Reviewer #1: In this work, the authors assessed the impact of stress on reinforcement  learning using a probabilistic reversal learning task, computational modelling and fMRI data.  
*A psychosocial stressor was used. The authors argue that the stressor slightly increased task performance. Model comparison tentatively suggested an effect of the stressor on the inverse temperature parameter, although parameter analyses did not show differences between stress and control conditions and the direction of the putative effect remains elusive. At the neural the level, the stressor might have increased the BOLD activity parametrically modulated by prediction error (PEs) signals during win trials. However, no interaction analyses with losses were conducted to assess if this effect was indeed specific to wins.*  
Although the combination of computational modelling with fMRI is valuable, I found that the analyses/data are difficult to reconcile across the different levels of analyses and can only provide tentative support for the authors' conclusions.  
  
For transparency and openness, I reviewed this manuscript elsewhere and, although I appreciate that the fMRI analyses have been slightly extended by looking separately at prediction errors at the time of wins and losses, the majority of my concerns remains unresolved. Below, I highlight my major concerns.  
  
1)      Data interpretation  
\*       I find difficult to reconcile the findings from the computational model with the behavioural and neural data. It is unclear whether and how the inverse temperature parameter relates with behavioural performance and neural data. Computational modelling should help to inform the behavioural/neural findings. However,  the integration of the findings across these levels of analyses is very poor, and as such it does not help resolving the question.  
\*       Control analyses should be provided to check the robustness of the findings and support the conclusions. Additionally, the authors refer to interindividual variability several times, but more analyses should be provided to support their arguments  (e.g., cortisol/stress levels, choice performance, individual model fits, etc., could be correlated with the BOLD activations)  
\*   Based on the the power analysis in the SI, the study is underpowered and this should be acknowledged and discussed.  
  
2) Stress induction  
It is questionable whether the stress manipulation worked as an effective and long-lasting acute stress induction inside the scanner. More detailed analyses should be conducted to assess the effectiveness of the stress induction and how it affects the reported brain activations (e.g., cortisol level/subjective stress ratings could be used as regressors)  
  
3) Computational modelling  
\*       It is unclear what does the inverse temperature parameter accounts for in terms of behaviour/brain. Further analyses and discussion are needed elucidate this.  
\*      The present results do not allow conclusions about how stress affects choice stochasticity. The vague conclusion that stress somehow might affect this parameter (without knowing the direction of the effect) is not enough and adds confusion.  
\*       Given that one of the main conclusions is based on model selection results, further checks, such as simulation of behavioural data and parameter recovery analyses, should be conducted to validate the model. Also, given the inter-individual variability I would like to see how well each model fits individual behaviour. Perhaps, fMRI parametric modulators could be built based on the PEs from the best-individual-fitting model to better capture individual learning processes and the differential impact of stress upon them.  
\*       The computational models should be properly introduced and justified throughout the text. The readership of the Journal may not necessarily have a strong computational background. More details about reinforcement learning mechanisms, computational models and the how their parameters can capture specific cognitive/neural mechanisms of interest should be provided.  
  
4)      fMRI  
\*       The onsets of stimuli (parametrically modulated by Q-values) should be modelled in the GLM to account for the shared variance between BOLD signals at the time of the stimuli and feedback. Alternatively, the justification for not including this regressor and control analyses should be provided.  
\*      Based on model selection, the authors conclude that stress might affect the use of choice values. Given this interpretation, which I don't entirely agree with , I wonder why the fMRI analyses are only focused on PEs, and not on on choice values (Q-values in the model)  
\*      It is critital to test the interaction between condition (stress/control) and valence of the trials(win/loss). The contrast estimates for PEs during loss trials in the two conditions should also be plotted.  
\*       Overall, I find that all analyses, but particularly fMRI analyses, need to be better justified by including more theoretical or model-driven background. For example, it is still not entirely clear why the authors first collapse PEs across valences and then conduct a second analysis with separate valences.  
  
