Reviewers' Comments:    
  
Reviewer #1:  
Remarks to the Author:  
Review of "The nature of decision noise in random exploration"  
  
\* Summary  
This manuscript reports on a study in which participants completed a variable number of two-armed (or two one-armed) bandit tasks, each preceded by 4 forced choice trials designed to provide equal or unequal information about the rewards of the arms, and with either a short (1 trial) or long (6 trial) horizon. The task has been previously used by Wilson et at (2014), but in this version, the same game (including the results of the forced choice trials) was repeated twice. Results replicate earlier evidence for both random and directed exploration in the task. The main focus is on consistency in the choices between the two repetitions of each game. The authors argue that if noise affecting decisions is external (stimulus-driven) then this noise, and hence also the resulting decisions, should be identical between the repeated games. If noise is internal, however, it should be independently drawn for each repetition, and hence the resulting decisions can be different between the  
repetitions. Results show a reasonable level of inconsistency between repeated games, providing evidence for internal noise. A model-based analysis relying on hierarchical Bayesian estimation shows that, while both internal and external noise are present, the estimated level of internal noise is substantially larger than the level of external noise. Both forms of noise increase with longer horizons.  
  
\* General evaluation  
There are many positives to this manuscript. It is interesting, clear, and well-written. The topic of internal/external noise in decisions will be of interest to many scientists working in the fields of (reinforcement) learning and decision-making. Including repeated games is a clever variation on the Horizon Task used previously. The analyses are appropriate and appear skilfully done. Also, including a parameter recovery study is exemplary. My main concerns regard the interpretation of the results. In particular, while simple, I'm not convinced the operationalisation of external noise captures the implied meaning of "external noise". As currently defined by the authors, I'm not surprised that they find evidence that internal noise is much larger than external noise. I'm almost more surprised they find any evidence for external noise.  
  
\* Major issues  
  
As acknowledged by the authors: "Perhaps the main limitation of this work is in the interpretation of the different types of noise as being internal and external". While I understand the rationale for defining one noise term as constant over all repetitions of a game, and the other to be variable and independent over all repetitions of a game, I'm not convinced this distinction necessarily reflects a distinction between "internal" and "external" noise.  
  
1. What is "external noise"?  
In areas such as perceptual decision making, there is a reasonably clear distinction (at least theoretically) between perceptual sources of noise and other, later, forms of noise, such as decision noise. Perceptual noise can be increased by changing aspects of the stimuli. Being able to manipulate different levels of external noise in such a way would have considerably strengthened the present experiment. For instance, the authors might have manipulated perceptual noise by varying the presentation time of the information, or decreasing the contrast to make it more difficult to read. While such manipulations might increase noise levels caused by external sources (the stimuli), I'm not sure this should be classified as purely "external" noise, as it is also clearly affected by the internal perceptual system. Moreover, I doubt that many people would believe perceptual noise to be identical over repeated presentations of the same stimulus. If the same stimulus is thought to produce  
exactly the same noise at each presentation, to what extent is this even "noise" instead of some form of consistent bias? In addition, while if internal noise is (partly) due to errors in computation (as suggested in the Discussion), it is plausible that this would increase with horizon, why would external noise also increase with horizon? What aspects of the stimulus induce more noise for longer horizons? In summary, I think the authors need to be clearer about what they believe constitutes external noise in tasks such as the Horizon Task, and why this would increase with horizon. At present, I think the authors have reasonably convincingly shown that "noise" which varies over repetitions is larger than noise which is identical over repetitions. But this is not the same as showing that internally generated noise is larger than externally generated noise. Also, I think the results pertain to the task at present, but in other tasks, where stimuli are not represented purely in terms  
of generated rewards, "external" noise might be more substantial.   
  
2. Inconsistency in choices in the absence of noise?  
I assume that in addition to the forced choices and the associated rewards, the position of the options was also held constant between repetitions (this is not obvious from the manuscript). While as such, the repeated games should perhaps look the same, there are many reasons (besides "internal noise") why they might not look the same. One thing is that during the experiment, participants may learn about the "meta-characteristics" of the games (e.g., that one mean is always 40 or 60, the level of reward variability around the mean, etc.). Or participants might even have learned that in an earlier repetition of the same game, things didn't turn out as expected, making them change their behaviour the next time round. If so, than the "information-state" of a participant is not the same between repetitions of the same game, which could result in inconsistent decisions in the absence of any form of noise, internal or external. As a thought-experiment, if the experiment would have  
consisted only of repetitions of the same game, would you expect the decision to be the same in all these games, even if there were no noise whatsoever? As such, I think it is important to rule out such between-game learning. I believe this was done in the 2014 paper, but I think it should be done here again.   
  
