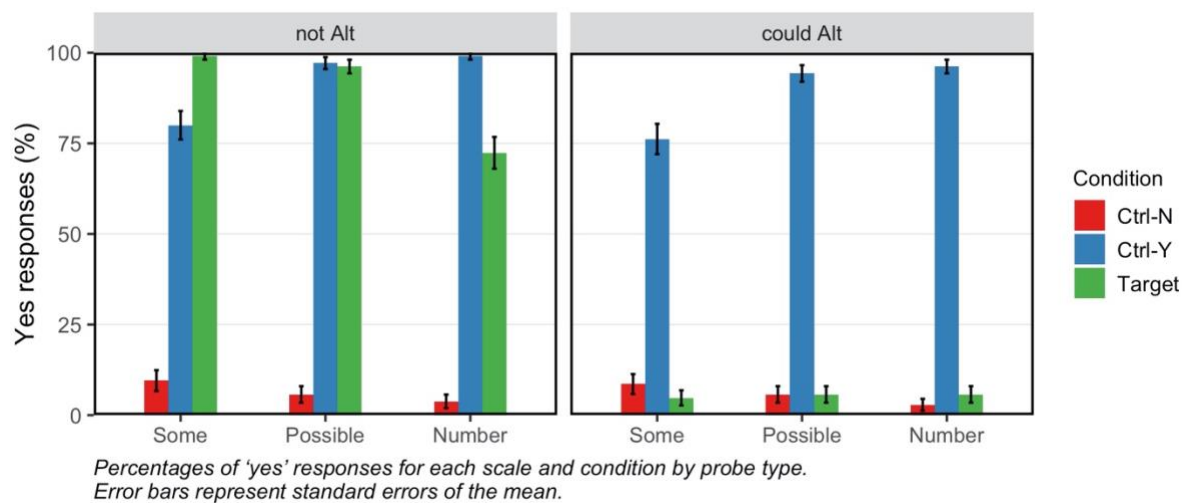


Partitive follow up results

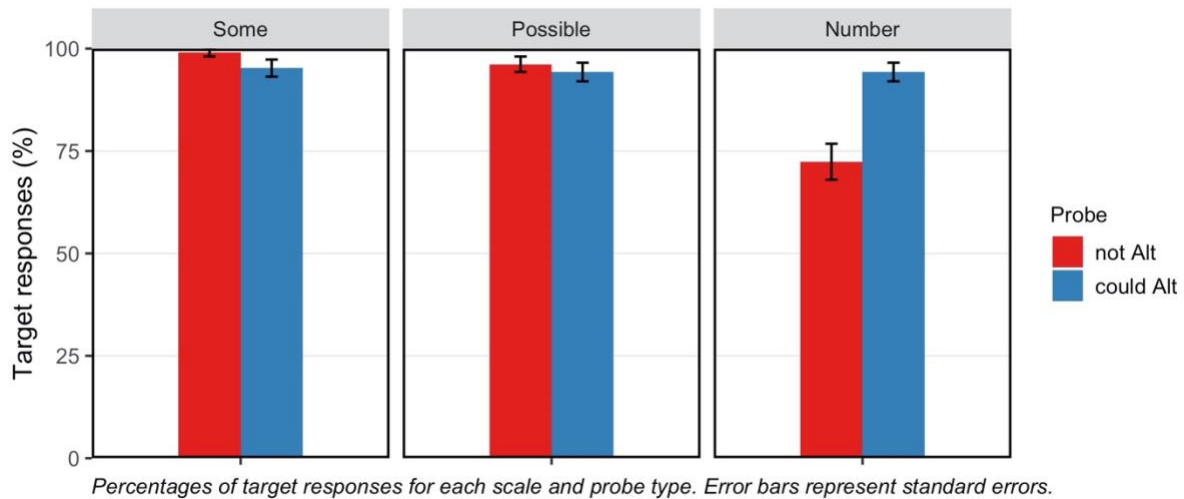
Five participants were removed because more than 20% of their answers to control items were incorrect. Figure 1 shows the percentages of ‘Yes’ responses for each scale and condition by probe type. The mean accuracy of the control items was 92% (control yes: 90%, control no: 94%).

