Vision Research 123 (2016) 26–32

Contents lists available at ScienceDirect

Vision Research

journal homepage: www.elsevier.com/locate/visres

Classification images in a very general decision model

Richard F. Murray

Department of Psychology and Centre for Vision Research, York University, 4700 Keele Street, LAS 0009, Toronto, Ontario M3J 1P3, Canada

ARTICLE INFO

Article history:
Received 30 January 2015
Received in revised form 24 March 2016
Accepted 21 April 2016

Keywords:
Classification images
Decision making
Modelling

ABSTRACT

Most of the theory supporting our understanding of classification images relies on standard signal detection models and the use of normally distributed stimulus noise. Here I show that the most common methods of calculating classification images by averaging stimulus noise samples within stimulus-response classes of trials are much more general than has previously been demonstrated, and that they give unbiased estimates of an observer's template for a wide range of decision rules and non-Gaussian stimulus noise distributions. These results are similar to findings on reverse correlation and related methods in the neurophysiology literature, but here I formulate them in terms that are tailored to signal detection analyses of visual tasks, in order to make them more accessible and useful to visual psychophysicists. I examine 2AFC and yes-no designs. These findings make it possible to use and interpret classification images in tasks where observers' decision strategies may not conform to classic signal detection models such as the difference rule, and in tasks where the stimulus noise is non-Gaussian.

1. Classification images in a very general decision model

Classification images have proven to be a useful tool for investigating visual processing in a wide range of tasks (Ahumada, 1996, 2002; Murray, 2011). In a classification image experiment we introduce many small fluctuations into a stimulus, and measure the influence of these fluctuations on observers' responses. One appealing feature of this approach is that it probes observers' strategies in a very open-ended way. Instead of using, say, proportion correct or reaction time measurements to choose between two or three candidate models, a classification image experiment gives a highly flexible description of how observers make visual judgements, and can reveal features of visual processing that may not have been anticipated by the experimenter (e.g., Ahumada, 1996; Gold, Murray, Bennett, & Sekuler, 2000; Neri & Heeger, 2002).

However, most of the theory for understanding classification images is based on a few standard models from signal detection theory (e.g., Abbey & Eckstein, 2002; Murray, Bennett, & Sekuler, 2002). As a result, this highly flexible method actually seems to depend on rigid assumptions about visual processing, such as the assumption that observers make 2AFC decisions by calculating a decision variable from each stimulus interval, and choosing the interval with the higher decision variable. Furthermore, there have long been doubts about whether these assumptions are always correct (e.g., Treisman & Leshowitz, 1969; Yeshurun, Carrasco, & Maloney, 2008), and this raises the question of what classification images tell us about observers' strategies when these assumptions fail.

In addition, some interesting results have come from studies where standard methods of calculating classification images are applied to new tasks that are not described well by the models that were originally used to justify the standard methods. For example, classification images have been measured in visual search tasks (Rajashekar, Bovik, & Cormack, 2006; Saiki, 2008), which are not instances of the yes-no or 2AFC tasks that underlie the justifications for standard classification image methods, and for which there is no broad agreement about the correct psychophysical model. Here classification images are used outside the domain where they are well understood theoretically, and so again there is room for questions about exactly what they tell us about observers' strategies.

Similarly, Pritchett and Murray (2015) used classification images to estimate observers' decision variables on individual trials, and then they used these estimates to study observers' decision rules in 2AFC tasks. The previous literature suggests that this approach is problematic, because the classification image methods that Pritchett and Murray used have been justified using a specific model of 2AFC decision making (the difference rule), whereas it is precisely the decision rule in 2AFC tasks that Pritchett and Murray are attempting to investigate.

The most widely used methods of calculating classification images are based on averages of Gaussian stimulus noise within stimulus-response classes of trials (Abbey & Eckstein, 2002;

E-mail address: rfm@yorku.ca

http://dx.doi.org/10.1016/j.visres.2016.04.003
0042-6989/

Ahumada, 2002; Murray et al., 2002). I will call these conditional average methods. (An example of a method outside this category is estimating classification images using the generalized linear model, e.g., Knoblauch and Maloney (2008).) Here I show that conditional average methods of calculating classification images are far more general than has been previously demonstrated, and that they give unbiased estimates of observers' templates for a wide range of decision rules and non-Gaussian stimulus noise distributions. The main assumption behind these results is simply that the observer's responses are mediated by the dot product of a template with the stimulus. To show that non-Gaussian noise can be used in classification image experiments, I also assume that each noise element has only a small influence on the observer's responses.

First I discuss classification images measured using Gaussian noise in a 2AFC task, and then using Gaussian noise in a yes-no task. Finally I discuss classification images measured using non-Gaussian noise in a yes-no task.

1.1. Classification images in a 2AFC task

I use upper case letters for matrices and random variables, and lower case letters for scalar constants. I use bold font for images and templates, which I represent as column vectors.

1.1.1. The task

In a 2AFC task there are two stimulus intervals, which I will label as $k = 1, 2$. In each interval the observer views a stimulus $\mathbf{s}_k + \mathbf{N}_k$, where $\mathbf{s}_k$ is a signal and $\mathbf{N}_k$ is noise. I assume that $\mathbf{N}_k$ is a linear transformation of independent and identically distributed (i.i.d.) Gaussian noise: $\mathbf{N}_k = A\mathbf{M}_k$, where $\mathbf{N}_k$ is an $n \times 1$ vector, A is an $n \times m$ matrix and $\mathbf{M}_k$ is an $m \times 1$ vector of Gaussian noise with each element an i.i.d. sample from $N(0, \sigma_M^2)$. In practice, A is usually a convolution, which can produce i.i.d. noise (if A is the identity matrix) or correlated noise. The two signals $\mathbf{g}_1$ and $\mathbf{g}_2$ appear in random order, and I represent the signal order with a random variable S that takes value 1 or 2 to indicate which stimulus interval $\mathbf{g}_1$ appeared in. Thus the signal in interval 1 is $\mathbf{s}_1 = \mathbf{g}_S$ and the signal in interval 2 is $\mathbf{s}_2 = \mathbf{g}_{3\text{-}S}$. The observer's task is to judge which stimulus order was shown. I represent the observer's responses with a random variable R that takes value 1 or 2 to indicate which stimulus interval the observer judged signal $\mathbf{g}_1$ to be in.

