LETTER Communicated by Adam Kohn

Visual Decisions in the Presence of Measurement and Stimulus Correlations

Manisha Bhardwaj
manisha@math.uh.edu
Samuel Carroll
carroll@math.utah.edu
Department of Mathematics, University of Houston, Houston, TX 77004, U.S.A.

Wei Ji Ma
weijima@nyu.edu
Center for Neural Science and Department of Psychology, New York University, NY 10003, U.S.A.

Krešimir Josić
josic@math.uh.edu
Department of Biology and Biochemistry and Department of Mathematics, University of Houston, Houston, TX 77004, U.S.A.

Humans and other animals base their decisions on noisy sensory input. Much work has been devoted to understanding the computations that underlie such decisions. The problem has been studied in a variety of tasks and with stimuli of differing complexity. However, how the statistical structure of stimuli, along with perceptual measurement noise, affects perceptual judgments is not well understood. Here we examine how correlations between the components of a stimulus—stimulus correlations—together with correlations in sensory noise, affect decision making. As an example, we consider the task of detecting the presence of a single or multiple targets among distractors. We assume that both the distractors and the observer's measurements of the stimuli are correlated. The computations of an optimal observer in this task are nontrivial yet can be analyzed and understood intuitively. We find that when distractors are strongly correlated, measurement correlations can have a strong impact on performance. When distractor correlations are weak, measurement correlations have little impact unless the number of stimuli is large. Correlations in neural responses to structured stimuli can therefore have a strong impact on perceptual judgments.

Samuel Carroll is now at the Department of Mathematics, University of Utah.

Neural Computation **27, 2318–2353** (2015) © 2015 Massachusetts Institute of Technology
doi:10.1162/NECO_a_00778

Downloaded from http://direct.mit.edu/neco/article-pdf/27/11/2318/952316/neco_a_00778.pdf by guest on 16 May 2021

1 Introduction

The perceptual system has evolved to extract ecologically meaningful information from sensory input. For example, in many mid- to high-level visual tasks, the brain has to make categorical, global judgments based on multiple stimuli where the identity of any individual stimulus is not of direct relevance. In a visual search task, the goal might be to detect whether a predefined target object is present in a scene that contains multiple objects. Complicating such tasks is the fact that noise corrupts sensory measurements, especially when observation time is short or many objects are present.

Much work has been devoted to modeling the decision processes by which the brain converts noisy sensory measurements of a set of stimuli into a judgment about a global world state such as the presence or absence of a target. These models often focus on various decision rules that can be applied to the measurements. By contrast, the measurements themselves are usually modeled in a rather stereotypical fashion namely, as independent and normally distributed (e.g., Peterson, Birdsall, & Fox, 1954; Nolte & Jaarsma, 1967; Pelli, 1985; Graham, Kramer, & Yager, 1987; Palmer, Ames, & Lindsey, 1993; Baldassi & Burr, 2000; Baldassi & Verghese, 2002; van den Berg, Vogel, Josić, & Ma, 2012; Ma, Navalpakkam, Beck, van den Berg, & Pouget, 2011; Mazyar, van den Berg, & Ma, 2012). Both the assumption of independence and the assumption of gaussianity can be questioned. Specifically, neural correlations can extend to distances as long as 4 mm in monkey cortex (Ecker et al., 2010; Cohen & Kohn, 2011). This suggests that sensory measurements can be strongly correlated (Rosenbaum, Trousdale, & Josić, 2010; Chen, Geisler, & Seidemann, 2006). Here we focus on the effects of violation of the assumption of independent measurements on performance in categorical, global perceptual judgments.

To make such perceptual judgments, an observer needs to take into account the statistical structure of the stimuli and the structure of measurements. Consider a search task where a subject is required to detect a target among distractors. The effects of measurement correlations and correlations between the distractors will be intertwined. If the distractors are identical on a given trial, then strong correlations between the measurements will help preserve their perceived similarity. Namely, an observer can group the distractor measurements and identify the target as corresponding to the outlying measurement. By contrast, when distractors are unstructured (independently drawn across locations), strong measurement correlations may have no effect on performance. Thus, measurement and distractor correlations should not be considered in isolation.

Here we examine how measurement and distractor correlations affect the strategy and the performance of an ideal observer in a target detection task. We assume that on half the trials, one or more target stimuli are

Downloaded from http://direct.mit.edu/neco/article-pdf/27/11/2318/952316/neco_a_00778.pdf by guest on 16 May 2021

presented along with a number of distractors, whereas on the other half of trials, only distractors are presented. The task is to infer whether targets are present or not. Importantly, we assume that the distractor stimuli are not drawn independently; for instance, in the extreme case, the targets could be identical. Our ideal observer infers target presence based on measurements of the stimuli. We assume that these measurements are corrupted by correlated noise. In an extreme case, this noise is perfectly correlated and all measurements are perturbed by the same random value.

We provide an analytical study of the optimal decision rule to show that the interplay of measurement and distractor correlations can be intricate. In general, if the distractors are strongly correlated, then measurement correlations can strongly affect the performance of an ideal observer. When distractors are weakly correlated, measurement correlations have a smaller impact.

We expect that these insights hold more generally. In natural search tasks, such as finding edible fruit in a bush or a car in a parking lot, distractors are heterogeneous but highly structured. Line segments belong to contours and form boundaries of shapes. Hence their orientations can be correlated depending on positions (Geisler & Perry, 2009). Thus the distributions corresponding to natural stimuli are concentrated along low-dimensional structures in stimulus space (Geisler, 2008). It is likely that humans and animals take this structure into account in visual search. Indeed in contour integration, observers seem to take into account natural co-occurrence statistics of line elements (Geisler & Perry, 2009), and in change detection, people incorporate knowledge about the large-scale statistical structure of a scene (Brady & Alvarez, 2011; Brady & Tenenbaum, 2013). In such situations, correlations in measurement noise could help in parameters of inference. Ultimately, we thus expect our results to be relevant to modeling perceptual decision making in natural scenes.

2 Model Description

To examine how decisions of an ideal observer are determined by the statistical structure of measurements and stimuli, we consider the following task. An observer is asked whether target stimuli are present among a set of distractor stimuli. Each stimulus, i, is characterized by a scalar, $s_i \in \mathbb{R}$. The set of N stimuli presented on a single trial is characterized by the vector $\mathbf{s} = (s_1, s_2, \cdots, s_N)$. For instance, stimuli could be pure tones characterized by their frequency, or ellipses characterized by the orientation of their major axis. A target is a stimulus with a particular characteristic, s_T. A target could be a vertical grating or a pure tone at 440 Hz.

For an example of this task with stimuli that are gratings characterized by their orientations, see Figure 1A. For simplicity, we assume that if i is a

Downloaded from http://direct.mit.edu/neco/article-pdf/27/11/2318/952316/neco_a_00778.pdf by guest on 16 May 2021

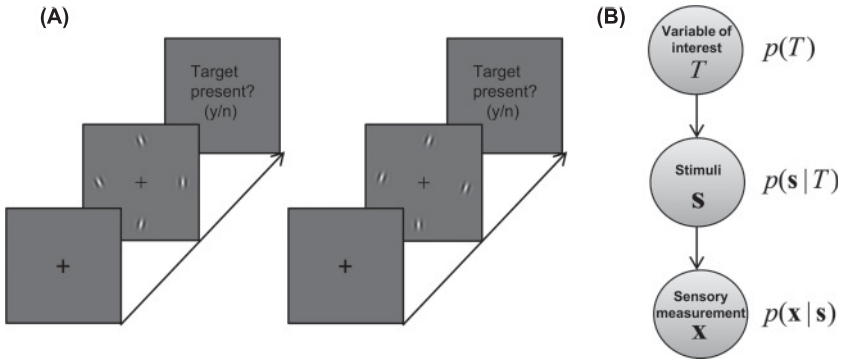

Figure 1: A schematic representation of the experimental task and Bayesian inference. (A) An example target detection task. The observer has to decide whether a target, in this case the vertically oriented Gabor patch, is present. The correlation between distractor stimuli can be weak (in the example on the left, the distractors all have different orientations and the target is on the right) or strong (in the example on the right, all distractors have equal orientation and the target is on the bottom). Note that the observer does not have to identify the target, only report whether it is present. (B) Information about a parameter of interest, T, is encoded in a stimulus vector, $\mathbf{s}$. The dimension of $\mathbf{s}$ could be much higher than the dimension of T. The observer makes a sensory measurement, $\mathbf{x}$, of the stimulus, $\mathbf{s}$, and must infer T from this measurement. This inference requires marginalization over $\mathbf{s}$.

target, then $s_i = s_T = 0$, and stimulus characteristics are measured relative to that of a target. In the Figure 1A example, we have a single vertical target, and we measure orientations relative to the vertical. Stimuli that are not targets are distractors. We determine the characteristics (orientations) of the distractors by a sample from a gaussian distribution. Marginally, each orientation is vertical on average, and the orientations can be pairwise correlated. Thus, the distractors are nonvertical with probability 1. If there are no targets, we refer to all stimuli as distractors, and their orientations are again determined by a sample from a multivariate normal distribution. We will consider situations with single and multiple targets. The number of targets and stimuli is assumed fixed during an experiment.

The following treatment parallels the one we used earlier for independent measurements (Mazyar et al., 2012; Mazyar, van den Berg, Seilheimer, & Ma, 2013; Bhardwaj, van den Berg, Ma, & Josić, 2015). We denote target presence by $T = 1$ and absence by $T = 0$. We assume that targets are present with probability 0.5. When $T = 0$, there are no targets, and all stimuli are distractors. Their characteristics are drawn from a multivariate normal distribution with mean $\mathbf{0}_N = (0, 0, \cdots, 0)$ and covariance matrix, $\mathbf{\Sigma}_\mathbf{s}$. The

Downloaded from http://direct.mit.edu/neco/article-pdf/27/11/2318/952316/neco_a_00778.pdf by guest on 16 May 2021

subscript denotes vector length, so $\mathbf{0}_N$ has N components. We can therefore write

$$p(\mathbf{s}|T=0) = \mathcal{N}(\mathbf{s}; \mathbf{0}_N, \boldsymbol{\Sigma}_\mathbf{s}), \tag{2.1}$$

where $\mathcal{N}(\mathbf{s}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes the density of the normal distribution with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$. For simplicity, we assume $\boldsymbol{\Sigma}_\mathbf{s}$ has constant diagonal and off-diagonal terms so that

$$\boldsymbol{\Sigma}_\mathbf{s} = \sigma_s^2 \begin{bmatrix} 1 & \rho_s & \cdots & \rho_s \\ \rho_s & 1 & \cdots & \rho_s \\ \vdots & & \ddots & \vdots \\ \rho_s & \rho_s & \cdots & 1 \end{bmatrix}. \tag{2.2}$$

The variance, σ_s^2, determines the variability of individual stimulus characteristics. When σ_s is smaller, the distractors are closer to each other, and the task becomes more difficult. The correlation coefficient, ρ_s, determines the relation between the distractors.

If $T = 1$, then $n \geq 1$ targets are present. The number of targets, n, and the number of stimuli, N, is fixed for the duration of the experiment. In this case, we assign, with uniform probability, the target characteristic to n out of the total of N stimuli. This subset of n targets is denoted by L. We denote by $\mathcal{L}$ the collection of all $\binom{N}{n}$ possible choices of the sets L of targets.

The characteristics of the remaining $N-n$ distractors are drawn from a multivariate normal distribution with mean $\mathbf{0}_{N-n}$ and covariance matrix $\boldsymbol{\Sigma}_{\mathbf{s}\setminus L}$ of dimension $(N-n) \times (N-n)$. Let $\mathbf{s}_L = (s_{i_1}, s_{i_2}, \cdots, s_{i_n}),\ i_l \in L$ denote the target stimuli and $\mathbf{s}_{\setminus L} = (s_{j_1}, s_{j_2}, \cdots, s_{j_{N-n}}),\ j_l \notin L$ the distractors. We can therefore write

$$p(\mathbf{s}_L|T=1) = \prod_{i \in L} \delta(s_i), \quad \text{and} \quad p(\mathbf{s}_{\setminus L}|T=1) = \mathcal{N}(\mathbf{s}_{\setminus L}; \mathbf{0}_{N-n}, \boldsymbol{\Sigma}_{\mathbf{s}\setminus L}),$$

where $\delta(\cdot)$ denotes the Dirac delta function. We assume that all pairs of distractors are equally correlated, so that the off-diagonal entries in $\boldsymbol{\Sigma}_{\mathbf{s}\setminus L}$ are identical. Moreover, the correlation between distractors is the same whether a target is present or not. We therefore refer to ρ_s as the distractor correlation.

