Supporting Information

Peterson and Eckstein 10.1073/pnas.1214269109

SI Text

A. Region of Interest Ideal Observer. Each grayscale face image in the stimulus set can be represented as a 500×500 matrix, with each element corresponding to a pixel whose value ranges from 0 (pure black) to 255 (pure white). We extracted corresponding square regions from each face of size 30×30 pixels, equivalent to a $1° \times 1°$ region of visual angle, given the experimental parameters encountered by the participants. We sampled all possible regions (overlap was allowed), leading to 221,841 unique calculations (471 regions in each of the horizontal and vertical directions).

For a given region, the input to the ideal observer on each independent Monte Carlo trial, $\mathbf{g}$, is a combination of a randomly sampled signal, $\mathbf{s}$, and zero-mean white Gaussian noise, $\mathbf{n}$, with SD, σ:

$$\mathbf{g} = \mathbf{s} + \mathbf{n}, \quad [S1]$$

$$\mathbf{n} \sim N(\mathbf{0}, \sigma^2\mathbf{I}), \quad [S2]$$

where $\mathbf{0}$ represents the zero vector and $\mathbf{I}$ is the identity matrix. The ideal observer takes this noisy observation and computes a posterior probability of each possible signal being present, given the observed data, $P(\mathbf{s}_i|\mathbf{g})$, where i is an integer ranging from 1 to 10 representing each possible signal or identity. Bayes' rule then states that the optimal decision is to select the maximum posterior probability. In this case, the probability of observing the data is the same for each possible signal and each signal is equally likely to be sampled; thus, $P(\mathbf{g})$ and $P(\mathbf{s}_i)$ can be omitted, leading to a rule in which the maximum likelihood, ℓ_i, is taken as the decision:

$$P(\mathbf{s}_i|\mathbf{g}) = \frac{P(\mathbf{s}_i)P(\mathbf{g}|\mathbf{s}_i)}{P(\mathbf{g})} \Rightarrow P(\mathbf{g}|\mathbf{s}_i) = \ell_i, \quad [S3]$$

$$decision = \arg\max_i (\ell_i). \quad [S4]$$

From this point on, we will treat the signals and noisy observations as 1D vectors for ease of notation. Because the additive noise is Gaussian-distributed and white, the likelihood metric is computed as:

$$\ell_i = \exp\left(\frac{2\mathbf{s}_i^T\mathbf{g} - \mathbf{s}_i^T\mathbf{s}_i}{2\sigma^2}\right), \quad [S5]$$

where T represents the transpose operator.

The identification condition has one signal for each identity; thus, the maximum likelihood is taken as an optimal decision rule. The gender, emotion, and happy versus neutral conditions each have signal uncertainty; that is, each class is represented by multiple signals. In the gender case, the two classes are male and female, with 40 unique face images in each class. The emotion condition contains seven classes (afraid, angry, disgusted, happy, neutral, sad, and surprised), with 20 unique images per class. The happy vs. neutral condition is another two-class situation, with 40 unique images per class. The possible signals, j, for each class, i, are now represented as $\mathbf{s}_{ij}$. Likelihoods are calculated in the same way; however, the likelihoods for each signal within a class are now summed to produce a new decision variable, L_i, the summed likelihood for class i (Eqs. **S6–S8**):

$$\ell_{ij} = \exp\left(\frac{2\mathbf{s}_{ij}^T\mathbf{g} - \mathbf{s}_{ij}^T\mathbf{s}_{ij}}{2\sigma^2}\right), \quad [S6]$$

$$L_i = \sum_j \ell_{ij}, \quad [S7]$$

$$decision = \arg\max_i (L_i). \quad [S8]$$

This calculation was done for each region 1 million times, with each trial sampling a new independent noise field. Performance was measured in terms of proportion of correct ideal observer decisions. The contrast of the signals was set such that the maximum performance across the regions equaled the maximum performance of the foveated ideal observer (FIO; see next section).

B. Spatially Variant Contrast Sensitivity Function. The visual field is not processed in a homogeneous manner. Due to photoreceptor density, ganglion cell convergence, cortical magnification, and a host of other factors, resolution and sensitivity are greatest at the point of fixation corresponding to the foveal region of the retina. The quality of processing falls off continuously with eccentricity. Thus, where in a visual scene one chooses to direct the eyes necessarily dictates the type and quality of visual information that enters the visual system. It is this front-end constraint that we aimed to model, thus allowing for a quantitative description of the loci of fixations that lead to maximal task-relevant information acquisition.

The sensitivity and resolution of the visual system are commonly formulated through the contrast sensitivity function (CSF). The CSF is a quantitative description of how well the visual system can process signals of varying spatial frequencies. In general, the CSF is measured using simple single-frequency sinusoidal patterns or Gabor patches in isolation. Most studies have looked at the CSF at the fovea, with the form being well-described as a band-pass filter with sensitivity peaking somewhere between three and six cycles per degree of visual angle. Various analytical estimations of the CSF have been proposed, each with similarly good agreement with empirical observations (1); here, we choose a simple form as represented by Eq. **S9** (2):

$$CSF(f) = c_0 f^{a_0} \exp(-b_0 f), \quad [S9]$$

where a_0, b_0, and c_0 are constants chosen to normalize the maximum contrast sensitivity at 1 and to set the peak at four cycles per degree of visual angle, and f is the spatial frequency in cycles per degree of visual angle.

