**Reviewer #2**

*The elements of this section seem a bit disconnected; transitions between paragraphs/major ideas are not as clearly spelled out as would be ideal, giving the impression of a series of related ideas, but not ones which clearly and directly lead to the current study. While the final framing in The Current Study section does present the information clearly, the rest of the introduction doesn't connect to this as coherently.*

We thank the reviewer for this feedback. We’ve tried to make the transitions and connections between ideas clearer in the literature review: for example, along with other smaller tweaks to the writing in this section, we have made our paragraph beginnings and endings on pp. 5 – 6 much more explicit.

*The information about the Akhtar et al study is more clearly provided here, but the connection to the current work is missing for me. When this study is raised again in the General Discussion, the meaning of it is still a bit unclear. Would this information make more sense moved entirely to the General Discussion?*

We appreciate this suggestion, and have now moved the Akhtar et al. discussion to the General Discussion (p. 20), where it anchors a discussion of temporal proximity-based heuristics as alternative explanations for our findings. We agree that this organization allows a cleaner link to the implications of our findings.

*Results: Experiment 1: On page 12, line 41, you indicated that "Children and adults were more likely…" but if I understood the other aspects of the results, this is true for only some of the children, not all as is implied by this statement.*

We’ve corrected this sentence to indicate that this statement is true for all age groups except for the youngest in our study, rather than the full sample.

*Appendix: I appreciated the inclusion of the objects and the description of two of them. Given that this is an appendix, is there any reason not to provide the details on each of the toys? It is certainly experimentally unnecessary, but for an appendix seems quite appropriate.*

We now include descriptions of each of the toys in Appendix A.

*Grammar/Language/APA style:*

*p. 4, line 39: "…we begin by considering…" Isn't the consideration of the role of discourse the primary focus of the introduction. This phrasing implies that you will go on to other purposes, which doesn't appear forthcoming.*

*p. 5, line 45: "…basic discourse roles." I wasn't sure if you meant roles or rules, this would be an easy one to miss.*

*p. 7, line 40: "But although…" One or the other, but not both.*

*p. 7, line 56: "…our example above…" I assumed you meant example (3), but all other examples are also above.*

*p. 11, line 41: "..Which one it the toma?" it should be is.*

*p. 12, Figure 2 caption: Should Embedded and After be italicized here?*

*p. 17, line 39-41: "…participants' performance in Experiment 1 could reliably be distinguished from their performance in Experiment 2, …" The use of their implies that these are the same participants, which they are not.*

*There are minor APA errors in the references for: Bakeman & Adamson, Bion et al., McMurray et al., Siskund, Song & Fisher (2007), and Tomasello & Barton.*

We thank Reviewer 2 for these comments and have corrected each.

*1. The first comment from both of the original reviewers was that the term "word learning" was being used where it perhaps should not be. The authors have addressed this concern and have changed the title (as requested by both reviewers). Reviewer 2 notes that the authors had used "word learning" in the sense of "meaning learning," but this is the standard way in which "word learning" is used. It is good that the authors are now clearer, but generally the onus is on those not studying word meaning to qualify what they mean, e.g., "phonetic word learning." Reviewer 3 makes a compelling argument for not using "word learning" when only one meaning is taught per pair. Both the reviewer and the authors may be interested in a recent paper reporting empirical evidence that only providing one meaning leads to ceiling effects (Axelsson & Horst, 2013 Acta Psychologica)—which could be useful for planning future studies.*

Thank you for pointing us to this paper – we agree that this is very relevant as we think about future work in this domain.

*2. Reviewer 2 recommended the authors revise the abstract so it was less "jargon-y." The abstract has been revised, but it is just as jargon-y: "discourse information," "ambiguous reference disambiguation," "simpler heuristic" etc. Why not simply change the one sentence to read "Experiment tested whether this inference relied on the timing of the utterances. A new group…." For search engine optimization the authors may want to replace some of the jargon-y phrases with synonyms such as "to connect a new word with its intended meaning." Also, the abstract does not provide any information about the ages of the participants. Personally, I think this is important for this paper because a lot of the research on "ambiguous reference disambiguation" involves children at the lower end of the authors' age-ranges (i.e., 2-year-olds) and infants. A potential reader looking through the abstract may appreciate knowing if the paper reports infant data, toddler or preschool data.*

We have revised the abstract to be clearer and more accessible to readers. We have also added in the ages of participants.