  
Minor concerns:  
\*       For completeness, please report the BOLD activations parametrically modulated by PEs in wins and losses, and their contrasts (win>loss; loss>win).  
\*      Please report the number of voxels in the clusters that survive FWE correction. Figure 6 is confusing; perhaps keep only the subfigures with the variable stress. Please double-check that the coordinates for the peak voxel in the legend are correct, because they do not match the ones in the text/ figure. The same applies for the  designation of the brain area (left striatum in the text, and ventral striatum in the legend)  
\*       The contrast estimates for PEs in win trials are near zero in the control condition, which I find a bit odd. Overall the values on the y axis also seem much larger than typical, so I suggest to double-check this. Connected individual dots to assess the intra and inter-individual variability of data (similar to the behavioural plots) should be displayed in Figure 6.  Additionally, please specify if the parametric modulator for PEs was built based on the PEs estimates by the selected computational model or by any other model.  
\*      The numbers in the Figures in the SI do not correspond to numbers in the manuscript.

Reviewer #2: *In this paper, Wieland and colleagues investigate the influence of acute stress on probabilistic reversal learning. They find an overall increase in accuracy following acute stress, but no changes in computational parameters of learning or neural signatures of learning. Although I agree that this is an interesting and timely question, I have some major concerns about the manuscript.*  
  
-       The inclusion of only male participants is not consistent with the standards of the stress field and limits the generalizability of these findings. This decision needs to be justified in the manuscript (there is currently only a brief discussion on p. 26) and the male population should be indicated in the title.  
  
-       More information about the justification and implementation of the included computational models is needed.  
o       Please describe the theoretical rationale behind the 8 tested models.  
o       Why were there 6 permutations of Rescorla-Wagner models but only 1 Pearce-Hall model? How were learning rates implemented in PH?  
o       The equations listed reflect basic RW updating (rather than helping differentiate models) and are a little confusing. Are we meant to understand that the learning rate for win is divided by the learning rate for loss (Eq 2-3)?  
o       Please provide the equations and the number of free parameters for each model tested. It is difficult to follow "RW-iDU-1al" (and determine how many free parameters are included in each model, which are then potentially modulated by stress) without having the equations written out.  
o       Relatedly, the authors later note that "choice temperature parameters were significantly higher after win trials compared to loss trials" (p. 19) - it was not clear that these were estimated separately. Was that true for all models? Only the models with separate win/loss learning rates?  
o       Please explain protected exceedance probabilities - assuming this came from the hierarchical Bayesian inference approach (p. 12) but would help to have more information.  
o       Determining the best model based on the control condition (while omitting the stress data from the same participants) leads to the possibility that the best model to describe performance under stress may have been missed. Can the authors explain why this step-wise model fitting structure was employed?  
o       Are stress weights multiplicative or additive?  
o       How was the within-subjects nature of the design incorporated into the model-fitting process? If hierarchical model fitting was used, it seems like data from each participant in both stress conditions could have been fitted concurrently  
o       Which model was used to derive the RPE that were included in the fMRI analyses? This was confusing as slightly different models were used for the stress and control conditions.  
o       The authors note in the discussion that there was "high stress-related interindividual variability in the use of learned values" (p. 23) - was this variance in beta values significantly different from the control condition?  
o       Please provide more information to facilitate understanding of Table S-D - which parameters were significant? What were the values of the other free parameters?  
  
-       Given the within-subjects nature of the design, how were practice effects incorporated into analyses? This was not included in the multilevel model and is likely to influence performance (as one example, Vo et al 2016 Psychopharmacology)  
  
-       I have several further concerns about the statistics:  
o       Please justify the use of one-tailed t-tests for stress efficacy (p. 9)  
o       Why were parameter estimates considered significant at "p < .05" (p. 10)?  
o       Stating that something is "numerically higher" when p = .6 seems like an over-interpretation (p. 19)  
o       Figures S5-S8 - are these correlations still significant when removing outliers?  
o       Please provide information about motion cutoffs and motion correction in fMRI analyses  
o       In the supplement, the authors estimate that they would have needed N = 71 participants for 80% power, yet the sample only included 28 participants. Please discuss the implications of this level of power.  
o       Were there any condition x phase interactions in the multilevel linear model? I understand that was not the best-fitting model, but this seems important for making the claim that stress altered reversal learning/cognitive flexibility rather than just probabilistic reinforcement learning  
  