1.1.2. The observer model

I assume that the observer's responses are based on decision variables D_1 and D_2 that are calculated from the two stimulus intervals. I assume that the stimulus affects the decision variable for stimulus interval k via a dot product of the stimulus with a template $\mathbf{t}_k$.

$$E_k = (\mathbf{s}_k + \mathbf{N}_k)^{\mathrm{T}}\mathbf{t}_k \tag{1}$$

Here T is matrix transposition. I call E_k the 'external component of the decision variable'. I allow different templates $\mathbf{t}_k$ for the two stimulus intervals. The decision variable D_k is some function of the random variable E_k and a multivariate random variable $\mathbf{V}_k$ that represents trial-to-trial fluctuations that are independent of the stimulus noise, such as internal noise that the observer adds to the external component of each decision variable.

$$D_k = f(E_k, \mathbf{V}_k) \tag{2}$$

To describe the observer's decision rule, I define the decision space as

$$H(x_1, x_2) = P(R = 2 | D_1 = x_1, D_2 = x_2) \tag{3}$$

This is the probability of the observer choosing response 2 given the values of the decision variables D_1 and D_2. I do not rely on the usual assumption that 2AFC decisions are based on the difference rule (Tanner & Swets, 1954), which says:

$$H(x_1, x_2) = \begin{cases} 1 & \text{if } x_2 \leqslant x_1 \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

Here the criterion $x_2 \leqslant x_1$ assumes that the templates $\mathbf{t}_k$ have a higher response to signal $\mathbf{g}_2$ than to $\mathbf{g}_1$; if they have a higher response to $\mathbf{g}_1$, then the criterion is $x_1 \leqslant x_2$. Instead of assuming Eq. (4), as previous studies have done (e.g., Abbey & Eckstein, 2002), I allow the decision space $H(x_1, x_2)$ to be an arbitrary function from $\mathbb{R}^2$ to $[0, 1]$.

In this observer model there is a redundancy between the internal variability $\mathbf{V}_k$ and the decision space H, because any randomness in the observer's responses caused by $\mathbf{V}_k$ could be absorbed into the decision space. However, I will keep this redundancy because it allows us to separately describe channel noise using $\mathbf{V}_k$ (e.g., additive Gaussian noise) and decision noise using H (e.g., randomness due to probability matching (Murray, Patel, & Yee, 2015)), and so it makes the observer model more easily relatable to common signal detection models.

1.1.3. The classification image

Conditional average methods of calculating a classification image in a 2AFC task are based on the average of stimulus noise samples within stimulus-response classes of trials. In Appendix A I derive the conditional expected value $E[\mathbf{M}_1 | S = 1, R = 2]$, where $\mathbf{M}_1$ is the Gaussian i.i.d. noise used to generate the stimulus noise $\mathbf{N}_1 = A\mathbf{M}_1$ in the first stimulus interval. I show that:

$$E[\mathbf{M}_1 | S = 1, R = 2] \propto A^T\mathbf{t}_1 \tag{5}$$

That is, the expected value of the noise vector $\mathbf{M}_1$ is either zero or proportional to the observer's template, transformed by the transpose of the matrix A used to generate the stimulus noise. If A^T is invertible, then the conditional expected value of $(A^T)^{-1}\mathbf{M}_1 = A^{-T}\mathbf{M}_1$ is either zero or proportional to the template $\mathbf{t}_1$. This means that the average of the samples of $A^{-T}\mathbf{M}_1$ on all trials where $S = 1$ and $R = 2$ gives an unbiased estimate of the template $\mathbf{t}_1$ that the observer uses in the first stimulus interval. Alternatively, if we wish to use the stimulus noise $\mathbf{N}_1 = A\mathbf{M}_1$ to estimate the template, then we can take the average of the samples of $A^{-\mathrm{T}}A^{-1}\mathbf{N}_1$ on all trials where $S = 1$ and $R = 2$.

The derivation in Appendix A shows that Eq. (5) is true regardless of the observer's decision space (i.e., the function H in Eq. (3)). Thus conditional average methods do not rely on specific assumptions about observers' decision rules, such as the difference rule in Eq. (4), but instead can be used whenever the stimulus affects the observer's responses via a dot product, as in Eq. (1).

This result can be explained simply and informally as follows. Starting with Eq. (1), the external component of the decision variable is

$$E_k = (\mathbf{s}_k + \mathbf{N}_k)^{\mathrm{T}}\mathbf{t}_k \tag{6}$$

$$= \mathbf{s}_k^{\mathrm{T}}\mathbf{t}_k + (A\mathbf{M}_k)^{\mathrm{T}}\mathbf{t}_k \tag{7}$$

$$= \mathbf{s}_k^{\mathrm{T}}\mathbf{t}_k + \mathbf{M}_k^{\mathrm{T}}(A^{\mathrm{T}}\mathbf{t}_k) \tag{8}$$

The shift in parentheses from Eqs. (7) to (8) shows that applying a template $\mathbf{t}_k$ to filtered noise $A\mathbf{M}_k$ gives the same result as applying a transformed template $A^{\mathrm{T}}\mathbf{t}_k$ to i.i.d. noise $\mathbf{M}_k$. If the stimulus affects the observer's responses only via a template, then i.i.d. noise that is orthogonal to the template can have no effect on the observer's responses, and the expected value of the i.i.d. noise in a stimulus-response class of trials can only be zero or proportional to the template. Thus the expected value of $\mathbf{M}_1$ in a stimulus-response

class of trials is either zero or proportional to $A^T\mathbf{t}_1$, no matter how the external components E_k of the decision variables enter into the observer's decision rule.