We further assume that an observer makes a noisy measurement, x_i, of each stimulus, s_i. This measurement can be thought of as the estimate of stimulus i obtained from the activity of a population of neurons that responded to the stimulus. We denote by $\mathbf{x} = (x_1, x_2, \cdots, x_N)$ the vector of N measurements. It is commonly assumed that these measurements are unbiased and corrupted by additive, independent, normally distributed noise,

Downloaded from http://direct.mit.edu/neco/article-pdf/27/11/2318/952316/neco_a_00778.pdf by guest on 16 May 2021

so that $p(\mathbf{x}|\mathbf{s}) = \prod_{i=1}^{n} \mathcal{N}(x_i; s_i, \sigma_x^2)$. Here we consider the more general situation where the measurements are unbiased, but noise could be correlated so that

$$p(\mathbf{x}|\mathbf{s}) = \mathcal{N}(\mathbf{x}; \mathbf{s}, \mathbf{\Sigma_x}). \tag{2.3}$$

Therefore, we assume that noise equally affects measurements of all stimuli, whether targets or distractors. Marginally, each measurement is additively perturbed by normally distributed noise with variance σ_x^2.

We consider the particular case when $\mathbf{\Sigma_x}$ has constant diagonal terms, σ_x^2, and off-diagonal terms, $\rho_x \sigma_x^2$, so that

$$\mathbf{\Sigma_x} = \sigma_x^2 \begin{bmatrix} 1 & \rho_x & \cdots & \rho_x \\ \rho_x & 1 & \cdots & \rho_x \\ \vdots & & \ddots & \vdots \\ \rho_x & \rho_x & \cdots & 1 \end{bmatrix}. \tag{2.4}$$

Thus, measurement noise is equally correlated between all pairs of stimuli.

Note that

$$p(\mathbf{x}|T = 1) = \int p(\mathbf{x}|\mathbf{s}) p(\mathbf{s}|T = 1) d\mathbf{s},$$

and similarly for $p(\mathbf{x}|T = 0)$. We must thus marginalize over $\mathbf{s}$ to obtain these distributions. A simple computation (see appendix A) shows that $p(\mathbf{x}|L, T = 1) = \mathcal{N}(\mathbf{x}; \mathbf{0}_N, \mathbf{\Sigma}^0_{\mathbf{s},L} + \mathbf{\Sigma_x})$ and $p(\mathbf{x}|T = 0) = \mathcal{N}(\mathbf{x}; \mathbf{0}_N, \mathbf{\Sigma_s} + \mathbf{\Sigma_x})$.

3 Results

Our goal is to describe how correlations between stimuli along with correlations between their measurements affect the decisions of an optimal observer in a target detection task. We examine how performance changes as both correlations between distractors and between measurements are varied.

The stimuli, $\mathbf{s}$, follow different distributions depending on whether $T = 0$ or $T = 1$. Measurement noise increases the overlap between the corresponding measurement distributions, $p(\mathbf{x}|T = 0)$ and $p(\mathbf{x}|T = 1)$. The higher the overlap is between these two distributions, the more difficult it is to tell whether a target is present. However, correlations in measurement noise can reduce such overlap (see Figure 2B) even when noise intensity is unchanged. Therefore, the estimate of a parameter from a neural response depends not only on the level, σ_x, but also on the structure of measurement noise (Abbott & Dayan, 1999; Sompolinsky, Yoon, Kang, & Shamir,

Downloaded from http://direct.mit.edu/neco/article-pdf/27/11/2318/952316/neco_a_00778.pdf by guest on 16 May 2021

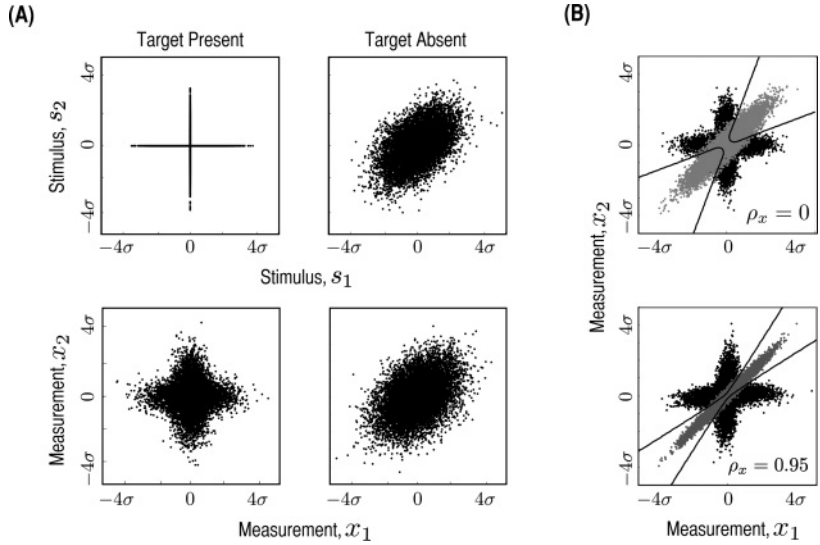

Figure 2: Samples from stimulus and measurement distributions in the single target detection task with $N = 2$ stimuli and $\sigma_s = 15$, $\sigma_x = 4$. (A) (Top) Samples from stimulus distributions on target present (left) and target absent (right) trials for $\rho_s = 0.5$. On target-present trials, the distribution is constrained to the axes, since one of the stimuli is a target. On target-absent trials, the two stimuli follow a bivariate normal distribution. (Bottom) Samples of measurements in the absence of measurement correlations, $\rho_x = 0$. On both the target-present (left) and target-absent cases (right), the measurement distribution inherits its shape from the stimulus distribution. (B) Overlap of measurement distributions for strongly correlated distractors, $\rho_s = 0.99$. Measurement samples on target-present (black) and target-absent (gray) trials are shown for $\rho_x = 0$ (top) and $\rho_x = 0.95$ (bottom). The overlap between the measurement distributions decreases with increasing measurement correlation, ρ_x. The decision boundary (solid line) therefore better separates the two distributions when ρ_s and ρ_x are both large. We used high correlations to bring out the difference between the distributions. Axes represent the total standard deviation of the measurements, $\sigma = \sqrt{\sigma_s^2 + \sigma_x^2}$.

2001; Averbeck, Latham, & Pouget, 2006; Josić, Shea-Brown, Doiron, & de la Rocha, 2009).

An ideal observer makes a decision based on the sign of the log posterior ratio,

$$d_{\mathrm{TD}}(\mathbf{x}) = \log \frac{p(T=1|\mathbf{x})}{p(T=0|\mathbf{x})} = \log \frac{p(\mathbf{x}|T=1)}{p(\mathbf{x}|T=0)} + \log \frac{p(T=1)}{p(T=0)}$$
$$= \log \frac{p(\mathbf{x}|T=1)}{p(\mathbf{x}|T=0)}. \tag{3.1}$$

Downloaded from http://direct.mit.edu/neco/article-pdf/27/11/2318/952316/neco_a_00778.pdf by guest on 16 May 2021

The last equality follows from our assumption that the prior probabilities, $p(T = 1)$ and $p(T = 0)$ are equal. If $d_{\mathrm{TD}}(\mathbf{x}) > 0$, the observer infers that a target is present. Note that

$$p(\mathbf{x}|T = 1) = \sum_{L \in \mathcal{L}} p(\mathbf{x}|L, T = 1)p(L).$$

If the prior probability that L is a set of targets is uniform, any subset of n stimuli is equally likely to consist of targets, and hence $p(L) = \binom{N}{n}^{-1}$, and

$$d_{\mathrm{TD}}(\mathbf{x}) = \log \frac{p(\mathbf{x}|T = 1)}{p(\mathbf{x}|T = 0)} = \log \left[\binom{N}{n}^{-1} \sum_{L \in \mathcal{L}} \frac{p(\mathbf{x}|L, T = 1)}{p(\mathbf{x}|T = 0)} \right]. \tag{3.2}$$

The decision variable thus depends on the sum of the normalized probabilities that a measurement $\mathbf{x}$ is made, given that the target set is L.

Note also that

$$p(L|\mathbf{x}, T = 1) \propto p(\mathbf{x}|L, T = 1).$$

Therefore, the decision variable can also be interpreted as a sum of likelihoods that L is a target set, given a measurement $\mathbf{x}$. Thus, the decision is directly related to the posterior distribution over the target sets L. The summands in equation 3.2 correspond to the evidence that L is a set of targets.

The distribution of measurements, $\mathbf{x}$, depends on whether a target or targets are present. The conditional distributions of measurements, $p(\mathbf{x}|T = 0)$ and $p(\mathbf{x}|L, T = 1)$, are gaussian with all means equal to 0 and covariances $\mathbf{C} = \mathbf{\Sigma_s} + \mathbf{\Sigma_x}$ and $\mathbf{C}_L = \mathbf{\Sigma}^0_{\mathbf{s},L} + \mathbf{\Sigma_x}$, respectively. We therefore have (see appendix A for details)

$$d_{\mathrm{TD}}(\mathbf{x}) = \log \left[\binom{N}{n}^{-1} \sqrt{\frac{|\mathbf{C}|}{|\mathbf{C}_L|}} \sum_{L \in \mathcal{L}} \exp\left(-\frac{1}{2} \mathbf{x}^T \left(\mathbf{C}_L^{-1} - \mathbf{C}^{-1} \right) \mathbf{x} \right) \right]. \tag{3.3}$$

Normalization of the two multivariate normal distribution requires division by the determinants $|\mathbf{C}|$ and $|\mathbf{C}_L|$, respectively. Their ratio therefore appears as a prefactor in the ratio of the likelihoods.

This decision variable depends on the model parameters and the measurement, $\mathbf{x}$. The total number of stimuli, N, number of targets, n, the variability, σ_s^2, and the correlation, ρ_s, between the distractors determine the structure of the stimulus, while the variability, σ_x^2, and correlation, ρ_x, describe the distribution of sensory measurements. An ideal observer knows all these parameters.

Downloaded from http://direct.mit.edu/neco/article-pdf/27/11/2318/952316/neco_a_00778.pdf by guest on 16 May 2021

Setting the right-hand side of equation 3.3 to 0 defines a nonlinear decision boundary in the space of measurements. This boundary separates measurements that lead to a target-present decision ($d_{\mathrm{TD}}(\mathbf{x}) > 0$) from those that lead to a target-absent decision ($d_{\mathrm{TD}}(\mathbf{x}) < 0$). For example, with a single stimulus, $N = n = 1$, the observer needs to decide whether a single measurement corresponds to a target. The decision variable has the form

$$d_{\mathrm{TD}}(\mathbf{x}) = \frac{1}{2}\log\frac{\sigma_s^2+\sigma_x^2}{\sigma_x^2} - \frac{1}{2\sigma_x^2}\frac{x_1^2}{1+\sigma_x^2/\sigma_s^2}.$$

If the measurement, x_1, differs sufficiently from 0, then $d_{\mathrm{TD}}(\mathbf{x}) < 0$. A measurement close to 0 gives $d_{\mathrm{TD}}(\mathbf{x}) > 0$.

With more stimuli, the observer needs to take into account the known correlations between the stimuli and between the measurements. The decision variable depends in a complicated way on the parameters that describe the structure of the stimulus and the response. The explicit form of equation 3.3 can be derived under the assumption of equal variances and equal covariances (see appendix A):

$$d_{\mathrm{TD}}(\mathbf{x}) = \log\Bigg(\frac{1}{M}\sqrt{\frac{1+(\rho_s\sigma_s^2+\rho_x\sigma_x^2)Nv}{\gamma_L}}\left(\frac{v_L}{v}\right)^n \sum_{L\in\mathcal{L}}$$
$$\exp\Bigg\{-\frac{1}{2}\Bigg(\underbrace{\sigma_s^2(1-\rho_s)v_L v\sum_{i\in L}x_i^2}_{\mathrm{I}} + \underbrace{\left(\beta v^2 - \frac{\rho_x\sigma_x^2\, v_L^2 g_L}{\gamma_L}\right)\sum_{i,j\in L}x_i x_j}_{\mathrm{II}}$$
$$\underbrace{+2\left(\beta v^2 - \frac{\rho_x\sigma_x^2\, v_L v}{\gamma_L}\right)\sum_{i\in L, j\notin L}x_i x_j}_{\mathrm{III}} + \underbrace{\left(\beta v^2 - \frac{v^2 f_L}{\gamma_L}\right)\sum_{i,j\notin L}x_i x_j}_{\mathrm{IV}}\Bigg)\Bigg\}\Bigg), \quad (3.4)$$

where $M = \binom{N}{n}$.

