In Press, Psychonomic Bulletin & Review

Visual noise reveals category representations

Jason M. Gold[1*], Andrew L. Cohen[2] and Richard Shiffrin[1]

[1]Departments of Psychology and Cognitive Science,

Indiana University, 1101 East 10th Street, Bloomington, Indiana, 47405.

[2]Department of Psychology, University of Massachusetts,

427 Tobin Hall, Amherst, Massachusetts, 01003

*To whom correspondences should be addressed. E-mail: jgold@indiana.edu.

Word Count: 3,335

Abstract

How are categories represented in human memory? Exemplar models assume that a category is represented by the individual instances experienced from that category. More generally, a category might be represented by multiple templates stored in memory. A new item is classified according to its similarity to these templates. Prototype models represent a category with a single summary abstraction (i.e., a single template), often the central tendency of the experienced items. A new item is classified according to its similarity to these category prototypes. Here, we show how correlating observers' responses with external noise can be used not only to distinguish single from multiple template representations, but also to induce the form of these templates. The technique is applied to two tasks requiring categorization of simple visual patterns and demonstrates that observers used multiple traces to represent their categories, highlighting the potential of the procedure for use in more complex settings.

Main Text

A central problem in cognitive psychology is the manner in which we represent perceptual categories in memory and the processes by which these representations are used to classify new perceptual inputs. For example, what memory representations and processes are used to classify a person we have not met before as a human, rather than as a toaster or a frog? Exemplar models assume that people represent categories by stored traces of the individual category instances that have been experienced (Medin & Schaffer, 1978; Nosofsky, 1986). A generalized form of this approach would represent a category by multiple templates, although not necessarily by a one-to-one mapping with specific instances. A new item is classified based on its combined similarity to the stored traces from a category in comparison to the similarity for each alternative category (Nosofsky, 1986). Prototype models assume that people form a single summary representation (or single template) for each category, often assumed to be the central tendency across all category instances. A new item is classified on the basis of its similarity to each category prototype (Minda & Smith, 2001, 2002). These model distinctions have arisen in category domains ranging from high-level concepts to low level perception, the latter domain often being modeled with Bayesian pattern recognition approaches (Duda & Hart, 1973; Graham, 1989; Knill & Richards, 1996).

Such Bayesian template-matching models often incorporate the strong assumption that category decisions are based on the statistically optimal use of information (Green & Swets, 1966; Tjan *et al*., 1995). For example, consider the simple visual pattern categorization task depicted in Figure 1. The observer is shown a white square (the ‘signal’) at one of four possible locations, randomly selected on each trial. Gaussian noise is added to each pixel in each of the four locations, with each location divided into a 4x4 grid of "pixels" (see Figure 1 for more stimulus details). The observer's task is to indicate whether the white square signal appeared above or below fixation. Thus, in this very simple categorization task there are two categories (Top, Bottom), each with two members (Left, Right). The contrast of the white square signal is placed at a level where performance is at threshold (e.g., 71% correct).

It turns out that the optimal strategy for this task is to represent each of the two categories by two templates and then classify test stimuli by their similarity to the templates using a decision rule that is remarkably similar to the decision rule utilized by several exemplar-based categorization models (Nosofsky, 1990). In particular, the ideal category templates are the four noise-free versions of the 'white square plus three grey squares' stimuli shown in Figure 1. The relative likelihood of the Top and Bottom categories in the presence of a noisy test stimulus, *S* (signal and noise at all locations), is given by

$$\frac{L(\text{Top} \mid S)}{L(\text{Bottom} \mid S)} = \frac{\sum\limits_{t \in \text{Top}} \exp\left(-\frac{1}{2\sigma^2} \sum\limits_{p} \left(S_p - T_{tp} \right)^2 \right)}{\sum\limits_{b \in \text{Bottom}} \exp\left(-\frac{1}{2\sigma^2} \sum\limits_{p} \left(S_p - T_{bp} \right)^2 \right)} \qquad (1)$$

where *t* and *b* range over the templates from the 'Top' and 'Bottom' categories, respectively, *p* ranges over the 8x8 grid of pixels that defines the set of potential stimulus locations, S_p is the p^{th} pixel of stimulus *S*, T_{xp} is the p^{th} pixel from the template for category member *x*, and σ is the standard deviation of the externally added noise (Green

& Swets, 1966; Tjan et al., 1995). Category Top or Bottom is selected if the likelihood ratio is greater than or less than 1, respectively. We term this a multiple template model rather than an exemplar model, because a ‘pure’ exemplar model would represent a category (Top, say) by each and every signal-plus-noise presentation that was accompanied by feedback. Using a multiple template model as a stand-in for an exemplar model seems reasonable for tasks like the present one in which the storage of exact noise patterns is implausible given known limitations on human memory, and in which the target signals are known exactly.