Fewer studies have looked at the effects of eccentricity on the CSF. We chose an analytical form developed by Peli et al. (3) as a starting point, using a simple decaying exponential factor as a function of distance from fixation (Eq. **S10**):

$$CSF(f, r) = c_0 f^{a_0} \exp(-b_0 f - d_0 r^{n_0} f), \quad [S10]$$

where r is the distance from fixation in degrees of visual angle, d_0 is termed the eccentricity factor, and n_0 is the steep roll-off factor.

Peli et al. (3) estimated values of d_0 from their own data, along with previous data sets, finding a range of between 0.03 and 0.06 with asymmetries in the direction from fixation (e.g., the common finding that sensitivity falls off more gradually in the horizontal than vertical direction). We found that using these previously

reported values for the eccentricity factor proved to be a poor predictor of fixation-dependent human performance in the face recognition tasks, with the performance profile varying to a much smaller extent than the behavioral observations. This is not surprising, given the complexity of more natural images, such as faces, compared with the basic grating stimuli used to measure CSFs in the past. To compensate for this difference, we allowed four free parameters, which were then used to fit the forced fixation performance profile for the identification condition. The term $d_0(\theta)$ was allowed to have a different value in the horizontal, upward, and downward directions from fixation. If θ is taken as the angle from the right horizontal axis, we have $dh = d_0(0) = d_0(\pi)$, $du = d_0\left(\frac{\pi}{2}\right)$, and $dd = d_0\left(-\frac{\pi}{2}\right)$. Eccentricity factors for intermediate directions were determined through a linear interpolation. The steep roll-off factor, n_0, was left free as well.

To simulate the effect of foveation for a given fixation point, k, the face image is passed through a set of spatially variant filters. Each combination of eccentricity (r) and direction (θ) from fixation is defined by its own CSF. Due to computational constraints, we spatially divided the image into small bins and assigned a single unique CSF to each bin. For the simulations presented here, we took 480 bins (30 eccentricities and 16 directions) because finer resolution proved to have little impact on the simulation results. For each bin, i, we multiplied on an element-by-element basis the fast Fourier transform (FFT) of the entire face image, $\mathbf{s}$, by the bin's rotationally symmetrical CSF, $\mathbf{CSF}_i$, before transforming back to the spatial domain using the inverse FFT (IFFT), resulting in a filtered image, $\mathbf{s}_i$:

$$\mathbf{s}_i = IFFT(FFT(\mathbf{s})\cdot\mathbf{CSF}_i). \qquad \textbf{[S11]}$$

We then created a composite image, $\tilde{\mathbf{s}}$, by extracting each bin's spatial region from $\mathbf{s}_i$ and placing it in the corresponding region of the new image (Fig. 4*B*). The complete set of CSFs used for a given fixation point is termed the spatially variant contrast sensitivity function (SVCSF) in the main text.

C. FIO. The FIO uses a variation of the nonprewhitening with an eye filter (NPWE) technique (2). The simulated foveation imposes spatial correlations on the additive white noise field. The NPWE does not attempt to correct for this correlation; instead, it applies the same filtering used for the input signal on the possible templates. These filtered images, called matched templates, are then used to formulate response variables as follows.

We designate the filtering process at fixation point k as the application of a linear operator, $\mathbf{E}_k$, termed the eye filter and described by $\mathbf{SVCSF}_k$, to the randomly sampled noisy face image. Zero-mean white Gaussian internal noise is then added to the filtered stimulus. Thus, the input signal, $\tilde{\mathbf{g}}$, is composed of a filtered combination of underlying signal, $\mathbf{s}_0$, and a Gaussian white noise process, $\mathbf{n}_{ex}$, along with an additive unfiltered internal noise process, $\mathbf{n}_{in}$, such that $\tilde{\mathbf{g}} = \mathbf{E}_k(\mathbf{s}_0 + \mathbf{n}_{ex}) + \mathbf{n}_{in}$. The FIO takes the dot product of this filtered input with the similarly filtered noise-free templates to arrive at a set of responses, $\mathbf{r}_k$:

$$\begin{aligned}\mathbf{r}_k = \{r_{1,k},\ldots,r_{m,k}\} &= \left\{(\mathbf{E}_k\mathbf{s}_1)^T(\mathbf{E}_k(\mathbf{s}_0+\mathbf{n}_{ex})+\mathbf{n}_{in}),\ldots,(\mathbf{E}_k\mathbf{s}_m)^T(\mathbf{E}_k(\mathbf{s}_0+\mathbf{n}_{ex})+\mathbf{n}_{in})\right\}\\ &= \{\mathbf{E}_k\mathbf{s}_1,\ldots,\mathbf{E}_k\mathbf{s}_m\}^T(\mathbf{E}_k(\mathbf{s}_0+\mathbf{n}_{ex})+\mathbf{n}_{in}).\end{aligned} \qquad \textbf{[S12]}$$