The variables v and v_L represent scaled inverse variances corresponding to distractor and target stimuli. The parameters β, g_L, f_L, and γ_L are given in equations A.5 and A.6 and are defined in terms of $\sigma_s^2, \rho_s, \sigma_x^2$, and ρ_x. equation 3.4 has a form that can be interpreted intuitively:

Downloaded from http://direct.mit.edu/neco/article-pdf/27/11/2318/952316/neco_a_00778.pdf by guest on 16 May 2021

1. Term I contains a sum of squares of individual measurements, x_i^2, over the putative set of targets, L. The smaller this sum is, the more likely that L contains targets.
2. Term II contains the sample covariance about the known target value, $s_T = 0$, that is, $\sum_{i,j\in L}(x_i - s_T)(x_j - s_T) = \sum_{i,j\in L} x_i x_j$. In the absence of measurement correlations, $\rho_x = 0$, covariability about the target value has vanishing expectation for target measurements. Therefore, the larger this sum, the less likely it is that the measurements come from a set of target stimuli. In the presence of measurement correlations, $\rho_x > 0$, covariability between target measurements is expected. Hence, the prefactor in term II decreases with ρ_x.
3. Term III is similar to term II, with the sum representing the sample covariance about the target value between putative target and nontarget stimuli. The larger this covariance is, the less likely it is that L or its complement contain targets.
4. Term IV contains the sample covariance about the mean of measurements outside the putative target set. If there are no targets, then all terms in the sum are expected to be large, regardless of the choice of L. However, if there are targets, then whenever the complement of L contains targets, some of the terms in the sum have expectation 0. Hence, the term again makes a smaller contribution if targets are present.

While this provides an intuitive interpretation of the sums in equation 3.4, the expression is complex, and it is difficult to understand precisely how an ideal observer uses knowledge of the generative model and the stimulus measurements to make a decision. We therefore examine a number of cases where equation 3.4 is tractable and all the terms can be interpreted precisely. We also numerically examine performance in a wider range of examples.

3.1 Single Target, $n = 1$. We start with the case when a single target is present at one of the N locations. This case was considered previously in the absence of correlations between the sensory measurements (Bhardwaj et al., 2015).

We observe in Figures 3A and 4A that the performance of an ideal observer is nearly independent of ρ_x when external structure is weak ($\rho_s < 1$). Performance depends strongly on ρ_x when distractors are strongly correlated, $\rho_s \approx 1$. An ideal observer performs perfectly when $\rho_s = \rho_x = 1$ (see Figure 3A).

Increased performance with increasing distractor correlations, ρ_s, accords with intuition that similar distractors make it easier to detect a target. However, correlations in measurement noise can play an equally important role and significantly improve performance when distractors are identical (see Figure 3B).

Downloaded from http://direct.mit.edu/neco/article-pdf/27/11/2318/952316/neco_a_00778.pdf by guest on 16 May 2021

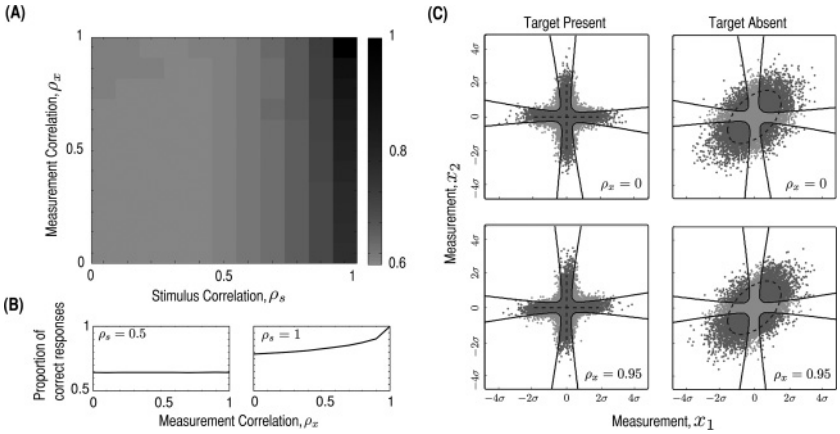

Figure 3: Performance of an optimal observer on a single target detection task with weak measurement noise, $\sigma_x^2 \ll \sigma_s^2$. (A) Proportion of correct responses as a function of stimulus correlation, ρ_s and measurement correlation, ρ_x, for $N = 4$ stimuli. (B) Proportion of correct responses as a function of measurement correlation, ρ_x, in the case of intermediate stimulus correlation, $\rho_s = 0.5$ (left) and strong correlation, $\rho_s = 1$ (right) for $N = 4$ stimuli. (C) Decision boundary, $d_{\mathrm{TD}}(\mathbf{x}) = 0$ (black solid line) and samples from the measurement distributions on target-present (left) and target-absent (right) trials for $N = 2$ and $\rho_s = 0.5$. Here and henceforth, dark gray dots correspond to measurements that result in a correct inference, and light gray dots correspond to measurements leading to an incorrect inference. The stimulus will result in a measurement that lies within the black dashed lines with probability 0.95. Other parameters used are $\sigma_s = 15$ and $\sigma_x = 4$.

Perfect performance when $\rho_s = \rho_x = 1$ can be understood intuitively. In this case measurements, x_i, of the stimuli are obtained by adding the same realization of a random variable, that is, identical measurement noise to each stimulus value, s_i. In target-absent trials, all measurements are identical. If the target is present, measurements contain a single outlier. An ideal observer can thus distinguish the two cases perfectly.

We examine in more detail the cases of weak measurement noise and then the case of comparable measurement noise and distractor variability.

3.1.1 Weak Measurement Noise, $\sigma_x^2 \ll \sigma_s^2$ with Highly Correlated Stimuli, $\rho_s \approx 1$. When measurement noise is weak, discriminability between the target-absent and target-present conditions is governed by external variability—trial-to-trial variability of the stimuli. As noted, measurement correlations improve discriminability in the presence of strong distractor correlations (see Figures 3A and 3C). When $\rho_s \approx 1$, equation 3.4 can be approximated as

Downloaded from http://direct.mit.edu/neco/article-pdf/27/11/2318/952316/neco_a_00778.pdf by guest on 16 May 2021

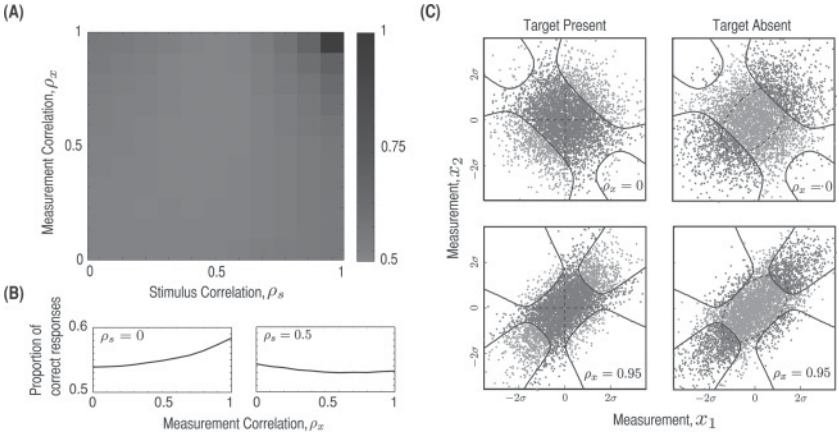

Figure 4: Performance of an optimal observer on a single target detection task with strong measurement noise, $\sigma_x^2 = \sigma_s^2$. (A) Proportion of correct responses as a function of ρ_s and ρ_x for $N = 4$ stimuli. (B) Performance as a function of measurement correlation, ρ_x, when $\rho_s = 0$ (left) and $\rho_s = 0.5$ (right) for $N = 4$ stimuli. (C) Decision boundary (black solid line) and samples from the measurement distributions on target-present (left) and target-absent (right) trials for $N = 2$ and $\rho_s = 0.5$. Other parameters used are $\sigma_s = \sigma_x = 15$.

(details in section B.1)

$$d_{\mathrm{TD}}(\mathbf{x}) \approx \log \left(\underbrace{\frac{1}{N}\sqrt{\frac{N(1-\rho_x)}{N-1}}}_{P(N,\rho_x)} \underbrace{\sum_{i=1}^{N} \exp\left\{ -\frac{1}{2N\sigma_x^2(1-\rho_x)} \left((1-N\rho_x)x_i^2 + 2x_i \sum_{j\neq i} x_j - \frac{1}{N-1}\left(\sum_{j\neq i} x_j\right)^2 \right) \right\}}_{\alpha_i(\mathbf{x},N,\rho_x,\sigma_x)} \right). \tag{3.5}$$

The ideal observer hence uses knowledge about the measurement correlations, ρ_x, in a decision.

We first confirm the intuitive observation that an ideal observer performs perfectly when measurement noise is highly correlated, $1 - \rho_x \ll 1$. In this

Downloaded from http://direct.mit.edu/neco/article-pdf/27/11/2318/952316/neco_a_00778.pdf by guest on 16 May 2021

case, the exponential term in equation 3.5 is, to leading order in $1/(1-\rho_x)$,

$$\alpha_i(\mathbf{x}, N, \rho_x, \sigma_x) \approx \exp\left[\frac{N-1}{2N\sigma_x^2(1-\rho_x)}\left(x_i - \bar{x}_{\hat{i}}\right)^2\right], \tag{3.6}$$

where $\bar{x}_{\hat{i}} = 1/(N-1)\sum_{j\neq i} x_j$ is the sample mean of the measurements excluding the putative target, i. Hence, to make a decision, the ideal observer subtracts the mean of the $N-1$ measurements of putative distractors from that of the putative target. In section B.1, we show that on target-absent trials, $d_{\mathrm{TD}}(\mathbf{x}) \to -\infty$, as $\rho_x \to 1$, and on target-present trials, $d_{\mathrm{TD}}(\mathbf{x}) \to \infty$, as $\rho_x \to 1$. Hence, performance improves as measurement correlations increase. This can also be seen in Figure 2B, as the overlap between the distributions $p(\mathbf{x}|T=0)$ and $p(\mathbf{x}|T=1)$ decreases with an increase in ρ_x. In the degenerate case when $\rho_s = \rho_x = 1$, the distractors' distribution on target absent trials is concentrated on the diagonal. Therefore, an ideal observer can infer that a target is present whenever $x_1 \neq x_2$, that is, the decision boundary also collapses to the diagonal.

When measurement correlations are absent, $\rho_x = 0$, an ideal observer performs perfectly only when measurement noise is vanishing. The exponential in equation 3.5 now equals

$$\begin{aligned}\alpha_i(\mathbf{x}, N, \rho_x, \sigma_x) &= \exp\left[-\frac{1}{2\sigma_x^2}\left(\frac{1}{N}\left(\sum_j x_j\right)^2 - \frac{1}{N-1}\left(\sum_{j\neq i} x_j\right)^2\right)\right] \\ &= \exp\left[-\frac{1}{2\sigma_x^2}\left(N\bar{x}^2 - (N-1)\bar{x}_{\hat{i}}^2\right)\right],\end{aligned} \tag{3.7}$$

where $\bar{x}$ is the sample mean of all observations. The ideal observer therefore compares the sample mean over all measurements, $\bar{x}$, with the sample mean over all observations excluding the putative target, $\bar{x}_{\hat{i}}$. Since measurement noise is uncorrelated, averaging over all measurements is beneficial. This is a very different strategy from when measurement noise is highly correlated (see equation 3.6), and the ideal observer compares a single putative target measurement, x_i, to the sample mean of the remaining target measurements. As shown in section B.1 at intermediate values of measurement correlations, $0 < \rho_x < 1$, the ideal observer uses a mixture of these two strategies.