An alternative, sub-optimal model can be posited that is akin to a prototype representation. Each category is represented by a single template *T*, consisting of two light squares up and two grey squares down (for Top) or two light squares down and two grey squares up (for Bottom). The relative likelihood of the Top and Bottom categories given a test stimulus *S* is

$$\frac{L(\mathrm{Top} \mid S)}{L(\mathrm{Bottom} \mid S)} = \frac{\exp\left(-\frac{\sigma^2}{2}\sum_p \left(S_p - T_{Tp}\right)^2\right)}{\exp\left(-\frac{\sigma^2}{2}\sum_p \left(S_p - T_{Bp}\right)^2\right)} \qquad (2)$$

where T_T and T_B are the single prototype templates for categories Top and Bottom, respectively.

Because only Equation 1 involves a sum of exponentials, these single and multiple template models differ in the way they incorporate non-linearities. It is easy to see that a logarithmic transformation reduces Equation 2 to a simple ratio of distances between the stimulus and each of the prototypes, but this simplification is not possible by taking a log of a sum as in Equation 1. This seemingly small distinction between the models is the key to the ability of the technique that we describe in this article to discriminate between single and multiple template category representations. We shall use an increasingly common approach for estimating the templates used by an observer, through formation of *classification images* (Ahumada, 2002; Ahumada & Lovell, 1971). This technique, known as *reverse correlation* (Ringach *et al.*, 1997) or *response classification* (Beard & Ahumada, 1998), involves computing the correlation between the noise that is added to each pixel in the stimulus and the observer's decisions across trials. It has been used to estimate observer templates in a wide variety of psychophysical tasks, ranging from simple detection of gratings (Ahumada & Beard, 1999) to face and object recognition (Gold *et al.*, 2000; Sekuler *et al.*, 2004). In this procedure, one sorts into separate bins the exact noise that had been added to the signal on each trial (only the noise is classified; the signal is discarded). In our extension of this procedure, there is one bin for each combination of signal presented and response given--the noise patterns within each bin are averaged. In the case of the square categorization task described in Figure 1, there are four possible signals (Top-Left, Top-Right, Bottom-Left, and Bottom-Right) and two possible responses (Top and Bottom), so the method produces eight average noise patterns called classification images. A classification image shows the relative weighting given to each pixel by the observer for a particular signal-response combination over the course of the experiment (Ahumada, 2002).

We first used simulations to verify the intuition that single and multiple template models produce distinctly and measurably different patterns of results when the external noise is analyzed in this fashion. For the Top-Bottom classification experiment described above, we used Equations 1 and 2 to classify 16,000 trials of signal plus noise. The templates used in the multiple template simulation were the ideal templates, that is, the four noise-free signals shown in Figure 1. The templates used in the single template model simulation were combined versions of the two signals within each category: The single ‘Top’ template was composed of two light top squares (with two grey bottom squares) and the single ‘Bottom’ template was composed of two light bottom squares (with grey top squares).

The eight classification images that were produced by averaging the noise patterns in each signal-response bin are shown for the single template (prototype) model in Figure 2a and for the multiple template (exemplar) model in Figure 2b. To understand these plots, consider the four small squares in the top left bin of the upper panel of Figure 2a. These squares show the correlation between the externally added noise and the prototype observer’s responses at each pixel in each square location for the trials where the stimulus was ‘Top Left’ and the observer responded ‘Top’. First, notice that the top two squares are lighter than the background, whereas the bottom two squares are darker than the background. Second, note that this pattern reversed when the observer responded ‘Bottom’. Whereas lighter noise in the top regions and darker noise in the bottom regions made the observer more likely to respond ‘Top’, the opposite pattern led the observer to respond ‘Bottom’. The key finding, however, is that the placement of the signal in the left or right position did not alter the classification image, that is, the classification images for signal-left and signal-right are the same. This is not the case, however, for the multiple template model as shown in Figure 2b. Light squares up and dark down still lead the multiple template observer to say ‘Top’, and vice versa, but the classification images differ depending on whether the signal was actually presented on the left or right.