The FIO then computes the posterior probability of the hypothesis that face f was shown, H_f, given the set of responses, $\mathbf{r}_k$, and chooses the maximum:

$$decision = \arg\max_f\left(P(H_f|\mathbf{r}_k)\right). \qquad \textbf{[S13]}$$

The prior probabilities of each face being shown are the same as is the probability of observing the data, allowing us to reduce the computation of the posterior to the computation of a simple likelihood:

$$P(H_f|\mathbf{r}_k) = \frac{P(H_f)P(\mathbf{r}_k|H_f)}{P(\mathbf{r}_k)} \Rightarrow P(\mathbf{r}_k|H_f) = \ell_{f,k}. \qquad \textbf{[S14]}$$

To compute the likelihood, the distribution of $\mathbf{r}_k$ must be known. For ease of notation, we note that $\mathbf{E}_k$ is a linear operator; thus, the responses can be rearranged to become the product of single-filtered templates, $\dot{\mathbf{s}}_i$, and double-filtered templates, $\ddot{\mathbf{s}}_i$:

$$\begin{aligned}r_{i,k} &= (\mathbf{E}_k\mathbf{s}_i)^T(\mathbf{E}_k(\mathbf{s}_0+\mathbf{n}_{ex})+\mathbf{n}_{in})\\ &= (\mathbf{E}_k^2\mathbf{s}_i)^T(\mathbf{s}_0+\mathbf{n}_{ex}) + (\mathbf{E}_k\mathbf{s}_i)^T\mathbf{n}_{in} = \ddot{\mathbf{s}}_{i,k}^T\mathbf{s}_0 + \ddot{\mathbf{s}}_{i,k}^T\mathbf{n}_{ex} + \dot{\mathbf{s}}_{i,k}^T\mathbf{n}_{ex}.\end{aligned} \qquad \textbf{[S15]}$$

Using the fact that the external and internal noise terms, $\mathbf{n}_{ex}$ and $\mathbf{n}_{in}$, are drawn from zero-mean distributions, the mean response of template i when face f is present is:

$$\begin{aligned}\mu_{i,f,k} = E[r_{i,f,k}] &= E\left[\ddot{\mathbf{s}}_{i,k}^T\mathbf{s}_f + \ddot{\mathbf{s}}_{i,k}^T\mathbf{n}_{ex} + \dot{\mathbf{s}}_{i,k}^T\mathbf{n}_{in}\right]\\ &= E\left[\ddot{\mathbf{s}}_{i,k}^T\mathbf{s}_f\right] + E\left[\ddot{\mathbf{s}}_{i,k}^T\mathbf{n}_{ex}\right] + E\left[\dot{\mathbf{s}}_{i,k}^T\mathbf{n}_{in}\right] = \ddot{\mathbf{s}}_{i,k}^T\mathbf{s}_f,\end{aligned} \qquad \textbf{[S16]}$$

where E[•] is the expected value operator. This leads to the mean vector when face f is present being $\boldsymbol{\mu}_{f,k} = \{\mu_{1,f,k},\ldots,\mu_{m,f,k}\}$. The covariance of the response distribution is such that the covariance between the ith and jth responses when face f is present is given by:

$$\begin{aligned}\sum_{i,j,k} &= \mathrm{cov}(r_{i,k},r_{j,k}) = \mathrm{E}\left[(r_{i,k}-\mathrm{E}[r_{i,k}])(r_{j,k}-\mathrm{E}[r_{j,k}])\right]\\ &= \mathrm{E}\left[\left(\ddot{\mathbf{s}}_{i,k}^T\mathbf{s}_f + \ddot{\mathbf{s}}_{i,k}^T\mathbf{n}_{ex} + \dot{\mathbf{s}}_{i,k}^T\mathbf{n}_{in} - \ddot{\mathbf{s}}_{i,k}^T\mathbf{s}_f\right)\right.\\ &\quad\left.\times\left(\ddot{\mathbf{s}}_{j,k}^T\mathbf{s}_f + \ddot{\mathbf{s}}_{j,k}^T\mathbf{n}_{ex} + \dot{\mathbf{s}}_{j,k}^T\mathbf{n}_{in} - \ddot{\mathbf{s}}_{j,k}^T\mathbf{s}_f\right)\right]\\ &= \ddot{\mathbf{s}}_{i,k}^T\ddot{\mathbf{s}}_{j,k}\mathrm{E}\left[\mathbf{n}_{ex}^T\mathbf{n}_{ex}\right] + \left(\ddot{\mathbf{s}}_{i,k}^T\dot{\mathbf{s}}_{j,k} + \dot{\mathbf{s}}_{i,k}^T\ddot{\mathbf{s}}_{j,k}\right)\mathrm{E}\left[\mathbf{n}_{ex}^T\mathbf{n}_{in}\right]\\ &\quad + \dot{\mathbf{s}}_{i,k}^T\dot{\mathbf{s}}_{j,k}\mathrm{E}\left[\mathbf{n}_{in}^T\mathbf{n}_{in}\right].\end{aligned} \qquad \textbf{[S17]}$$