Intuitively, an increased number of distractors makes it more difficult to detect the target. An analysis of the decision variable in the limit $N \to \infty$ shows that this is indeed the case (see section B.3). We comment on this further in section 4.

3.1.2 Weak Measurement Noise, Nonidentical Distractors, $\rho_s < 1$. Measurement correlations have little effect on performance when distractor

Downloaded from http://direct.mit.edu/neco/article-pdf/27/11/2318/952316/neco_a_00778.pdf by guest on 16 May 2021

correlations are weaker (see Figure 3A). Consider again $\rho_x \approx 1$, so that measurements are corrupted by adding a random but nearly identical perturbation to the stimuli. An ideal observer uses the knowledge that all measurements are obtained by adding an approximately equal value to the stimulus. However, when distractor correlations are weak, the target measurement is no longer an outlier. Correlations in measurement noise provide little help in this situation.

These observations are reflected in the structure of the decision boundaries ($d_{\mathrm{TD}}(\mathbf{x}) = 0$) and the distributions of the measurements (see Figure 3B). In the target-absent (left column) and target-present trials (right column), the distributions of measurements, $p(\mathbf{x}|T)$, is shaped predominantly by variability in the stimulus. Measurement correlations have little effect on this shape, and the decision boundary therefore changes little with an increase in ρ_x. In contrast, when $\rho_s \approx 1$, measurement correlations have a significant impact on the overlap between the distributions $p(\mathbf{x}|T = 1)$ and $p(\mathbf{x}|T = 0)$, as shown in Figure 2B.

We can confirm this intuition about the role of measurement correlations by approximating the decision criterion when $\sigma_x^2 \ll \sigma_s^2$ and $\rho_s < 1$ (see the details in section B.4),

$$d_{\mathrm{TD}}(\mathbf{x}) \approx \log\left(\frac{1}{N}\sqrt{\frac{\sigma_s^2(1-\rho_s)(1+(N-1)\rho_s)}{\sigma_x^2(1+(N-2)\rho_s)}}\sum_{i=1}^{N}\exp\left[-\frac{x_i^2}{2\sigma_x^2}\right]\right). \tag{3.8}$$

The strength of noise correlations, ρ_x, does not affect the decisions of an ideal observer at highest order in (σ_x^2/σ_s^2). This explains the approximate independence of performance on ρ_x observed in Figure 3B. In this case, an ideal observer considers stimuli at each location separately and weighs each measurement individually by its precision, $1/\sigma_x^2$ (Green & Swets, 1966; Palmer, 1999).

When the number of distractors becomes larger, equation 3.8 is no longer valid. The observer compares measurements of all stimuli to make a decision. We return to this point below.

3.1.3 Strong Measurement Noise, $\sigma_x^2 = \sigma_s^2$. Increasing measurement noise trivially degrades performance. However, in the limit of perfect stimulus and measurement correlations, an ideal observer still performs perfectly for the reasons described earlier.

Measurement correlations affect performance differently than in the case of weak measurement noise (see Figures 4A and 4C). Even with uncorrelated stimuli, $\rho_s = 0$, performance increases slightly (approximately 6%) with ρ_x. Surprisingly, for intermediate values of distractor correlations, (e.g., $\rho_s = 0.5$), measurement correlations have a negative impact on performance. If measurement correlations are fixed at a high value, then the worst performance is observed at an intermediate value, $0 < \rho_s < 1$. The

Downloaded from http://direct.mit.edu/neco/article-pdf/27/11/2318/952316/neco_a_00778.pdf by guest on 16 May 2021

reason for this unexpected behavior is unclear, as equation 3.4 is difficult to analyze in this case.

Generally when measurement noise is strong, measurement correlations will change the shape of the measurement distributions $p(\mathbf{x}|T=0)$ and $p(\mathbf{x}|T=1)$ and hence have an impact on decisions and performance. Note that when measurement correlations increase, the region corresponding to $d_{\mathrm{TD}}(\mathbf{x}) > 0$ is elongated along the diagonal to capture more of the mass of the distribution $p(\mathbf{x}|T=1)$ (see Figure 4C). However, when measurement noise is high, the interactions between measurement and distractor correlations are intricate.

3.2 Multiple Targets, $n > 1$. When multiple targets are present, they are all identical and hence perfectly correlated. Thus, regardless of the value of ρ_s, on half the trials, the stimuli will be strongly structured and the density $p(\mathbf{s}|T=1)$ concentrated on a low-dimensional subspace. As a consequence, measurement correlations always have an impact on performance.

Regardless of distractor correlations, an ideal observer performs perfectly when $\rho_x = 1$ (see Figure 5A). Even when $\rho_s < 1$, performance increases with ρ_x (see Figure 5A). When $\rho_x = 1$, all target measurements are identical. Hence, an ideal observer performs perfectly by checking whether n of the measurements, x_i, are equal. We analyze only the case $\rho_s < 1$, since the case of perfectly correlated distractors is similar.

With weak measurement noise, equation 3.4 can be approximated as (see section B.4)

$$d_{\mathrm{TD}}(\mathbf{x}) \approx \log\left(\frac{1}{M}\sqrt{\frac{(1-\rho_x)(1+(N-1)\rho_s)}{(1+(N-n-1)\rho_s)(1+(n-1)\rho_s)}\left(\frac{\sigma_s^2(1-\rho_s)}{\sigma_x^2(1-\rho_x)}\right)^n}\right.$$
$$\left.\times \sum_{L\in\mathcal{L}} \exp\left\{-\frac{n}{2\sigma_x^2(1-\rho_x)}\underline{\left(\frac{1}{n}\sum_{i\in L}x_i^2 - \frac{n\rho_x}{(1+(n-1)\rho_x)}\left(\frac{1}{n}\sum_{i\in L}x_i\right)^2\right)}\right\}\right). \tag{3.9}$$

An ideal observer takes into account measurement correlations for all values of ρ_s even when measurement noise is low. Distractor correlations appear only in the prefactor and are not used in the comparison of the measurements inside the exponential.

Interestingly, decisions in this case are based only on measurements of stimuli within the set of putative targets, L. In the absence of measurement correlations, a decision is based solely on the sample second moment of the n stimulus measurements about the target characteristic, $s_T = 0$, that is, $1/n\sum_{i\in L} x_i^2$ (the underlined term in equation 3.9). A low value of this sample moment indicates that L contains targets.

Downloaded from http://direct.mit.edu/neco/article-pdf/27/11/2318/952316/neco_a_00778.pdf by guest on 16 May 2021

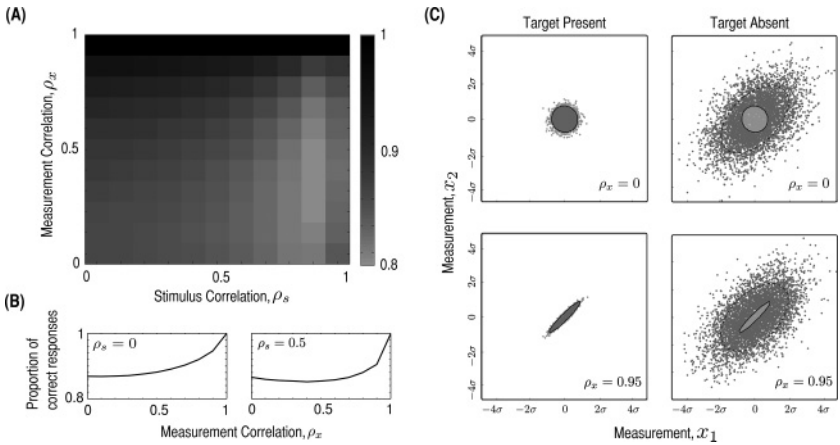

Figure 5: Performance of an optimal observer in a multiple target detection task with weak measurement noise, $\sigma_x^2 \ll \sigma_s^2$. (A) Proportion of correct responses as a function of stimulus correlation, ρ_s, and measurement correlation, ρ_x, with $N = 4$ stimuli and $n = 3$ targets. (B) Proportion of correct responses as a function of ρ_x for $\rho_s = 0$ (left) and $\rho_s = 0.5$ (right) when $N = 4$ and $n = 3$. (C) Samples from the measurement distributions and the decision boundary on target-present (left) and target-absent (right) trials for $\rho_s = 0.5$ and $N = n = 2$.

In the limit of $\rho_x \to 1$, the underlined term in equation 3.9 approaches the sample variance, $s^2_{i\in L}$, that is, the sample second moment about the sample mean,

$$\frac{1}{n}\sum_{i\in L} x_i^2 - \frac{n\rho_x}{1+\rho_x(n-1)}\left(\frac{1}{n}\sum_{i\in L} x_i\right)^2$$
$$\to \frac{1}{n}\sum_{i\in L} x_i^2 - \left(\frac{1}{n}\sum_{i\in L} x_i\right)^2 = s^2_{i\in L}. \tag{3.10}$$

If measurement correlations are strong, the measurements of a set of target stimuli will be approximately equal (regardless of ρ_s). The ideal observer makes use of this knowledge by comparing the value of putative targets in the set L. If the sample variance of the measurements $\{x_i\}_{i\in L}$ is small, then it is likely that L is a set of targets. When $\rho_x = 1$, an ideal observer performs perfectly (see section B.5 for details).

When $0 < \rho_x < 1$, the underlined term in equation 3.9 shows that the ideal observer takes an intermediate strategy by computing a second moment about a point between the target characteristic, s_T, and the sample mean. Interestingly, the larger the number of targets, the larger the weight

Downloaded from http://direct.mit.edu/neco/article-pdf/27/11/2318/952316/neco_a_00778.pdf by guest on 16 May 2021

on the sample mean, since the prefactor $n\rho_x/(1+(n-1)\rho_x)$ increases with n for fixed ρ_x.

These observations are reflected in the distributions shown in Figure 5C. The distribution of measurements, $p(\mathbf{x}|T=1)$, moves closer to the diagonal ($x_1 = x_2$) as $\rho_x \to 1$, and the overlap with the distribution of measurements, $p(\mathbf{x}|T=0)$ decreases. In higher dimensions, for N stimuli and n targets, the measurement distribution, $p(\mathbf{x}|T=1)$, is concentrated on the union of $(N-n+1)$-dimensional subspaces when $\rho_x = 1$. The target measurements lie on a line, while the $N-n$ distractor measurements are distributed along the remaining directions.

To conclude, when there are multiple targets part of the stimulus set is always perfectly correlated. When measurement correlations are high, the observer checks whether the measurements are similar to each other to make a decision. When measurement correlations are low, the observer compares the measurements to the known target value. Measurement correlations can again decrease the overlap between the conditional distributions of measurements and have a significant impact on decisions and performance. For finite N, decisions are based on the comparison of measurements within a putative set of stimuli, L. We show next that when N is large, this is no longer the case.

3.3 Larger Number of Targets and Stimuli. If we assume a fixed proportion, K, of the stimuli consists of targets, so that $n = KN$, then equation 3.4 simplifies considerably in the limit of large N. If we let $c_x = \sigma_x^2(1-\rho_x)$ and $c_s = \sigma_s^2(1-\rho_s)$, the exponential in equation 3.4 has the form (see section B.6)

$$\alpha_L(\mathbf{x}, N, c_x, c_s) \approx \exp\left(-\frac{1}{2}\left[\underbrace{\frac{NK}{c_x}s^2_{i\in L}}_{I} - \underbrace{\frac{N}{c_s+c_x}s^2}_{II} + \underbrace{\frac{N(1-K)}{c_s+c_x}s^2_{i\notin L}}_{III}\right]\right), \tag{3.11}$$

where $s^2_{i\in L}$, $s^2_{i\notin L}$, and s^2 are the sample variances of measurements from the putative target set, L, outside the putative target set and over all N measurements, respectively.

