Supplementary Methods

Stimulus construction

Shapes were generated using custom software written in MATLAB (The MathWorks, Inc., Natick, MA, USA). The overall procedure was to first produce a sufficient number of acceptable prototype shapes (that were not excluded according to the pre-specified criteria described below), and then to generate multiple acceptable exemplars associated with each prototype shape. To construct each of the 400 initial prototypes, four points were randomly placed on each side of a bounding square (2 arbitrary units in width). Each adjacent pair of points (end points) in addition to two randomly placed control points within the bounding square were used to construct a Bezier curve, the four corresponding curves constituting one prototype shape.

Next, we quantified the degree to which spatial distortion of each prototype shape would result in a qualitative perceptual difference; such spatial distortion was a critical component in the subsequent production of the experimental exemplar shapes (as described below). Each prototype was spatially distorted to produce a series of successively more distinct exemplars of that prototype – an increasing magnitude of gaussian distributed noise (with 0.05, 0.10, …0.70 standard deviations, using the same random number seed for each prototype) was added to the end points and the control points, with the constraint that all points fell on or within the bounding square. Progressively increasing the magnitude of noise had the perceptual effect of systematically increasing the degree of stretching or compressing particular components of each prototype. After this, 5 independent observers demarcated the level of noise at which exemplars in each set began to appear qualitatively different from the corresponding prototype (i.e. a 'different' rating). In addition, observers indicated whether the prototype shape looked like an animal or object.

The first two exclusion criteria used to reject prototype shapes as candidates for the study included: 1) shapes where at least 2 observers indicated similarity to an animal or object (to avoid verbal encoding strategies) and 2) shapes with ratings at least 2 standard deviations from the mean 'different' rating, computed across all observers and all shapes (such prototypes were considered to vary either too much or too little with the addition of noise, which would have contributed to non-uniform prototype-exemplar similarity). To ensure prototypes were sufficiently different from one another, the third and final exclusion criterion used the similarity between each prototype and its corresponding exemplar at the 'different' rating as a threshold for distinctiveness. That is, each prototype was cross-correlated with all other prototypes (via two-dimensional spatial cross-correlation) where the correlation ranged from 0-1, scaled from a minimum of 0 by the correlation between the first prototype and randomly oriented line segments and scaled to a maximum of 1 by the first prototype with itself (i.e. autocorrelation). Prototypes were rejected if their cross-correlation with any other prototype was greater than the average cross-correlation between each prototype and its corresponding exemplar at the 'different' rating. After such prototype refinement, the 313 prototypes that remained were sufficient to conduct the experiment.

To ensure prototypes and their corresponding exemplars were distinct, exemplars were generated according to the 'different' ratings. For each of the surviving prototypes, ten new exemplars were produced at the corresponding level of noise at the 'different' rating (with unique random number seeds); the mean cross-correlation between the prototype and these ten exemplars constituted a 'different' threshold that was subsequently used to produce the experimental exemplar set associated with a given prototype. Exemplars for each prototype were then sequentially generated using noise at the 'different' rating, but were rejected if the cross-correlation with either their corresponding prototype or any previously generated exemplars of that set was greater than the 'different' threshold (to enforce prototype-exemplar and exemplar-exemplar distinctiveness within each prototype-exemplar set). This was conducted until 10 exemplars were produced that could be used in the experiment. Study sets were comprised of the first through ninth exemplars. At test, two old items were randomly selected from this set, while related items consisted of the tenth nonstudied exemplar and the prototype; the latter pair also serving as new items for counterbalancing purposes. To make the shapes more memorable, the shapes within each prototype-exemplar set were filled-in using the same color and line orientation (Fig. 1), with the constraint that these features were not repeated within a study-test phase. For the follow-up experiment, in addition to the shapes used in the main experiment, 200 additional shapes were created and selected (from an initial 600) using the identical procedures described above.

Event-related timecourse undershoot

To illustrate that the initial activity decrease in the event-related timecourse is due to a return to baseline of the previously occurring event, a simulation was conducted by assuming there were 50 trials with varying inter-trial-intervals, convolving this protocol with the HRF described by Equation 2 to yield the corresponding hemodynamic response, adding random noise of uniform distribution to each timepoint (with a minimum amplitude of 0 and a maximum amplitude equal to the hemodynamic response maximum amplitude), and calculating the event-related timecourse with baseline correction from 0 to 2 s preceding stimulus onset. Three simulations of 1,000 iterations each were conducted using an inter-trial-interval of 4-12 s (as in the present study), 8-16 s, and 12-20 s. As was observed in our empirical data, the 4-12 second inter-trial-interval was associated with a significant decrease in event-related activity 2 s following stimulus onset (–0.25 ± 0.0018; t = 135.1, *P* < .001, paired t-test). A similar result was obtained at this timepoint for the 8-16 second inter-trial-interval (–0.13 ± 0.0017; t = 74.5, *P* < .001, paired t-test), but there was no effect for the 12-20 second inter-trial-interval (0.0012 ± 0.0016; t < 1). There was a highly significant linear trend at this timepoint across the three inter-trial-intervals ($F_{1,2998}$ = 10,535, MS_{error} = 0.0030, *P* < 0.001) supporting the notion that the empirically observed undershoots can be attributed to the short inter-trial-interval employed in the present study.