Figure 1



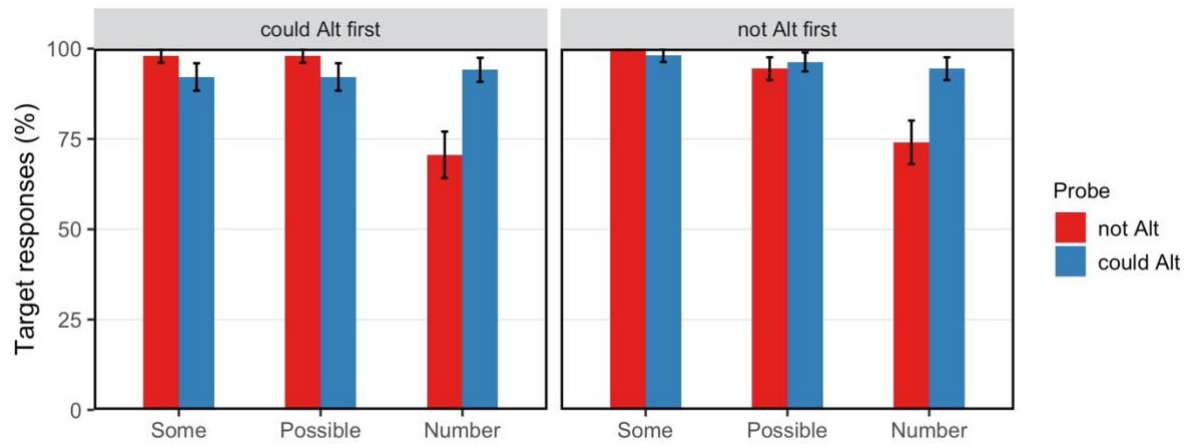
We coded the ‘Yes’ response in the ‘not Alt’ condition and the ‘No’ response in the ‘could Alt’ condition as target response. Figure 2 shows the percentages of target responses for each scale and probe type. We fitted a mixed effects logistic regression model predicting response (target or non-target) from probe type, scale, block order (‘not Alt’ first or ‘could Alt’ first), and their interactions, including random intercepts for participants¹. Random slopes were dropped due to non-convergence or singularity.

¹ Mixed-effect analyses were conducted in R (R Core Team, 2020) using the ‘lme4’ package (Bates et al., 2015) and the ‘lmerTest’ package (Kuznetsova et al., 2017). Scale was dummy-coded, probe type and block order were deviation coded. Model comparisons were conducted to test the significance of fixed effects with more than two levels, using likelihood ratio tests. Significant interactions were followed up by conducting analyses on subsets of data defined by the levels of relevant factors.



There was a main effect of scale ($\chi^2(2) = 35.61, p < .001$) and a significant interaction between scale and probe type ($\chi^2(2) = 19.71, p < .001$). Planned comparisons within the levels of scale revealed that the probability of target responses was higher in the ‘not Alt’ condition compared to the ‘could Alt’ condition for ‘some’ ($\beta = -2.6, SE = 1.46, p = .07$), no difference was found for ‘possible’ ($p = .44$). This effect was reversed for NNPs: the probability of target responses was significantly higher in the ‘could Alt’ condition compared to the ‘not Alt’ condition ($\beta = 2.84, SE = 0.70, p < .001$). Planned comparisons within the levels of probe type revealed that the probability of target responses was higher for scalars than for NNPs in the ‘not Alt’ condition (some: $\beta = -4.07, SE = 1.07, p < .001$; possible: $\beta = -2.61, SE = 0.62, p < .001$). In contrast, in the ‘could Alt’ condition, there was no difference between the three conditions.

There was no significant three-way interaction. See ‘by-block’ plot in the next page.



*For each block order, the percentages of target responses for each scale and probe type.
Error bars represent standard errors.*