In Appendix A I also find the magnitude of the conditional expected value in Eq. (5), which allows one to combine stimulus noise from different stimulus-response classes so as to maximize the signal-to-noise ratio of the resulting classification image (Murray et al., 2002). Similar derivations give the conditional average for other stimulus-response classes of trials.

1.2. Classification images in a yes-no task

The decision rule in the yes-no task has not been as controversial as the difference rule in the 2AFC task. Nevertheless, we can use a similar framework to examine classification images in a yes-no task under a very general decision model.

In a yes-no task there is one stimulus interval, and on each trial one of two signals, $\mathbf{g}_1$ and $\mathbf{g}_2$, is shown in external noise $\mathbf{N}$, so the stimulus is $\mathbf{s}+\mathbf{N}$. I assume that $\mathbf{N}=A\mathbf{M}$, where $\mathbf{M}$ is zero-mean Gaussian i.i.d. noise. I also assume that the observer's responses are based on a decision variable $D=f(E,\mathbf{V})$ that is a function of the dot product $E=(\mathbf{s}+\mathbf{N})^T\mathbf{t}$, where $\mathbf{t}$ is the observer's template and a random variable $\mathbf{V}$ that represents internal variability. In Appendix B I show that, regardless of the observer's decision space (which is now one-dimensional),

$$E[\mathbf{M}|S=1,R=2] \propto A^T\mathbf{t} \tag{9}$$

This parallels the result for 2AFC tasks, and for similar reasons: if the stimulus noise enters into the observer's decision rule via a dot product, then noise components that are orthogonal to the transformed template $A^T\mathbf{t}$ can have no influence on the observer's responses, and so the conditional expected value of the noise can only be zero or proportional to the transformed template.

1.3. Non-Gaussian noise in a yes-no task

Under some conditions, conditional average methods *cannot* be used with non-Gaussian stimulus noise. This is illustrated in Fig. 1, which shows simulated classification image experiments in a yes-no task where there are just two stimulus pixels, and the observer uses a template matching strategy. In Fig. 1a the stimulus noise is i.i.d. and Gaussian. Here the circular symmetry of the noise distribution ensures that expected value of the noise in each stimulus-response class of trials is proportional to the template (cf. Chichilnisky, 2001). In Fig. 1b the stimulus noise is i.i.d. and uniform. Here the noise distribution is not circularly symmetric, and the expected value of the noise in each stimulus-response class of trials is not necessarily proportional to the template.

However, if each i.i.d. stimulus noise element has only a small influence on the observer's responses, then the expected value of the noise in each stimulus-response class of trials is nevertheless approximately proportional to the template. To show this I assume that the stimulus noise $\mathbf{N}$ in a yes-no task is i.i.d. noise, where the noise elements share a probability density function $p_N(x)$ (not necessarily Gaussian) that has a mean of zero and a standard deviation σ_N. In Appendix C I show that if each i.i.d. noise element of the stimulus has only a small influence on the observer's responses, then the observer's response probabilities are an approximately linear function of each stimulus noise element, with a slope that is proportional to the corresponding template element. I show that we can use a linear approximation for the effect of an element n_i of the stimulus noise $\mathbf{N}$ on the observer's response probabilities when the signal is $S=1$:

$$P(R=2|S=1,n_i=x) \simeq u_1 + v_1 t_i x \tag{10}$$

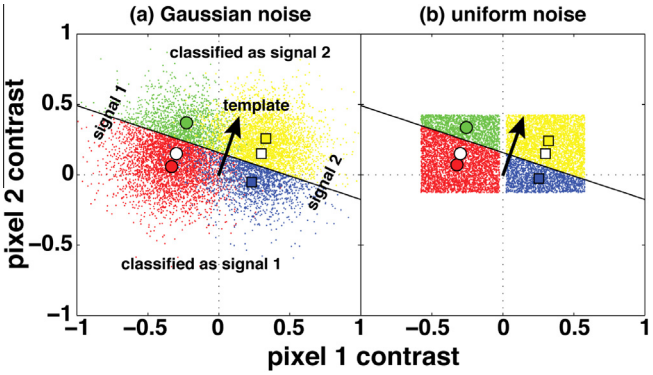

Fig. 1. Illustration of a simulated two-pixel classification image experiment with a template matching observer in a yes-no task. In panel (a) the stimulus noise is Gaussian and in panel (b) it is uniform. The white circle is the mean of the stimuli containing signal 1, and the white square is the mean of the stimuli containing signal 2. The small, coloured data points represent stimuli on individual trials. Small red points are stimuli that contained signal 1 and were identified as signal 1, and small green points are stimuli that contained signal 1 and were identified as signal 2. Small yellow points are stimuli that contained signal 2 and were identified as signal 2, and small blue points are stimuli that contained signal 2 and were identified as signal 1. The large coloured symbols are the averages of the corresponding small coloured points, e.g., the red circle is the average of the small red points. The black arrow is the hypothetical observer's template, and the oblique black line is the decision line that the observer used to decide whether to identify a stimulus as signal 1 or signal 2. In panel (a) the noise distributions are circular, and so the conditional stimulus means (large coloured symbols) are displaced from the unconditional means (large white symbols) parallel to the template. That is, the conditional means of the stimulus noise are proportional to the template. In panel (b) the noise distributions are not circular, and so the conditional stimulus means are not necessarily displaced parallel to the template. (This figure and caption are modelled after Figure 2 in Murray (2011).) (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Here t_i is the template element that corresponds to noise element n_i. The constants u_1 and v_1 depend on the signal being shown ($S=1$ or $S=2$), so I subscript them u_s as v_s and, where s is the signal number. In Appendix C I also show that from this approximation, it follows that the expected value of noise element n_i on trials where $S=1$ and $R=2$ is

$$E[n_i|S=1,R=2] = \frac{t_i v_1 \sigma_N^2}{P(R=2|S=1)} \tag{11}$$

Thus the conditional expected value of each noise element n_i is proportional to the corresponding template element t_i, regardless of whether the noise density $p_N(x)$ is Gaussian. This result justifies measuring classification images with non-Gaussian i.i.d. noise when each noise element has only a small influence on the observer's responses, and also gives a new explanation of how classification images recover an observer's template from Gaussian i.i.d. noise.