3. Issues with identifying external noise in a model-free analysis  
In the model-free analysis, the level of external noise is not estimable and the prediction under pure external noise is just p(inconsistent) = 0. Clearly, there are other reasons why p(inconsistent) could be 0, such as participants following a deterministic decision strategy. So the absence of inconsistency cannot be taken as a sign of external noise, while any inconsistency is as sign of internal noise. In some sense, I think this stacks the analysis against external noise. As the observed inconsistency falls between that predicted by pure internal and pure external noise, the authors conclude that "both external and internal noise are present in driving this choice inconsistency". But as any inconsistency cannot be explained by external noise (according to the operationalisation), I think that conclusion is itself inconsistent with the operationalisation, unless the view is taken that on some trials, decisions are driven by external noise, and on some trials, internal noise  
(i.e., the observed proportions of inconsistency are a mixture of consistent and inconsistent trials).  
  
4. Definition of noise in the model-based analysis  
In the "model-based" analysis, it is somewhat clearer how the two types of noise are defined: again as terms which are either identical between repetitions of a game, or independent (are likely different) between repetitions of a game, but in addition, both are deviations from model predictions defined by differences between the options in mean reward, information state, and position bias. Given that these "fixed effects" in the model capture all consistent influences on people's decisions, it would be plausible to call the "random effects" (the noise terms) noise. But this rests upon the model indeed capturing all consistent aspects of choices, and while the model is plausible, there is no convincing evidence that this is the case. There is no indication how well the model fits in an absolute sense. Such measures are generally difficult to obtain, but it is important to provide some further evidence for model fit. For instance, one thing that could be done is posterior-predictive  
checks. Another additional analysis to strengthen the results would be to repeat the model analysis, but instead of using an actual repetition of a game, pick a random other game. Then there should be no consistent external noise, and hence the analysis should estimate the variance of external noise at 0, whilst the variance of internal noise should be the same as in the reported analysis.  
  
I believe the model put forward in Equation 2 is identical to a model with a single noise term, but where this single noise term is drawn for each repetition of a game from a bivariate distribution that allows those draws to be correlated (e.g., drawing both noise terms from a bivariate Gaussian distribution, in which case a Probit regression-type model would be better than a logistic regression-type model). In the present formulation, the external noise term is drawn once for both repetitions (which is equivalent to drawing two terms with a correlation of 1) and the internal noise term is drawn twice and independently for both repetitions (identical to drawing two noise terms with a correlation of 0). As the noise terms are additive, this is the same as drawing two noise terms from a correlated bivariate distribution. While I'm not suggesting this is a better model (as the models are statistically equivalent), it puts the focus more directly on the correlation in noise terms (is  
the correlation 0, 1, or somewhere in between?) which potentially might be more reliable to estimate than the two separate variance terms for internal and external noise. In addition, it is less suggestive of two sources of noise than the current formulation, which again points to my general issues with the current interpretation in terms of two separate types of noise.  
  
5. Bias in parameter estimates?  
The results of the parameter recovery study indicate that the correlation between simulated and estimated parameters is generally reasonably high. However, in Figure 6, it does look like there is some bias in estimates which such correlations can hide. For instance, the bias term appears to be estimated on average as smaller than the true (simulated) value. More importantly, it looks like there is a similar bias for the variance of external noise, while this bias appears absent for the variance of internal noise. This could partly explain why the results show that external noise is smaller than internal noise. To see more clearly whether there is bias in the estimates, it would be useful to provide histograms for the difference between estimated and true values for all parameters. If there is evidence of estimation bias, then either this should be removed by changing aspects of the model, or it should at least be explicitly noted how this affects the results.  
  
\* Minor issues  
  
- p.9 "to come from logistic distribution with mean 0." -> "to come from a logistic distribution with mean 0."  
- p.10 "n\_ext denotes the external, external noise, which..." -> "n\_ext denotes the external noise, which..."   
- p.13 "The number of games participants played depended on how well they performed, which acted as the primary incentive for performing the task." How was performance determined and how did it affect number of games?  
- p.15 In hierarchical models, there is usually quite a strong correlation between MCMC samples obtained with e.g. Gibbs sampling, resulting in slow mixing. For these models, Hamiltonian Monte Carlo as implemented in Stan can provide better results. If there is large correlation between samples, it would be good to run the chains for longer than 4000 iterations (whether combined with larger thinning or not).  
- p.15 "Convergence of the Markov chains was confirmed post hoc by eye." While visual inspection is generally a good way to assess convergence, it would be useful to supplement this with more objective measures such as the Gelman-Rubin convergence diagnostic, etc.  
  