-       The discussion of background literature could be clarified:  
o       The example at the beginning of the paper is a little confusing - what are the stable and predictable costs and reward for different modes of transit? Is the weather forecast uncertain because it may not be accurate or because it varies across days? What does stress have to do with this example?  
o       The consideration of cognitive flexibility and decision-making is a bit muddled. For example, when the authors state that "Studies have found mixed results for the influence of stress on decision-making" (p. 3), the referenced papers are actually examining cognitive flexibility.  
o       The authors state that "habitual decision-making appears unaffected [by stress] at the behavioral as well as neural level" - this claim appears to be contradicted by findings from rodents (e.g., Packard & Goodman 2012 Front Behav Neurosci, Siller-Perez et al 2017 Neurobiol Learn Mem) and humans (e.g. Schwabe & Wolf 2012 J Neuro)  
o       What is the basis for "physiological paradigms lead to more immediate stress"? (p. 4)  
o       Some additional references that would be useful to discuss are human research on how stress influences reward learning in an unstable environment (mentioned on p. 4; Lenow et al 2017 J Neuro) as well as rodent studies of stress effects on reversal learning (e.g. Graybeal et al 2011 Nat Neurosci)  
  
Minor comments:  
-       More information is needed about these participants. How old were they? What were inclusion/exclusion criteria? How were they determined to be "healthy"?  
-       Understanding stressor efficacy is important; please move from supplement to main text  
-       How many times did reversals repeat during middle experimental phase?  
-       Did participants earn different amounts of $ in the stress vs control condition?  
-       Did degree of cortisol responsivity correlate with learning?

Reviewer #3: This study investigated the effects of acute psychosocial stress on participants' performance in a probabilistic reversal learning task. In addition, the authors used computational modeling and fMRI to find computational and neural correlates of potential stress effects. The findings suggest that acute stress may slightly increase correct responses in the task and that this effect may be driven by altered choice stochasticity under stress. Furthermore, the authors found that reward prediction error (RPE) signals in the striatum, specifically for win trials, were stronger under stress than control conditions.  
  
Overall, I think this study is well-written and tackles a timely topic that will be of interest to the broad readership of the journal. While the effect sizes appear to be rather small, I think both the design and the analysis are well done and the results deserve to be published. I mostly have a few conceptual and methodological questions that I am sure can be addressed by the authors. I list my comments below, separately for each section of the paper.  
  
Introduction:  
-       On page 5, the authors state  
  
"To our knowledge, only two studies combine a within subject design with computational modeling and functional neuroimaging (fMRI) to elucidate underlying cognitive mechanisms (Carvalheiro et al., 2021; Robinson et al., 2013). However, type of stressor and reward-learning paradigm differ from our study design."  
I like concise introductions but given that these two studies seem to be the closest to what the authors investigated here, it may be beneficial to provide a few more details as to how they were different, what they found, and what the current experiment will add to the literature beyond these earlier findings.   
  
Methods:  
-       One obvious limitation of the current study is the sample size. The authors tested 38 people but only 28 made it into the final sample, which is a large drop from their target of n = 71 stated in the supplement. I do not think that this is necessarily a problem for their behavioral findings due to the within-subject design of the study and their logistic regression analysis strategy, but I fear the sample may be way too small to provide reliable results regarding their fMRI investigations and may also explain why there were only very few significant differences between stress conditions. Although having more participants would be ideal, I think it is usually unreasonable to request additional testing. However, the authors should probably discuss the implications of their small sample size with respect to their imaging findings in the limitations section at the end of the manuscript.  
-       The authors mention that stress and control sessions were counterbalanced, is that still true for the final 28 people? How many of those started with stress or control, respectively?  
-       Although there is some justification for it, it is always a bit disappointing to see stress studies being conducted only in men. Under the assumption that male participants produce cleaner data (a claim I feel is a bit exaggerated), such exclusions reduce generalizability by neglecting half of the population and make it very difficult to detect meaningful sex differences. This is not a specific criticism of the current study, but I would urge the authors to appropriately plan future work to include women in their research as well.  
-       In the reversal learning paradigm, feedback would indicate either a monetary gain of 10 cents or a loss of 10 cents. Given that losses usually loom larger than gains (Tversky & Kahneman, 1985), some other decision making paradigms have implemented asymmetrical reward/punishment magnitudes to counteract this framing effect (e.g., see Proudfit, 2015). I was wondering why the authors chose symmetrical values and whether this may relate to their finding of lower learning rates in win compared to loss trials in their computational modeling analysis.  
-       The description of the reversal learning task on page 8 could benefit from adding a few more details, namely how many reversals there were and whether all participants experienced the same number of reversals (I assume they did but it would be nice to state this directly). It would also be nice if the authors could provide an explanation as to why they designed their paradigm the way it is (i.e., what is the reason for the including the stable phases before and after the reversal phase and why do they have a different number of trials? Did reversals appear after a fixed number of trials, i.e., was it possible for participants to form some kind of expectation about when reversals could occur?).  
-       The authors write that six participants were excluded due to having "different task environments". What exactly does that mean?  
-       The authors describe how they decided on the exact statistical model to use to analyze choice behavior (i.e., a model comparison procedure that yielded the best fitting model). Did the tested models also include variations in the random effect structures? If not, what was the reason for only including a subject specific intercept but not random slopes for the other predictors? Given the within-subject nature of the study, each fixed effect (i.e., task phase and stress condition) could be included as a random effect as well and the trial count of the task seems high enough to not run into model conversion problems.  
-       On page 12, the authors mention all eight different RL models they compared for best fit. Although they are very related to the basic equations on the previous page, it would be great if the authors could provide the exact formulation somewhere, either as a table in the main text or in the supplement, so that there is all the information needed for other researchers who wish to implement similar models in their work.  
  
