3-arm | 9 lesion group, 13 control | 16 session per rat, ~80 trials per session | 20 min/session | 33280 trials total |
| ﻿Chen, Knep, Han, Ebitz,  & Grissom, 2021 | Box | 16+16 | 8 session, 276 trial | 2 hours/session | 70656 trials |
| ﻿Cinotti et al., 2019 | Box/3-arm | 23 rats | 8 session, 12 blocks, 24 trials |  | 52992 trials  (for testing sessions) |
| ﻿Verharen, den Ouden, Adan,  & Vanderschuren, 2020 |  | 60 rats | 196 sessions total x ~250 trials per session |  | ~50000 trials |

For gender differences,

1) this is not a NIH study,

2) we agree that the issue is important, but there is no reasons to think that there should be sex differences in this task (not our hypothesis)

I can show that in our 50 participants, we don’t see a gender difference.

No gender differences were found in either directed or random exploration measures in a recent study using a similar task to ours. (Smith et al 2021)

3) this would add 6 more rats (2 years of work). We will leave that for later study…etc.

- [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.

1. The fact that rats adapted the percent of switching in the first “uncued” choice based on rewards gained in the “cued” responses, means that rats carried the learned value from the “cued” phase to the “uncued” phase. This suggests that rats did not treat them as irrelevant conditions.
2. I think our main contribution was to:
   1. Design this task that can properly assess directed exploration in rats
   2. Show that rats do adapt their behavior to horizon, but in a different way to humans

Volatility and time horizon are related constructs, long time horizon is correlated with low volatility in the environment. In this paper, it’s not our intention to disprove that rats do not make decisions based on total volatility. Instead, we want to show that rats used a “lazy” strategy to adapt exploration to the time horizon.

Jean-Marc, I need to think a bit more about this and about how to respond. There may be some analysis we can do on this point.

~~there are two ways to deal with them:~~

~~Explore more in a low volatility/long horizon environment. This is what optimal agents or humans do. It’s more beneficial to explore more as there is more time to use the information gained through exploration.~~

~~Explore more in a high volatility/short horizon environment. This is what rats do. With limited computing resources/control, then they should focus on get long horizon right, and loosely speaking can just guess on short horizons.~~

- [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.

1. Experiment 1 can be viewed as pre-training for experiment 2. The only differences between Experiment 1 and 2 is whether both horizon conditions occur within the same session or between sessions. I don’t think the learned expectation from previous ones will bias our conclusions in any way.
2. Experiment 1 can be viewed as a pre-training for experiment 2. In order to train our rats to do both horizon conditions in the same session (Experiment 2), we think the most natural way to train them is to do a single horizon condition (Experiment 1) before doing both horizon conditions. Naïve rats are unlikely to be able to do Experiment 2 without going through a phase of Experiment 1.
3. Experiment 3 is a pilot study in this study and is not essential to our conclusions, and we view all the results from Experiment 3 as preliminary. We also acknowledged the carry-over effects from Experiment 1 and 2 as limitations for Experiment 3.

- 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.

The essential difference we observe is that humans increase directed exploration with horizon, whereas rats decrease directed exploration with horizon. The reviewer has concern that this difference arises from differences between the two tasks as opposed to between species.

1. There is no perfect way of matching the way rewards are consumed in humans vs rats. We cannot ask rats to do the task without immediate reward, and it’s not a simple thing to find an equivalent of water to give to humans. We do not have any hypothesis how this particular mismatch of rewards can influence the horizon dependence of exploration.
2. Rats showed this irrational dependence of horizon. I am not sure we can perfectly explain why rats decrease exploration with horizon. We know that rats and humans use different algorithms, but it’s not an easy thing to identify what rats did. I can try to test this… do a simulation to illustrate the following two algorithms:
3. The optimal human one
4. Have a term that adds a cost to compute, so that “like rats”, it’s more beneficial for the algorithm to focus on getting the long horizon games correct and give up (pick randomly) on the short horizon games.

~~We know from previous literature that humans (with both paid/unpaid studies) explore more with horizon. In our experiment, we just confirmed that by using only 1-5 drops and deterministic rewards, we are still able to get the desired behavior in humans and we did.~~

~~The contrast between species is not done directly by comparing measures in the task across species. However, the contrast is done by performing a horizon comparison within species, and then compare the qualitative difference across species.~~

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

1. Figure 4 does not measure exploration, and simply shows that rats are able to learn the best option significantly above chance. As in Chen et al 2021, exploration strategies can vary even though the accuracy remains the same.
2. Figure 5 does not indicate the boundary effect. The legend low and high refers to whether the guided option has a relative low/high value compared to the unguided option. So if the guided value is 2, and the unguided value is 3, then in this case the rat is guided to the “low” option.
3. Figure 7 and 8 are for the between-session experiment. Figure 7 is behavior and Figure 8 is model fits. We showed a qualitative increase of threshold in humans and decrease of threshold in rats. But the decrease in rats is not significant. But here, the human experiment is within session and the rat experiment is between session. To maximize the contrast between horizons, we performed Experiment 2 (the within-session experiment for rats).
4. In the within-session experiment (Figure 9), rats show a significant decrease in directed exploration.
5. This difference should not arise from the number of subjects. Given the difference in data acquisition between species, we have to deal with that we have more subjects in humans, and more trials per subject in rats. Our analytic choice hierarchical Bayesian analysis accounts for this difference in the number of subjects. Hierarchical Bayesian analysis considers both variances across trials within a subject and variations of subjects within a species.   
     
   - 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 can be answered by parameter recovery. I will run this analysis and show that sigma does not influence theta.