Here is a todo list, where I have put either your name or mine. Do you agree on the partition (even if there are more marios than mehdis)? :-)

Reviewer #1: This was an excellent article. It was a great mixture of literature review of fundamental OFC behavior and modelling, followed by interesting newer behavioral data with corresponding model simulations. I thought it was comprehensive and well written. The short sub sections made the article very tractable and easy to digest. I only have minor thoughts and comments.  
**MARIOS:**-I understand the point made that the state of the field generally thinks that OFC is not that involved in simple initial learning, but I thought it was a little overstated. My impression from the previous work was that it was involved initially but demonstrating it behaviorally was difficult and could only be seen in behavior post learning, and that the lack of behavioral effects might reflect function associated with parallel systems.

The reviewer is correct that this point is a little overstated. However, we believe there is a disconnect between a nuanced understanding of the field of OFC research and the way the effect is often discussed in modeling and review papers. For example, the modeling performed by Wilson et al (2014; Neuron) portrays initial acquisition in simple tasks as being entirely identical between lesion and control animals. Clearly this is a computational simplification, however it can easily mislead readers into believing that OFC lesions cause no deficits in simple acquisition. Take as a further example, the description provided in the final paragraph of the introduction of their paper where the phrasing is rather ambiguous:

“an OFC-lesioned animal can still learn and perform basic tasks using RL, albeit using only observable (stimulus-bound) states based on current perceptual information. As a result, basic learning and decision making are possible without the OFC, but behavior becomes more and more impaired as tasks become abstract, and more of their states are partially observable.”

Here the reader is required to also acknowledge that learning even in putatively “simple” Pavlovian task might not simply involve only “observable (stimulus bound) states”. This is not helped by the fact that all simulations of simple tasks that are subsequently presented show that OFC lesions have no effect on simple task acquisition because the underlying task representation in Lesion and Control animals is identical i.e. no unobservable/latent states. Note that we are not trying to single out this particular paper. However, it is representative of the potential misinterpretation that we believe often occurs when the effects of OFC lesions on simple acquisition are discussed.

Indeed, as anecdotal evidence of the actual persistence of this belief even among OFC researchers, we have recently had a paper that was rejected as a result of all 3 reviewers taking issue with an effect of acquisition following OFC inactivation in a control cue. Rather than the reviewers considering that task acquisition might have involved unobservable states, their assumption was that the findings did not fit other reports of null results and there must be something wrong with the experimental manipulation instead.

Therefore, while we entirely agree with the reviewer that the point is overstated, this was also our intention. We also believe that paragraphs 2 and 3 of Section 2.3 clearly indicate this mismatch between computational and linguistic simplification and the nuanced view that is actually expressed in these computational models.

**MEHDI:**-I thought latent states could be defined better.  
**MEHDI:**-Why does initial learning rely so much on MB? If there is no model how could it be MB?  
**MEHDI:**-Could some of the OFC inactivation results reflect an altered exploration system? **(Here I am tempted to say that if a task was so complex that initially EXP is more selected than MB and MF [because MB and MF still require time before becoming efficient], then ofc inactivation would impair the selection of EXP and lead to a [1/3 1/3 1/3] selection of [EXP MB MF]. Thus it would produce an impaired exploration which might prevent the agent from exploring enough to learn appropriate representations in MB and MF systems. But it is just a prediction from the computational principles adopted in the model. One would need to think of such a complex task to illustrate this principle.)**

**MARIOS:**-Overall, I was surprised that so much of the behavior was explained by MB is such a simple task.

We share the reviewers initial surprise that a putatively simple task would be heavily dependent upon a relatively complex MB learning system. However, as we highlight throughout the manuscript, there are a myriad of complex processes that underlie learning/behaviour in even the most basic behavioural tasks (e.g. perceptual discrimination, working memory, attention, behavioural competition, timing, uncertainty etc…). From the perspective of the subject, particularly at the start of training, there is no reason to believe that the environment will contain a single fixed duration, easily detectable stimulus that will deterministically predict the delivery of food into the same receptacle. Indeed, most environments in the natural world involve substantial variability and uncertainty. It is therefore quite likely that subjects experience this simple Pavlovian procedure is uncertain and ambiguous for an extended period of time, and therefore keeps a MB system engaged.

Magazine approach behavior itself is also quite an ambiguous readout of the underlying psychological processes which drive it (e.g. Killcross & Blundell, 2002; Mackintosh, 1974; Timberlake, 1994). Indeed, a procedure such as outcome devaluation suggests that initial acquisition behavior involves learning about a causal/sensory model of the world, which is only revealed during the final devaluation test. Given that performance to a devalued Pavlovian CS can be significantly suppressed (e.g. a single Light-Pellet CS-US protocol employed by (Pickens et al., 2003)), it is not surprising that an MB system can be heavily engaged in such a simple task.

We thank the reviewer for raising this point, as it is worth highlighting in the discussion. We have added the following paragraph to the discussion section:

These modifications make explicit the implicit understanding that there are numerous psychological processes that underlie even simple learning procedures that are often implicitly acknowledged by researchers. Indeed, this is reflected in the models we have considered by the surprisingly strong MB contributions to behaviour in simple Pavlovian tasks, consistent with other RL models of Pavlovian approach behaviour (Dayan & Berridge, 2014; Lesaint et al., 2014; Zhang et al., 2009).

