SUPPORTING ONLINE INFORMATION

Comparison between Object and Scene Memory Experiments:

In the object memory experiment, only one object was tested per category and it was always the first object presented. However, in the scene memory experiment, we tested four items per category. In order to ensure comparability with the object experiment, we purposefully used the scene presented first in its category as one of the tested items, while the other three test scenes were randomly placed in the stream. Furthermore, in the test phase, the first test trials were always the scenes presented first in the category, making these trials identical in procedure to the object experiment.

We found there was no difference in memory performance between any of the 4 test pairs over any of the 3 interference conditions: 4x3 ANOVA, no main effect of test pair: $F(3,287)<1$, $p=0.67$; main effect of exemplar condition: $F(2, 287)= 14.6$, $p<0.001$; no interaction between test pair and interference condition: $F(6,287)=1.3$, $p=0.26$. This is further supported by a direct comparison of the first and fourth test pair, collapsing over exemplar conditions: $t(23)=0.65$, $p=.45$. Thus, while on average there were fewer subsequent interfering scene exemplars than object exemplars, the interference effect was not specific to retroactive interference, and thus the scenes do not have an advantage over objects.

These results are consistent with our previous observations that category interference effects are driven by both proactive and retroactive processes: in the object memory task, the 2-AFC memory tests examined the influence of retroactive interference, while the repeat detection task examined the influence of proactive interference, and both tasks showed systematic interference with increasing exemplars (Konkle et al. 2010).

Creating the Scene Stimulus Database:

We used the following procedure to build the scene database. The database was started from an existing database of scene categories (Oliva & Torralba, 2001), and then multiple individuals generated ideas for scene categories not represented in the database, and conducted searches using Google Images to gather as many examples from a given category as possible (up to 64 exemplars). We attempted to avoid images that were of the same place as another exemplar but from a slightly different camera position (e.g. translation, zoom, rotation transformations). We also tried to avoid images which were not clearly instances of the scene category.

The ideal scenario would be to sample from the natural variability of the scene (or object) category. While we could not sample from the entire set of variability in the world, we sampled relatively broadly from the web and from scene databases to attempt to capture the entire range of variation within categories. Thus, while selecting images we did not, for example, pick 64 images of amusement parks that all had a large central Ferris wheel.

The entire database of scene and object categories can be browsed and downloaded on the web at http://cvcl.mit.edu/MM.

We think it is unlikely that the converging data shown in Figure 3 of the manuscript arose out of a coincidence of stimulus selection criteria. However, it will be important in future work to manipulate different kinds of scene exemplar similarity and distinctiveness to see how dramatic an effect such manipulations have on memory interference slopes.

References:

Konkle, T., Brady, T. F., Alvarez, G. A. and Oliva, A. (2010). Conceptual distinctiveness supports detailed visual long-term memory for real-world objects. *Journal of Experimental Psychology: General*.

Oliva, A., & Torralba, A. (2001). Modeling the Shape of the Scene: a Holistic Representation of the Spatial Envelope. *International Journal in Computer Vision*, 42, 145-175.