The different terms in this expression can be interpreted as earlier. Term I is the sample variance of measurements of the putative targets. If this variance is large, then the set L is unlikely to contain targets. Term II is the sample variance among all terms. If this term is large, then all stimuli are dissimilar, and there is evidence that targets are present. For example, when distractors are correlated and $c_s \ll 1$, then the sample variance of all stimulus measurements is small only in the absence of targets. Finally, term III

Downloaded from http://direct.mit.edu/neco/article-pdf/27/11/2318/952316/neco_a_00778.pdf by guest on 16 May 2021

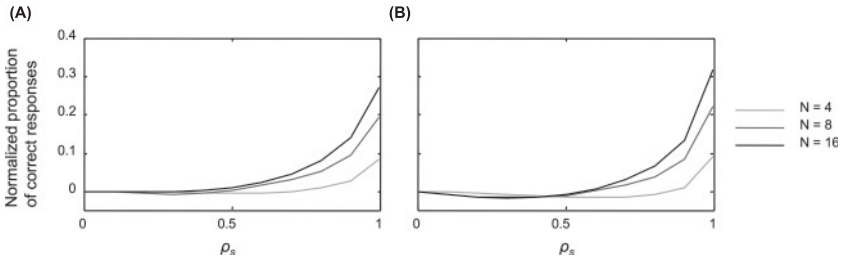

Figure 6: Normalized performance of an optimal observer with respect to $\rho_s = 0$ as a function of stimulus correlation, ρ_s, for different values of N in case of (A) $\rho_x = 0$ and (B) $\rho_x = 0.5$. Intermediate correlations between distractor measurements become increasingly important as the number of distractors increases from 4 to 16. Other parameters used are $\sigma_s = 15$, $\sigma_x = 4$, and $n = 2$.

is the sample variance among putative distractors. If distractors are correlated, this term will be small if L contains targets, and hence stimuli outside L are distractors. The sign of the three terms agrees with this interpretation: terms I and III are negative and term II is positive.

The main difference between the cases of large N and the examples discussed previously is that an observer takes into account putative distractor measurements—measurements outside the putative target set L. An exception is the case when measurement correlations are much stronger than distractor correlations, $c_x \ll c_s$. In this case, the putative targets are more strongly structured, and hence only their measurements are used in a decision. When $c_s \ll 1$, or, equivalently, $\rho_s \approx 1$ and distractors are strongly correlated, all three terms in equation 3.11 are comparable. In this case, ideal observers base their decision on the similarity, as measured by sample variance, of both putative distractor and target measurements.

Importantly, the decision is made using distractor measurements, even when distractors are not perfectly correlated. Figure 6 shows that intermediate distractor correlations increasingly have an impact on decisions with an increase in distractor number. Indeed, the higher the fraction of distractors, $(1 - K)$, the more weight is assigned to their sample variance (term III). This is unlike the case of small N, where distractor measurements are used only when they are perfectly correlated.

4 Discussion

We have shown that in a simple task, the statistical structure of the stimulus as well as that of noise in perceptual measurements determine the strategy and performance of an ideal observer. Correlations in measurement noise can have a significant impact on performance, particularly when distractor correlations are high. When the distribution of stimuli conditioned on the

Downloaded from http://direct.mit.edu/neco/article-pdf/27/11/2318/952316/neco_a_00778.pdf by guest on 16 May 2021

parameter of interest is concentrated in a small volume of stimulus space, the statistical structure of measurement noise can be particularly important (Mazyar et al., 2012, 2013; Bhardwaj et al., 2015).

The impact of noise correlations in neural responses on the inference of a parameter has been studied in detail (Averbeck et al., 2006; Averbeck, 2009; Latham & Nirenberg, 2005; Perkel, Gerstein, & Moore, 1967; Schneidman, Bialek, & Berry, 2003; Sompolinsky et al., 2001). Frequently the parameter of interest was identified with the stimulus, and both were univariate, although more complex and realistic cases have been examined (Montemurro & Panzeri, 2006; Mathis, Herz, & Stemmler, 2013). The estimation of the orientation of a bar in the receptive field of a population of neurons has been a canonical example.

Reality is far more complex. Natural visual and auditory scenes are high dimensional and highly structured. Moreover, only some of the parameters are typically relevant. Intuitively, if noise perturbs measurements along relevant direction (i.e., along the directions of the parameters of interest), then estimates will be corrupted. Perturbations along irrelevant directions in parameter space have little effect (Moreno-Bote et al., 2014). Measurement correlations can channel noise into irrelevant directions without decreasing overall noise magnitude, and thus improve parameter inference.

This is difficult to study using general theoretical models without putting some constraints on the structure of measurement noise. We therefore considered a relatively simple, analytically tractable example where both measurement and stimulus structure are characterized by a small number of parameters. We have used a similar setup to examine decision making in controlled search experiments (Mazyar et al., 2012, 2013; Bhardwaj et al., 2015).

We assumed that measurement noise and measurement correlations can be varied independently. This is not realistic. For instance, it is known that changes in the mean, variability, and covariability of neural responses can be tightly linked (Cohen & Kohn, 2011; de la Rocha, Doiron, Shea-Brown, Josić, & Reyes, 2007; Rosenbaum & Josić, 2011). It is thus likely that the statistics of measurement noise also change in concert. However, this relationship has not yet been well characterized.

More important, noise from the periphery of the nervous system will limit the performance of any observer. It is therefore not possible that a simple change in measurement correlations can lead to perfect performance (Moreno-Bote et al., 2014). To address this question, it would be necessary to provide a more accurate model of both the noise correlations in a recurrent network encoding information about the stimuli (Beck et al., 2011), as well as the resulting measurement correlations. This is beyond the scope of this study.

We also made strong assumptions about the structure of measurement and distractor correlations. We chose to restrict our analysis to positive correlations. The reason is that the requirement that a covariance matrix

Downloaded from http://direct.mit.edu/neco/article-pdf/27/11/2318/952316/neco_a_00778.pdf by guest on 16 May 2021

is positive definite implies restrictions on the range of allowable negative correlations between measurements and stimuli (Horn & Johnson, 2012). These restrictions depend on the number of stimuli, N, and complicate the analysis. To make the model tractable, we also assumed that all off-diagonal elements in the stimulus and measurement covariance matrices are identical. While we did not examine it here, heterogeneity in the correlation structure can strongly affect inference (Shamir & Sompolinsky, 2006; Chelaru & Dragoi, 2008; Berens, Ecker, Gerwinn, Tolias, & Bethge, 2011).

In practice, human and animal subjects are unlikely to behave like optimal observers. Indeed suboptimal inference may be a dominant cause of trial–to–trial variability (Beck, Ma, Pitkow, Latham, & Pouget, 2012). However, understanding the computations of an optimal observer still provides valuable information. First, understanding optimal inference provides a bound on what is possible; it is not possible to say whether an observer is suboptimal without it. Second, while computations required to achieve optimal performance may be expensive, biological systems may have evolved to efficiently approximate them. While the performance of humans and animals is likely suboptimal in all but the simplest tasks, in most cases we do not have sufficient information to infer what models they use to make an inference. Approximations and perturbations of the optimal model can serve as plausible models in these cases. However, the optimal model may not explain important aspects of visual information processing in human observers, such as pop-out (Wolfe, 1998).

Visual search is often used to investigate how stimulus number and homogeneity have an impact on inference. Detecting a target among distractors generally becomes more difficult as the number of stimuli increases. Here we followed the assumption made in a number of Bayesian and signal detection models that the precision with which individual stimuli are represented does not depend on stimulus number (Ma et al., 2011; Nolte & Jaarsma, 1967; Palmer et al., 1993; Palmer, Verghese, & Pavel, 2000; Rosenholtz, 2001; Verghese, 2001; Vincent, Baddeley, Troscianko, & Gilchrist, 2009). In this case, increased set size leads to an increase in noise, and makes it more difficult to detect the target unless stimulus correlations are strong. However, attentional resources may be limited (Townsend, 1974). An increase in the number of stimuli may therefore affect the precision with which each stimulus is encoded, and some stimuli may not be detected at all (Palmer, 1990; Shaw, 1980). Recent evidence suggests that precision decreases with set size (Mazyar et al., 2012, 2013). This effect could affect the details but not the general ideas we present here.

To end, we provide another illustration of the fact that measurement correlations can affect the performance of an ideal observer in different ways depending on the task. Suppose an observer is presented with N-oriented stimuli, such as Gabor patches. The stimuli and measurements follow the same gaussian distributions introduced earlier in this study. The observer is asked to perform one of the following two tasks: (1) report

Downloaded from http://direct.mit.edu/neco/article-pdf/27/11/2318/952316/neco_a_00778.pdf by guest on 16 May 2021

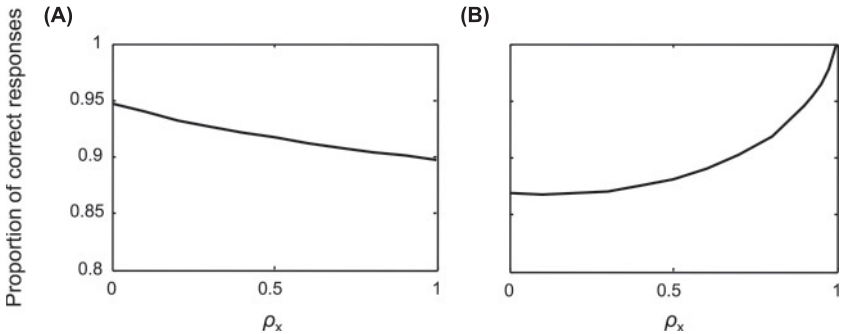

Figure 7: Performance comparison of an optimal observer as a function of measurement correlations in (A) a mean left/right discrimination task with $N = 4$ stimuli and (B) a target detection task with $n = 3$ targets. Other parameters used are $\sigma_s = 15$, $\sigma_x = 4$, and $\rho_s = 0.5$.

whether the mean orientation of the stimuli is to the left or right of vertical or (2) report whether a vertically oriented target is present or absent. In this case, a subset of the stimuli has vertical orientation on half the trials.

The first task is a discrimination task, and the observer needs to integrate information from different sources. When measurement correlations are high, it is more difficult to average out the noise between the stimulus measurements (Sompolinsky et al., 2001; Zohary, Shadlen, & Newsome, 1994). The estimate of the average orientation is therefore degraded and performance decreases with an increase in measurement correlations (see Figure 7A and appendix C). The second is a detection task that requires extracting information that is buried in a sea of distractors. As discussed above, in this case, measurement correlations can increase performance if there is more than one target or if the distractors are strongly correlated (see Figure 7B).

In this example, the stimuli and the measurements have the same statistical structure on target-absent trials for the two tasks. However, the parameter of interest differs. In the first task, the observer needs to estimate the average stimulus orientation, and in the second determine whether a target is present. The distributions of measurements conditioned on these parameters are therefore also different and are differently affected by measurement noise.

The question of how measurement correlations affect decision making and performance does not have a simple answer (Hu, Zylberberg, & Shea-Brown, 2014). Correlations in measurement noise can have a pronounced effect when the stimuli themselves are highly correlated (i.e., when they occupy a small volume in stimulus space). We have illustrated how in this case, measurement correlations can help in separating the distribution of

Downloaded from http://direct.mit.edu/neco/article-pdf/27/11/2318/952316/neco_a_00778.pdf by guest on 16 May 2021

measurements conditioned on a parameter of interest. Similar considerations will be important whenever we try to understand how information can be extracted from the collective responses of neural populations to high-dimensional, and highly structured stimuli.

Appendix A: Derivation of Equations 3.3 and 3.4

Here we present the details of some of the calculations leading to the results presented in the main text. We follow the assumptions made in section 2. The following derivations involve rearrangement of somewhat complicated algebraic expressions. These steps are best checked using a computer algebra system.

The computations rely on the assumption of gaussianity. We therefore make repeated use of the fact that the density $\mathcal{N}(\mathbf{x}; \boldsymbol{\mu}\boldsymbol{\Sigma})$ of the normal distribution with mean $\boldsymbol{\mu}$, and covariance $\boldsymbol{\Sigma}$ is

$$\mathcal{N}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{\sqrt{(2\pi)^N |\boldsymbol{\Sigma}|}} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} (\mathbf{x}-\boldsymbol{\mu})\right), \tag{A.1}$$

where division by $|\boldsymbol{\Sigma}|$, the determinant of the matrix $\boldsymbol{\Sigma}$, is required for normalization.