Consider the case where the signal is ‘Top-Left’ (top left bins in Figures 2a and 2b). For the prototype observer, the top square locations are weighted equally, and this pattern is the same when the signal is ‘Top- Right’. However, the multiple template observer shows a greater influence of lightness in the top left location when the signal is present in the top left (row 1) and vice versa when the signal is present in the top right location (row 2). Notice that this kind of asymmetry is present in all of the other signal-response bins for the multiple template observer; it is always the case that the location where the signal was actually present is weighted more than the adjacent location within the same category. This difference is made even more apparent in lower panels of Figures 2a and 2b, which summarize the data for each observer by collapsing across all of the bins. These summaries were computed by flipping and/or contrast reversing each bin to make it consistent with the ‘signal = Top-Left / response = Top’ bin (e.g., the ‘signal = Top-Left / response = Bottom’ bin was contrast reversed, the ‘signal = Top-Right / response = Top’ bin was flipped about the vertical midline). The leftmost bin in each summary figure is the raw summary data computed by simply collapsing across bins as described above. The rightmost bin in each summary figure is a smoothed version of the raw summary figure, computed by replacing all of the values within each square region with the mean value across pixels.

These simulations show that, in contrast to a prototype observer, a multiple template observer will give more weight to the location where the signal was present on a trial. This difference between models is caused by the exponential non-linearities in

Equations 1 and 2 that applies singly for the prototype model (Equation 2) and summed for the multiple template model (Equation 1). To aid intuition, consider that for the multiple template model an incorrect classification requires a great deal of opposite polarity noise to overcome the signal that is present, an effect that is magnified by the non-linearity in the models. The prototype model does not sum the exponentiated locations separately, so noise in either location has an equal effect. It is important to note that this difference between the models is independent of the specific choice of prototype and only depends on the use of a single template per response category.

We next applied this same analysis to the classification data for four human observers in the same experiment. Each human observer participated in 4,000 trials. The results of this analysis for the data combined across all four observers are shown in Figure 2c (the individual observer patterns were similar to that shown in Figure 2c, but more noisy). These images were computed from the human data in the same fashion as in Figures 2a and 2b. The data clearly show that human observers exhibited the same differential pattern of stimulus location weighting as the multiple template model observer (Figure 2b), demonstrating that the pure prototype model is not adequate: An adequate model must include more than one template per category. We quantified this effect by computing the ratio of the leftmost to the rightmost top locations in each of the smoothed summary plots for the simulated and human observers. The results of this analysis are shown in Figure 3. As expected, the correlation ratio for the prototype observer was exactly 1, indicating the weighting of the two locations was the same for this observer. In contrast, the correlation ratios for the exemplar and human observers were much greater than 1 (~2.5), and were nearly identical.

Note that the true model used by observers might well be a complex mixture of multiple template representations and decision rules, rather than either of the pure models we have simulated here. Finding the true model used by our observers, or testing which ideal model better approximated the true model, was not the aim of this article. Nonetheless we carried out a number of comparison analyses to see which ideal model best predicted the trial-by-trial responses made by the human observers. In all cases, the multiple template model fared better than the single template model. The strongest test (and simplest to understand) involves looking at the most diagnostic trials among the 4,000 experienced by each observer. These are trials on which the two models strongly predict opposite responses. Because the two models generally make highly correlated predictions, such trials are relatively rare, occurring on only 135 trials across the data for all four human observers. However, on these highly diagnostic trials, the human observers' responses matched the predictions of the 'ideal' multiple template model on 113 of the 135 trials.

Thus, our results provide an important proof of concept: namely, that the response classification technique can be used to discriminate between single and multiple template models of categorization. To the extent that the multiple template model is a reasonable approximation to the exemplar model, this technique has potentially wide-ranging applications for distinguishing prototype and exemplar theories of perceptual categorization.

