**Reviewer comments relating to analysis in the Dots versus Numerosity manuscript**

**Reviewer 1**

**Point 2)**

Mean RTs in figures should be included within the results. This can be done for each condition. Is it better in text or within a table (as descriptives?).

Analysis script should be reproducible (currently working on this). Valter, I think it would be good for you to run the analysis scripts on your machine.

**Point 3)**

Significant effect of congruency needs mean RTs reported to judge magnitude of the effect.

**Point 4)**

"Investigate whether symbolic and non-symbolic numerals interact…" was apparently not tested. Wasn’t this tested via the 3-way ANOVA? If so, perhaps this needs spelling out a bit more in text?

**Reviewer 2**

**Point 3) (Running analysis for errors)**

Due to the limited amount of errors, we have decided to omit this analysis?

**Reviewer 3**

**Point 2)**

The sample/power estimation being based on a 3-way interaction. From a quick glance, the paper the reviewer cites seems to only relate to single factor designs with 3 levels, or two factor designs. Regardless of that, our power calculation specifies a large effect size. For small and medium effect sizes, we would need sample of 786 and 127 respectively (and thus I agree with the reviewer). We might have to outline that our study is only suitably powered to find large effects?

We can further discuss the moderate evidence of the Bayesian analysis but to my eyes it just supports the findings from the frequentist analysis. The ANOVA had a medium effect partial eta squared (0.129) and the Bayes factor was moderate. These findings corroborate with each other?

**Point 3)**

Analysing SNARC within time bins. Idea being that it might take more time for non-symbolic numerosity’s to mapped onto spatial codes. This seems like an interesting analysis. Couple of questions from me

1. For this analysis, we should probably include outliers. If this is the case, we should probably keep outliers for the other analyses (I have also been in favour of keeping outliers within the dataset).
2. How many time bins should we have? I assume each bin would have to have a similar number of observations for the analysis to be successful?

**Future analysis (end of July)**

* Means and SDs table
* Hand-number congruency \* Presentation-Congruency \* Symbolic
* Fully congruent vs partially congruent
* Make analysis script reproducible

**Parity data (end of August)**