Here are the comments of the reviewers (if applicable):  
----------------------------------------  
Reviewer: 1  
  
Comments to Author  
In these experiments, Panayi and Killcross directly test the role of the lateral region of OFC (lOFC) in the formation of conditioned inhibitory associations. The experiments are thoughtful and well-designed however not all conclusions are supported by the data. While the authors provide clear evidence that inactivation of lOFC does not impair the acquisition of conditioned inhibitory associations, I am not convinced that the results show that lOFC inactivation disrupts the expression of conditioned inhibition. Instead, the authors have shown a non-specific effect of lOFC inhibition on Pavlovian responding (see Major comment below).  
  
Major  
The authors present two studies. Experiment 2 is largely a replication of the authors’ previous work. The results nicely replicate Panayi & Killcross (2014) showing that lOFC inhibition impairs between session extinction and, in this study, it also appears that within-session extinction is enhanced by lOFC inactivation. The current experiment goes slightly further to show that the training parameters used here (and in Panayi & Killcross, 2014) do not promote the formation of conditioned inhibition. Together, this indicates that the impairment in extinction learning cannot be attributed to impaired conditioned inhibitory learning. By contrast, I am less convinced by the interpretation provided for Experiment 1. The major conclusion of this Experiment is that “OFC inactivation disrupts the expression but not acquisition of conditioned inhibition.” This conclusion does not fully capture the results of this study. It was actually shown that OFC inactivation disrupted learning and expression of Pavlovian excitatory associations as well as the expression of conditioned inhibitory learning. Indeed, OFC inactivation led to a non-specific decrease in magazine entries for all stimuli, including the control cue Z. Moreover, there may be a potential floor effect in muscimol treated rats (Figure 2B) with these rats showing max 1 entry per minute. Could this mask potential differences in responding to A+ versus AX- in muscimol infused rats? The framing of the results of this experiment as selective to conditioned inhibition is misleading (this occurs in the title, abstract, results and discussion). The authors even state in the Discussion that “Despite the abolition of selective inhibitory behavioural control following OFC inactivation, OFC function was not necessary for the underlying learning of an inhibitory association” (page 20, lines 9-10). However, there is no evidence at all that this effect was selective to inhibition. These results are particularly difficult to interpret given that many previous studies have shown no effect of OFC inhibition on the acquisition of Pavlovian associations, although this may be explained by variations in the subregion of OFC that was targeted. Nevertheless, given this non-specific effect, the authors’ argument that these results are inconsistent with latent state models of OFC function seems weak. In fact, the most convincing result is that lOFC inactivation does not interfere with learning about an explicitly signaled conditioned inhibitor (as seen during the summation and retardation tests of Experiment 1), which is consistent with this model.  
  
Minor  
1.      Page 6, line 49: “…the rate of acquisition to cue X will be lower than acquisition to a novel…” I believe “cue” is missing from the end of this sentence.  
2.      It should be made clear in the title, abstract and introduction that this work targets the lateral OFC.  
  
Reviewer: 2  
  
Comments to Author  
In two experiments, Panayi and Killcross examine the contribution of the orbitofrontal cortex (OFC) to behavioral expression of conditioned inhibition, as well as inhibitory learning via summation and retardation tests. In experiment 1, OFC inactivation with muscimol impaired both excitatory conditioning to the control cue Z and also impaired behavioral expression of conditioned inhibition, with muscimol rats showing equivalent responding to A+ and AX-. Despite this deficit, rats receiving muscimol went on to demonstrate summation and retardation at levels equivalent to controls, though overall reductions in the responding were observed. In experiment 2, muscimol infusions accelerated within session extinction to C, impaired extinction retrieval to C in the following session, but did not impair subsequent summation and retardation to the conditioned inhibitor X. The authors conclude with a thorough discussion of the relevance of the results to existing theories of OFC function.  
  
The behavioral design is appropriate and the general use of muscimol to inhibit OFC activity is justified. The results are thoroughly analyzed and correctly interpreted.  My primary concerns are the effects of muscimol on the acquisition of appetitive conditioning and the logic in combining these two studies in a single paper.  
  
As the authors are aware, it is not commonly observed that disrupting OFC activity disrupts the acquisition of appetitive conditioning or instrumental responding. This is a fairly consistent finding across labs (Balleine, Schoenbaum, Gourley, etc.) and method of inactivation (baclofen/muscimol, DREADD, optogenetic inhibition). The authors suggest an anatomical explanation. They targeted more anterior OFC, while prior work targeted posterior OFC or a longer anterior-posterior axis. Yet, Burke et al. (2008) performed neurotoxic lesions that mostly targeted anterior OFC and saw no impairments in the acquisition of appetitive responding. A more parsimonious explanation seems to be that the dose/volume of muscimol produced deficits in responding. A better companion experiment would have been to muscimol a distinct OFC subregion not hypothesized to impair responding OR to disrupt OFC activity in this subregion with another technique.  
  