The four-square task, however, is admittedly simple. To test the generalizability of our results we applied this technique to a task that requires classifications along dimensions that are more abstract than the spatial position of the stimuli. Figure 4a-c illustrates a task in which observers must classify patterns based upon their spatial frequency (i.e., bar width) while the stimuli vary in both frequency and orientation. In this task, there are two categories: high frequency (5 cycles/deg; left column of Figure 4a) and low frequency (2 cycles/deg; right column of Figure 4a). Within each category, one grating is oriented 45° left of vertical (bottom row of Figure 4a) and the other 45° right of vertical (top row of Figure 4a). As in the square categorization task, the stimulus is corrupted by white Gaussian pixel noise (e.g., the top of Figure 4c), and the observer's task is to classify a stimulus as belonging to one of the two possible categories (High or Low frequency; see Figure 4 for more stimulus details).

Figure 4b shows the stimuli described in Figure 4a represented in Fourier frequency space. As illustrated by the bottom of Figure 4c, spatial frequency in these plots is represented as the distance from the center of the image and orientation is represented as the angle made relative to the horizontal axis. The amount of relative power at each constituent frequency component is represented by the contrast at each location in each plot. For clarity, only frequencies below 8 cycles per image are shown in the figures. Figure 4b shows that the stimuli in Figure 4a can be equivalently represented as localized 'bumps' in Fourier space. In addition, because white Gaussian contrast noise in the spatial domain introduces white Gaussian amplitude noise in the spatial frequency domain, our response classification analyses can be equivalently carried out in Fourier space (Ahumada *et al.*, 1975).

Despite the differences in stimuli, the multiple template (exemplar) and single template (prototype) model decision rules are the same in both the grating and square discrimination tasks. Just as for the square categorization task, we would expect the multiple template model to produce differential classification images conditional on the within category signal presented, that is there should be differences conditional on signal orientation. Figures 4d (spatial domain) and 4e (spatial frequency domain) show the results of this experiment. The top row gives the simulated results for the single template (prototype) model and the second row for the multiple template (exemplar) model, each based on 45,000 simulated trials. The bottom row gives the combined results for three human observers (each received 15,000 trials, the greater number of trials being necessary due to the greater number of pixels in the stimuli). As in the summary plots described at the bottom row of Figure 2, the symmetries across the various signal-response bins allowed us to produce a single summary image for each spatial frequency response type (i.e., one image for 'high frequency' and one image for 'low frequency'). This figure shows that the prototype model produces equal classification images dependent on stimuli with different orientations, whereas the multiple template model and human observers show large orientation dependent differences in the classification images. We quantified this effect by computing the ratio of the similarities (cross-correlations) of the Fourier signal plots shown in Figure 4b within a given category ('High' or 'Low') to the corresponding Fourier classification plots in shown Figure 4e. The results of this analysis are shown in Figure 4f. As with the square categorization task, the performance of the exemplar model matches the performance of the human observers far more closely than the performance of the prototype model. Thus, we can conclude that any adequate model of human performance in this task must include multiple templates.

As with the square categorization task, we carried out a series of tests to determine which ideal model better predicted the trial-by-trial responses made by the human observers. Unlike the results for the square categorization task, the performance of the multiple template model was only slightly better than the single template model. Furthermore, it is clear from the data that the idealized multiple template model described above does not fully capture the pattern of human data. In particular, as can be seen in Figure 4, the multiple-template model predicts a much wider range of frequency and orientation influence than is seen in the human data. These results raise an important cautionary note and highlight a benefit of the current technique: Although our data do not allow us to distinguish between the idealized single and multiple template models, the response classification technique can still be used to make inferences about the form of the templates used by observers as well as demonstrate that multiple templates must be a part of any adequate model. It is also worth noting that, although both of the examples we have presented here involve adding noise at the level of individual pixels, it is also possible to restrict the added noise to a stimulus sub-space (Ringach et al., 1997) and add noise along higher order stimulus dimensions, such as size, curvature and aspect ratio (e.g., Neri *et al.*, 1999). Adding noise along these kinds of higher order dimensions could greatly reduce the size of stimulus perturbation space and may lend itself more naturally to more standard categorization tasks (e.g., Nosofsky, 1986)

