Prototypicality predicts the mention of state information in referring expressions

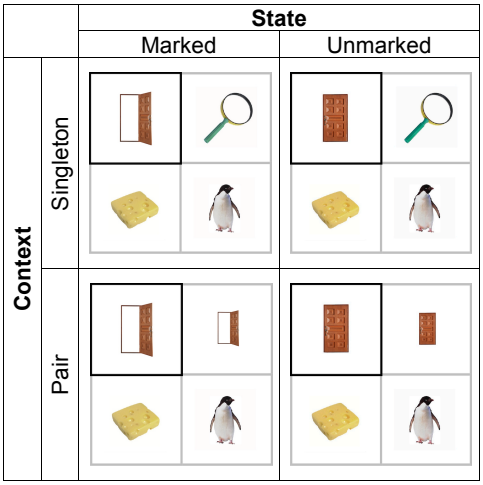
Megan Parker and Daphna Heller (University of Toronto)
megan.parker@mail.utoronto.ca

When speakers formulate a referring expression for an object, they may include different amounts of information and types of properties. For example, when referring to a door that is big, brown, and open, a speaker can say *the door*, *the big brown door*, or *the open door*; all could potentially allow an addressee to identify the intended object. It has been observed that speakers generally follow Grice's (1975) maxim of quantity and include just the information needed to differentiate the intended referent from other objects in the context [1]. However, in many cases speakers nonetheless include properties that are not necessary for differentiation, a phenomenon known as over-specification [2]. Some modifiers, such as scalar adjectives (e.g., *long*, *small*), are rarely overspecified, whereas others, most notably color adjectives, are overspecified at a high rate [1,3]. Interestingly, overspecification of color occurs at a low rate when the color of the referent is predictable (e.g., a yellow banana) [4]. One account suggests that speakers tend to overspecify properties that are *atypical* (e.g., a blue banana) [5]. Here we ask whether this explanation extends to state modifiers (e.g., *dirty/clean*, *open/closed*). A secondary goal is to examine whether speakers are more likely to include unnecessary modifiers when the context requires producing a (different) modifier to differentiate the referent, and thus speakers are already planning a modified noun phrase, or if overspecified modifiers are actually produced in contexts where there are no *required* modifiers.

Method. In a referential communication task in a lab setting, participants performed the role of director, with the experimenter as the addressee. On each trial, four images appeared (see Figure) and one was highlighted; participants were asked to instruct the addressee to click on the highlighted image. We used a 2x2 design.

- STATE manipulated the state of the target object (e.g., an open vs. a closed door). States were classified as (un)marked by 7 individuals (see Table). Note that state modifiers are *never* necessary and thus are all cases of overspecification.
- CONTEXT manipulated whether the target was the only object of this nominal category (singleton), or whether it appeared alongside a second, identical image contrasting in size (pair). This was done to examine whether the requirement of another modifier affects the likelihood of overspecification.

The experiment had 148 experimental items and 74 fillers.



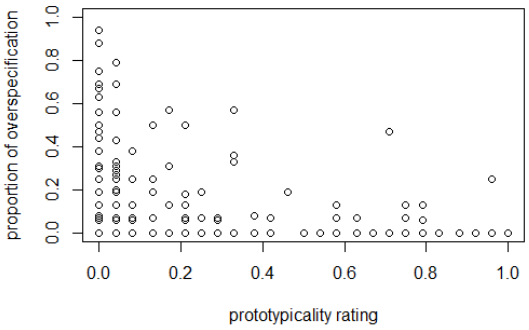
Marked	Unmarked	Number of items
Broken/ ripped	Whole	22
Sliced	Unsliced	9
Rolled	Unrolled	4
Peeled	Unpeeled	6
Dirty	Clean	19
Folded	Unfolded	8
Bent/ curved	Straight	18
Empty	Full	6
Open	Closed	50
Tied	Untied	2
Wet	Dry	4

Results (n=32). Trials in which the participant was under-informative or incorrectly labeled the object were eliminated (6%), resulting in a total of 4434 usable trials. Our first analysis examines the rate of state modifiers across the four conditions:

	Marked	Unmarked	
Singleton	24%	1%	13%
Pair	18%	0.5%	9%
	21%	1%	11%

A mixed-effects logistic regression model with state (marked vs. unmarked) and context (singleton vs. pair) as predictors revealed, first, a significant main effect of state (z = -9.69, p < .001); speakers produced significantly more overspecifications for marked states (21%) than unmarked states (1%). The main effect of context was also significant (z = -3.25, p = .001), indicating that overspecification was more likely when speakers did *not* need to produce another modifier that is required for referential success (the interaction was not significant, p = .58).

Our second analysis examined markedness by individual object. To this end, we obtained prototypicality ratings of the same images from new participants (n=24): Participants were shown two images of the object in both states (e.g., an open and a closed door), and had to indicate which one was more prototypical (we introduced prototypicality by showing pairs of images that contrasted in shape, color and material). We then correlated, for each object and state, the prototypicality of an object in a given state with the rate of overspecification in production. Pearson's product-moment correlation revealed a significant negative correlation of intermediate-high strength (r = -0.5237, t = -10.542, *df* = 294p < .0001). For example, *peeled* garlic was judged to be rather prototypical (58%), and speakers uttered *"peeled garlic"* at a rate of only 13%. In contrast, a *peeled* orange was judged as less prototypical (4%), and speakers uttered *"peeled orange"* at a much higher rate (56%). Note that these ratings also support our original markedness classification: For *most* objects, the marked state was overspecified, but for some, it was the state classified as unmarked that was mentioned. For example, *open* was mentioned for several objects, but for *laptop* and *eye*, *closed* was mentioned. Similarly, for some objects both states of *empty* and *full* were mentioned (e.g., *jar*, *wine glass*), and for others only one of the two modifiers was ever used (e.g., *empty* for *egg carton*).



Conclusion. First, we obtained a novel finding regarding the production of modifiers: speakers prefer not to use more than one modifier (because the overspecification of state information was less likely when another modifier had to be included to achieve referential success). For state modifiers, we find that objects in states that are considered marked elicit more overspecification of state modifiers. This pattern depends not just on the markedness of the state itself, but rather on the markedness of a state for a *particular object*: when an object appears in a less prototypical state, state information is more likely to be included in the referring expression than when the object appears in a more prototypical state. This finding supports the idea that overspecification is tied to prototypicality, as has been proposed for color [5]. It is important to note, however, that the *rate* of overspecification of state found here (21%) is much lower than what has been previously observed for color modifiers (Sedivy: 2005, ~40%; Brown-Schmidt & Konopka, 2011: 78%). While this may simply reflect differences in the properties of the stimuli, it may relate to findings from visual memory [6], where an object's color was forgotten more quickly than its state, suggesting that color properties are not initially bound to the object representation, whereas state information is more unified with the object identity; the same representations may play a role in the properties that are encoded in referring expressions.

References. [1] Brown-Schmidt & Konopka (2011). *Information*. [2] Pechmann (1989). *Linguistics*. [3] Sedivy (2005). *Approaches to Studying World-situated Language Use*. [4] Sedivy (2003). *Journal of Psycholinguistics*. [5] Kreiss et al. (2017) Poster retrieved from https://osf.io/8vpj5/ [6] Brady et al. (2012). *JEP: General.*