*3. Reviewer 2 asked the authors to elaborate on how the Samuelson and Smith (1998) paper critiquing Akhtar et al., 1996 was "relevant to the present study." The authors have not done this. They have "elaborated in a more substantial discussion of Akhtar et al., (1996) on page 6" but this is not what Reviewer 2 asked them to do. Reviewer 2 asked for more information about how the "role of attention and memory factors" noted by Samuelsson and Smith related to the current study. Instead, the authors summarized Akhtar et al., in more detail and then were more explicit about Samuelson and Smith's critique. Then, instead of noting how that critique informed their own study (or possibly using it to set-up the importance of sequence of events, which is highly relevant to the current study), the authors write "we remain agnostic…." Personally, I think the introduction would be stronger WITHOUT this new Akhtar et al., paragraph. I think this Akhtar et al., reference should go behind*

*the other one on Page 5, line 60, then with a "but see Samuelson & Smith, 1998." The revised version doesn't address the reviewer's concern and is now distracting. (Also, we should only be granting authorship for those who made a substantial contribution. A corollary is that all authors made a substantial contribution. Thus, writing "Akhtar's experiment" and ignoring the other authors (Page 6, line 14) is poor referencing.)*

We thank the reviewer for these comments. We have removed the paragraphs discussing the Akhtar et al. and Samuelson & Smith studies from the introduction, and taken the suggestion to mention both references on p. 6. We have also included more details about the connection between our study and Akhtar et al.’s in the General Discussion. Thank you also for noting our error in referencing the study by only the first author. This citation error was unintentional and we have corrected it.

*4. Reviewer 2 noted that the original Figure 1 and Table 1 (now Table 4) were redundant. The reviewer is correct. The information is redundant and per the APA manual only one should be included. The authors have now placed the table in an Appendix because they wanted to keep both in the paper "to help clarify a complicated design." This is complicated and placing the information in the appendix will not be nearly as helpful for the reader as completing Figure 1 to depict all 4 conditions and then removing Table 1 (now Table 4). The authors cannot rely on the reader to look at the appendix and information that is vital to understanding the methods should be in the paper proper. (Pictures of stimuli, while nice, are not necessary to understanding a paper that is not about stimuli features and should remain in the appendix.) Revising Figure 1 will address Reviewer 2's concern, the authors' concern about helping the reader and follow the APA manual guidelines. Revise Figure 1 to*

*depict all 4 conditions (the dotted line is nice). Note, Reviewer 3 was also confused about the design and how many conditions there were (as was I when I first read the paper, despite the figure caption).*

We have removed Table 4 and modified Figure 1 so that all conditions are now depicted visually.

*5. Reviewer 2 questioned the decision to depict Experiments 1 and 2 together in the Figures and Tables. The authors make a compelling justification of keeping the paper as is. Now that there are direct statistical comparisons and given data was collected simultaneously for both experiments, I would rename them Experiments 1A and 1B, which may further elevate the reviewer's concern. There is precedent for this, see e.g., O'Doherty et al., 2011 Child Development (That's just the first paper that came to my mind).*

We have relabeled the experiments: They are now referred to as Experiments 1A and 1B rather than Experiments 1 and 2.