In sum, our results demonstrate that the response classification technique is an effective tool for making inferences about the number and form of templates used to make category judgments, for assessing the adequacy of single template (prototype) models, and equivalently, for demonstrating the existence of multiple templates within each category. Our two experimental demonstrations were carried out with relatively simple tasks, but ones requiring quite different visual representations. However, it worth emphasizing that the experiments presented here were explicitly designed to be simple, easy to analyze and to provide a straightforward proof of concept. Although the results of our experiments happen to support an exemplar-type representation, they in no way offer a systematic and thorough comparison of prototype and exemplar models. In future research we hope to use this technique to explore the kinds of representations human observers use in more complex and realistic categorization tasks, especially ones in which there are many more items within each category and the templates of the alternative models are not obvious a priori.

Acknowledgements

This research was supported by NEI Grant 1 R03 EY015787 to J.M.G. and NIMH Grants 1 R01 MH12717 and 1 R01 MH63993 to R.S.

Figure Legends

Figure 1. Stimuli used in the square categorization task. The top four images show the noise-free versions of the square stimuli (signals) from each category (Top and Bottom). The bottom image shows an example of a noisy 'Bottom-Right' stimulus. Each square subtended 0.65° of visual angle from a viewing distance of 130 cm. The distance of the center of each square from the center of the display was 1.39°. Each square was coarsely divided into a 4 x 4 grid of pixels. The screen-pixels within each grid location were set to the same contrast value (with contrast of a pixel being defined as the difference between pixel and background luminance, normalized by background luminance). A 4x4 screen-pixel fixation point remained at the center of the display for the duration of the experiment. On each trial, the stimulus was shown for approximately 500 ms. The background luminance was 81.4 cd/m^2. The contrast variance of the Gaussian noise added to the square grids was 0.04. Accuracy feedback was provided in the form of a high or low beep.

Figure 2. Response classification results for (a) a prototype model observer, (b) an exemplar model observer and (c) human observers in the square categorization task. Each block of four squares in the top row shows the resulting classification image for the corresponding signal-response combination. The bottom row of images summarizes the data for each observer by collapsing across all of the bins in the top image (left side) and smoothing the data (right side).

Figure 3. Ratios of the mean values obtained for the left and right top square locations in the smoothed classification images shown at the bottom of Figure 2. Error bars correspond to +/- 2 s.d.

Figure 4. (a) Signals used in the Gabor categorization task. (b) The same signals represented Fourier space (amplitude spectra only). Each plot shows frequencies <= 8 cycles/image. (c) Example of a noisy, high frequency right-oriented Gabor stimulus in the spatial domain (top) and a description of how frequency f and orientation θ are represented in Figure 3b (bottom). The Gabor stimuli were 64 x 64 pixels in size, which subtended 1.05° of visual angle from a viewing distance of 130 cm. Fixation was maintained by a dark box that surrounded the stimulus region for the duration of the experiment. The spatial frequencies of the Gabors were 5 and 2 cycles/degree of visual angle, and the orientations were ± 45° to the left and right of vertical. The stimulus duration was approximately 500 ms, and the Gaussian noise added to each pixel had a contrast variance of 0.04. Accuracy feedback was provided in the form of a high or low beep. (d) Summary spatial classification images for both model observers and the human observers in each category (High and Low frequency). (e) Frequency-space summary classification images for both model observers and the human observers in each category. (f) Ratios of cross-correlation values obtained for each observer type in each condition (see text for details). Error bars correspond to +/- 2 s.d.

