# The denotation of tall: A categorization-based approach

It is well-known that the lower bound of scalar properties such as tallness is both context-dependent and vague (Kennedy 2007). Less well established is how this lower bound or threshold is determined. It is often assumed that thresholds are based on the height of members of the comparison class (Kennedy 2007) which is confirmed in some experimental works (Barner and Snedeker, 2008; Syrett et al., 2010). Based on three experiments, we suggest that when confronted with distinct sets of new objects, varying in ways other than mean or maximum height, speakers use percentiles of numbers of objects to set the threshold of tallness and choose a threshold that minimizes categorization errors.

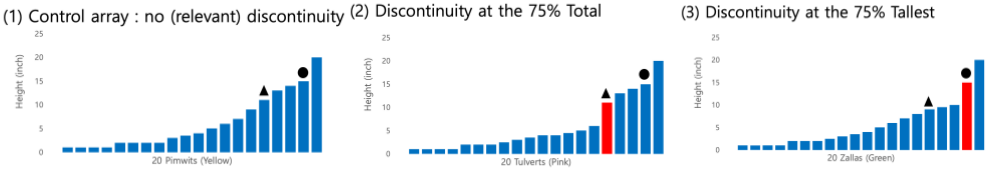
All three experiments we report included three arrays of 20 objects. Critically, mean, median, and maximum height were the same across arrays. The three arrays differed on where a large discontinuity in height was located. In the first two arrays, the large discontinuity in height was at either the object that corresponds to the 75th percentile of heights in an array (henceforth, 75% of Total) or to 75% of the height of the tallest object in an array (henceforth, 75% of Tallest); in the control array, there was no relevant discontinuity in the heights of the objects (Figure 1). In Experiment 1, the 20 objects were all randomly and simultaneously displayed on a table for each participant and for each of three arrays of 20 objects. The objects remained on the table while participants placed objects they deemed tall in a new location. In Experiments 2 and 3, each object in each array was shown successively and removed from view. After all 20 objects had been presented, 5 objects from the array were presented again one by one (the 65th, 70th, 75th, 80th, 85th, 90th percentile objects). Participants had to decide for each object whether it was tall.

The goals of our experiments were (1) to ascertain whether the maximum height or the total number of objects is used as a landmark to compute the threshold of tallness, (2) to determine whether a discontinuity in height at the 75% of Total or 75% of Tallest affects the choice of threshold, and (3) whether a mode of presentation (simultaneous or sequential) affects participants' judgments. In Experiment 1, a comparison of the number of participants that chose as their threshold objects in intervals centered around either the 75% of Total object or the 75% of Tallest object showed that, in the absence of a discontinuity, participants prefer to set the category boundary at the 75% of Total object ($\chi^2$ $p$<0.0001). This suggests that participants used percentiles of the total number of objects as a landmark rather than percentages of the maximum height (Previous researches, Barner and Snedeker's as well as our own, could not distinguish between these two possibilities since the two criteria converged on the same object in their and our arrays). A comparison of the number of participants that chose as threshold the 70%, 75%, or 80% of Total object in the control array with no large discontinuity in height versus the array where there was a large discontinuity in height at the 75% of Total object showed that the spread of the distribution around the 75% of Total object was smaller when a discontinuity occurred at the 75% of Total object (19, 14, 8 participants vs. 2, 38, 3 participants respectively, Fisher's exact test $p$<0.0001). A comparison of the number of participants that chose as threshold objects the intervals around the 75% of Tallest object in the control array vs. in the array in which the large discontinuity in height occurred at the 75% of Tallest object showed that a discontinuity significantly affected participants' choices (4 vs. 12, $\chi^2$ $p$<0.0001). But the presence of a discontinuity at the 75% of Tallest object did not override the overall preference for the 75% of Total object as a threshold for tallness. More participants still chose an object within the intervals centered around the 75% of Total object than the 75% of Tallest object as the threshold (N = 26 vs. N= 13, $\chi^2$ $p$<0.05).