We now use two properties of random variables. First, for any random variable X, $\mathrm{Var}(X) = \mathrm{E}[X^2] - (\mathrm{E}[X])^2$. Second, for any two independent random variables, X and Y, $\mathrm{E}[XY] = \mathrm{E}[X]\mathrm{E}[Y]$. Using these two properties, Eq. **S17** simplifies as:

$$\ddot{\mathbf{s}}_{i,k}^T\ddot{\mathbf{s}}_{j,k}\mathrm{E}\left[\mathbf{n}_{ex}^T\mathbf{n}_{ex}\right] = \ddot{\mathbf{s}}_{i,k}^T\ddot{\mathbf{s}}_{j,k}\left(\mathrm{Var}(\mathbf{n}_{ex}) + (\mathrm{E}[\mathbf{n}_{ex}])^2\right) = \sigma_{ex}^2\ddot{\mathbf{s}}_{i,k}^T\ddot{\mathbf{s}}_{j,k}, \qquad \textbf{[S18]}$$

$$\left(\ddot{\mathbf{s}}_{i,k}^T\dot{\mathbf{s}}_{j,k}+\dot{\mathbf{s}}_{i,k}^T\ddot{\mathbf{s}}_{j,k}\right)\mathrm{E}\left[\mathbf{n}_{ex}^T\mathbf{n}_{in}\right]=\left(\ddot{\mathbf{s}}_{i,k}^T\dot{\mathbf{s}}_{j,k}+\dot{\mathbf{s}}_{i,k}^T\ddot{\mathbf{s}}_{j,k}\right)\mathrm{E}[\mathbf{n}_{ex}]\mathrm{E}[\mathbf{n}_{in}]=0, \quad \textbf{[S19]}$$

$$\dot{\mathbf{s}}_{i,k}^T\dot{\mathbf{s}}_{j,k}\mathrm{E}\left[\mathbf{n}_{in}^T\mathbf{n}_{in}\right]=\dot{\mathbf{s}}_{i,k}^T\dot{\mathbf{s}}_{j,k}\left(\mathrm{Var}(\mathbf{n}_{in})+(\mathrm{E}[\mathbf{n}_{in}])^2\right)=\sigma_{in}^2\dot{\mathbf{s}}_{i,k}^T\dot{\mathbf{s}}_{j,k}, \quad \textbf{[S20]}$$

$$\Rightarrow\sum_{i,j,k}=\sigma_{ex}^2\ddot{\mathbf{s}}_{i,k}^T\ddot{\mathbf{s}}_{j,k}+\sigma_{in}^2\dot{\mathbf{s}}_{i,k}^T\dot{\mathbf{s}}_{j,k}. \quad \textbf{[S21]}$$

The response covariance matrix does not depend on which face is shown and will be designated as $\sum_k$.

On any given trial, when face f is present, the response vector is a random sample from a multivariate normal distribution, $\mathbf{r}_k\sim\mathrm{MVN}(\boldsymbol{\mu}_{f,k},\sum_k)$. The likelihood of observing the response vector, given the hypothesis that face f is shown, is:

$$\ell_{f,k}=\exp\left(-\frac{1}{2}(\mathbf{r}_k-\boldsymbol{\mu}_{f,k})^T\sum_k^{-1}(\mathbf{r}_k-\boldsymbol{\mu}_{f,k})\right). \quad \textbf{[S22]}$$

Performance in terms of proportion correct is given by the probability that the likelihood for the presence of the true face is greater than the likelihoods for each other face. Each possible face then has its own predicted performance:

$$PC_{f,k}=\Pr(\ell_{f,k}>\ell_{f',k},\forall f'\neq f). \quad \textbf{[S23]}$$

The overall performance is the weighted sum of each face's predicted performance:

$$PC_k=\sum_{f=1}^{m}\pi_f PC_{f,k}. \quad \textbf{[S24]}$$

The term π_f is the probability of face f being present (the prior) and normalizes the overall performance. In the identification condition, $\pi_f=\frac{1}{10}$ for each face.

We fit the five free SVCSF parameters (dh, du, dd, n_0, and σ_{in}) by matching the overall group performance from the forced fixation condition at each of the four fixation locations. We then used these parameters to simulate 100,000 trials for each possible fixation point in the stimulus, leading to a 2D predicted performance map. The emotion, gender, and happy vs. neutral tasks were simulated using the same eccentricity factor values from the identification fit while allowing contrast to vary so as to match the overall group performance at the eye fixation in the forced fixation condition.

