Distractor heterogeneity in visual arrays may affect referring expression production

Two psycholinguistic studies examined how homogeneity among distractors objects affects which features are mentioned when describing a target. The results show no effect of homogeneity, the possible reasons for which are discussed briefly below.

Current referring expression generation (REG) models (Dale & Reiter 1995) incrementally select features to incorporate into referring expressions, but do not account for speakers' decisions between multiple features when each feature uniquely identifies the target (van Deemter, Gatt, van der Sluis & Power, 2012). We hypothesized that this choice may be driven by the relative visual accessibility of the different features. A target is harder to find when surrounded by visually heterogeneous distractors (Duncan & Humphreys, 1989; Wolfe, 2007). This effect is hypothesized to extend to REG, such that a feature which varies heterogeneously among distractors is less likely to be mentioned than one which is homogeneous among distractors.

Prior visual search experiments on homogeneity use simplified stimuli, e.g., arrays of matching-orientation rectangles with a target rotated 20 degrees (Poiese, Spalek & Di Lollo, 2008) or arrays of different-sized circles (Ariely, 2001). These stimuli lack ecological validity, and the current visual search task seeks to verify the effect of distractor homogeneity when stimuli are more complex. Our referential communication task then seeks to explore any effects of homogeneity on reference.

Together these two studies offer the REG literature the first test of how heterogeneity among distractors influences the salience of a target's features during selection. Our results did not support our hypotheses, but future work should explore this area further in order to inform modifications to REG algorithms as well as to enhance models of human integration of visual and linguistic information.

Experiment 1: Visual Search

Experiment 1 used a visual search task to test whether the reported target-distractor similarity effects from prior work extend to distractor homogeneity, specifically for contexts such as Figure 1. Stimuli were images of real world objects with two obvious categorical features, expressed such that the target was always unique on both features, and the distractors were unique in their conjunction of the two (see Figure 1). Participants were presented with a cue depicting a single feature of the target, and then saw an array – either the homogeneous-flavour version or the homogeneous-topping version of the trial. Importantly, the cue consisted of only one feature, the flavour or the topping, that pertained to the whole target item: the ice cream cone.

If distractor homogeneity affects search efficiency with these complex stimuli, then participants should find a target item faster when cued by a feature that is homogeneous among the distractors than a cue that is heterogeneous among the distractors. For example,

the ice cream flavour in Figure 1A is homogeneous among the targets – all of the cones have vanilla ice cream – so the flavour cue should facilitate finding the target, compared with the same cue in Figure 1B, which should take longer as the ice cream flavour is heterogeneous among the many-flavoured ice creams. Conversely, a topping cue should facilitate search in the right hand array, where topping is homogeneous, and take longer in the left hand array, where it is heterogeneous.

Experiment 2: Referential Communication

Experiment 2 uses the materials from Experiment 1 for a referential communication task. New participants were presented with the arrays and asked to describe the target such that a hypothetical listener could find it quickly. The position of the target was indicated before the onset of the array, with a bounding square encompassing the space the upcoming target would take up. This eliminated any search element of the referential communication task, disentangling the two processes.

If referring expression generation is sensitive to homogeneity, participants should mention the feature that is homogeneous among distractors more often than the feature that is heterogeneous among distractors. In the case that participants mentioned both features, we predicted that the homogeneous feature would be mentioned first.

Results and Discussion

Both experiments were analysed with multi-level modelling. Models specifying random effects for participants and images were not improved by specifying a fixed effect of homogeneity. Therefore the current experiments did not find an effect of homogeneity on search efficiency or reference production.

Due to a lack of resources, it was not possible to compute saliency maps of the stimuli. Thus, there may have been a confound that resulted in the targets being more salient than the distractors. It is also possible that homogeneity did affect participants' behaviour but the units of measurement in the current study were not fine-grained enough to detect this. Future studies should use computer mice with higher sampling rates, and may wish to count the number of words speakers used to describe each feature (Clarke et al., 2013) and incorporate analysis of the pitch and length of words (as in Engelhardt & Ferreira, 2014).

Key references

Ariely, D. (2001). Seeing sets: Representation by statistical properties.Psychological Science, 12(2), 157 - 162.

Clarke, A. D. F., Elsner, M. & Rohde, H. (2013). Where's Wally: The influence of visual salience on referring expression generation. Frontiers in Psychology, 4, 329.

Dale, R. & Reiter, E. (1995). Computational interpretation of the Gricean maxims in the generation of referring expressions. Cognitive Science, 19(8):233–263.

