**Reviewer comments to be addressed:**

**Review 1:**

(2) If possible, it'd be great to provide more intuition for why the  
   authors chose the stick breaking process to generate cue weights. Figure  
   1 seems to suggest that there is a fixed number of 4 cues, in which case,  
   the use of stick breaking seems an overkill.

* I ignored this one. We used the stick breaking process because it is useful and novel in general.

 (3) Figure 3 shows that people actually exhibited a U-shape function  
   during the test phase. Although understandably this is not the focus of  
   the current paper, the authors could recognize this pattern, as it is  
   fairly common throughout many areas of learning (especially as a function  
   of the entropy of stimuli). And that would make intuitive sense - if a  
   system is of high-entropy, it would overload the learning capacity and  
   thus leads to poor learning. The current wording makes it sound like  
   performance during test is monotonically related to compensatoriness.

* I changed that section slightly and explained where the peak is

(5) Figure 4: I suspect that there is no statistical difference between  
   the performance of the two models. If there is, the authors should  
   present the test to backup the statement "it can be seen that  
   take-the-best was better ..." and etc.

🡪 I don’t think we found a sign. difference

**Review 2:**

 The description of the take-the-best algorithm is confusing.  I think  
   that given K cues the algorithm generates all K! cue orders and keeps  
   weights with each that indicate the number of times they generated the  
   correct prediction about information gathering.  I'd like to see a much  
   more formal description of the algorithm that makes precise the prose.  
   The LR-based algorithm is more clear, but a description of the training  
   data would help immensely.

* This still needs to be addressed a little. Maarten made several points in this section too.

**Review 3:**

   Imagine the non−compensatory case with one dominate cue, most of the  
   comparison without difference on this cue will have 50% chance to win,  
   and subjects' judgement play a minimal role here. The same situation may  
   happen when all the cues have equal weight on the complete compensatory  
   case.  
🡪 yes, we know that and I made that point more clear now.

   So I suggest authors plot a curve of test performance of an ideal tester  
   with all the weights known. If that is that curve has a same shape of  
   Figure 3, which I believe so. I think the experiment can be better  
   designed to balance the testing bias.

* We could look at that for the journal article, but no need now.
* I don’t think we have any “testing bias” here, as the change in informative comparisons naturally comes along when you change the degree of compensatoriness?

It is better to have more explanations in figure captions.

And figure references are misaligned.

* I took care of that.