D. White Noise vs. Contrast Degradation. The FIO uses internal noise to degrade performance to comparative levels as humans. Although this is arguably the most common method in the literature (4–8), we also wanted to test the robustness of our model to alternative performance degradation techniques, namely, modifying signal contrast. We thus implemented models that were subject to either a decrease in the signal contrast (c_0) without the inclusion of internal noise or to both decreased contrast and internal noise addition. For either of these cases, the derivations remain the same as above, except there is now a possible sixth fitting parameter, c_0. The results show modest changes in the profile shape but no change in the predicted peak performance location (Fig. S3*A*).

E. Spatial Uncertainty. There is well-documented uncertainty in the human visual system regarding the estimated spatial position of visual signals (9–11). This leads to the interesting question of how this might affect optimal fixation strategies. We ran our model on the 1-of-10 identification condition while adding spatial uncertainty to the stimuli. To achieve this, we created variants of each template by shifting the images up, down, left, and right by a small distance (0.25° of visual angle). The ideal observer now sums likelihoods for each variant, arriving at summed likelihoods for each identity. The maximum summed likelihood was then used as the decision rule (analogous to summing within class likelihoods for the emotion, gender, and happy/neutral tasks). The results show that overall performance is degraded somewhat but that the prediction of the maximum performing fixation location is unaltered (Fig. S3*B*).

F. White Noise Ideal Observer with Foveation. The FIO makes decisions in an optimal manner, given the joint distribution of template responses. However, the most common ideal observer model in the literature is the traditional white noise ideal observer (WNIO) (2, 12, 13). The WNIO assumes the image noise is additive zero-mean Gaussian and white with variance $\sigma^2=\sigma_{ex}^2+\sigma_{in}^2$, whereas the underlying face image, $\tilde{\mathbf{s}}_{f,k}$, is filtered by the SVCSF. The likelihood calculation is then:

$$\tilde{\ell}_{f,k}=\exp\left(-\frac{\left(\mathbf{g}-\tilde{\mathbf{s}}_{f,k}\right)^T\left(\mathbf{g}-\tilde{\mathbf{s}}_{f,k}\right)}{2\sigma^2}\right)=\exp\left(\frac{-\mathbf{g}^T\mathbf{g}+2\tilde{\mathbf{s}}_{f,k}^T\mathbf{g}-\tilde{\mathbf{s}}_{f,k}^T\tilde{\mathbf{s}}_{f,k}}{2\sigma^2}\right)$$
$$\Rightarrow\exp\left(\frac{2r_{f,k}-\tilde{E}_{f,k}}{2\sigma^2}\right), \quad \textbf{[S25]}$$

where $r_{f,k}$ is the matched template response of face image f with foveation at k to the noisy data and $\tilde{E}_{f,k}$ is the energy of face f with foveation k. With signal uncertainty, the WNIO sums within-class likelihoods.

This model is suboptimal for the current task due to the spatial correlation of the additive noise. However, as seen in Fig. S3*C*, the performance of the WNIO is only slightly lower than that of the FIO for the identification task, while maintaining a similar spatial profile. The deviations from optimality grow with increasing signal uncertainty, yet the spatial profile and predicted point of maximum performance remain steady.

G. Task-Specific Modifications. The derivations were for the identification case in which each decision category, f, was represented by a single image; thus, the decision rule was to choose the maximum likelihood. The other tasks each have signal uncertainty, wherein each decision category can be represented by one of many images. The following sections describe the changes to the FIO algorithm necessary to account for this uncertainty.

The emotion task stimulus set consisted of 140 face images evenly divided into seven categories: afraid, angry, disgusted, happy, neutral, sad, and surprised. The FIO calculates likelihoods for each possible image, $\mathbf{s}_{e,f}$, in the same manner as in Eq. **S19**. The term e is an integer from 1 to 7 specifying the emotion. Now, the FIO's task is to choose the most likely emotional state, given the observed set of responses. Because each emotional state can be represented by any of 20 face images, the FIO must sum the likelihoods for each image within each emotion category to arrive at a summed likelihood, $L_{e,k}$, and then choose the maximum:

$$\ell_{e,f,k}=\exp\left(-\frac{1}{2}(\mathbf{r}_k-\boldsymbol{\mu}_{e,f,k})^T\sum_k^{-1}(\mathbf{r}_k-\boldsymbol{\mu}_{e,f,k})\right), \quad \textbf{[S26]}$$

$$L_{e,k}=\sum_{f=1}^{m}\ell_{e,f,k}, \quad \textbf{[S27]}$$

$$decision = \arg\max_{e} \left(L_{e,k}\right). \quad \text{[S28]}$$

The gender task used 80 face images evenly divided into two categories: male and female. The FIO computation is the same as for the emotion discrimination, where the likelihoods for each individual face within a category, $\ell_{g,f,k}$, are summed and the maximum category summed likelihood, $L_{g,k}$, is taken as the decision.