2. Discussion

Classification images are a useful tool for investigating observers' strategies in perceptual tasks, but most of the theory supporting them has made strong assumptions about observers' decision rules. This has made it difficult to use classification images in cases where observers' decision rules are in question. The present findings show that we do not need to make such strong assumptions in order to estimate observers' templates using classification images. In the visual search tasks mentioned in the introduction, for instance, we can calculate classification images using conditional averages, even if we know very little about how observers actually carry out visual search tasks, as long as we are willing to assume that visual processing begins with a template matching operation. Similarly, Pritchett and Murray (2015) are able to

estimate decision variables on individual 2AFC trials using classification images, even in tasks where the standard model of 2AFC decision making that underlies previous analyses of classification images is in doubt.

Reverse correlation methods in neurophysiology have tended to make fairly weak modelling assumptions about neurons' response properties, compared to the assumptions that classification image methods have made about human observers' decision rules. Some results in this paper, which generalize previous results on classification images, can be seen as instances of results in the reverse correlation literature. For example, Chichilnisky (2001) assumes only that a neuron's mean spike rate is some nonlinear function of the dot product of a template with the stimulus, and that the probability distribution of stimuli is spherically symmetric, e.g., multivariate Gaussian white noise. Chichilnisky shows that under these conditions, the expected value of the spike-triggered average (i.e., the average of all stimuli immediately preceding spikes) is proportional to the template. This is similar to the present paper's results concerning yes-no tasks with Gaussian noise and an arbitrary decision space (i.e., Eq. (9)). Much like Chichilnisky, I have assumed that the observer's response probability is some nonlinear function (i.e., the decision space) of the dot product of a template with the stimulus, and shown that the expected value of the white noise $\mathbf{M}$ (which is used to generate the stimulus noise $\mathbf{N} = A\mathbf{M}$) on trials where the observer makes a particular response (e.g., $R = 2$) is proportional to the template. I allow the stimulus noise to be non-white, but as Eqs. (6)–(8) show, this is a superficial difference, because a trial with non-white noise $A\mathbf{M}$ and a template $\mathbf{t}$ is equivalent to a trial with white noise $\mathbf{M}$ and a transformed template $A^T\mathbf{t}$. Paninski (2003) generalizes Chichilnisky's results on the spike-triggered average, allowing for a wider range of stimulus distributions, and he also considers problems that can arise when the response nonlinearity is such that the spike-triggered average has an expected value of zero, as with complex cells.

In addition to showing that classification images are asymptotically unbiased estimators of the observer's template, it is also useful to consider their rate of convergence. Murray et al. (2002) defined the signal-to-noise ratio (SNR) of a classification image as the sum-of-squares of its expected value, divided by its pixelwise variance. They showed that given white Gaussian stimulus noise and a standard linear model of the observer, the SNR of a classification image is proportional to the number of trials used to measure it. In the physiological literature, Paninski (2003) used weaker assumptions, similar to those I have made in this paper, to derive the convergence rate of the spike-triggered average. In this paper I do not examine convergence properties of classification images in the more general decision space model. However, it seems likely that Murray et al.'s methods can easily be used to establish the SNR in this more general case, since their methods only rely on knowing the expected value and pixelwise variance of the stimulus noise in each stimulus-response class of trials, which I derive in Appendices A and B. It also seems likely that Paninski's methods can be used to examine convergence of classification images in the more general model, since the assumptions he makes about neural responses and the assumptions I make here about human response probabilities are similar, i.e., in both cases they are a nonlinear function of the dot product of a template and the stimulus.

The observer model that I have relied on is very general, but there are mechanisms that it does not cover. For example, intrinsic uncertainty models effectively use multiple templates (Pelli, 1985), and energy models use two templates in quadrature (Adelson & Bergen, 1985), so the class of models that use a single template in each stimulus interval does not include these mechanisms; linear subspace methods that allow multiple templates may be useful in these cases (Paninski, 2003). Similarly, models with an early pointwise nonlinearity (e.g., Chubb, Econopouly, & Landy, 1994; Neri, 2010) do not meet the model's assumption that the template is applied directly to the stimulus.

The generalized linear model can be used to calculate classification images (Knoblauch & Maloney, 2008), and this model makes even weaker assumptions about the stimulus noise distribution than the model I have used here. This has been exploited by some researchers to measure classification images using only the natural variability of the stimulus set instead of adding stimulus noise (Macke & Wichmann, 2010), and to measure classification images in tasks where there are more than two possible responses (Knoblauch & Maloney, 2008; Murray, 2011). The present model is more general, though, in allowing arbitrary maps from the dot product of the stimulus and template to the decision variable, and in allowing arbitrary decision spaces. Furthermore, template estimates from conditional average methods are statistically more tractable in some ways than those from the generalized linear model, e.g., it is easy to show that the expected value of the noise field in a stimulus-response class of trials is proportional to the template, and to find its signal-to-noise ratio. Each method has its advantages and will be useful in different circumstances.

Author note

Correspondence may be sent to Richard Murray (rfm@yorku.ca).
Parts of this work appear in the supporting information to Pritchett and Murray (2015). This work was funded by grants from the Natural Sciences and Engineering Research Council of Canada and the Canada Foundation for Innovation.

Appendix A

Here I find the conditional expected value $E[\mathbf{M}_1|S = 1, R = 2]$, where $\mathbf{M}_1$ is the i.i.d. noise used to generate the stimulus noise $\mathbf{N}_1 = A\mathbf{M}_1$ in the first stimulus interval. I use the notation and observer model introduced in the section 'Classification images in a 2AFC task' in the main text.

