I reviewed this paper previously, and I appreciate the chance to weigh in again. I have now read the response to the reviews and the revised paper. Let me begin by reiterating a key point from my previous review—this is an important paper, it should absolutely be published, and it merits publication in this journal. My goal in this review, however, is to ensure that this exemplary science is backed up with rigorous and careful writing to let the science shine. Simply put, the science here is strong enough that the authors don't have to do a lot of interpreting. Just let the data speak.

What do the data say? I see three facts:
1. The basic SC effect is replicated in the 1-3 order regardless of the stimulus presentation details
2. The basic SC effect is not replicated in the 3-1 order *with the effect primarily impacting the one exemplar trials.*
3. The SC effect is reversed in the 3-1 ordering as a function of the presentation timing with broader generalization on the 3-sub trials with sequential presentation.

These three facts are taken directly from the paper. Only points 1 and 2 are in the results section. Point 3 is raised in the general discussion. I would recommend that this be included in the results section as this analysis follows directly from the previous literature (i.e., it examines whether a key effect in the literature was replicated).

If we agree on the facts, the next question is what do these facts mean? I will be honest and say that my read of the paper suggests that the authors are happy with a Bayesian status quo interpretation—that the data are generally consistent with XT's account, and that the data en masse don't support the SPSS account. I suspect the authors will agree with my assessment. But is this interpretation warranted by the facts?

I think not. My conclusion stems from the answer to a central question: what is the SC effect about? At the heart, it's about the size principle (see equation 5 from XT 2007 Psych Review), also called 'strong sampling' in XT 2007 Developmental Science (and here). In their own words, "hypotheses with smaller extensions assign greater probability than do larger hypotheses to the same data, and *they assign exponentially greater probability as the number of consistent examples increases*." [emphasis added]. Critically, the primary data that test the size principle are from the 3-sub trials—the size principle doesn't modulate anything if there is only one exemplar—and how generalization narrows as a function of the number of exemplars at a particular generalization level. This logic is evident in, for instance, XT 2007 Dev Science where they only had children generalize with multiple subordinate items (i.e., there were no 'one exemplar' trials).

If I look at the data here, the 3-sub trials do move around, *increasing* in basic level generalization the sequential 3-1 order condition (see point 3 above). This effect is small but significant. It replicates the key effect from SPSS. *It is not explained by XT's model.*

The biggest effect, of course, is on the one exemplar trials. There, generalization narrows or broadens based on the trial order. But here's the rub—*XT's model and strong sampling doesn't predict this effect AT ALL.* The authors have come up with a reasonable explanation for this effect based on pragmatics, but scientifically, the authors should acknowledge that

this effect is not consistent with XT's model. There is nothing in the size principle or the equations in XT's Bayesian model that says that generalization on the one exemplar trials should bop around as a function of trial order.

Concretely, if we agree on the facts above and if we agree on how equation 5 from XT's model creates the SC effect, then here are the conclusions I see…
1. The basic SC effect is replicated in the 1-3 order regardless of the stimulus presentation details [*XT's model explains this quite well*; the data here provide an important qualification on the SPSS interpretation]
2. The basic SC effect is not replicated in the 3-1 order with the effect impacting the one exemplar trials only. [*This is not explained by XT's mod*el; SPSS showed the same effect but did not offer an interpretation]
3. The SC effect is reversed in the 3-1 ordering as a function of the presentation timing with broader generalization on the 3-sub trials with sequential presentation. [*This is not explained by XT's model*; this replicates the effect reported by SPSS]

In summary, then, the data reported here do not support XT's model—2 of the 3 effects are not consistent with the model. I would like to see the authors at least acknowledge this.

Beyond that, the authors are, of course, free to offer new explanations for the data. Their pragmatic interpretation seems reasonable. I thought the explanation of the reversal to be a bit weaker.

Below I provide several specific comments. But let me be clear about the key revisions I would like to see:
1. Move the stimulus presentation timing result from the GD to the results (see point 3).
2. Acknowledge that findings 2 and 3 from the paper are not consistent with the size principle (strong sampling) from the XT model.

Detailed comments (note: page numbers are PDF page numbers…)

- Abstract: "Yet both children and adults successfully learn noun meanings at the correct level of abstraction from similar evidence." What's the 'correct' level in this example? It depends. It could be Dalmatian, could be dog, could be animal. And what's the 'true' underlying category? There isn't one. Finally, 'making certain patterns of examples more consistent with a subordinate meaning than others'…this is a really loose summary of the SC effect. Please be more precise.
- P5, line33: should be basic memory and comparison processes (or perceptual processses) [the SPSS account did not focus solely on memory]
- P6, line 16: should be 'children and adults' [since the experiments reported here included adults]
- P6, line 16: not sure why Lawson goes against SPSS—if I highlight commonalties and all three exemplars are chili peppers, wouldn't that lead to subordinate level generalization which is what we all find?

- P7, line 57: something is off here. Perhaps it should be: 'participants are aware of the exemplars from the previous trial and therefore do not interpret the single exemplar as the only observed exemplar from the target category."
- P 9, line 42: suggested edit: 'while in the key conditions from SPSS, participants saw…' This is important because we did replicate XT in the other experiments.
- P 10, line 56: typo (an extra 'not')
- P 14, line 10: the SC effect suggests a 'powerful mechanism" by which learners might overcome… 'Powerful' seems like a stretch to me since the effect is pushed around across conditions, but ok…
- P 15, first full paragraph: the authors replicate our effect, but offer a new interpretation. This is ok, but perhaps the authors could at least acknowledge our explanation of the same effect?