**MARIOS (?):**-I know that this was a review about OFC, but I was left wondering how OFC arbitrates between MB and MF (ie, the mechanism) and what larger circuit was involved.  
  
  
  
Reviewer #3: This manuscript presents data and model simulations suggesting that the rodent lateral OFC is an arbitrator between model-based (MB) and model-free (MF) learning systems. The manuscript starts with a review of the literature on the effects of OFC lesions on behavior in reversal learning and devaluation tasks. This is followed by an overview of RL models, especially considering MB and MF learning. Next, the results from four rodent Pavlovian learning experiments are presented, along with model simulations suggesting that the effects of OFC lesions can be explained by changes in MB learning, or the arbitration between different learning systems.  
  
Understanding the role of OFC in learning and behavior is an important and timely goal. Relating empirical results and computational models can help us to better decipher and understand the complex ways in which the OFC contributes to these functions. Thus, this manuscript makes an important contribution toward this goal.  
  
However, there are several issues with the way this is done here. Most importantly, there seems to be a high degree of flexibility in model assumptions and how "OFC lesions" are implemented across the different models. For instance, it appears that different versions of the model can explain opposite effects of OFC lesions on initial learning performance. I am not sure what we learn from these models in this case.  
  
Essential comments  
**MEHDI:**1. It appears the model simulations for experiment 1 (Fig 1) hinge on the assumption that arbitration occurs not only for MB vs. MF, but also includes exploration (EXP) as a third option. Moreover, there is the additional assumption that when the arbitration system is blocked, the model always settles for either MB or MF, but never for EXP. How can both of these assumptions be true? These assumptions are somewhat unusual and because they appear critical for the overall conclusion that OFC is an arbitrator, it would be important to discuss the importance of these assumptions for ability of the model to account for the data.  
  
**BOTH:**2. More generally, there is a lot of change in the model assumptions across the four experiments. For instance, for experiment 2 (Fig 2B) disruption of the arbitration system is assumed to push the model toward EXP instead of only using either MB or MF, as in experiment 1. Why are OFC lesions not modeled in the same way across all experiments? What happens when the same assumptions as in Exp 1 are used for Exp 2? Similarly, model assumptions are again fundamentally changed to account for data from Exp 4. Here, OFC inactivation is assumed to disable a MB compensation for forgetting. I am really worried about modeling the data so flexibly because with enough flexibility, almost any model can account for almost any data. In the light of this, a method section that includes details of the model assumptions and implementations for each of these experiments would be important. In addition, it would be best to also publish the code for all models and simulations along with this manuscript.  
  
**MARIOS:**3. How do you reconcile results from Experiment 1 with the results of Experiment 2 and 4? Why is acquisition impaired in the muscimol group (Exp 4 and 2) but not in Exp 1, where OFC lesions enhance learning? This is an illustrative example of the effects of model flexibility. The different OFC lesion models predict enhanced initial acquisition in Exp 1 and disrupted acquisition in Exp 4. Shouldn't the same lesion produce the same alteration in performance across experiments?  
  
  
Minor comments  
**MARIOS:**1. Page 5: "if the OFC is necessary for representing the identity of expected outcomes, OFC lesions would disrupt only reversal learning and not initial acquisition because outcome identity is only relevant to task performance at the point of reversal." I don't understand this argument. Outcome identity is also relevant during initial learning (nothing->sucrose). And you could argue that outcome identity is not necessary for performance at the point of reversal, but that only outcome value is required for reversal. The authors should rethink their argument of why encoding of outcome identity in OFC would predict that OFC lesions only disrupt reversal learning but not initial acquisition. Moreover, it is unclear how why this argument is relevant for the current paper which seems to suggest that OFC is the arbitrator between different learning systems rather than the location where outcome identity is represented.  
  
**MARIOS:**2. More generally, the introduction and the discussion about outcome identity does not fit well with the rest of the manuscript which is about the arbitration between different learning systems.

The consideration of expected outcome identity within the OFC literature has been a dominant feature of understanding OFC function, and is a key point of contact between associative learning theory and RL models (particularly for MB systems). Therefore, we believe that discussion of outcome identity was important for the review of how understanding of OFC function has developed historically, and when considering the contribution of MB systems.

We have added the following sentence at the start of the discussion to reemphasize the context in which the arbitration model is being considered:

Here we briefly reviewed the developments and changes in our understanding of OFC function which have closely followed developments in our understanding of associatively learning theory and refined further by recent progress in RL modelling. OFC dysfunction has been successfully modelled as an impairment in MB inferences resulting from disruption of the formation of latent states necessary for a detailed cognitive map of task space (Wilson et al., 2014).

**MARIOS: (yes the labeled are switched)**3. Are the labels in Fig 2A switched? Otherwise the text does not match what is shown. Namely, only the muscimol group but not the saline group continued to learn in sessions 12-15.

We thank the reviewer for correctly pointing out this labeling error. Fig 2A has now been updated with the labeling corrected. Fig 1A has also been updated to keep the colour/symbol coding consistent across all four figures.  
  
**BOTH:**4. The authors focus on rodents and non-human primates. It would be important to also cite and discuss examples from the human literature.  
  
**MARIOS:**5. The data modeled here are currently only published as preprint on biorxiv. The fact that they have not been peer-reviewed should be clearly stated in the current manuscript.