I represent $\mathbf{M}_1$ in an orthonormal basis where the first basis vector is parallel to $A^T\mathbf{t}_1$. To do this I define the orthogonal matrix

$$Q = \left[\frac{A^T\mathbf{t}_1}{|A^T\mathbf{t}_1|} \quad \mathbf{u}_2 \quad \dots \quad \mathbf{u}_m\right] \tag{A1}$$

where $\mathbf{u}_i$ are mutually orthogonal unit vectors that are also orthogonal to $A^T\mathbf{t}_1$. The specific choice of $\mathbf{u}_i$ is not important. In this basis, the conditional expected value of $\mathbf{M}_1$ on trials where the signal order is $S = 1$ and the observer's response is $R = 2$ is:

$$E[Q^T\mathbf{M}_1|S = 1, R = 2] = \int_{\mathbb{R}^m} \mathbf{x}P(Q^T\mathbf{M}_1 = \mathbf{x}|S = 1, R = 2)d\mathbf{x} \tag{A2}$$

Bayes' theorem shows that this is equal to

$$= \int_{\mathbb{R}^m} \mathbf{x}\frac{P(R = 2|Q^T\mathbf{M}_1 = \mathbf{x}, S = 1)P(Q^T\mathbf{M}_1 = \mathbf{x}|S = 1)}{P(R = 2|S = 1)}d\mathbf{x} \tag{A3}$$

In order to treat the first coordinate and the remaining coordinates separately, I define u as the first component of $\mathbf{x}$, and $\mathbf{v}$ as the remaining components, so that $\mathbf{x} = (u, \mathbf{v})^T$. I also define $W_{||}$ as the first component of $\mathbf{W} = Q^T\mathbf{M}_1$, and $\mathbf{W}_\perp$ as the remaining components, so that $Q^T\mathbf{M}_1 = (W_{||}, \mathbf{W}_\perp)^T$. To shorten the notation I define $p_{12} = P(R = 2|S = 1)$. Using these definitions, and the fact that $\mathbf{M}_1$ is independent of the signal order S, we can write Eq. (A3) as

$$= \frac{1}{p_{12}}\int_{-\infty}^{\infty}\int_{\mathbb{R}^{m-1}} (u, \mathbf{v})^T P(R = 2|W_{||} = u, \mathbf{W}_\perp = \mathbf{v}, S = 1)P(W_{||} = u, \mathbf{W}_\perp = \mathbf{v})d\mathbf{v}du \tag{A4}$$

The observer's responses are independent of $\mathbf{W}_\perp$. To see this, note that Eq. (1) shows that the noise in the first stimulus interval affects the observer's responses only through the dot product

$$\mathbf{N}_1^{\mathrm{T}}\mathbf{t}_1 = (A\mathbf{M}_1)^T\mathbf{t}_1 = \mathbf{M}_1^T(A^T\mathbf{t}_1) \tag{A5}$$

The matrix Q is orthogonal, so this is equal to

$$= (Q^T\mathbf{M}_1)^T(Q^TA^T\mathbf{t}_1) \tag{A6}$$

By definition, $Q^T\mathbf{M}_1 = (W_{||}, \mathbf{W}_\perp)^T$, and Eq. (A1) shows that the basis represented by Q is chosen so that $Q^TA^T\mathbf{t}_1 = (|A^T\mathbf{t}_1|, \mathbf{0})^T$, so Eq. (A6) can be written as

$$= (W_{||}, \mathbf{W}_\perp)(|A^T\mathbf{t}_1|, \mathbf{0})^T = W_{||}|A^T\mathbf{t}_1| \tag{A7}$$

which does not depend on $\mathbf{W}_\perp$.

Having shown that the observer's responses are independent of $\mathbf{W}_\perp$, we can rewrite Eq. (A4) as

$$E[Q^T\mathbf{M}_1|S=1,R=2] = \frac{1}{p_{12}}\int_{-\infty}^{\infty}\int_{\mathbb{R}^{m-1}}(u,\mathbf{v})^T P(R=2|W_{||}=u,S=1)P(W_{||}=u,\mathbf{W}_\perp=\mathbf{v})d\mathbf{v}du \tag{A8}$$

$\mathbf{M}_1$ is isotropic Gaussian i.i.d. noise, so $Q^T\mathbf{M}_1 = (W_{||}, \mathbf{W}_\perp)^T$ is also isotropic Gaussian i.i.d. noise, and this means that $W_{||}$ and $\mathbf{W}_\perp$ are independent.

$$= \frac{1}{p_{12}}\int_{-\infty}^{\infty}\int_{\mathbb{R}^{m-1}}(u,\mathbf{v})^T P(R=2|W_{||}=u,S=1) P(W_{||}=u)P(\mathbf{W}_\perp=\mathbf{v})d\mathbf{v}du \tag{A9}$$

$$= \frac{1}{p_{12}}\int_{-\infty}^{\infty}(u,\mathbf{0})^T P(R=2|W_{||}=u,S=1)P(W_{||}=u)du \tag{A10}$$

Eq. (A10) shows that the conditional expected value of $Q^T\mathbf{M}_1$ is $(u^*, \mathbf{0})$ for some value of $u*$, so the conditional expected value of $\mathbf{M}_1$ is either zero or proportional to the transformed template $A^{\mathrm{T}}\mathbf{t}_1$. As noted in the main text, this means that if A^T is invertible, then the conditional expected value of $(A^T)^{-1}\mathbf{M}_1 = A^{-T}\mathbf{M}_1$ is either zero or proportional to the template $\mathbf{t}_1$.