*6. Reviewer 2 asked the authors to make their rationale and predictions for Experiment 2 clearer. Because the authors' response did not include page numbers for where they made any changes, I could not find much evidence that they did this. The text from the end of the results to the start of the methods of Experiment 2 appeared largely identical between the original and revision with minor sentence breaks in the first paragraph after the Experiment 2 heading. The paragraph with the actual predictions (Page 15, line 32 on, "This design allows…") has not been changed at all. My understanding is that this is the paragraph the reviewer struggled with. I struggled with it too. I think the problem is the sentence: "If children recognize discourse continuity as a cut to reference, they should infer that new information contacted within a single topic is likely to also refer to that topic." What is this sentence doing here? Upon re-reading this section several times, I now understand that the authors are arguing they believe children recognize discourse continuity as a cut to reference but there is an alternative possibility, namely temporal proximity. Following from this, the authors should first state the predictions for that explanation (which they do on page 15, lines 39-45) and THEN remind the reader that if temporal proximity is not the driving force, that they should expect no difference/chance performance for both trial types, which is what they found.*

We have revised this paragraph on p. 14. We first state the predictions for a temporal associations account, and then provide the predictions under discourse continuity account. We have also removed the sentence listed above and hope that the explanation is now clearer.

*7. Reviewer 2 asked the authors to elaborate on their logic in the "driven by familiarity" paragraph in the general discussion (originally page 17 line 54 to page 18 lies14, now page 19 lines 37-54). The authors have added a helpful definition ("if [familiarity] were the case, children should map a new lab to an item already introduced by the experimenter"), but I do not believe this addresses the reviewer's concern for more explanation of the LOGIC here. Personally, I find this paragraph troubling because there are so many studies demonstrating that familiarity with items cause children to map a new item to something else even when all items are equally likely to be referents of the given name (see for example, Horst, Samuelson, Kucker & McMurray, 2011, Cognition; Kucker & Samuelson, 2012 Infancy; Mather & Plunkett, 2012 Cognitive Science and every single result in the Akhtar Diesendruck vs Sameulson Smith back-and-forth). When placed in the context of the wider literature,*

*I must agree with the reviewer that the logic as to why children would map a name to familiar item does need clarification.*

We have modified this paragraph on p. 19 to clarify the alternative account: that children may assume a speaker’s use of a novel label refers to an item that has already been discussed. We agree that this does not seem to us to be a plausible strategy when placed in the context of the broader literature – indeed we included it only because reviewers on our Cognitive Science Proceedings paper suggested this alternative explanation. We have retained it (in modified form) in this revision because we believe other readers might have this same concern as well, despite its disconnection from the broader literature. In the revision, we have tried to make our logic clearer: Linking novel naming events with the most proximate previous discourse referents could be a logically possible strategy but we don’t see evidence for this strategy in our data.

*8. Reviewer 3 explained that it was not entirely clear how the current paper was inspired from the literature reviewed in the introduction. The authors respond by noting "previous research has provided evidence for this" without supplying any references. Then, the authors explain—quite beautifully—that this paper is the first to isolate the role of discourse continuity. But, the reviewer clearly felt that this information was lacking from the introduction. Simply adding these sentences to the introduction would have addressed this concern. Instead, the authors claim to have clarified this in the discussion, but I am unable to find where they do this because no page numbers are provided in the response to reviewers.*

On p. 5, we now state that “This body of work [on discourse information] as a whole suggests that children's production and comprehension shows substantial evidence of understanding basic discourse structure, though no studies to our knowledge link this knowledge to word learning specifically. Making this link is one of the goals of the current studies.”

*9. Reviewer 3 notes the description of Table 3 (now Table 2) was confusing regarding how proportions were calculated. The authors claim to have addressed this concern, but they have, in fact, only added "and the second toy in a Second Toy trial" to the text, which again was not what the reviewer was seeking clarification on. The authors should realize that if a reviewer, who is reading very closely and carefully, is having a difficult time understanding something, then a "regular reader" who might be skimming or reading quickly before a lab meeting or before trying to cite this paper in his/her own paper is likely to also be confused or have a difficult time understanding. The authors should want to make their paper as clear and easy to understand as possible so others cite the paper and do not cite it incorrectly, for example, thinking the authors calculated proportion incorrectly.*

Reviewer 3 wrote, “I was slightly confused by the proportion value in Table 3 [2 in the current ms]. Is this the mean of the proportion correct over the presentation orders?” We thought that we had clarified this source of confusion by specifying that the mean proportion correct over presentation orders refers to the mean proportion of selecting the first toy in first toy trials and the second toy in second toy trials. We appreciate the comment that we want our work to be accurately interpretable across readers and have tried to clarify this issue further in the caption.