The happy vs. neutral task used 160 faces evenly divided into two categories: happy and neutral expressions. The FIO computations and decision mechanism are the same as with the gender discrimination task.

The FIO results for the happy vs. neutral task did not fit the human forced fixation data nearly as well as the other three tasks (although the saccade prediction was just as consistent; Fig. 7 *A* and *C*). We speculate that this is a consequence of the task's relative lack of prevalence in day-to-day social interactions. Identifying individuals, determining gender, and recognizing a general emotional state are all ubiquitous tasks that humans monitor and accomplish continuously and automatically. However, noting whether somebody is in a single emotional state (e.g., happiness) would seem to be a more situation-specific task that may recruit different mechanisms and perceptual strategies. Whereas the other three common tasks can be accomplished optimally or near optimally using a single foveation strategy of fixating near the eyes, differentiating a happy vs. neutral expression requires a significant shift in strategy. This difference in familiarity with the task using drastically different foveation strategies could lead to the observed functional effects.

1. Watson AB, Ahumada AJ, Jr. (2005) A standard model for foveal detection of spatial contrast. *J Vis* 5(9):717–740.
2. Burgess AE (1994) Statistically defined backgrounds: Performance of a modified nonprewhitening observer model. *J Opt Soc Am A Opt Image Sci Vis* 11(4):1237–1242.
3. Peli E, Yang J, Goldstein RB (1991) Image invariance with changes in size: The role of peripheral contrast thresholds. *J Opt Soc Am A* 8(11):1762–1774.
4. Burgess AE, Colborne B (1988) Visual signal detection. IV. Observer inconsistency. *J Opt Soc Am A* 5(4):617–627.
5. Najemnik J, Geisler WS (2005) Optimal eye movement strategies in visual search. *Nature* 434(7031):387–391.
6. Legge GE, Kersten D, Burgess AE (1987) Contrast discrimination in noise. *J Opt Soc Am A* 4(2):391–404.
7. Tjan BS, Braje WL, Legge GE, Kersten D (1995) Human efficiency for recognizing 3-D objects in luminance noise. *Vision Res* 35(21):3053–3069.
8. Eckstein M, Bartroff J, Abbey C, Whiting J, Bochud FO (2003) Automated computer evaluation and optimization of image compression of x-ray coronary angiograms for signal known exactly detection tasks. *Opt Express* 11(5):460–475.
9. Davis ET, Kramer P, Graham N (1983) Uncertainty about spatial frequency, spatial position, or contrast of visual patterns. *Percept Psychophys* 33(1):20–28.
10. Burgess AE, Ghandeharian H (1984) Visual signal detection. II. Signal-location identification. *J Opt Soc Am A* 1(8):906–910.
11. Pelli DG (1985) Uncertainty explains many aspects of visual contrast detection and discrimination. *J Opt Soc Am A* 2(9):1508–1532.
12. Bochud FO, Abbey CK, Eckstein MP (2000) Visual signal detection in structured backgrounds. III. Calculation of figures of merit for model observers in statistically nonstationary backgrounds. *J Opt Soc Am A Opt Image Sci Vis* 17(2):193–205.
13. Solomon JA, Pelli DG (1994) The visual filter mediating letter identification. *Nature* 369(6479):395–397.