We can continue from Eq. (A10) to find the magnitude of the conditional expected value of $Q^T\mathbf{M}_1$. To find $P(R=2|W_{||}=u,S=1)$ in Eq. (A10) we use the law of total probability, and partition over the internal variability $\mathbf{V}_1$ and the second decision variable D_2.

$$E[Q^T\mathbf{M}_1|S=1,R=2] = \frac{1}{p_{12}}\int_{-\infty}^{\infty}(u,\mathbf{0})^T\int_{\mathbb{R}^p}\int_{-\infty}^{\infty}P(R=2|W_{||}=u, S=1,\mathbf{V}_1=\mathbf{v},D_2=y)P(\mathbf{V}_1=\mathbf{v},D_2=y|W_{||}=u, S=1)dyd\mathbf{v}P(W_{||}=u)du \tag{A11}$$

To go further we need a model of the internal variability $\mathbf{V}_k$. For illustration I will assume that the variability $\mathbf{V}_k$ is limited to independent, normally distributed random variables representing internal noise, $I_k \sim N(0,\sigma_I^2)$, and that the internal noise is added to the external component of the decision variable, so $D_k = f(E_k, I_k) = E_k + I_k$. In the case we are considering where $S=1$, the decision variables D_k are normally distributed with mean $\mu_{Dk} = \mathbf{g}_k^{\mathrm{T}}\mathbf{t}_k$ and variance $\sigma_{Dk}^2 = |A^T\mathbf{t}_k|^2\sigma_M^2 + \sigma_I^2$.

Continuing from Eq. (A11), I use the fact that the internal and external noise samples in the two stimulus intervals are mutually independent.

$$= \tfrac{1}{p_{12}}\int_{-\infty}^{\infty}(u,\mathbf{0})^T\int\int_{-\infty}^{\infty}P(R=2|W_{||}=u,S=1,I_1=v,D_2=y) P(I_1=v)P(D_2=y|S=1)dydvP(W_{||}=u)du \tag{A12}$$

On trials where $S=1$, the decision variable for the first stimulus interval is

$$D_1 = (\mathbf{g}_1 + A\mathbf{M}_1)^{\mathrm{T}}\mathbf{t}_1 + I_1 \tag{A13}$$

$$= \mathbf{g}_1^{\mathrm{T}}\mathbf{t}_1 + W_{||}|A^{\mathrm{T}}\mathbf{t}_1| + I_1 \tag{A14}$$

Eq. (A14) and the decision space $H(x_1, x_2)$ let us evaluate the first factor inside the double integral in Eq. (A12).

$$E[Q^T\mathbf{M}_1|S=1,R=2] = \frac{1}{p_{12}}\int_{-\infty}^{\infty}(u,\mathbf{0})\int\int_{-\infty}^{\infty}H(\mathbf{g}_1^{\mathrm{T}}\mathbf{t}_1+u|A^{\mathrm{T}}\mathbf{t}_1|+v,y)P(I_1=v) P(D_2=y|S=1)dydvP(W_{||}=u)du = \frac{1}{p_{12}}\int_{-\infty}^{\infty}(u,\mathbf{0})\int\int_{-\infty}^{\infty}H(\mathbf{g}_1^{\mathrm{T}}\mathbf{t}_1+u|A^{\mathrm{T}}\mathbf{t}_1|+v,y) \tag{A15}$$

$$\phi(v,0,\sigma_I)\phi(y,\mathbf{g}_2{}^{\mathrm{T}}\mathbf{t}_2,\sigma_{D2})dydv\phi(u,0,\sigma_M)du \tag{A16}$$

Here $\phi(x,\mu,\sigma)$ is the normal probability density function. We can rewrite Eq. (A16) more clearly as

$$= \frac{1}{p_{12}}\int_{-\infty}^{\infty}(u,\mathbf{0})g(u)\phi(u,0,\sigma_M)du \tag{A17}$$

where

$$g(u) = \int\int_{-\infty}^{\infty}H(\mathbf{g}_1{}^{\mathrm{T}}\mathbf{t}_1 + u|A^{\mathrm{T}}\mathbf{t}_1| + v,y)\phi(v,0,\sigma_I)\phi(y,\mathbf{g}_2^{\mathrm{T}}\mathbf{t}_2,\sigma_{D2})dydv \tag{A18}$$

The function $g(u)$ is a weighted sum over the decision space $H(x_1, x_2)$, with weights given by an ellipsoidal Gaussian centered at $(\mathbf{g}_1^{\mathrm{T}}\mathbf{t}_1 + u|A^{\mathrm{T}}\mathbf{t}_1|, \mathbf{g}_2^{\mathrm{T}}\mathbf{t}_2)$.

To complete the expression for $E[Q^T\mathbf{M}_1|S=1,R=2]$ in terms of the decision model, we can also evaluate p_{12}.

$$p_{12} = P(R=2|S=1) \tag{A19}$$

$$= \int\int_{-\infty}^{\infty}P(R=2|S=1,D_1=x,D_2=y) P(D_1=x,D_2=y|S=1)dxdy \tag{A20}$$

$$= \int\int_{-\infty}^{\infty}H(x,y)\phi(x,\mathbf{g}_1{}^{\mathrm{T}}\mathbf{t}_1,\sigma_{D1})\phi(x,\mathbf{g}_2{}^{\mathrm{T}}\mathbf{t}_2,\sigma_{D2})dxdy \tag{A21}$$

The first component of the vector $(u*, \mathbf{0})$ that is the result of the integral in Eq. (A17) is the magnitude of $E[Q^T\mathbf{M}_1|S=1,R=2]$. Corresponding expressions, derived in the same way, give the magnitudes of the expected value of the stimulus noise in other stimulus-response classes of trials. To find the expected value of the noise in the second interval we can adopt a basis Q_2 where the first basis vector is parallel to $A^T\mathbf{t}_2$ instead of $A^{\mathrm{T}}\mathbf{t}_1$, and evaluate $E[Q_2\mathbf{M}_2|S=s,R=r]$.

Appendix B

I use the notation and observer model introduced in the section 'Classification images in a yes-no task' in the main text. The decision space is the function $H(x) = P(R=2|D=x)$.