*10. Reviewer 3 suggested a helpful condition to control for a confound in the current study (the number of comments before the naming events). I understand the authors not collecting more data, but they authors did decide to collect additional data with adults (but this is not mentioned in the paper). However, for some reason the authors did not run the conditions the reviewer recommended (with and without a third comment) and instead ran two other conditions (toy 1 or toy 2 named, all with a third comment). It does look like the new data support the claim "we found that responding was identical whether one or two comments were provided before naming" when compared to the adult data in the paper (Experiment 1). These data should be added to the paper, perhaps at least to the appendix.*

We apologize for the confusion – we actually did run both conditions of this study, but did not make the experimental design clear in our response. We have included both conditions now as a control study in the Appendix.

*A. First, Figure 2 must be revised. At the moment the data are depicted using line-graphs. This is only appropriate if the variable on the x-axis is continuous. Age is continuous but only for within-subjects designs, which this is not. The authors MUST use bar-graphs. I know the line-graphs look really great, but the bar-graphs will too. The contrast will still be apparent.*

We have complied with the Reviewer’s suggestion, since it is true that the comparison to adult data on a continuous horizontal axis is awkward. Please see the revised Figure 2 on p. 12.

Nevertheless, we respectfully disagree with the Reviewer that line graphs are always inappropriate for displaying between-subjects developmental data. Age is a continuous variable and should be represented this way when the authors are interested in depicting trends across that continuous variable (Zacks & Tversky, 1999). This plotting convention is followed in a large number of well-respected and highly-cited papers that use between-subjects designs, including (the first three that come to mind): Davidson et al. (2006), *Neuropsychologia*; Gathercole (1999), TiCS; Deloache (1987), *Science.* In fact, some theorists of visualization believe that the use of bar plots should be discouraged more broadly.

*B. Second, once Figure 2 is changed to a bar-graph, Table 2 (previously Table 3) will be redundant with Figure 2 except for the p-values. Therefore, Table 2 should be removed and asterisks added to Figure 2 to denote which means are different from chance (which is not what the p-values in the table are: these should also be depicted as well). Asterisks will add beauty to the Figure to make up for the loss of the lines.*

We respectfully disagree with the Reviewer’s suggestion that we add asterisks to Figure 2. As part of a broader move away from emphasizing Null Hypothesis Significance Testing in psychology (e.g., Cumming, 2013, *Psychological Science*), we believe that it is not appropriate to give prominence to *p* values rather than to means and confidence intervals (which allow for statistical inference “by eye”; Cumming & Finch, 2005, *American Psychologist*).

For purposes of meta-analysis, means and standard deviations are important, non-redundant information that the APA explicitly mandates for reporting (APA Publications and Communications Board Working Group, 2008, *American Psychologist*). In addition, providing test statistics is also recommended. We thus retained Table 2 in the revised text.

*C. While we're on this topic: the participants are described in the text as "2-year-olds" (page 10, line 16) but in the Figures/Tables as "2-3." This wording should match. It's easier for the reader to focus on the main aspects of the paper when there is a lot of consistency.*

We have changed the labels so that ages are referred to by year of age.

*D. FYI: In my region it would violate our privacy acts to identify participants by their birthdates. Might be worth thinking about an alternative way to ID participants in case the laws in the authors' area change.*

As our lab transitions to making all data publicly available, we have been developing policies to address questions like this one. We have been in consultation with our IRB regarding the specific issue of posting birth dates. While there is no issue in posting dates of birth alone, in combination with other information (e.g., date and location of test), they could be considered identifiable information. We have hence removed birthdates from our posted files and now only identify age at time of test to one significant digit.