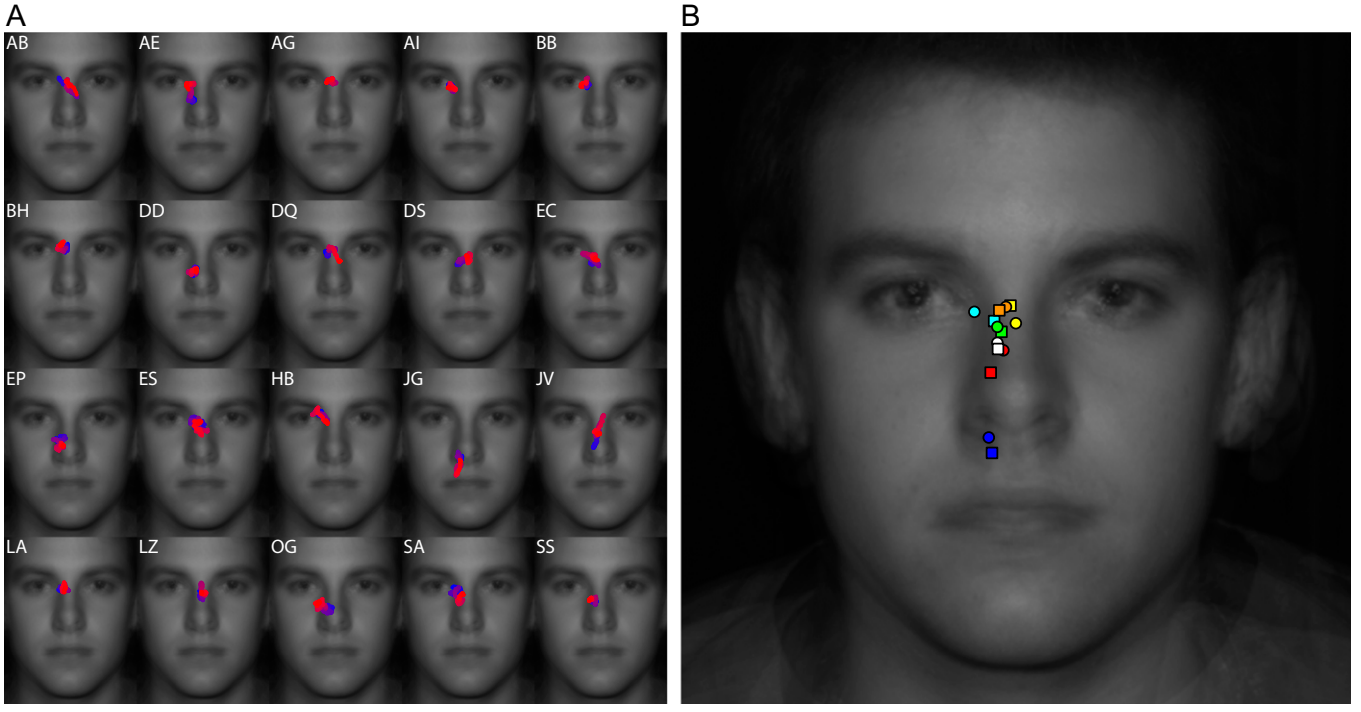

Fig. S1. Effects of feedback. (*A*) Mean saccade landing position is shown for each of the 20 observers in the identification condition. Means were taken across a sliding window of 50 trials, with blue indicating the initial trials and bright red indicating the later trials (500 trials in total). Eye movements generally remain stable across the testing session, indicating a lack of sensitivity to feedback. (*B*) Comparison of eye movement results between feedback conditions. Different colors represent individual observers, with white representing the group average. Observers used the same strategy regardless of whether feedback was provided (circles) or not (squares).

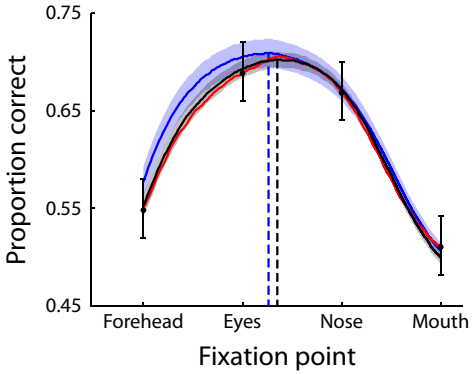

Fig. S2. Comparison of FIO model predictions across databases. Results for the 10 faces used in the human study are shown in red. The solid black line represents predictions for our entire 150 face in-house database, whereas the solid blue line is the prediction for the 850 faces gathered from the Internet. Shaded regions are the SEs for each database (gray for the in-house set and light blue for the Internet set). Dashed lines indicate the location of each database's predicted maximum performance.

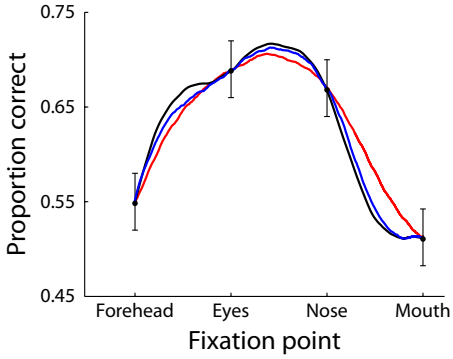

Fig. S3. FIO model predictions for different forms of performance degradation. A model that attenuates signal strength solely through lowering global contrast is shown in black. A model that leaves contrast unaltered from the human experiment but adds white Gaussian internal noise is shown in red. A hybrid model, which attenuates signal contrast and adds internal noise, is shown in blue. The overall shape of the performance profile varies moderately with choice of decreasing contrast vs. internal noise, yet the prediction for the maximum performing fixation location remains largely unaltered.

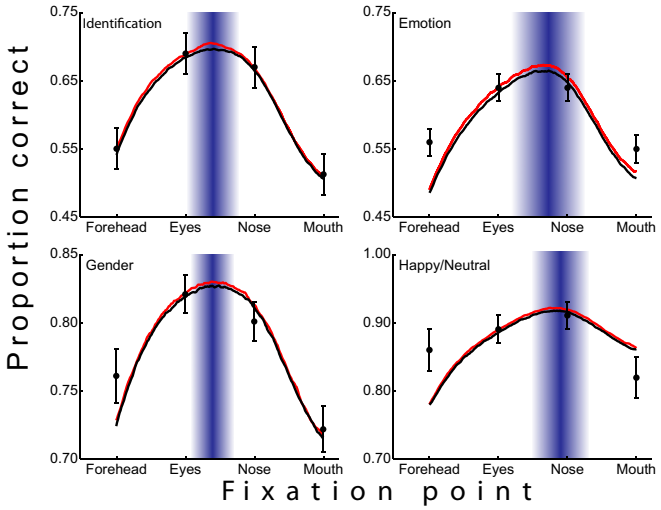

Fig. S4. Foveated WNIO results are shown in black. In general, this suboptimal model yields similar predictions but with slightly lower performance at the same internal noise levels as the FIO (red). This difference becomes more pronounced as signal uncertainty increases. As in the main text, human saccade distributions are shown in blue, whereas human performance in the forced fixation condition is represented, with the SEM, by black dots.