Since the density $p(\mathbf{s}|T=1, L)$ is singular, we introduce the auxiliary covariance,

$$(\boldsymbol{\Sigma}^{\eta}_{\mathbf{s},L})_{i,j} = \begin{cases} (\boldsymbol{\Sigma}_{\mathbf{s}\setminus L})_{i,j}, & \text{if } i, j \notin L, \\ \eta, & \text{if } i = j \in L, \\ 0, & \text{if } i \in L, \text{ or } j \in L, \text{ and } i \neq j, \end{cases} \tag{A.2}$$

where $\eta > 0$. Therefore,

$$p(\mathbf{s}|T=1, L) = \lim_{\eta \downarrow 0} \mathcal{N}(\mathbf{s}; \mathbf{0}_N, \boldsymbol{\Sigma}^{\eta}_{\mathbf{s},L}).$$

Let $M = \binom{N}{n}$ denote the cardinality of the set $\mathcal{L}$ of all possible sets L. We compute $p(\mathbf{x}|T=1)$ in equation 3.1 by marginalizing over $\mathbf{s}$,

$$p(\mathbf{x}|T=1) = \int p(\mathbf{x}|\mathbf{s}) p(\mathbf{s}|T=1) d\mathbf{s}.$$

We note that

$$p(\mathbf{s}|T=1) = \sum_{L \in \mathcal{L}} p(\mathbf{s}|T=1, L) p(L) = \frac{1}{M} \sum_{L \in \mathcal{L}} p(\mathbf{s}|T=1, L).$$

Downloaded from http://direct.mit.edu/neco/article-pdf/27/11/2318/952316/neco_a_00778.pdf by guest on 16 May 2021

Therefore,

$$\begin{aligned}
p(\mathbf{x}|T=1) &= \frac{1}{M}\int p(\mathbf{x}|\mathbf{s})\sum_{L\in\mathcal{L}} p(\mathbf{s}|T=1,L)d\mathbf{s} \\
&= \frac{1}{M}\sum_{L\in\mathcal{L}}\int p(\mathbf{x}|\mathbf{s})p(\mathbf{s}|T=1,L)d\mathbf{s} \\
&= \frac{1}{M}\lim_{\eta\to 0}\sum_{L\in\mathcal{L}}\int \mathcal{N}(\mathbf{x};\mathbf{s},\mathbf{\Sigma_x})\mathcal{N}(\mathbf{s};\mathbf{0}_N,\mathbf{\Sigma}^{\eta}_{\mathbf{s},L})d\mathbf{s} \\
&= \frac{1}{M}\sum_{L\in\mathcal{L}}\mathcal{N}(\mathbf{x};\mathbf{0}_N,\mathbf{C}_L).
\end{aligned}$$

Similarly,

$$\begin{aligned}
p(\mathbf{x}|T=0) &= \int p(\mathbf{x}|\mathbf{s})p(\mathbf{s}|T=0)d\mathbf{s} \\
&= \int \mathcal{N}(\mathbf{x};\mathbf{s},\mathbf{\Sigma_x})\mathcal{N}(\mathbf{s};\mathbf{0}_N,\mathbf{\Sigma_s})d\mathbf{s} \\
&= \mathcal{N}(\mathbf{x};\mathbf{0}_N,\mathbf{C}),
\end{aligned}$$

where $\mathbf{C}_L = \mathbf{\Sigma_x} + \mathbf{\Sigma}^0_{\mathbf{s},L}$ and $\mathbf{C} = \mathbf{\Sigma_x} + \mathbf{\Sigma_s}$. We therefore obtain

$$\begin{aligned}
d_{\mathrm{TD}}(\mathbf{x}) &= \log\frac{p(\mathbf{x}|T=1)}{p(\mathbf{x}|T=0)} = \log\left(\frac{1}{M}\sum_{L\in\mathcal{L}}\frac{\mathcal{N}(\mathbf{x};\mathbf{0}_N,\mathbf{C}_L)}{\mathcal{N}(\mathbf{x};\mathbf{0}_N,\mathbf{C})}\right) \\
&= \log\left(\frac{1}{M}\sqrt{\frac{|\mathbf{C}|}{|\mathbf{C}_L|}}\sum_{L\in\mathcal{L}}\exp\left(-\frac{1}{2}\mathbf{x}^T\left(\mathbf{C}_L^{-1}-\mathbf{C}^{-1}\right)\mathbf{x}\right)\right). \qquad \text{(A.3)}
\end{aligned}$$

The determinants of the covariance matrices, $|\mathbf{C}|$ and $|\mathbf{C}_L|$, appear in the definition of the multivariate normal distribution, and thus their ratio appears in the likelihood ratio. We note that the determinant of $\mathbf{C}_L$ does not depend on the set L since all matrices $\mathbf{C}_L$ can be obtained from each other by permuting appropriate rows and columns.

In the case that variances and covariances are equal, we can invert $\mathbf{C}_L$ and $\mathbf{C}$. In general, the matrix $\mathbf{C}_L$ has the following form:

$$(\mathbf{C}_L)_{i,j} = \begin{cases} \sigma_x^2, & \text{if } i=j\in L, \\ \sigma_s^2+\sigma_x^2, & \text{if } i=j\notin L, \\ \rho_x\sigma_x^2, & \text{if } i\neq j, \text{ and } i \text{ or } j\in L, \\ \rho_s\sigma_s^2+\rho_x\sigma_x^2, & \text{if } i\neq j, \text{ and } i,j\notin L. \end{cases}$$

Downloaded from http://direct.mit.edu/neco/article-pdf/27/11/2318/952316/neco_a_00778.pdf by guest on 16 May 2021

We use the Sherman-Morrison-Woodbury formula (Meyer, 2000),

$$(A + UEV)^{-1} = A^{-1} - A^{-1}U\left(E^{-1} + VA^{-1}U\right)^{-1}VA^{-1}, \tag{A.4}$$

to obtain the inverses of $\mathbf{C}_L$ and $\mathbf{C}$. To do so, we first rewrite $\mathbf{C}_L$ as $A_L + U_L E V_L$: When $L = \{1, 2, \ldots, n\}$ (targets placed at the first n out of N possible locations), we have

$$A_L = \begin{bmatrix} (\sigma_x^2 - b)I_n & 0_{n\times(N-n)} \\ 0_{(N-n)\times n} & (\sigma_s^2 + \sigma_x^2 - a)I_{(N-n)} \end{bmatrix}_{(N\times N)},$$

$$U_L = \begin{bmatrix} b & \cdots & b & a & \cdots & a \\ b & \cdots & \cdots & \cdots & \cdots & b \end{bmatrix}^{\mathrm{T}}_{(N\times 2)},$$

$$V_L = \begin{bmatrix} 0 & \cdots & 0 & 1 & \cdots & 1 \\ 1 & \cdots & 1 & 0 & \cdots & 0 \end{bmatrix}_{(2\times N)}, \quad \text{and} \quad E = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}_{(2\times 2)},$$

where $a = \rho_s\sigma_s^2 + \rho_x\sigma_x^2$ and $b = \rho_x\sigma_x^2$.

Using equation A.4, we find that a lengthy but straightforward calculation gives

$$(\mathbf{C}_L^{-1})_{i,j} = \begin{cases} v_L - \dfrac{bv_L^2 g_L}{\gamma_L}, & \text{if } i = j \in L, \\ v - \dfrac{v^2 f_L}{\gamma_L}, & \text{if } i = j \notin L, \\ -\dfrac{bv_L^2 g_L}{\gamma_L}, & \text{if } i \neq j, \text{ and } i, j \in L, \\ -\dfrac{v^2 f_L}{\gamma_L}, & \text{if } i \neq j, \text{ and } i, j \notin L, \\ -\dfrac{bvv_L}{\gamma_L}, & \text{if } i \neq j, i \in L, j \notin L \text{ or } i \neq j, i \notin L, j \in L, \end{cases}$$

where

$$\begin{aligned} v &= \frac{1}{\sigma_s^2(1-\rho_s) + \sigma_x^2(1-\rho_x)}, \quad v_L = \frac{1}{\sigma_x^2(1-\rho_x)} \\ f_L &= a + n\rho_s\sigma_s^2\rho_x\sigma_x^2 v_L, \quad g_L = 1 + \rho_s\sigma_s^2\,(N-n)v, \quad \text{and} \\ \gamma_L &= 1 + a(N-n)v + n\rho_x\sigma_x^2\, v_L g_L. \end{aligned} \tag{A.5}$$

Downloaded from http://direct.mit.edu/neco/article-pdf/27/11/2318/952316/neco_a_00778.pdf by guest on 16 May 2021

Using the matrix determinant lemma (Harville, 1997), we can obtain the determinant of $\mathbf{C}_L$:

$$\det(\mathbf{C}_L) = \frac{\gamma_L}{v_L^n\, v^{N-n}}.$$

Similarly, we compute the inverse and determinant of matrix $\mathbf{C}$,

$$\mathbf{C}^{-1} = \begin{bmatrix} v-\beta v^2 & -\beta v^2 & \cdots & -\beta v^2 \\ -\beta v^2 & v-\beta v^2 & \cdots & -\beta v^2 \\ \vdots & & \vdots & \\ -\beta v^2 & -\beta v^2 & \cdots & v-\beta v^2 \end{bmatrix},$$

where

$$\beta = \frac{a}{1+aNv} \tag{A.6}$$

and

$$\det(\mathbf{C}) = \frac{1+aNv}{v^N}.$$

The prefactor in equation A.3 is therefore

$$\sqrt{\frac{|\mathbf{C}|}{|\mathbf{C}_L|}} = \sqrt{\frac{1+aNv}{\gamma_L}\left(\frac{v_L}{v}\right)^n},$$

and by straightforward multiplication, we can compute

$$\mathbf{x}^T\mathbf{C}^{-1}\mathbf{x} = (v-\beta v^2)\sum_{i=1}^{N} x_i^2 - \beta v^2 \sum_{i\neq j}^{N} x_i x_j,$$

$$\begin{aligned}\mathbf{x}^T\mathbf{C}_L^{-1}\mathbf{x} = &\left(v_L - \frac{1}{\gamma_L} b v_L^2 g_L\right)\sum_{i\in L} x_i^2 + \left(v - \frac{1}{\gamma_L} v^2 f_L\right)\sum_{i\notin L} x_i^2 \\ &- \frac{1}{\gamma_k} b v_L^2 g_L \sum_{\substack{i,j\in L\\ i\neq j}} x_i x_j - \frac{2}{\gamma_L} b v v_L \sum_{\substack{i\in L\\ j\notin L}} x_i x_j - \frac{1}{\gamma_k} v^2 f_L \sum_{\substack{i,j\notin L\\ i\neq j}} x_i x_j.\end{aligned}$$

Downloaded from http://direct.mit.edu/neco/article-pdf/27/11/2318/952316/neco_a_00778.pdf by guest on 16 May 2021

A slight rearrangement of the terms therefore shows that when variances and covariances are equal, equation A.3 is equivalent to

$$d_{\mathrm{TD}}(\mathbf{x}) = \log\left(\frac{1}{M}\sqrt{\frac{1+aNv}{\gamma_L}\left(\frac{v_L}{v}\right)^n}\sum_{L\in\mathcal{L}}\exp\left\{-\frac{1}{2}\left(\sigma_s^2(1-\rho_s)v_L v\sum_{i\in L}x_i^2\right.\right.\right.$$
$$+\left(\beta v^2-\frac{v_L^2 b g_L}{\gamma_L}\right)\sum_{i,j\in L}x_i x_j + 2\left(\beta v^2-\frac{v_L v b}{\gamma_L}\right)\sum_{i\in L,\, j\notin L}x_i x_j$$
$$\left.\left.\left.+\left(\beta v^2-\frac{f_L v^2}{\gamma_L}\right)\sum_{i,j\notin L}x_i x_j\right)\right\}\right).$$

Appendix B: Asymptotic Analysis of Equation 3.4

Here we present some asymptotic results for the decision variable $d_{\mathrm{TD}}(\mathbf{x})$ given in equation 3.4. The main results are obtained for small measurement noise, σ_x^2. Equivalent results can be obtained for large external variability, σ_s^2. We let $\epsilon = \frac{\sigma_x^2}{\sigma_s^2}$ and find the leading terms in equation 3.4 assuming $\epsilon \ll 1$. We consider only terms with larger contribution—terms of the orders of $\mathcal{O}(\frac{1}{\epsilon})$ to obtain approximations for $d_{\mathrm{TD}}(\mathbf{x})$ in different parameter regimes.

