

1 Appendix A: Statistical Density Calculations

1.1 Statistical Density Formulae

Statistical density is the method that Sloutsky and colleagues use to define categories (?, ?). Dense categories have multiple intercorrelated features, while sparse categories have few relevant features. Statistical density can vary between 0 and 1. Higher values (closer to 1) are dense, while lower values (closer to 0) are sparse. We calculate statistical density (D) with the following formula, where H_{within} is the entropy within the category and $H_{between}$ is the entropy between the category and contrasting categories.

$$D = 1 - \frac{H_{within}}{H_{between}}$$

To find total entropy(H), we sum entropy due to varying dimension and entropy due to varying relations among dimensions.

$$H = H^{dim} + H^{rel}$$

This equation is the same whether you are calculating within-category entropy or between-category entropy. To find entropy due to dimensions, you use the following formulas, where M is the total number of varying dimensions, w_i is the attentional weight of a particular dimension (assumed to be 1), and p_j is the probability of value j on dimension i .

$$H_{within}^{dim} = \sum_{i=1}^M w_i \left[\sum_{j=0,1} within(p_j \log_2 p_j) \right]$$

$$H_{between}^{dim} = \sum_{i=1}^M w_i \left[\sum_{j=0,1} between(p_j \log_2 p_j) \right]$$

To find entropy due to relations, you use a similar set of formulas, where O is the total number of possible dyadic relations among the varying dimensions, w_k is the attentional weight of a relation (assumed to be 0.5), and p_{mn} is the probability of the co-occurrence of values m and n on dimension k .

$$H_{within}^{rel} = - \sum_{k=1}^O w_k \left[\sum_{\substack{m=0,1 \\ n=0,1}} within(p_{mn} \log_2 p_{mn}) \right]$$

$$H_{between}^{rel} = - \sum_{k=1}^O w_k \left[\sum_{\substack{m=0,1 \\ n=0,1}} between(p_{mn} \log_2 p_{mn}) \right]$$

All categories have 7 dimensions. For dense categories, 6 of these dimensions are correlated. The seventh dimension is allowed to vary randomly. For sparse categories, 6 of the dimensions vary randomly. The seventh dimension is category-relevant and defines the category. All dimensions have two levels (e.g., for hair shape in aliens – curly and straight).

1.2 Statistical Density Calculations – Sparse