

One pattern in the correlation table is particularly prominent. The correlations amongst the first 6 tasks have markedly high absolute magnitudes. All of these tasks are related to centrality, either conceptual centrality (mutability) or category centrality, and all correlations with the mutability tasks are in the direction predicted by a common underlying mutability scale.²

The four mutability tasks and counterfactual naming are similar in a sense because they all ask people to imagine instances without a feature, so some correlation amongst them should have been expected. The tasks are not identical of course; neither are they paraphrases of one another. Surprise, imagining, similarity, goodness-of-example, and naming are not obviously related on their own terms, although commonalities have often been demonstrated amongst the latter three (reviewed in Goldstone, 1994), but so have differences (e.g., Carey, 1985; Gelman & Markman, 1986; Keil, 1989; Rips, 1989). All five tasks, as well as the variability task, do seem to draw on a common resource, the degree to which a concept seems normal. This is consistent with our position: People can quickly and easily evaluate the extent to which a feature of an object is transformable. The tasks became similar merely by asking people to consider an object with a feature removed.

The four mutability tasks and counterfactual naming are also similar in that they were all preceded by lists of objects' features. However, this common procedural element cannot explain their high correlations because they were also correlated with variability judgments, which were not preceded by a list of features, and they were not highly correlated with inferential potency judgments, which were preceded by a list of features.

The correlations obtained in Study 1 did not distinguish conceptual centrality from category centrality. Tasks that required an inside view were highly correlated with tasks that required an outside view. Studies 4 and 5 below focus on this issue.

The two measures of diagnosticity, cue validity and inferential potency, correlated with each other (0.68), but not with any other task. Cue validity judgments did not correlate with other tasks in part because they were consistently low, the highest average judgment for any feature was 34% (the percentage of things that have a red breast that are robins). No feature was judged to pick out one and only one category (even "has a seat" is not unique to chairs). Solitary features tend to give little evidence in favor of any single, specific category which contributes to their inability to predict mutability (or salience). However, the low correlations cannot be dismissed as an artifact of range attenuation because cue validity did correlate with inferential potency and because inferential potency also did not correlate with other tasks, even though its range was not attenuated (normalized for scale, inferential potency had a larger range than 5 of the other tasks).

A surprising finding is the low correlation between prominence and availability (-0.11). In part, this was because features were verified of the category a majority of the time so availability (feature verification) was not a sensitive measure. Hence, mutability must measure some conceptual property beyond mere knowledge about the presence or absence of features.³