Dear Dr. Baillargeon,

Thank you for providing more thoughtful comments on our manuscript, “Social cues modulate the representations underlying cross-situational learning.” Please accept our final submission. We have addressed your comments and the comments of the reviewers, and we believe that the manuscript is improved. Please find below a point-by-point response to the comments.

Please contact us if you have any questions or concerns. We look forward to your consideration of this revision.

Sincerely,

Kyle MacDonald

**Editor**

*Figure 1: As a non-expert in this field, it took me some time to understand Figure 1. Would it be possible to have both sides of the figure use the schematic face, so that the figure is specific to Experiment 1? You could also mention in the caption that the number of referents varied (2, 4, 6, 8). If you wish, you could add another, similar figure for Experiment 2, with the 4 referents used.*

We appreciate your point about the clarity of the figure. We decided not to change the figure, but we updated the caption, and we think the new version will make it easier to interpret. The caption now reads,

“Screenshots of exposure and test trials from Experiments 1-4. The top left panel shows an exposure trial in the No-gaze condition using the schematic gaze cue (Experiment 1). The top right panel shows an exposure trial in the Gaze condition using the video gaze cue (Experiments 2-4). Participants saw either Gaze or No-gaze exposure trials depending on condition assignment, and all participants saw both types of test trials: Same (bottom left panel) and Switch (bottom right panel). On Same trials, the object that participants chose during exposure appeared with a new novel object. On Switch trials the object that participants did not choose appeared with a new novel object. Participants either saw 2, 4, 6, or 8 referents on the screen depending on condition assignment.”

*p. 4: "The goal of our current investigation was to ask...could change", or would change.*

Fixed.

*p. 10: The sentence "And we limited...interactions" does not appear necessary given what you stated on p. 7. Footnote 4 seems sufficient here.*

Fixed.

*p. 10, end of footnote 3: "this filtering criterion".*

Fixed.

*p. 16: If I understand correctly, the results reported are solely for the trials following the 16 familiarization trials. Please make this clear to the readers.*

You are correct. We changed the first sentence of the Test Results section to be clearer. It now reads, “Next, we tested whether the reliability manipulation altered the strength of participants' memory for alternative word-object links in the second block of test trials that followed the initial familiarization phase.”

*p. 17: "people store"*

Fixed.

*p. 19, footnote 7: "exploratory data analyses"*

Fixed.

*p. 20, bottom line: Some care is needed here, because the fact that the objects were shown for a fixed amount of time does not mean that the participants looked at them for that entire time (e.g., they could have been looking at the timer displayed, at their watch, and so on). So here and elsewhere, please be careful in the phrasing used, e.g., "participants were shown the objects for a fixed amount of time".*

Thank you for pointing this out. We have updated the phrasing in several places in the paper to make it clear that participants were shown the objects for the same amount of time.

*p. 22: I was a bit confused here, did you mean that you initially used 3 and 6 s but increased to 6 and 9 s following pilot testing?*

Thank you for pointing this out. We changed the text and think it is now much clearer: “Initially, the length of the inspection times was based on participants' self-paced inspection times in the Gaze and No-Gaze conditions in Experiment 2 (Short = 3 seconds; Long = 6 seconds). However, after pilot testing, we added three seconds to each condition to ensure that participants had enough time to respond before the experiment advanced (Short = 6 seconds; Long = 9 seconds).”

*p. 24: "even when inspection times were equivalent" . See the point above on p. 20, and rephrase as needed here and elsewhere.*

Fixed.

*p. 27: "we used a video of an actress"*

Fixed.

*p. 28: "learners tend to store"*

Fixed.

**Reviewer 1**

*This revision addressed the points in the original review.*

Thank you.

**Reviewer 2**

*- p. 6-7: in the gaze condition, and in Same trials: is the object that is kept constant the one they selected, or the one that was gazed at? (definition of what is a correct answer at test is now much clearer).*

For all Same trials the object at test was the object that the participant selected, which could have been different than the gaze target.

*p. 7: you say that you “examined accuracy and response time on exposure trials to provide evidence that learners were sensitive to your experimental manipulation”; but accuracy on exposure trials cannot be compared across conditions since it is not defined for no-gaze trials, right?*

Yes, we are unable to compare accuracy on Exposure trials across conditions, but we can compare response times. To make this clearer, we updated the text to read, “ First, we examined accuracy on exposure trials in the Gaze condition and then we compared response times on exposure trials across the Gaze and No-Gaze conditions.”

*- p 8: ‘there was variability across conditions in the mean proportion of gaze cue’: what do you mean here? Selecting the gazed-at object?*

Thank you for pointing this out. We have fixed the sentence and it now reads, “However, the mean proportion of gaze following varied across conditions.”

*- p11: ‘we found a subset of participants who did not reliably use the gaze cue at all, potentially reducing the effect of gaze on cross-situational learning in the experiment’: but you got rid of those subjects, no? as you mentioned in footnote 3.*

The sentence now reads, “Moreover, we found a subset of participants who did not reliably use the gaze cue at all.”

*- p. 18: ‘we added accuracy on exposure trials as a predictor’ (plus the parenthesis to explain why); I think using the phrase ‘followed gaze’ as you did in your figure, was much easier to understand.*

We agree. We changed the sentence to: “First, we added the number of exposure trials on which participants' chose to use the gaze cue as a predictor in our model.”