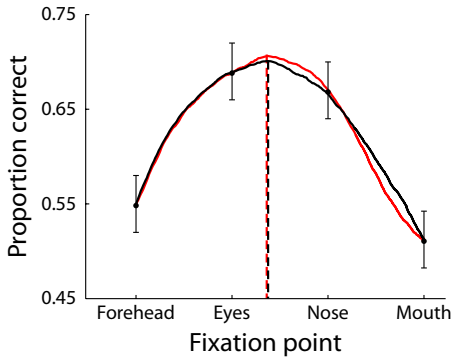

Fig. S5. Model that incorporates spatial uncertainty (solid black line) alters the shape of the performance profile slightly while leaving the predicted maximum fixation point (dashed lines) unchanged compared with a model with no uncertainty (solid red line).

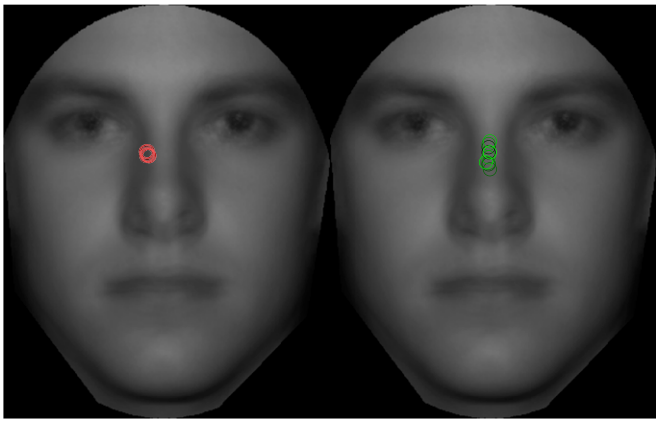

Fig. S6. (*Left*) Mean saccade landing points are shown, averaged across observers, for the identification task as a function of which of the 10 faces was displayed (shown by different shades of red). The dense overlap shows that humans used a single foveation strategy. (*Right*) Maximum performance points according to the FIO are shown, again as a function of the actual face that was displayed. Here, we see a slight differentiation, yet the grouping remains strong. Thus, even if good-quality information was available before saccade from the far periphery, the fixation strategy would remain stable.

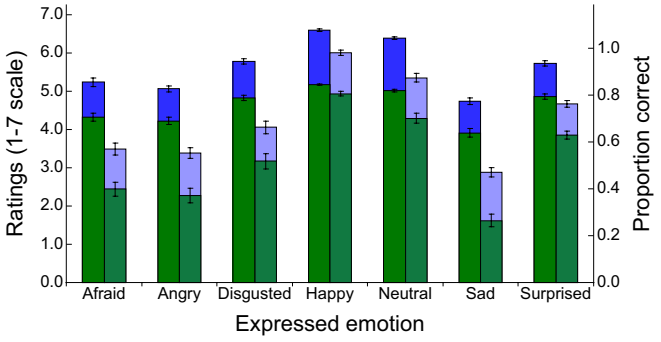

Fig. S7. Ratings of the emotional faces used in the study taken from our in-house database. Average ratings (on a 1–7 scale) are shown for the top 10 rated faces in each gender group (dark blue) and the faces from the database that were deemed unsuitable for experimental use (light blue). The corresponding proportion correct for the same face groups (defined as being rated a 4 or above, on average) are shown in dark and light green.

Table S1. Results of *t* tests for how well the FIO predicts human fixations compared with the other models for each task

Task	ROI	Visible face	Frame	Head
Identification	**5.32**	**5.20**	**2.78**	1.58
df = 19	2.0e-5	2.5e-5	0.006	0.065
Emotion	**6.20**	**2.39**	–0.65	**2.89**
df = 19	3.0e-6	0.014	0.739	0.005
Gender	**8.44**	**4.81**	**2.04**	**2.14**
df = 19	3.8e-8	6.1e-5	0.028	0.023
Happy/neutral	**10.39**	0.40	1.68	**7.94**
df = 19	1.4e-9	0.348	0.055	9.3e-8
All tasks	**13.00**	**6.34**	**3.50**	**6.43**
df = 79	1.3e-21	6.6e-9	3.9e-4	4.4e-9

The upper number for each task is the *t* statistic, and the lower number is the *P* value. The *t* statistics in bold are significant at the 0.05 level, false discovery rate-corrected. ROI, region of interest.