Following the derivation in Appendix A for the 2AFC task, but more briefly,

$$E[Q^T\mathbf{M}|S=1,R=2] = \int_{\mathbb{R}^n}\mathbf{x}P(Q^T\mathbf{M}=\mathbf{x}|S=1,R=2)d\mathbf{x} \tag{B1}$$

$$= \frac{1}{p_{12}}\int_{\mathbb{R}^n}\mathbf{x}P(R=2|Q^T\mathbf{M}=\mathbf{x},S=1)P(Q^T\mathbf{M}=\mathbf{x}|S=1)d\mathbf{x} \tag{B2}$$

$$= \frac{1}{p_{12}}\int_{-\infty}^{\infty}\int_{\mathbb{R}^{m-1}}(u,\mathbf{v})^T P(R=2|W_{||}=u,\mathbf{W}_\perp=\mathbf{v},S=1) P(W_{||}=u,\mathbf{W}_\perp=\mathbf{v})d\mathbf{v}du \tag{B3}$$

$$= \frac{1}{p_{12}} \int_{-\infty}^{\infty} (u, \mathbf{0}) P(R = 2 | W_{\|} = u, S = 1) P(W_{\|} = u) du \tag{B4}$$

Thus the conditional expected value of **M** is proportional to the transformed template $A^{\mathrm{T}}\mathbf{t}$.

Appendix C

Here I describe conditions under which the observer's response probabilities in a yes-no task are an approximately linear function of each element of i.i.d. stimulus noise, with a slope that is proportional to the corresponding template element. I assume that the observer's responses are based on a decision variable D that is calculated from the stimulus as described in the section 'Classification images in a yes-no task':

$$E = (\mathbf{s} + \mathbf{N})^{\mathrm{T}} \mathbf{t} \tag{C1}$$

$$D = f(E, \mathbf{V}) \tag{C2}$$

I define the 'phenomenal decision space' as

$$\bar{H}(x) = P(R = 2 | E = x) \tag{C3}$$

The phenomenal decision space gives the observer's response probability as a function of the external component of the decision variable, E.

I will use the following non-standard notation. If $\mathbf{v}$ is an $n \times 1$ vector, then $\mathbf{v}_{[\sim i]}$ is $\mathbf{v}$ with the ith element removed: $\mathbf{v}_{[\sim i]} = [v_1, \ldots, v_{i-1}, v_{i+1}, \ldots, v_n]^T$, where T is matrix transposition.

We wish to find the probability of the response $R = 2$ on trials where a given signal is shown ($S = 1$ or $S = 2$), as a function of the value of a single noise element n_i. We can expand the probability $P(R = 2 | S = 1, n_i = x)$ using the law of total probability, partitioning over the stimulus noise **N**.

$$P(R = 2 | S = 1, n_i = x) = \int_{\mathbb{R}^{n-1}} P(R = 2 | S = 1, n_i = x, \mathbf{n}_{[\sim i]} = \mathbf{y}) P(\mathbf{n}_{[\sim i]} = \mathbf{y} | S = 1, n_i = x) d\mathbf{y} \tag{C4}$$

The noise is i.i.d. and independent of the signal, so this simplifies to

$$= \int_{\mathbb{R}^{n-1}} P(R = 2 | S = 1, n_i = x, \mathbf{n}_{[\sim i]} = \mathbf{y}) P(\mathbf{n}_{[\sim i]} = \mathbf{y}) d\mathbf{y} \tag{C5}$$

The first factor inside the integral is given by the phenomenal decision space.

$$= \int_{\mathbb{R}^{n-1}} \bar{H}((\mathbf{g}_{1[\sim i]} + \mathbf{y})^T \mathbf{t}_{[\sim i]} + (g_{1i} + x) t_i) P(\mathbf{n}_{[\sim i]} = \mathbf{y}) d\mathbf{y} \tag{C6}$$

If x has a small influence on the external component of the decision variable, then we can make a first-order approximation to the phenomenal decision space as a function of x, using a Taylor series expansion around $x = 0$:

$$\simeq \int_{\mathbb{R}^{n-1}} [\bar{H}((\mathbf{g}_{1[\sim i]} + \mathbf{y})^T \mathbf{t}_{[\sim i]} + g_{1i} t_i) + x t_i \bar{H}'((\mathbf{g}_{1[\sim i]} + \mathbf{y})^T \mathbf{t}_{[\sim i]} + g_{1i} t_i)] P(\mathbf{n}_{[\sim i]} = \mathbf{y}) d\mathbf{y} \tag{C7}$$

Here $\bar{H}'$ is the first derivative of $\bar{H}$. Even if the decision space H has sharp boundaries, the phenomenal decision space $\bar{H}$ will usually be smooth because the observer has internal noise, and so the observer's response probability will change smoothly as a function of the external component E of the decision variable. (For a formal treatment of the relationship between H and $\bar{H}$, see Pritchett and Murray (2015), Supporting Information, section 'Convolution of the decision space'.) Under these conditions the derivative $\bar{H}'(x)$ is well defined. We can rewrite Eq. (C7) as

$$= \int_{\mathbb{R}^{n-1}} \bar{H}((\mathbf{g}_{1[\sim i]} + \mathbf{y})^T \mathbf{t}_{[\sim i]} + g_{1i} t_i) P(\mathbf{n}_{[\sim i]} = \mathbf{y}) d\mathbf{y} + x t_i \int_{\mathbb{R}^{n-1}} \bar{H}'((\mathbf{g}_{1[\sim i]} + \mathbf{y})^T \mathbf{t}_{[\sim i]} + g_{1i} t_i) P(\mathbf{n}_{[\sim i]} = \mathbf{y}) d\mathbf{y} \tag{C8}$$

$$= u_{1i} + v_{1i} t_i x \tag{C9}$$

Here u_{1i} stands for the first integral in Eq. (C8) and v_{1i} stands for the second integral. The constants u_{1i} and v_{1i} in this linear approximation depend on which signal is being shown ($S = 1$ or $S = 2$) and on which noise element n_i we are considering, which is why they need subscripts as in u_{si} and v_{si}, where s is the signal number and i is the noise element number. However, if each noise element has only a small effect on the observer's responses, then $u_{1i} \simeq u_{1j}$ and $v_{1i} \simeq v_{1j}$ for all i and j, so we can drop the second subscript and write

$$\simeq u_1 + v_1 t_i x \tag{C10}$$

That is, the response probability is an approximately linear function of the noise element n_i, with a slope that is proportional to the template element t_i. To choose specific values for u_1 and v_1, we can assign them the integrals in Eq. (C8), modified to integrate over all noise elements instead of omitting element n_i.