Results:  
-       The caption of figure 2 should state what error bars represent (I assume SEM?)  
-       Differences in cortisol response between conditions look very striking, yet according to the supplement, the t-test results in a relatively modest p-value of .02, which was surprising to me. Is this maybe due to the specific measure (AUCg) used to compare the cortisol curves? Or are there maybe a few participants who do not show a large difference between conditions? It would be interesting to see the authors' thoughts on this.  
-       Panel labels in figure 3 are a little confusing. I understand the authors' intention (Figure 1 being split into three sub panels) but I think the labels should stay consistent (i.e., use A, B, C etc. instead of 1, 2…).  
-       Looking at the results presented in figure 3, it appears as if the slightly larger proportion of correct responses under stress stems mostly from participants' behavior in the first stable phase, whereas the pattern actually seems to be opposite in the reversal and late stable phase. Given that the logistic regressions presented by the authors do not include interaction terms, it is difficult to assess this properly, so I was wondering if the authors tried to model these different phases separately. If the significant differences really only appear in the first phase, I think this may call into question the conclusion that stress affects reversal learning and really only suggests a difference in probabilistic reinforcement learning in somewhat stable environments, which would still be interesting but different from what is suggested here.  
-       Related to the last point, it would be great if the authors could provide a few more details about the participants' behavior specifically at and around the point of reversals in the task. Are there any differences between stress conditions in terms of recovery after reversals (i.e., how fast participants realize that a reversal has indeed occurred)?  
-       A more minor point: Have the authors tested whether there are any carry-over effects between testing sessions (e.g., participants performing better on day 2)?   
  
Discussion:  
-       On page 24, the authors write:  
  
"When differentiating RPE signals during win and loss trials, we found stronger coding of positive RPEs in the ventral striatum during the stress compared to the control condition in our sample of healthy male participants. This increased neural activation following acute social stress corresponded to better performance in the stress condition"  
I think even if we believe the RPE effect (which only appeared when looking at win and loss trials separately instead of in the full model) is valid, the conclusion in the second sentence would only be justified if the authors could show a correlation between the strength of win-RPEs and task performance. There is nothing wrong to speculate about how these findings may relate to each other but saying they correspond is not supported by the presented data.  
-       The authors discuss stress-related changes in dopamine levels as a potential mechanism for their findings, which makes sense given the known role of dopamine in coding RPEs. Yet, in the experiment, only cortisol levels were measured and related to reversal learning performance. Why did the authors not consider analysing alpha amylase levels that are more related to catecholaminergic activity instead of or in addition to cortisol? Similarly, if stress effects on reversal learning are expected to be caused primarily by affecting dopamine pathways, it might have been advantageous to test behavior closer to stress onset, as the effects of stress on the sympathetic nervous system and the HPA axis are very time sensitive (Hermans et al., 2014). I would like to hear the authors' thoughts on these issues.      
  
Typos:  
-       Page 13: In the sentence starting with "On the first level individual subject level the feedback onsets…" the word "level" should be deleted after "first".    
-       Page 24: The sentence "No whole-brain correctable stress effects on RPE representation were observed when assessing win and loss trials together." appears twice.  
-       Page 25: The part "while no difference between positive and negative PE signals were observed in the safe condition in a condition of threat (potential of electric shock)" also appears twice.