B.1 Small Measurement Noise and Identical Distractors, $\rho_s = 1$. We start with the case $n = 1$. The prefactors of the exponential terms in equation 3.4 simplify to

$$\sigma_s^2(1-\rho_s)v_L v = 0,$$

$$\beta v^2 - \frac{\rho_x\sigma_x^2 v_L^2 g_L}{\gamma_L} = \frac{\sigma_s^2+\rho_x\sigma_x^2}{\sigma_x^2(1-\rho_x)[N\sigma_s^2+\sigma_x^2(1+(N-1)\rho_x)]}$$
$$-\frac{\rho_x[\sigma_s^2(N-1)+\sigma_x^2(1-\rho_x)]}{\sigma_x^2(1-\rho_x)[\sigma_s^2(N-1)+\sigma_x^2(1-\rho_x)(1+(N-1)\rho_x)]}$$
$$= \frac{1-N\rho_x}{N\sigma_x^2(1-\rho_x)} + O(1),$$

$$\beta v^2 - \frac{\rho_x\sigma_x^2 v_L v}{\gamma_L} = \frac{\sigma_s^2+\sigma_x^2\rho_x}{\sigma_x^2(1-\rho_x)[N\sigma_s^2+\sigma_x^2(1+(N-1)\rho_x)]}$$
$$-\frac{\rho_x}{\sigma_s^2(N-1)+\sigma_x^2(1-\rho_x)(1+(N-1)\rho_x)}$$
$$= \frac{1}{N\sigma_x^2(1-\rho_x)} + O(1),$$

Downloaded from http://direct.mit.edu/neco/article-pdf/27/11/2318/952316/neco_a_00778.pdf by guest on 16 May 2021

$$\beta v^2 - \frac{v^2 f_L}{\gamma_L} = \frac{\sigma_s^2 + \sigma_x^2 \rho_x}{\sigma_x^2(1-\rho_x)[N\sigma_s^2 + \sigma_x^2(1+(N-1)\rho_x)]} - \frac{\sigma_s^2 + \rho_x \sigma_x^2(1-\rho_x)}{\sigma_x^2(1-\rho_x)[\sigma_s^2(N-1) + \sigma_x^2(1-\rho_x)(1+(N-1)\rho_x)]}$$
$$= -\frac{1}{N(N-1)\sigma_x^2(1-\rho_x)} + O(1).$$

And the leading determinant term becomes

$$\sqrt{\frac{1+aNv}{\gamma_L}\frac{v_L}{v}} = \sqrt{\frac{(1-\rho_x)[N\sigma_s^2 + \sigma_x^2(1+(N-1)\rho_x)]}{\sigma_s^2(N-1) + \sigma_x^2(1-\rho_x)(1+(N-1)\rho_x)}}$$
$$= \sqrt{\frac{N(1-\rho_x)}{N-1} + O\left(\sigma_x^2\right)}.$$

Therefore, the decision variable becomes approximately,

$$d_{\mathrm{TD}}(\mathbf{x}) \approx \log\left(\frac{1}{N}\sqrt{\frac{N(1-\rho_x)}{N-1}}\sum_{i=1}^{N}\exp\left\{\frac{-1}{2N\sigma_x^2(1-\rho_x)}\right.\right.$$
$$\left.\left.\times\left((1-N\rho_x)x_i^2 + 2x_i\sum_{j\neq i}x_j - \frac{1}{N-1}\left(\sum_{j\neq i}x_i\right)^2\right)\right\}\right).$$

After rearranging terms, this can be rewritten as

$$d_{\mathrm{TD}}(\mathbf{x}) \approx \log\left(\frac{1}{N}\sqrt{\frac{N(1-\rho_x)}{N-1}}\sum_{i=1}^{N}\exp\left[\rho_x\frac{N-1}{2N\sigma_x^2(1-\rho_x)}\left(x_i - \bar{x}_{\hat{i}}\right)^2\right.\right.$$
$$\left.\left. - \frac{1}{2\sigma_x^2}\left(N\bar{x}^2 - (N-1)\bar{x}_{\hat{i}}^2\right)\right]\right), \tag{B.1}$$

where $\bar{x}$ is the sample mean of all measurements and $\bar{x}_{\hat{i}}$ is the sample mean of the measurements excluding the putative target, i. In the limiting cases $\rho_x = 0$ and $(1-\rho_x) \ll 1$, we obtain the exponents given in equations 3.6 and 3.7 discussed in the text.

B.2 Perfect Performance When $\rho_x = 1$. We show that when $\rho_s = 1$, an ideal observer performs perfectly in the limit of identical measurement noise. For a fixed number of stimuli, N, on target-absent trials, $x_i - (N-1)^{-1}\sum_{j\neq i}x_j = O(\sqrt{\epsilon})$, and hence $\alpha(\mathbf{x}, N, \rho_x, \sigma_x) = O(1)$. On the

Downloaded from http://direct.mit.edu/neco/article-pdf/27/11/2318/952316/neco_a_00778.pdf by guest on 16 May 2021

other hand the prefactor, $P(N, \rho_x) = O(\sqrt{\epsilon})$, and hence $d_{\mathrm{TD}}(\mathbf{x}) \to -\infty$, as $\epsilon \to 0$, that is, as $\rho_x \to 1$. On target-present trials, when stimulus i is the target, then $x_i - (N-1)^{-1} \sum_{j \neq i} x_j = O(1)$, and $\alpha_i(\mathbf{x}, N, \rho_x, \sigma_x) = O(1/\epsilon)$. In this case, the prefactor is still $P(N, \rho_x) = O(\sqrt{\epsilon})$, and the exponential term dominates. A similar argument works for the summands for which i is not a target.

B.3 Single Target with Increasing Number of Distractors. We still work under the assumption that measurement noise is relatively weak so that we can use equation B.1. Note that on target-absent trials, $\bar{x} = s + O(\sigma_x^2/\sqrt{N})$, where s is the true value of the (identical) distractors. We also have $\bar{x}_{\hat{i}} = s + O(\sigma_x^2/\sqrt{N-1})$. Hence, the first term in the exponential of equation B.1 is $O(1)$, while the second term is $O(\sqrt{N})$. A similar argument holds in target-present trials.

On target-absent trials,

$$-\frac{1}{2\sigma_x^2}\left(N\bar{x}^2 - (N-1)\bar{x}_i^2\right) = \exp\left[-\frac{1}{2\sigma_x^2}s^2 + O(\sqrt{N}\sigma_x)\right] \qquad \text{for all } i,$$

to leading order in N. We abuse notation slightly and use only order notation on the terms that include measurement noise. As stimuli become more dissimilar to the target (i.e., as s^2 increases), α_i decreases exponentially, $d_{\mathrm{TD}}(\mathbf{x})$ becomes more negative, and it is hence easier to infer that a target is absent. However, the $O(\sqrt{N}\sigma_x)$ terms can be both positive and negative. Thus, performance decreases with the number of stimuli. Similarly we can see that when a target is present,

$$\alpha_i(\mathbf{x}, N, \sigma_x^2) = \exp\left[\frac{1}{2\sigma_x^2}\frac{N-1}{N}s^2 + O(\sqrt{N}\sigma_x)\right] \qquad \text{when } i \text{ is the target,} \tag{B.2}$$

$$\alpha_i(\mathbf{x}, N, \rho_x) = \exp\left[-\frac{1}{2\sigma_x^2}\frac{1+N-N^2}{N-N^2}s^2 + O(\sqrt{N}\sigma_x)\right] \qquad \text{when } i \text{ is not a target,} \tag{B.3}$$

again to leading order in N. As measurement noise decreases or s^2 increases, the first term in equation B.2 diverges exponentially, and the terms in equation B.3 approach 0 exponentially. As a result, $d_{\mathrm{TD}}(\mathbf{x})$ increases. However, the $O(\sqrt{N}\sigma_x)$ noise term increases with N, and an increase in the number of stimuli again decreases performance.

If $(1 - \rho_x) \ll 1/N$, (i.e., measurement noise is strongly correlated), the first term in the exponential of equation B.1 dominates. Thus, when correlations increase faster than the inverse of the number of distractors, performance increases with the number of distractors.

Downloaded from http://direct.mit.edu/neco/article-pdf/27/11/2318/952316/neco_a_00778.pdf by guest on 16 May 2021

B.4 Weak External Structure, $\rho_s < 1$, Arbitrary Number of Targets. We approximate each term in the exponential of equation 3.4 assuming $\sigma_x^2 \ll 1$:

$$\begin{aligned}
&\sigma_s^2(1-\rho_s)v_L v \\
&\quad = \frac{\sigma_s^2(1-\rho_s)}{\sigma_x^2(1-\rho_x)[\sigma_s^2(1-\rho_s)+\sigma_x^2(1-\rho_x)]} \\
&\quad = \frac{1}{\sigma_x^2(1-\rho_x)} + O(1), \\
&\beta v^2 - \frac{\rho_x \sigma_x^2 v_L^2 g_L}{\gamma_L} \\
&\quad = \frac{\rho_s\sigma_s^2 + \rho_x\sigma_x^2}{[\sigma_s^2(1-\rho_s)+\sigma_x^2(1-\rho_x)][\sigma_s^2(1+(N-1)\rho_s)+\sigma_x^2(1+(N-1)\rho_x)]} \\
&\qquad - \frac{\rho_x[\sigma_s^2(1+(N-n-1)\rho_s+\sigma_x^2(1-\rho_x)]}{\sigma_x^2(1-\rho_x)[\sigma_s^2(1+(n-1)\rho_x)(1+(N-n-1)\rho_s)} \\
&\qquad\quad + \sigma_x^2(1-\rho_x)(1+(N-1)\rho_x))] \\
&\quad = -\frac{\rho_x}{\sigma_x^2(1-\rho_x)(1+(n-1)\rho_x)} + O(1), \\
&\beta v^2 - \frac{\rho_x \sigma_x^2 v_L v}{\gamma_L} \\
&\quad = \frac{\rho_s\sigma_s^2 + \rho_x\sigma_x^2}{[\sigma_s^2(1-\rho_s)+\sigma_x^2(1-\rho_x)][\sigma_s^2(1+(N-1)\rho_s)+\sigma_x^2(1+(N-1)\rho_x)]} \\
&\qquad - \frac{\rho_x}{[\sigma_s^2(1+(N-n-1)\rho_s)(1+(n-1)\rho_x)+\sigma_x^2(1-\rho_x)(1+(N-1)\rho_x)]} \\
&\quad = O(1),
\end{aligned}$$

and

$$\begin{aligned}
\beta v^2 - \frac{v^2 f_L}{\gamma_L} &= \frac{\rho_s\sigma_s^2 + \rho_x\sigma_x^2}{[\sigma_s^2(1-\rho_s)+\sigma_x^2(1-\rho_x)][\sigma_s^2(1+(N-1)\rho_s)} \\
&\qquad + \sigma_x^2(1+(N-1)\rho_x)] \\
&\quad - \frac{\rho_s\sigma_s^2(1+(n-1)\rho_x)+\rho_x\sigma_x^2(1-\rho_x)}{[\sigma_s^2(1-\rho_s)+\sigma_x^2(1-\rho_x)][\sigma_s^2(1+(N-n-1)\rho_s)(1+(n-1)\rho_x)} \\
&\qquad + \sigma_x^2(1-\rho_x)(1+(N-1)\rho_x)] \\
&= O(1).
\end{aligned}$$

Downloaded from http://direct.mit.edu/neco/article-pdf/27/11/2318/952316/neco_a_00778.pdf by guest on 16 May 2021

We also approximate the leading coefficient of the exponential term in equation 3.4 as

$$\sqrt{\frac{1+(\rho_s\sigma_s^2+\rho_x\sigma_x^2)Nv}{\gamma_L}\left(\frac{v_L}{v}\right)^n} =$$

$$\sqrt{\frac{(1-\rho_x)[\sigma_s^2(1+(N-1)\rho_s)+\sigma_x^2(1+(N-1)\rho_x)]}{[\sigma_s^2(1+(N-n-1)\rho_s)(1+(n-1)\rho_x)+\sigma_x^2(1-\rho_x)(1+(N-1)\rho_x)]} \times \left(\frac{\sigma_s^2(1-\rho_s)+\sigma_x^2(1-\rho_x)}{\sigma_x^2(1-\rho_x)}\right)^n}$$

$$= \sqrt{\frac{(1-\rho_x)(1+(N-1)\rho_s)}{(1+(N-n-1)\rho_s)(1+(n-1)\rho_x)}\left(\frac{\rho_s(1-\rho_s)}{\sigma_x^2(1-\rho_x)}\right)^n} + O(1).$$