Duncan, J. & Humphreys, G. W. (1989). Visual search and stimulus similarity. Psychological Review, 96(3), 433 – 458.

Engelhardt, P. E. & Ferreira, F. (2014). Do speakers articulate over-described modifiers differently from modifiers that are required by context? Implications for models of reference production. Language, Cognition and Neuroscience, 29(8), 975 – 985.

Poiese, P., Spalek, T. M. & Di Lollo, V. (2008). Attentional capture by a salient distractor in visual search: The effect of target-distractor similarity. Canadian Journal of Experimental Psychology, 62(4), 233-236.

van Deemter, K., Gatt, A., van der Sluis, I. & Power, R. (2012). Generation of referring expressions: Assessing the Incremental Algorithm. Cognitive Science, 36(5), 799 - 836.

Wolfe, J. M. (2007). Guided Search 4.0: Current Progress with a model of visual search. In W. Gray (Ed.), Integrated Models of Cognitive Systems (pp. 99-119). New York: Oxford.

Appendix

Figure 1 A

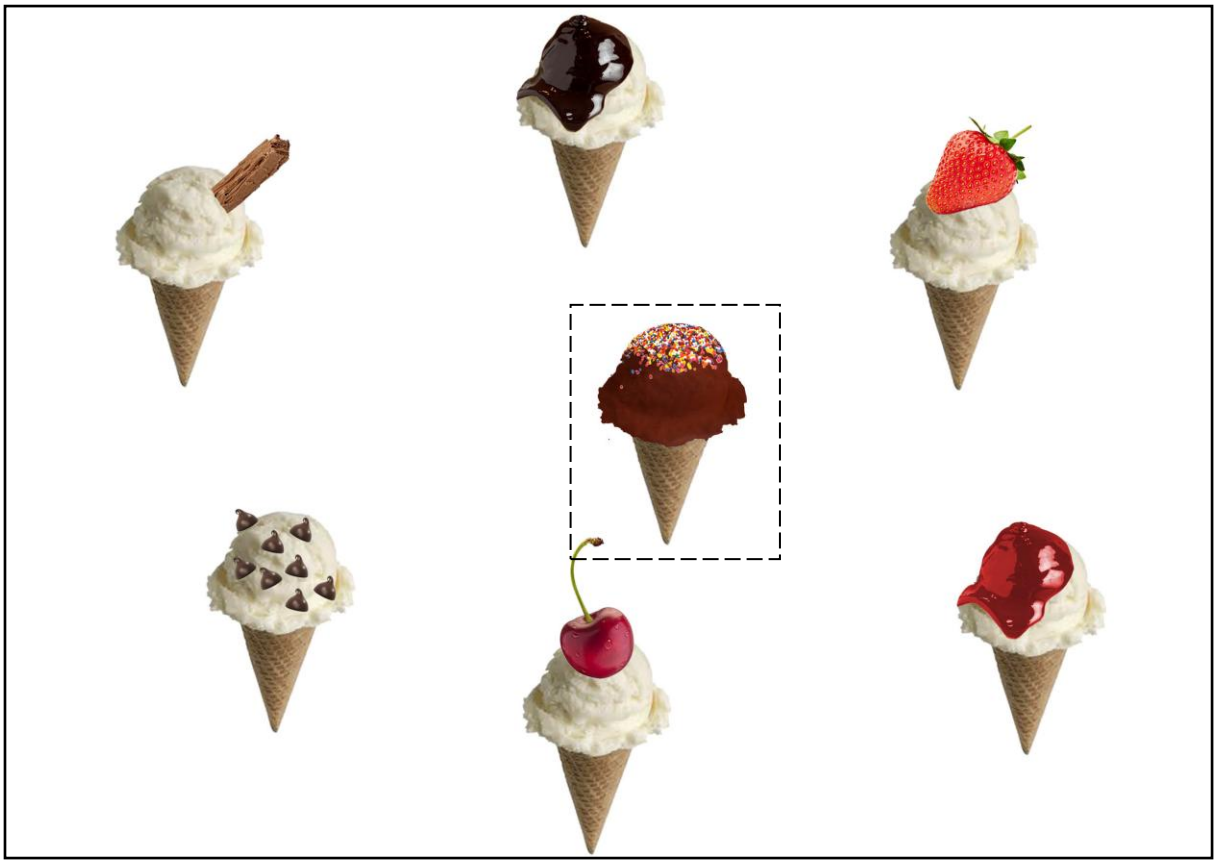

Figure 1B