The relevance of discontinuities in setting the threshold of tallness in Experiment 1 could be an artifact of the fact that objects were presented together and remained in view throughout the task, as it allowed for an explicit visual comparison between stimuli. We therefore conducted two additional experiments in which objects were presented one by one and removed from view after presentation. The results of Experiment 2 confirmed the results of Experiment 1: participants still preferred to set the threshold of tallness at the intervals centered around the 75% of Total object rather than the intervals centered around the 75% of Tallest object and the presence of a discontinuity increased the choice of the 75% of Total object as the threshold and increased the choice of the intervals centered around the 75% of Tallest object, without overriding the preference for the intervals centered around the 75% of Total object. The influence of discontinuities in height on choice of threshold is thus not due to visual grouping or visual comparison.

Interestingly, since the presentation order was random in Experiment 2, participants could not determine whether a discontinuity was presented or not until all objects were presented. Also, they must have kept all objects' height in memory for discontinuities in height to affect their judgements. Experiment 3 examined the possible effect of the presentation order on the threshold for tallness. Presentation of objects was sequential in Experiment 3 as in Experiment 2, but the ten shortest objects were presented first, followed by the ten tallest objects for half of the participants, while the ten tallest objects were presented first and the ten shortest objects second for the other half. The order of presentation within the ten shortest and ten tallest subsets was randomized. The results of Experiment 3 replicate those of Experiments 1 and 2, except for Array 3 in Figure 1 (where there was a discontinuity in height at the 75% of Tallest object), as there was an association between order of presentation and choice of a low (65% of Total) or high (90% of Total) threshold for tallness (21 vs. 9 and 7 vs.12, respectively, $\chi_2$ $p$<0.04), suggesting that the influence of discontinuity was modulated by whether participants saw short or tall objects first. Participants who saw short objects first were more likely to set a lower threshold than participants who saw tall objects first.

Overall, the results of our three experiments show that participants used discontinuities in height ***and*** % of the total number of objects rather than % of maximum heights when determining the thresholds of tallness, and that they preferred thresholds that ensured that no more than about 25% of objects are categorized as tall. These results are not affected by whether objects are presented as a group, or one by one (i.e., whether participants can rely on visual input or must access their memory of objects that were presented successively). The consistency of a preference for setting a threshold *around* the 25th percentile is striking. Now, reliance on % of objects (rather than a % of maximum height) is not, *a posteriori*, surprising since few scalar properties are associated with metrics. But why is a threshold at the 25% of objects the sweet spot for this boundary? Inspired by Huttenlocher et al. (2000), we propose that the preference for the 25% of objects boundary is best explained by a goal of minimizing the miscategorizing an object as tall/not-tall. Thresholds are set so that, on average, there are as few errors as possible given "true" membership in the category. This amounts to maximizing the harmonic mean *F* between the goal of including in the category only objects that are actually tall (a.k.a. precision) and the goal of including as many actually tall objects as possible in the category (a.k.a. recall). We tested this hypothesis by computing which proportion of objects selected as tall would on average maximize the harmonic mean *F* for all possible subsets of the top 50% of objects that may constitute the denotation of tall. Selecting the top 27.5% of objects, on average, results in the highest harmonic mean, a proportion very similar to what we observed empirically. Our model suggests that the fact that pairs of scalar adjectives like *short*/*tall* do not cover the entire scale may be the result of people's attempt to minimize errors when determining which of a new category of objects are tall.


(1) Control array : no (relevant) discontinuity

(2) Discontinuity at the 75% Total

(3) Discontinuity at the 75% Tallest

Figure 1. Arrays for all three Experiments. Objects were shown in random order, but are graphed by height for presentation purposes. A red bar indicates the object that follows the large discontinuity in height, ● indicates 75% of the total number of objects, ▲ indicates 75% of the height of the tallest object.