The second concern is rationale for combining these two experiments. Experiment 2 demonstrates that OFC activity is not necessary during feature negative sessions in order to subsequently show summation and retardation to the conditioned inhibitor. At the same time, OFC contributes to within and between-session extinction. This result is consistent with a prior report. As I pointed above, my concern is that Experiment 2 does not do much to extend or address the weakness in Experiment 2. Instead, Experiment 2 repeats a prior result specific to extinction and provides another demonstration the OFC is not necessary for conditioned inhibitors to acquire their inhibitory properties.  
  
Reiterating my prior point. A more compelling Experiment 2 would have been to manipulate activity is a distinct OFC subregion that might not have disrupted behavioral responding but did disrupt acquisition of the inhibitory property of the conditioned inhibitor. Alternatively, manipulate the same anterior region but using another inactivation approach to determine if effects similar to muscimol are observed.  
  
Reviewer: 3  
  
Comments to Author  
Panayi and Killcross tested the role of Orbitofrontal Cortex (OFC) in the learning and expression of Pavlovian conditioned inhibition. In the first experiment, authors found that muscimol inactivation of OFC abolished differences in discriminative (CS-PreCS) responding during the learning phase (A+ AX-), but later off-drug summation and retardation tests revealed intact learning about the conditioned inhibitor (X-). Inactivation also impaired learning of control cue Z that remained unchanged in terms of its reward prediction. Additionally, in a second experiment, authors found OFC inactivation did not disrupt Pavlovian extinction learning by impairing acquisition of conditioned inhibition. Cue extinction learning was attenuated between-session, but actually enhanced within-session following OFC inactivation. The results are discussed in the context of OFC’s involvement in ‘inhibitory control,’ outcome expectancy, representation of latent states, and generally proposed to be critical in the modulation of behavior when contingencies change.  
  
The methods, experimental design, and results were clearly-presented and easy-to-follow. The control experiments were outstanding. The supplemental figures show convincingly that learning and expression results cannot be accounted for changes in appetite or vigor of reward approach changes, changes in acquisition of conditioned responding to discrete cues, or changes in locomotor activity/exploration. My main concern centers around whether the findings, as presented, provide a clear role for OFC in conditioned inhibition. My main criticism is in the interpretation of the data and its couching in the existing literature. Suggestions for essential revisions center mainly around clarifying and distilling the findings in a narrower context.  
  
1)      As mentioned above, authors taking on the major theories of OFC function, some in the Introduction but mostly in the Discussion. Though clearly a valiant attempt to explain their findings in the context of very diverse experimental contexts, I thought it strayed too far from their actual data and what the data reveal. In the end, I was more confused about the role for OFC in inhibitory control. So I suggest to reign in the Intro and Discussion to perhaps inhibitory control. I found the latent state discussion the most speculative/reaching. Perhaps condense that part or leave out?  
  
2)      A relevant example of branching off-topic is the mention of how their findings are consistent with the ‘updating expected outcome values’ and how translating this can be dissociated within OFC subregions. How are the present experiments informing this functional heterogeneity if authors inactivate only lateral OFC?  
  
3)      Related to differences in subregions of OFC, much has been written on the highly-related topic of reversal learning (Hervig et al. 2019; Verharen et al. 2020). It is particularly relevant as authors describe single-unit activity studies where OFC has been found sensitive to the direction of responding, requiring suppression of an alternative response (resolving competition) rather than control of new behavior. But there is no link proposed between extinction and reversal, which seems like a missed opportunity.  
  
4)      There is the interesting, yet puzzling, finding that within-session extinction is facilitated, but between-session learning is attenuated following OFC inhibition. Authors write that responding at the start of the session may be driven by stimulus-response associations, protecting the responding from extinction learning. Together with the authors’ proposal that OFC is involved in appropriate response selection, I am unclear if authors are concluding and indeed if they are claiming to dissociate OFC involvement in cue- vs. action- outcome learning? In favor of a balanced evaluation, there is some evidence of specific OFC pathways involved the latter, that can be cited- Fresno et al. 2019 eLife.  
  
More Minor:  
1)      It appears that authors are analyzing the saline group separately from the OFC muscimol group (page 10 lines 52-60, page 11 lines3-8), until I read that there was a more inclusive analysis, revealing Group\*Cue\*Day interaction and a Group\*Day interaction. I suggest to move report of the omnibus analysis first, to justify probing of the subsequent effects.  
2)      In the Intro, Authors state “in extinction procedures when an expected reward is no longer delivered the expected value of the cue should be updated to reflect this new state of affairs, a process that is predicted to involve OFC function.” The following could also be cited as evidence of precisely this in nonhuman primates: Izquierdo & Murray, 2005 EJN.  
3)      Suggest a more descriptive title for Figure 1.