Figure 1
Production # R 485-B

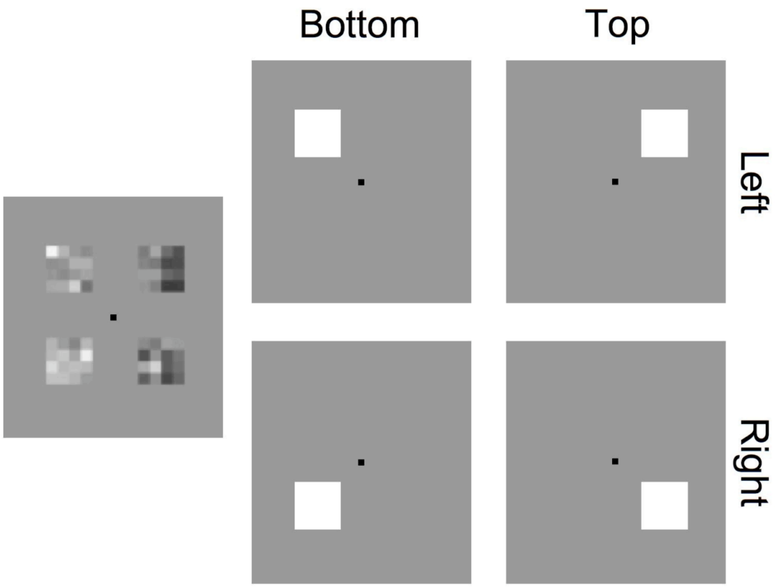


Figure 2
Production # R 485-B

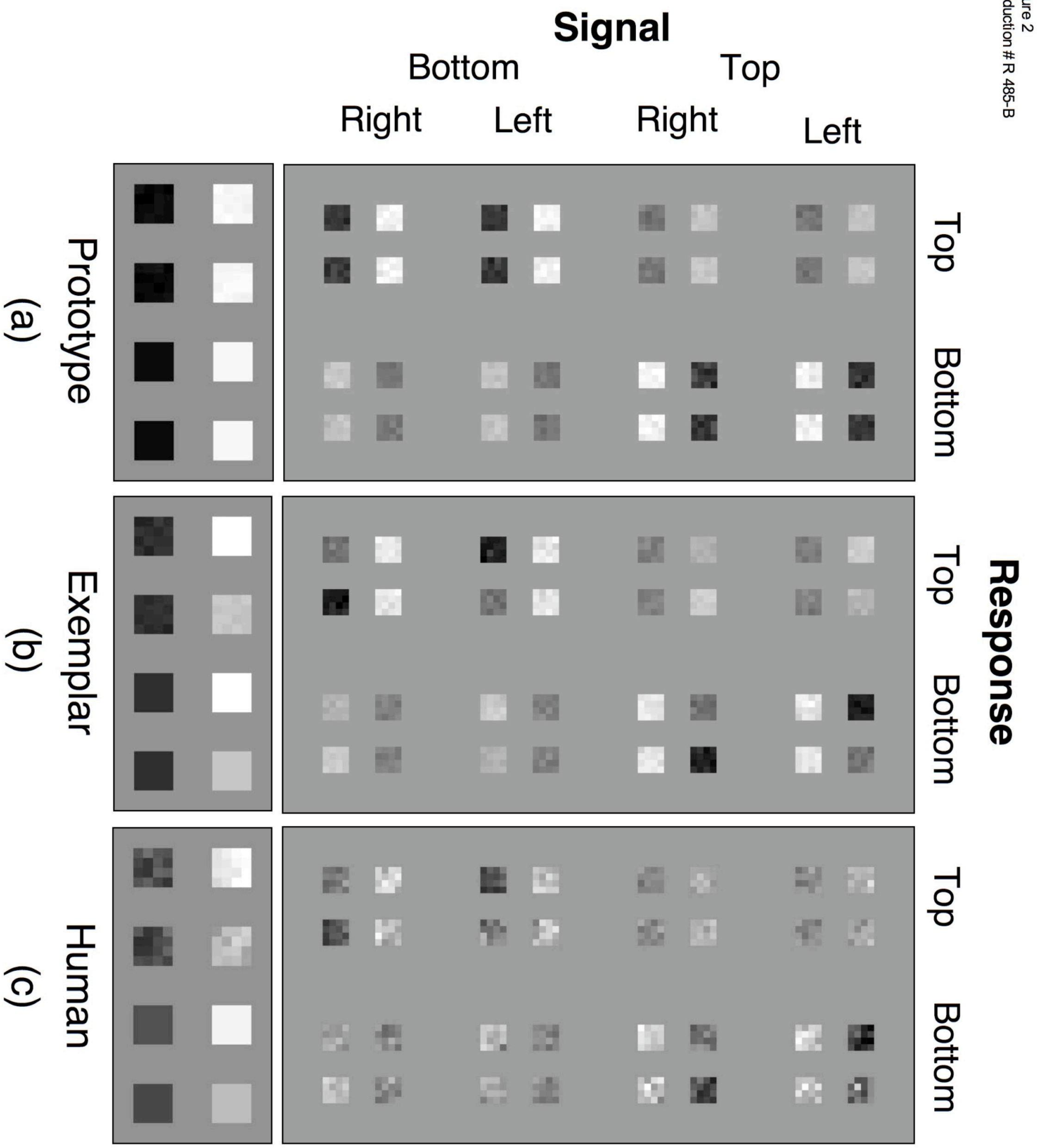