*-p. 18: ‘participants’ use of gaze may be a stronger predictor of the effect of reliability…’: stronger than what?*

We clarified this sentence. It now reads, “This reduction suggests that participants' tendency to use the gaze cue is a stronger predictor of learners' memory for alternative word-object links compared to our reliability manipulation.”

*-p. 19, top: Panel A of fig. 5 (not panel B)*

Fixed.

*-p. 20: ‘mean proportion of gaze following was still 0.82’: I don’t think we got that information before, did we?*

We included the mean proportion gaze following for all 5 reliability conditions in our analysis of Exposure trials, section 4.2.1.

*- p. 20: ‘thus, we do not know whether there is an independent effect of referential cues ON the representations…’ (missing ‘on’)*

Fixed.

*- figure 5 and 6: legend for the no-gaze and gaze: you forgot to empty the blue dot.*

Fixed.

*- p. 24: ‘ and not the strength OF THE learners’ candidate hypothesis’ (missing ‘of the’)*

Fixed.

**Reviewer 3**

*First, at the end of the abstract, the authors say “our data suggest that the representations underlying cross-situational word learning are quite flexible: In conditions of greater uncertainty, learners store a broader range of information.” It seems given their acknowledgment that this finding might be very different under conditions where the speech isn’t believed to be about the co-present visual world, some clarification is needed. Something like: “At least when hearing speech that is about co-present objects, greater referential uncertainty leads learners to store a broader range of information.” My point is that If speech isn’t about the here-and-now (and a learner knows this), it seems unlikely they would store a broader range of object information when gaze cues are neutral, and would instead store little or no information. This is an empirical question, but such an important one that it bears mentioning with some emphasis in this paper. If on the other hand, the authors believe this is a general fact (i.e., with “greater uncertainty, learners store a broader range of information”), then they should defend this point more clearly. I suspect though that one could make the following reasonable claim, namely that the learner is certain about some things and uncertain about others: they know that the speech is about the co-present objects but they don’t know which object is the right object – that is, the learner knows what she doesn’t know (“known unknowns”, to quote our (ridiculous) former Secretary of Defense Donald Rumsfeld -- this term though, along with “unknown unknowns” does have a somewhat interesting history in psychology and philosophy).*

We appreciate this point, and we changed the sentence in the abstract to make the limits on the generalization of our findings clearer. It now reads, “Taken together, our data suggest that the representations underlying cross-situational word learning of concrete object labels are quite flexible: In conditions of greater uncertainty, learners store a broader range of information.”

*Second, and perhaps more importantly, if the authors believe that any physical cue (such as an arrow) will generate the same effects in the same way, i.e., reduce the number of referents considered, I’m feeling a little more uneasy about the theoretical significance of the present findings. Since we know that attentional cues focus attention, and we know that focused attention restricts encoding, and we know that encoding leads to learning, I’m now not sure how this is not simply additional evidence supporting common assumptions. I’m not trying to be a mischievous reviewer here, I’m quite honestly puzzled about the contribution and theoretical position that the authors wish to make in this regard. For instance, Nappa & Arnold (2014, Cognitive Psychology) made a significant, relevant contribution by showing that physical cues to pronoun reference to objects depended on listeners believing that the cue had referential intent (the cue was intentional). Why not take this stance here? Or provide experimental evidence that replicates Nappa & Arnold in the word learning domain? Otherwise, it seems that the authors are saying that restricted attention restricts encoding, etc. I recognize that the authors attempt to address these concerns (and cite some of their own work that I believe dovetails nicely with Nappa & Arnold, 2014). But, my sense is that by bending over backwards to acknowledge that a number of interpretations are valid (as they do on page 26), the paper has lost some of its teeth. --- It is also worth noting that Waxman has had much to say about referential intent and word learning, which would be relevant to discuss here. I don’t think additional experimental evidence is needed, but rather, again, a clearer statement of their own position relative to the existing literature that seems to support it.*

We appreciate these points and we do think they are important for understanding how cross-situational word learning unfolds in real world contexts. However, our results do not provide direct evidence in favor of either account: social cues are special vs. social cues are one of many cues. So we felt it was important to leave the paragraph (on p. 26) where we discuss these alternatives in the paper.

*Since my first review, I’ve spent some time reading and discussing with others the paper by Griffiths, Lieder & Goodman (2015), and this has changed my reaction to the paragraph about that work in the present paper (page 25). It seems that a wide range of findings could have been viewed as adaptive – that is, if the authors of the present paper had instead found that a decrease in memory for alternatives actually increased the strength of learners’ memory for their candidate hypothesis, this too could be interpreted as adaptive – given available resources, etc. So I am not sure exactly what this paragraph on page 25 is trying to say.*

We agree that this paragraph is speculative. We included it as an attempt to explain why we did not see strong evidence for a predicted result: that participants’ memory for a single hypothesis would be stronger in the Gaze conditions since they did not have to spread attention/memory across multiple alternatives. We still think that the results of Experiments 2-3 could be explained by the cognitive system trying to adaptively conserve cognitive resources for subsequent learning problems. So we chose to leave this paragraph in our final submission since we think it adds to the breadth/depth of the General Discussion.