Perceptual Difficulty Differences Predict Asymmetry in Overmodification with Color and Material Adjectives

When referring to objects, speakers are often more specific than they need to be for establishing unique reference (e.g., "the green plastic bottle" instead of "the plastic bottle" in Fig. 1). Adjectival overspecification patterns are not random, but structured: color adjectives are produced redundantly more often than size or material adjectives [1-5]; color adjectives are more likely to be produced redundantly with increasing scene variation [5-7]; and adjectives are more likely to be produced redundantly, the more atypical the property denoted by the adjective is for the object under discussion 5;8-9]. The only current computational model of referring expression production that accounts jointly for all of these patterns is couched within the Rational Speech Act framework [9] and assumes that adjectives differ in how noisy, and consequently, how useful they are for the purpose of establishing reference [5]. One hypothesis about the nature of this noise is that it reflects the perceptual difficulty of establishing whether the property denoted by the adjective holds of the contextually relevant objects. Here, we take a first step towards testing the prediction that systematic differences in the overmodification patterns observed for color and material adjectives can be explained by a difference in perceptual difficulty of establishing whether objects are of a particular color or material. In Exp.1, we norm the perceptual difficulty associated with establishing whether an object exhibits a color or material and select objects with highest and lowest perceptual difficulty for testing in Exp. 2. In Exp. 2, we test in a reference game wheter adjectives that denote more perceptually difficult properties indeed are less frequently produced redundantly.

Exp. 1 (n=120): we collected perceptual difficulty norms for 12 objects. Participants saw pictures of objects and adjectives and were asked to indicate whether the object had the property denoted by the adjective or not by pressing a key as quickly as possible. Each object occurred in two or three different materials and in three different colors. All resulting 81 images were separately normed for object nameability and feature nameability. Material words resulted in higher error rates (β= 0.40, SE=0.09, p<.0001) and greater response times (β=5.46, SE=4.73, t=11.55, p<.0001) than color words. We grouped the 8 image-material adjective pairs with the highest error rate and response times into a *high difficulty group*, and the 8 image-color adjective pairs with the lowest error rate and response times into a *low difficulty* group.

Exp. 2 (n=50 pairs) was an interactive reference game used to elicit production probabilities of redundantly mentioning color and material adjectives for the high- and low-difficulty items normed in Exp. 1. Participants used a chat box to communicate the target object to the listener, who in turn selected an object. On critical trials (see example in Fig. 1), participants saw a display with 4 objects of the same type in which either *color* or *material* was redundant for distinguishing the target (marked by a green border in the speaker's display). On high-difficulty trials, mentioning the material was redundant; on low-difficulty trials, color was redundant. We classified the produced utterances as 'color-and-material' (redundant), 'only-color' or 'only-material'.
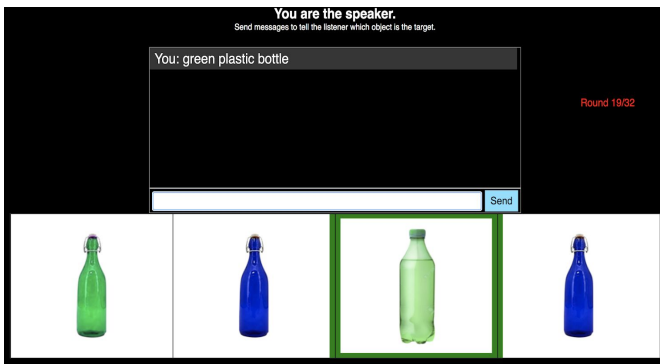

Fig1: Example display from the speaker's perspective on a *low-difficulty* (*color redundant*) trial.
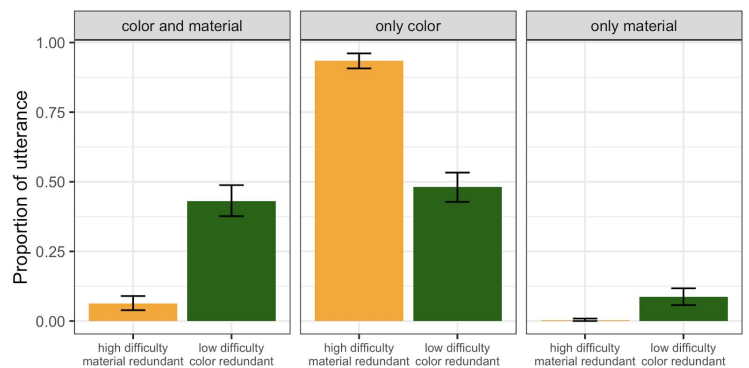

Fig. 2: Proportion of redundant "color and material" utterances vs non-redundant utterances in high and low difficulty trials.

Results. Speakers were more likely to redundantly mention color than material (β= 2.11, SE=0.62, p<.0001, see Fig. 2), replicating the previously observed asymmetry between overmodification with color and material adjectives on a new set of items. Given the norms from Exp. 1, this suggests that the more difficult it is to judge whether an object has a feature, the less likely speakers are to redundantly mention that feature, providing initial support for the perceptual difficulty hypothesis of [4]. Future work includes testing the strong version of this hypothesis by manipulating perceptual difficulty within property type as well as testing further hypotheses about the adjectival noise source, including the semantics of the adjective and the past communicative success in using the adjective.

References
[1] Pechmann (1989). [2] Sedivy (2003). [3] Gatt, van Gompel, Krahmer & van Deemter (2011). [4] Rubio-Fernandez (2016). [5] Degen et al. (2020). [6] Davies & Katsos (2013). [7] Koolen, Goudbeek & Krahmer (2013). [8] Westerbeek, Koolen & Maes (2015). [9] Mitchell (2013). [10] Goodman & Frank (2016).