Signed,  
Maarten Speekenbrink  
  
  
  
Reviewer #2:  
Remarks to the Author:  
The present paper dissociates the contributions of internal and external noise to random exploration in decision making under uncertainty.   
  
In its present form, the work and manuscript are missing crucial information.   
  
First, according to the authors, the “ ‘repeated games’ are the key manipulation”. Yet, the only information about this key manipulation is that the two games were “separated by several minutes in time so as to avoid detection”. How did the authors decide when to present the game again? Was there randomization? If the number of trials depended on subject performance, where some games not repeated? Is the distance between the two games fixed? Did the authors verify that detection was avoided?  
  
Second, there is no clear definition (beyond a statistical one) of what constitutes external and internal noise in the present experiment. From what I understand, there is an assumption that the external noise is constant based on the first and second presentation of the stimulus being identical. On the one hand, it’s somewhat of a post-hoc definition. On the other hand, the authors provide no evidence for the claim of identical external noise across presentations.   
  
Third, if that’s the definition (consistent vs. inconsistent), then why this task as opposed to another (maybe simpler or more established) decision making task under uncertainty with repeated trials?   
  
More specifically:  
What is “external noise” in this specific task?   
Can the authors exclude that external noise/factors of the first presentation affected external noise/factors of the second presentation (e.g, through learning)?  
  
For instance, if I apply the authors’ definition and data analysis to the example of someone walking into a restaurant, then all I see is that the person walks into the restaurant the first time but not the second time. They are inconsistent. What I don’t necessarily know is whether there was an external factor or not. For instance, the first time, the person saw a friend, followed them but the food turned out to be bad. So the next time, they don’t follow the friend. The person’s choice was inconsistent although the situation was identical and they followed an external noise both times. The problem is that I wasn’t aware of the additional information available to the person I observed, nor what impact it had on the person.  
  
One question the author’s could ask the present data is whether (in)consistency is somehow linked to the reward received on the first occurrence of the trial.   
Another option would be an explicit manipulation of external noise to its impact on choice.   
I think some of the answer to this might be in figure 4 but the authors address this only very briefly. It might help to figure out what the question is that figure 4 answers.  
  
The quality of the data and data analysis is very high.  
(Note that I am somewhat familiar with model-fitting using MCMC but not an expert (i.e., have never implemented it myself) so I wouldn’t be able to spot anything but obvious errors.)  
  
The authors point out in the discussion that it is difficult to interpret the external and internal noise but do not offer much to resolve this. Again, I think some of the answers are in their data and the argument about why the present results constitute an upper and lower limit on the noise estimates would benefit from being clarified and expanded.   
  
Minor points:   
  
- In the methods section, it says that this is a “modified version of the Horizon task”. Please state clearly what the modification is. Is it only that trials are presented twice or where other aspects altered?  
- How was the number of games affected by performance?  
- The formula on page 8 is presented without further explanation and the notation is ambiguous. If I understand correctly, what the authors call “theoretical values” of p(consistency) is the expectation (or expected value) of the probability of consistency. Even with this in mind, it’s a bit of a leap to go straight to the squared probabilities. Please clarify.  
- The following two statements seem inconsistent: In Figure 1, it says “Overall participants play 160 such games”. In the methods (page 13) it says “On average, participants played 151.6 games …” Please clarify.  
- I would refrain from referring to participants as “badly performing subjects”. I recommend choosing a more neutral phrase such as “participants whose performance was below a pre-defined threshold”. After all, participants are volunteering their time for science and our goal is to study, not to judge their performance.  
- “internal noise increases dramatically” is a subjective evaluation unless the authors provide a (measurable) definition of a dramatic increase :-)   
- Figure 1. Explain in caption that green box indicates choice option(s)  
  
  
  
Reviewer #3:  
Remarks to the Author:  
The authors analyze data from a 2-armed bandit task, using a model that decomposes random exploration into internal and external noise. Internal noise is described as arising from inherent stochastic processes in the brain and defined in the model as the level of self-inconsistency when presented with the same data, whereas external noise is defined as stimulus driven and is modeled jointly for all repeat versions of each game. Thus, there is a noise parameter that is consistent for all repeats of the same setting (external), and a noise parameter that is independent (internal). The authors estimate non-zero values for both parameters, and show that internal (independent) noise has a greater contribution than the jointly defined noise (external).   
  
