
The simulations reported in Dougherty, Franco-Watkins, and Thomas (in press Psychological Review) used sample-size as the basis of defining the reference class: the 83 largest German and American cities. An equally appropriate way to define the reference class is on the basis of the criterion variable: all cities with populations above 100,000. There are 83 German and 195 American cities over 100,000. Does changing the basis of how the reference class is defined affect the predictions of the recognition heuristic+TTB model, the familiarity model, and the frequency-sampling model?

We examined this question by comparing a recognition+TTB model, a familiarity model, and a frequency sampling model, and defining the reference class either on the basis of the number of cities included in the reference class (the largest 83 American cities), or on the basis of population size (all American cities with populations above 100,000). To demonstrate the sensitivity of the models to changes in the reference class, we compared predictions of our models for predicting American city size. We sampled city names to include in the simulation from the distribution of citations of each city as it occurred in the Chicago tribune, and varied the sample size from 1 to 1,195,000. All details of these simulations are identical to those included in Dougherty, Franco-Watkins, and Thomas (in press, psychological review), with the exception that the distribution of city names included either the largest 83 American cities (USA 83), or the largest 195 American cities (USA 195) that had populations over 100,000. Note that by selecting the 83 largest American cities, we equate the sample size of American cities with the sample size of German cities reported in Gigerenzer & Goldstein (1996). In contrast, by selecting all cities
over 100,000 we equate the American and German cities on the value of the criterion variable
(Gigerenzer & Goldstein, 1996 defined their reference class as ‘all German cities over 100,000’).
The figures below plot the results of our simulations.

Not only do simulations based on this definition of the reference class yield qualitatively
different predictions (see Figures below), but the ecological correlation between city size and
citation rates also changes: Tau = .49 (Pearson r = .39) when computed over the 195 American
largest cities, but is Tau = .39 (Pearson r = .38) when computed over the 83 largest cities. The
idea that there are multiple bases for defining the reference class raises the question of the
importance of the definition of the reference class for making predictions in the city-size task
used by Goldstein and Gigerenzer (2002; Gigerenzer & Goldstein, 1996). Given this result, it is
clear that how one defines the reference class affects the predictions of the models. In some
ways, the definition of the reference class is a free parameter.

Figure. Predicted proportion correct for the Recognition-heuristic model (top), Familiarity model
(middle), and the Frequency model (bottom), for two different means of defining the American
reference class: “all cities over 100,000” (denoted USA 195) versus “the 83 largest US cities”
(denoted USA 83).