Figure 3
Production # R 485-B

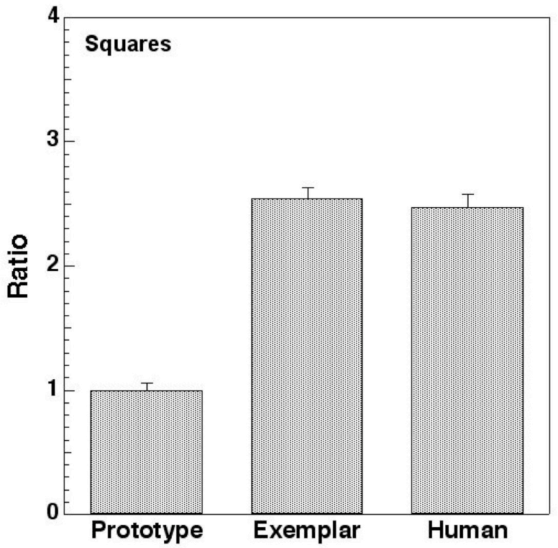
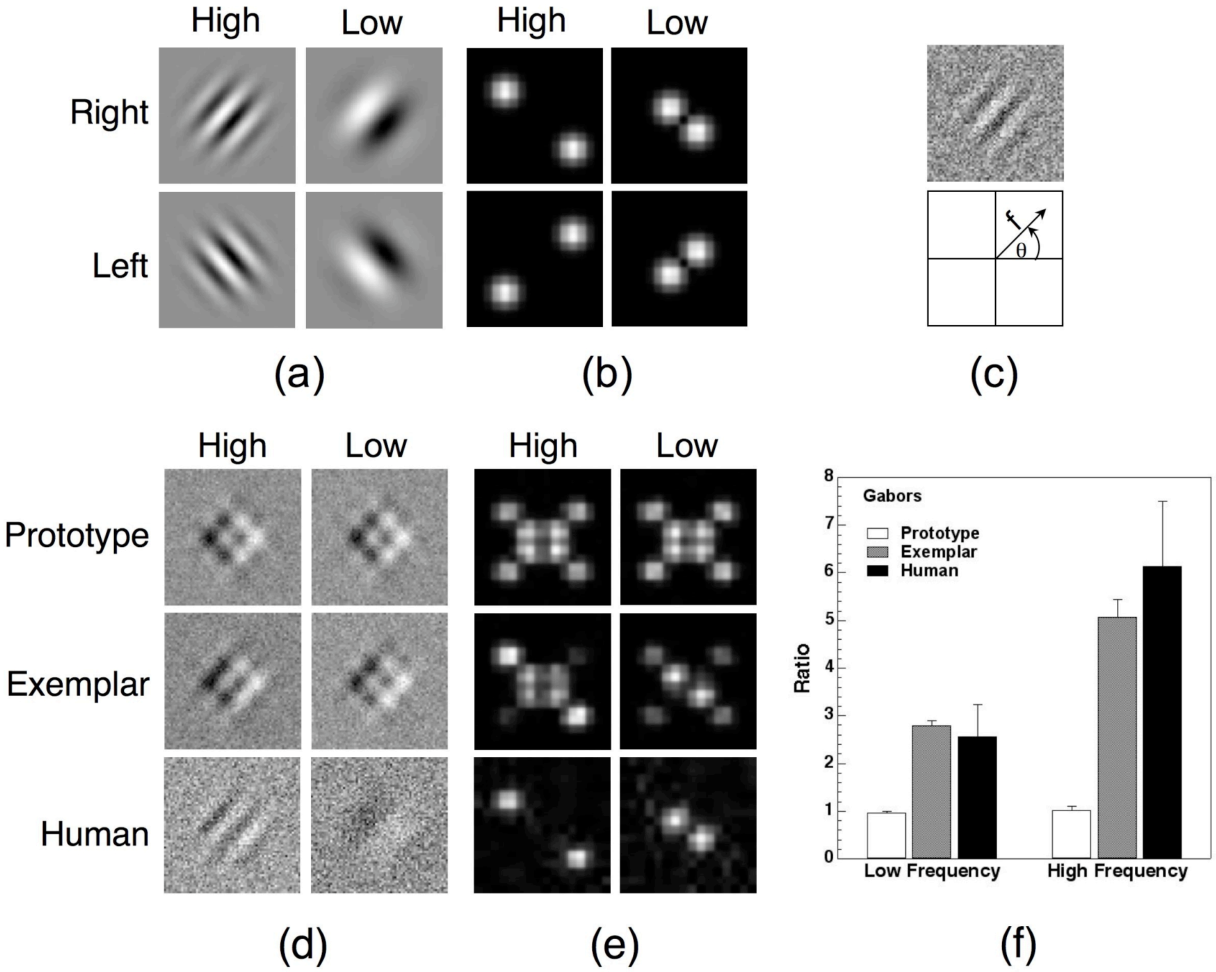