To the credit of the authors, the manuscript is clearly written and they make an effort to assess the robustness of their models by performing parameter recovery. However, there are considerable theoretical and methodological concerns that I address below.  
  
First of all, I have problems with the theoretical setup of random noise being decomposed into internal and external noise. The author’s themselves highlight some of the limitations in the discussion (i.e., that the experiment is not perfectly suited for disentangling these two notions), but they also explicitly make the point that their results suggest that “random exploration is driven by intrinsic variability in the brain” (pg. 12). If this were the case, then why are there differences in random exploration as a function of horizon, including a change in both internal and external noise? Rather, the data suggests that random exploration is modulated by the strategic goal of acquiring information and not just intrinsic variability.   
  
The authors concede that due to limitations of the experimental design, their estimate of external noise may be a lower bound and their estimate of internal noise may be an upper bound. One way to take these limitations seriously, would be to attempt a counterfactual parameter recovery, where internal and external noise parameter values are swapped, and then seeing if they still recover (see Schulz et al., 2018 for an example of this procedure). This is to say that if external noise was in fact lower and internal noise was higher, could the model identify it?   
  
Secondly, and maybe most important, I find that this paper represents an example of a well-meaning approach to cognitive modelling, but one that doesn’t actually teach us anything about cognition. Another interpretation of the results is that that an independently estimated noise term (internal noise) has a bigger contribution than a jointly estimated noise term (external noise). Isn’t this just a feature of statistics? You could look at any dataset with some level of randomness, and the jointly estimated noise is always going to be a smaller component than independently estimated noise. The connections between internal and external noise to brain variability and adaptive decision noise are greatly exaggerated and I think rather unwarranted here, when much simpler explanations are also available.   
  
Additionally, the overall goodness of fit for the model is not discussed, but the authors could at least compare to some standard reinforcement learning models, which learn the reward distributions of each option (e.g,. Rescorla-wagner, q-learning, kalman filter, etc…). I also suspect that if one were to replace the simplistic information value component of the model (A \delta I) with the posterior uncertainty of a RL model, the influence of both noise components would be reduced. That is to say, the overall variance in the draws from each arm probably influences choice in a more systematic way than just the {+1,-1,0} values used in the model (see Gershman, 2018).   
  
This brings me to my last point, which is that substantial relevant literature is ignored, to the detriment of the theoretical position being advanced by this paper. A recent paper by Gershman (2018) looks at the contributions of relative and absolute uncertainty in a 2-armed bandit task (i.e., difference in uncertainty between the two arms and total uncertainty of the arms) and how they influence the trade-off between random exploration (proceduralized using thompson sampling) and directed exploration (using upper confidence bound sampling). This seems to be the predominant model in the current literature, and has a very high level of overlap with this current manuscript. The authors should be aware of this literature, and should consider comparing it to their own model.   
  
There is also the recent Findling et al. (2018) manuscript, which provides an exemplary account of how random exploration can be explained by noise during the learning process. Indeed, the bandit task described in the manuscript is also a learning task, where noisy samples from the mean of each arm require the decision-maker to adaptively learn the future expected payoffs of each arm. A quick comparison between the Findling et al. (2018) manuscript and this one, reveals a large difference in scope and contributions, but with a similar level of conviction in the conclusions.   
  
  
Other issues:  
-No effect sizes are provided despite the authors indicating so in the reporting summary.   
-Error bars in figures 2 and 3 are not explained  
-Stars in figure 2 are not explained  
-The degrees of freedom of the ANOVA on page 7 seems to be wrong, if N=66 participants and k = 2 samples, then df2 = 66-2 = 64 as opposed to 196  
-How many data points were modelled for each participant?  
-How good of a fit is your model?   
  
References:  
  
Findling, C., Skvortsova, V., Dromnelle, R., Palminteri, S., & Wyart, V. (2018). Computational noise in reward-guided learning drives behavioral variability in volatile environments. bioRxiv, preprint. Retrieved from <https://www.biorxiv.org/content/early/2018/10/11/439885.full.pdf+html>  
  
Gershman, S. J. (2018). Deconstructing the human algorithms for exploration. Cognition, 173, 34-42. <https://doi.org/10.1016/j.cognition.2017.12.014>  
  
Schulz, E., Wu, C. M., Ruggeri, A., & Meder, B. (2018). Searching for rewards like a child means less generalization and more directed exploration. bioRxiv preprint. Retrieved from <https://www.biorxiv.org/content/biorxiv/early/2018/09/13/327593.full.pdf>