$$u_1 = \int_{\mathbb{R}^n} \bar{H}((\mathbf{g}_1 + \mathbf{y})^T \mathbf{t}) P(\mathbf{n} = \mathbf{y}) d\mathbf{y} \tag{C11}$$

$$v_1 = \int_{\mathbb{R}^n} \bar{H}'((\mathbf{g}_1 + \mathbf{y})^T \mathbf{t}) P(\mathbf{n} = \mathbf{y}) d\mathbf{y} \tag{C12}$$

We can use the approximation in Eq. (C10) to find the expected value of noise element n_i on trials where $S = 1$ and $R = 2$. The expected value is

$$E[n_i | S = 1, R = 2] = \int_{-\infty}^{\infty} x P(n_i = x | S = 1, R = 2) dx \tag{C13}$$

Using Bayes' theorem, this becomes

$$= \int_{-\infty}^{\infty} x \frac{P(R = 2 | S = 1, n_i = x) P(n_i = x | S = 1)}{P(R = 2 | S = 1)} dx \tag{C14}$$

Now we can use the linear approximation in Eq. (C10) and the noise density $p_N(x)$.

$$\simeq \frac{1}{P(R = 2 | S = 1)} \int_{-\infty}^{\infty} x (u_1 + v_1 t_i x) p_N(x) dx \tag{C15}$$

$$= \frac{1}{P(R = 2 | S = 1)} \left[u_1 \int_{-\infty}^{\infty} x p_N(x) dx + v_1 t_i \int_{-\infty}^{\infty} x^2 p_N(x) dx \right] \tag{C16}$$

$$= \frac{1}{P(R = 2 | S = 1)} \left[u_1 E[n_i] + v_1 t_i E[n_i^2] \right] \tag{C17}$$

I have assumed that $E[n_i] = 0$, so this is

$$= \frac{t_i v_1 \sigma_N^2}{P(R = 2 | S = 1)} \tag{C18}$$

Thus the conditional expected value of each noise element n_i is proportional to the corresponding template element t_i, regardless of whether the noise density $p_N(x)$ is Gaussian.

References

Abbey, C. K., & Eckstein, M. P. (2002). Classification image analysis: Estimation and statistical inference for two-alternative forced-choice experiments. *Journal of Vision, 2*, 66–78.

Adelson, E. H., & Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America A, 2*, 284–299.

Ahumada, A. J. Jr., (1996). *Perceptual classification images from Vernier acuity masked by noise. Perception*, **25**, ECVP abstract supplement.

Ahumada, A. J. Jr., (2002). Classification image weights and internal noise level estimation. *Journal of Vision, 2*, 121–131.

Chichilnisky, E. J. (2001). A simple white noise analysis of neuronal light responses. *Network: Computation in Neural Systems, 12*, 199–213.

Chubb, C., Econopouly, J., & Landy, M. S. (1994). Histogram contrast analysis and the visual segregation of IID textures. *Journal of the Optical Society of America A, 11*, 2350–2374.

Gold, J. M., Murray, R. F., Bennett, P. J., & Sekuler, A. B. (2000). Deriving behavioural receptive fields for visually completed contours. *Current Biology, 10*, 663–668.

Knoblauch, K., & Maloney, L. T. (2008). Estimating classification images with generalized linear and additive models. *Journal of Vision, 8*(16). 10, 1-19.

Macke, J. H., & Wichmann, F. A. (2010). Estimating predictive stimulus features from psychophysical data: The decision image technique applied to human faces. *Journal of Vision, 10*(5). 22, 1-24.

Murray, R. F. (2011). Classification images: A review. *Journal of Vision, 11*(5). 2, 1-25.

Murray, R. F., Bennett, P. J., & Sekuler, A. B. (2002). Optimal methods for calculating classification images: Weighted sums. *Journal of Vision, 2*, 79–104.

Murray, R. F., Patel, K., & Yee, A. (2015). Posterior probability matching and human perceptual decision making. *PLOS Computational Biology, 11*(6), e1004342.

Neri, P. (2010). Visual detection under uncertainty operates via an early static, not late dynamic, non-linearity. *Frontiers in Computational Neuroscience, 4*, 151.

Neri, P., & Heeger, D. J. (2002). Spatiotemporal mechanisms for detecting and identifying image features in human vision. *Nature Neuroscience, 5*, 812–816.

Paninski, L. (2003). Convergence properties of three spike-triggered analysis techniques. *Network: Computation in Neural Systems, 14*, 437–464.

Pelli, D. G. (1985). Uncertainty explains many aspects of visual contrast detection and discrimination. *Journal of the Optical Society of America A, 2*, 1508–1532.

Pritchett, L. M., & Murray, R. F. (2015). Classification images reveal decision variables and strategies in forced choice tasks. *Proceedings of the National Academy of Sciences of the U.S.A., 112*(23), 7321–7326.

Rajashekar, U., Bovik, A. C., & Cormack, L. K. (2006). Visual search in noise: Revealing the influence of structural cues by gaze-contingent classification image analysis. *Journal of Vision, 6*, 379–386.

Saiki, J. (2008). Stimulus-driven mechanisms underlying visual search asymmetry revealed by classification image analysis. *Journal of Vision, 8*(4). 30, 1-19.

Tanner, W. P., & Swets, J. A. (1954). A decision-making theory of visual detection. *Psychological Review, 61*, 401–409.

Treisman, M., & Leshowitz, B. (1969). The effects of duration, area, and background intensity on the visual intensity difference threshold given by the forced-choice procedure: Derivations from a statistical decision model for sensory discrimination. *Perception & Psychophysics, 6*, 281–296.

Yeshurun, Y., Carrasco, M., & Maloney, L. T. (2008). Bias and sensitivity in two-interval forced choice procedures: Tests of the difference model. *Vision Research, 48*, 1837–1851.