Figure 4
Production # R 485-B



References

Ahumada, A. (2002). Classification image weights and internal noise level estimation. *J Vis, 2*(1), 121-131.

Ahumada, A., & Beard, B. L. (1999). Classification images for detection. *IOVS, 40*(4), 3015.

Ahumada, A., & Lovell, J. (1971). Stimulus features in signal detection. *JASA, 49*(6-2), 1751-1756.

Ahumada, A., Marken, R., & Sandusky, A. (1975). Time and frequency analyses of auditory signal detection. *JASA, 57*(2), 385-390.

Beard, B. L., & Ahumada, A. J. (1998). *Technique to extract relevant image features for visual tasks*. Paper presented at the SPIE, San Jose, CA.

Duda, R. O., & Hart, P. E. (1973). *Pattern classification and scene analysis*. New York: John Wiley & Sons.

Gold, J. M., Murray, R. F., Bennett, P. J., & Sekuler, A. B. (2000). Deriving behavioral receptive fields for visually completed contours. *Current Biology, 10*, 663-666.

Graham, N. V. S. (1989). *Visual pattern analyzers*. New York: Oxford University Press.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: John Wiley and Sons.

Knill, D. C., & Richards, W. (1996). *Perception as bayesian inference*. Cambridge, U.K.; New York: Cambridge University Press.

Medin, D. L., & Schaffer, M. M. (1978). Context theory pf classification learning. *Psychological Review, 85*, 207-238.

Minda, J. P., & Smith, J. D. (2001). Prototypes in category learning: The effects of category size, category structure, and stimulus complexity. *J Exp Psychol Learn Mem Cogn, 27*(3), 775-799.

Minda, J. P., & Smith, J. D. (2002). Comparing prototype-based and exemplar-based accounts of category learning and attentional allocation. *J Exp Psychol Learn Mem Cogn, 28*(2), 275-292.

Neri, P., Parker, A. J., & Blakemore, C. (1999). Probing the human stereoscopic system with reverse correlation. *Nature, 401*(6754), 695-698.

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *J Exp Psychol Gen, 115*(1), 39-61.

Nosofsky, R. M. (1990). Relations between exemplar-similarity and likelihood models of classification. *Journal of Mathematical Psychology, 34*(4), 393-418.

Ringach, D. L., Sapiro, G., & Shapley, R. (1997). A subspace reverse-correlation technique for the study of visual neurons. *Vision Res, 37*(17), 2455-2464.

Sekuler, A. B., Gaspar, C. M., Gold, J. M., & Bennett, P. J. (2004). Inversion leads to quantitative, not qualitative, changes in face processing. *Curr Biol, 14*(5), 391-396.

Tjan, B. S., Braje, W. L., Legge, G. E., & Kersten, D. (1995). Human efficiency for recognizing 3-d objects in luminance noise. *Vis. Res., 35*(21), 3053-3069.