In this revised registered report, the authors have addressed some of my previous concerns. However, the link between the study and theory, and what is mentioned in the introduction, remain unclear, and some of the analyses are not described in sufficient detail.

My main concern is that the introduction still reads as if it the study will ask a question about whether bilinguals’ language balance relates to CSSL, but this has now become only exploratory. Throughout the introduction, whether bilinguals were balanced or not in previous studies is mentioned inconsistently, which leads the reader to think that this study might look at balanced bilingualism as a reason for possible incongruent findings to date. But now it is only an exploratory analysis, and only for one of the hypotheses (though I could imagine that potential effects of last competitor accuracy on mutual exclusivity could be related to language balance). To be clear, I think looking at whether balance across languages matters should be exploratory, since it isn’t clear what the theory behind why it would matter is, but it feels like the introduction was only minimally edited in response to previous concerns. Below are more comments related to this issue.

* Pg 3-4, why split out that the bilingual advantage for word learning was found in balanced and unbalanced bilinguals, seems like it is just found in bilinguals
* Pg 4 – “the goal of the study is twofold – understand if there is a bilingual advantage and to look trial by trial”, so why mention whether participants are balanced bilinguals or not sporadically?
* Pg 7 – new paragraph starts with “Hypotheses 1-4”, I don’t think I necessarily agree that more balanced bilinguals = better for the regulation hypothesis, or the learning adaptation hypothesis. I’m not sure what the mechanism would be? If the L1 regulation hypothesis has to do with more regulatory experience in their L1, I’m not sure that has anything to do with their “balance” in experience/exposure.
  + Relatedly, this paragraph calls these “Hypotheses 1-4” but they are not called that in text when they are described, and then the study itself has H1—H5, so I would recommend not calling them this.

While Hypotheses 1, 2, and 3 (for the RR) are clear, H4 and H5 are not particularly well described.

* For one, it isn’t clear how last-competitor accuracy is calculated. Since there are always two competitors per trial, is it averaged accuracy across the two from the last time each one was the target? And for 2:1 mappings, how is accuracy on the other mapping considered in this, because it seems like it could be a competitor right?
* Another concern is that I don’t fully understand how these are different from finding an effect of mapping type in H1 or an interaction between mapping type and language group in H2. If the interpretation is about mutual exclusivity, then finding that bilinguals do better on the 2:1 mappings would already tell you that they rely less on mutual exclusivity. What will this analysis add?
* It’s also not clear how your effects can be interpreted. I think that for 2:1 mappings maybe the expectation is that there is no effect of last-competitor accuracy, because it shouldn’t matter whether you got that right or wrong if it needs to be remapped, but lack of clarity on how that’s calculated for 2:1 trials which are not blocked and could be the target or the competitor on other trials makes it hard to understand. The language in the table at the end says something like “if bilinguals show less of an effect of last-competitor accuracy” but what is less? Just not significant, not as significant, smaller betas?

Some additional comments

* Interlingual homographs like pie in English and Spanish don’t seem super related towhat is described at this point (in spoken language they sound differe) but seem more related to this specific paradigm, but that isn’t clear
* Edit on page 6 – “advance over monolinguals is due to having more regulatory skills in their L1 (first acquired and often most proficiency language)…” – it seems like a citation would be necessary here, there are so many paths of bilingualism (which was brought up in the first draft), and first language is not necessarily the most proficient for a lot of people, particularly depending on age of acquisition of the second language and the context (e.g. schooling).
* Pg 6 – “therefore, bilinguals may have better cognitive control than monolinguals, which may in turn facilitate word learning” – what is the link between cognitive control and word learning? There is also an example “(e.g. higher inhibition may reduce interference from competitor referents)”, but that seems like maybe it’s just for bilinguals? So is it the better cognitive control or is it the experience of being bilingual?
* Are participant exclusions included in the sample size? E.g. will you recruit 150 and then exclude participants, or is 150 the non-excluded number?
* Bayes Factor – on Page 27 is says a BF larger than 3 is moderate support for the alternative hypothesis but a BF smaller than 1/3 is moderate support for the null, but on page 19 it said you would use 6 or 1/6? Is the 6 or 1/6 specific to H4?
* Can the authors spell out how low/high entropy maps to balance? I assume that high entropy scores = high diversity of outcomes = more use of both languages (e.g. less predicable use of each language), but I am not sure this is correct
* Pg 30 – how does only including participants that reach 40% accuracy during the last block ensure that you have enough usable trials?