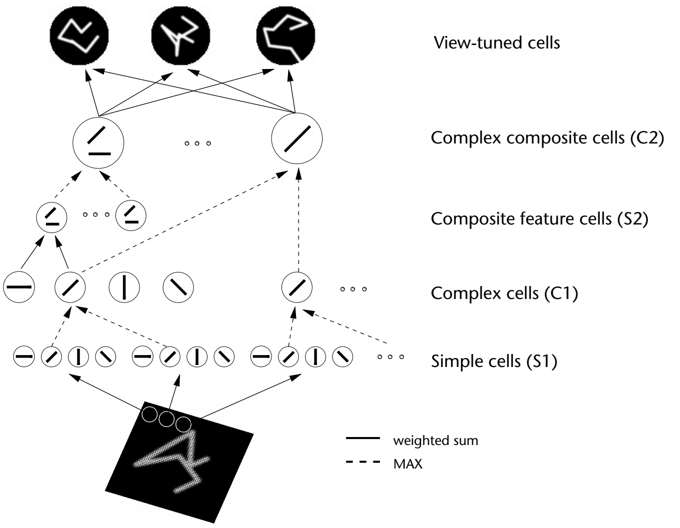
Fig. 2. Sketch of the model. The model was an extension of classical models of complex cells built from simple cells[4], consisting of a hierarchy of layers with linear ('S' units in the notation of Fukushima[6], performing template matching, solid lines) and non-linear operations ('C' pooling units[6], performing a 'MAX' operation, dashed lines). The nonlinear MAX operation—which selected the maximum of the cell's inputs and used it to drive the cell—was key to the model's properties, and differed from the basically linear summation of inputs usually assumed for complex cells. These two types of operations provided pattern specificity and invariance to translation, by pooling over afferents tuned to different positions, and to scale (not shown), by pooling over afferents tuned to different scales.



lus (unless the afferents showed no overlap in space or scale); consequently, excitation of the 'complex' cell would increase along with the stimulus size, even though the afferents show size invariance! (This is borne out in simulations using a simplified two-layer model[25].) For the MAX mechanism, however, cell response would show little variation, even as stimulus size increased, because the cell's response would be determined just by the best-matching afferent.

These considerations (supported by quantitative simulations of the model, described below) suggest that a nonlinear MAX function represents a sensible way of pooling responses to achieve invariance. This would involve implicitly scanning (see Discussion) over afferents of the same type differing in the parameter of the transformation to which responses should be invariant (for instance, feature size for scale invariance), and then selecting the best-matching afferent. Note that these considerations apply where different afferent to a pooling cell (for instance, those looking at different parts of space), are likely to respond to different objects (or different parts of the same object) in the visual field. (This is the case with cells in lower visual areas with their broad shape tuning.) Here, pooling by combining afferents would mix up signals caused by different stimuli. However, if the afferents are specific enough to respond only to one pattern, as one expects in the final stages of the model, then it is advantageous to pool them using a weighted sum, as in the RBF network[15], where VTUs tuned to different viewpoints were combined to interpolate between the stored views.

MAX-like mechanisms at some stages of the circuitry seem compatible with neurophysiological data. For instance, when two stimuli are brought into the receptive field of an IT neuron, that neuron's response seems dominated by the stimulus that, when presented in isolation to the cell, produces a higher firing rate[24]—just as expected if a MAX-like operation is performed at the level of this neuron or its afferents. Theoretical investigations into possible pooling mechanisms for V1 complex cells also support a maximum-like pooling mechanism (K. Sakai and S. Tanaka, *Soc. Neurosci. Abstr.* 23, 453, 1997).

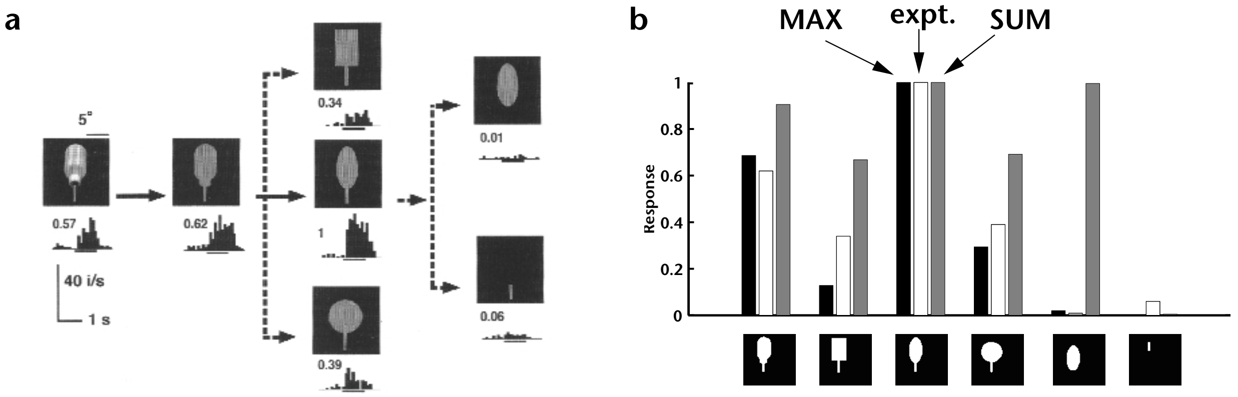


Fig. 3. Highly nonlinear shape-tuning properties of the MAX mechanism. (**a**) Experimentally observed responses of IT cells obtained using a 'simplification procedure'[26] designed to determine 'optimal' features (responses normalized so that the response to the preferred stimulus is equal to 1). In that experiment, the cell originally responded quite strongly to the image of a 'water bottle' (leftmost object). The stimulus was then 'simplified' to its monochromatic outline, which increased the cell's firing, and further, to a paddle-like object consisting of a bar supporting an ellipse. Whereas this object evoked a strong response, the bar or the ellipse alone produced almost no response at all (figure used by permission). (**b**) Comparison of experiment and model. White bars show the responses of the experimental neuron from (**a**). Black and gray bars show the response of a model neuron tuned to the stem-ellipsoidal base transition of the preferred stimulus. The model neuron is at the top of a simplified version of the model shown in Fig. 2, where there were only two types of S1 features at each position in the receptive field, each tuned to the left or right side of the transition region, which fed into C1 units that pooled them using either a MAX function (black bars) or a SUM function (gray bars). The model neuron was connected to these C1 units so that its response was maximal when the experimental neuron's preferred stimulus was in its receptive field.