Combining the above terms, equation 3.4 reduces to the following expression under the assumption of $\sigma_x^2 \ll \sigma_s^2$, $\rho_s < 1$,

$$d_{\mathrm{TD}}(\mathbf{x}) \approx \log\left(\frac{1}{M}\sqrt{\frac{(1-\rho_x)(1+(N-1)\rho_s)}{(1+(N-n-1)\rho_s)(1+(n-1)\rho_s)}\left(\frac{\sigma_s^2(1-\rho_s)}{\sigma_x^2(1-\rho_x)}\right)^n} \times \sum_{L\in\mathcal{L}}\exp\left\{-\frac{1}{2\sigma_x^2(1-\rho_x)}\left(\sum_{i\in L}x_i^2-\frac{\rho_x}{1+(n-1)\rho_x}\sum_{i,j\in L}x_ix_j\right)\right\}\right).$$

B.4.1 Special Case: $n = 1$. In case of a single target, set L has only one element and $\mathcal{L} = \{1, 2, \ldots, N\}$. In this case, equation 3.9 reduces to a much simpler expression

$$d_{\mathrm{TD}}(\mathbf{x}) \approx \log\left(\frac{1}{N}\sqrt{\frac{\sigma_s^2(1-\rho_s)(1+(N-1)\rho_s)}{\sigma_x^2(1+(N-2)\rho_s)}}\sum_{i=1}^{N}\exp\left\{-\frac{x_i^2}{2\sigma_x^2}\right\}\right).$$

This expression is independent of ρ_x. Hence, the decision boundary and the performance of an ideal observer are unaffected by measurement correlations.

B.5 Near-perfect Performance with $n > 1$, and $\rho_x \approx 1$. We first assume that $T = 1$. From equation 3.10, if L_T is the set of targets, then this expression is approximately zero and the exponential is approximately unity. When L is a set not consisting of all targets, then the expression in equation 3.10 has

Downloaded from http://direct.mit.edu/neco/article-pdf/27/11/2318/952316/neco_a_00778.pdf by guest on 16 May 2021

expectation greater than zero. Then

$$\sum_{L\in\mathcal{L}} \exp\left\{-\frac{n}{2\sigma_x^2(1-\rho_x)}\left(\frac{1}{n}\sum_{i\in L} x_i^2 - \left(\frac{1}{n}\sum_{i\in L} x_i\right)^2\right)\right\}$$
$$= 1 + \sum_{L\in\mathcal{L}\setminus L_T} \exp\left\{-\frac{n\alpha_L^2}{2\sigma_x^2(1-\rho_x)}\right\} \to 1. \tag{B.4}$$

The prefactor in equation 3.9 diverges since the exponential dominates. If $T = 0$, then equation 3.10 will be greater than zero for all sets, L. Thus, $d_{TD}(\mathbf{x}) \to -\infty$ with $\rho_x \to 1$.

B.6 Asymptotics for Large *N*. Here we develop asymptotic results for the decision variable when the number of stimuli, N, is large. We assume that a fraction K of the stimuli are targets, so that there are NK targets and $(1-K)N$ distractors. To simplify notation, we write $c_x = \sigma_x^2(1-\rho_x)$ and $c_s = \sigma_s^2(1-\rho_s)$. We find that

$$(1-\rho_s)v_L v = \frac{c_s}{c_x(c_s+c_x)}$$
$$v^2\left(\beta - \frac{1}{\gamma_L}\left(\frac{v_L}{v}\right)^2 \rho_x\sigma_x^2 g_L\right) = \frac{1}{N}\frac{1}{c_s+c_x} - \frac{1}{NK}\frac{1}{c_x} + O(\frac{1}{N^2})$$
$$v^2\left(\beta - \frac{1}{\gamma_L}\frac{v_L}{v}\rho_x\sigma_x^2\right) = \frac{1}{N}\frac{1}{c_s+c_x} + O(\frac{1}{N^2})$$
$$v^2\left(\beta - \frac{1}{\gamma_L}f_L\right) = \left(\frac{1}{N} - \frac{1}{(1-K)N}\right)\frac{1}{c_x+c_s} + O(\frac{1}{N^2})$$
$$\frac{1+(\rho_s\sigma_s^2+\rho_x\sigma_x^2)Nv}{\gamma_L}\left(\frac{v_L}{v}\right)^n = \left(1+\frac{c_s}{c_x}\right)^{KN}\left(\frac{1}{N}\frac{(\rho_x-1)(\rho_s\sigma_s^2+\rho_x\sigma_x^2)}{(K-1)K\rho_s\rho_x\sigma_s^2}\right).$$

We can therefore group the terms in the exponential of the decision variable given by equation 3.4 as

$$-\frac{1}{2}\left(\frac{c_s}{c_x(c_s+c_x)}\sum_{i\in L} x_i^2 + \frac{1}{N}\frac{1}{c_x+c_s}\sum_{i,j}^{N} x_i x_j - \frac{1}{NK}\frac{1}{c_x}\sum_{i,j\in L} x_i x_j\right.$$
$$\left. - \frac{1}{(1-K)N}\frac{1}{c_x+c_s}\sum_{i,j\notin L} x_i x_j\right).$$

A simple reorganization of the terms yields equation 3.11.

Downloaded from http://direct.mit.edu/neco/article-pdf/27/11/2318/952316/neco_a_00778.pdf by guest on 16 May 2021

Appendix C: Mean Stimulus Orientation: Left or Right Discrimination Task

We next present details of the analysis of the task presented in section 4: an observer is again presented with N stimuli on every trial. For concreteness, we can think of the stimuli as Gabor patches, with orientations $\mathbf{s} = (s_1, s_2, \cdots, s_N)$. The task is to decide whether the mean orientation of the set is to the left (a condition we denote by $C = -1$) or right ($C = 1$) of the vertical. The observer makes a decision based on the measurements, $\mathbf{x} = (x_1, x_2, \cdots, x_N)$. Stimulus orientations are drawn from a multivariate normal distribution with mean vector, $\mathbf{0}_N$, and covariance matrix, $\mathbf{\Sigma_s}$,

$$p(\mathbf{s}) = \mathcal{N}(\mathbf{0}_N, \mathbf{\Sigma_s}), \tag{C.1}$$

with $\mathbf{\Sigma_s}$ defined in equation 2.2. As in the target detection task, we assume the measurements follow a multivariate normal distribution with mean vector and covariance matrix specified in equation 2.3.

An ideal observer performs the task by making a decision based on the log posterior ratio,

$$\begin{aligned} d_{\mathrm{MOD}}(\mathbf{x}) &= \log \frac{p(C=1|\mathbf{x})}{p(C=-1|\mathbf{x})} = \log \frac{p(\bar{s}>0|\mathbf{x})}{p(\bar{s}<0|\mathbf{x})} \\ &= \log \frac{p(\mathbf{x}|\bar{s}>0)}{p(\mathbf{x}|\bar{s}<0)} + \log \frac{p(\bar{s}>0)}{p(\bar{s}<0)}, \end{aligned} \tag{C.2}$$

where $\bar{s} = \sum_{i=1}^{N} s_i$ denotes the mean stimulus orientation on a trial. If $d_{\mathrm{MOD}}(\mathbf{x}) > 0$, the observer infers $\hat{C} = 1$, that is, the mean stimulus orientation is to the right of the vertical. We compute equation C.2 by marginalizing over $\mathbf{s}$ and applying Bayes' rule,

$$\begin{aligned} p(\mathbf{x}|\bar{s}>0) &= \int p(\mathbf{x}|\mathbf{s})p(\mathbf{s}|\bar{s}>0)d\mathbf{s} \\ &= \int p(\mathbf{x}|\mathbf{s})p(\bar{s}>0|\mathbf{s})\frac{p(\mathbf{s})}{p(\bar{s}>0)}d\mathbf{s} \\ &= \frac{1}{p(\bar{s}>0)}\int_{\bar{s}>0} p(\mathbf{x}|\mathbf{s})p(\mathbf{s})d\mathbf{s} \\ &= \frac{1}{p(\bar{s}>0)}\int_{\bar{s}>0} \mathcal{N}(\mathbf{x};\mathbf{s},\mathbf{\Sigma_x})\mathcal{N}(\mathbf{s};\mathbf{0}_N,\mathbf{\Sigma_s})d\mathbf{s} \\ &= \frac{z_c}{p(\bar{s}>0)}\int_{\bar{s}>0} \mathcal{N}\left(\mathbf{s};\left(\mathbf{I}+\mathbf{\Sigma_x}\mathbf{\Sigma_s}^{-1}\right)\mathbf{x},\left(\mathbf{\Sigma_x}^{-1}+\mathbf{\Sigma_s}^{-1}\right)^{-1}\right)d\mathbf{s}, \end{aligned}$$

Downloaded from http://direct.mit.edu/neco/article-pdf/27/11/2318/952316/neco_a_00778.pdf by guest on 16 May 2021

where z_c is a normalization constant. Similarly, we compute $p(\mathbf{x}|\bar{s} < 0)$ and obtain

$$d_{\mathrm{MOD}}(\mathbf{x}) = \log\left(\frac{\int_{\bar{s}>0} \mathcal{N}\left(\mathbf{s}; \left(\mathbf{I} + \mathbf{\Sigma_x}\mathbf{\Sigma_s}^{-1}\right)^{-1}\mathbf{x}, \left(\mathbf{\Sigma_x}^{-1} + \mathbf{\Sigma_s}^{-1}\right)^{-1}\right) d\mathbf{s}}{\int_{\bar{s}<0} \mathcal{N}\left(\mathbf{s}; \left(\mathbf{I} + \mathbf{\Sigma_x}\mathbf{\Sigma_s}^{-1}\right)^{-1}\mathbf{x}, \left(\mathbf{\Sigma_x}^{-1} + \mathbf{\Sigma_s}^{-1}\right)^{-1}\right) d\mathbf{s}}\right). \quad \text{(C.3)}$$

By symmetry, it is easy to see that the decision boundary in the space of measurements is given by the hyperplane $\bar{x} = 0$, where $\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$ is the sample mean of the measurement. An ideal observer therefore bases the decision only on the sample mean. Therefore, we have $\bar{x} > 0 \Rightarrow \bar{s} > 0$.

In order to understand the negative impact of noise correlations on the performance of an optimal observer on this task, consider $\mathbf{x} = \mathbf{s} + \vec{\xi}$, where $\vec{\xi} = (\xi_1, \xi_2, \cdots, \xi_N) \sim \mathcal{N}(0, \mathbf{\Sigma_x})$, so that

$$\bar{x} = \frac{1}{N}\sum_i (s_i + \xi_i) = \bar{s} + \bar{\xi}.$$

In this case, the variance of the mean noise scales with increasing noise correlation strength, ρ_x:

$$\mathrm{Var}(\bar{\xi}) = \frac{\sigma_x^2(1-\rho_x)}{N} + \rho_x\sigma_x^2.$$

As the variance increases, the overlap between the conditional distributions $p(\bar{x}|\bar{s} > 0)$ and $p(\bar{x}|\bar{s} < 0)$ increases. It is therefore more difficult to tell which condition the measurement comes from, and thus the performance deteriorates.

Acknowledgments

K.J. was supported by NSF award DMS-1122094. W.J.M. was supported by award R01EY020958 from the National Eye Institute and award number W911NF-12-1-0262 from the Army Research Office.

References

Abbott, L., & Dayan, P. (1999). The effect of correlated variability on the accuracy of a population code. *Neural Computation*, *11*(1), 91–101.

Averbeck, B. B. (2009). Noise correlations and information encoding and decoding. In K. Josić, J. Rubin, M. Matias, & R. Romo (Eds.), *Coherent behavior in neuronal networks* (pp. 207–228). New York: Springer.

Averbeck, B. B., Latham, P. E., & Pouget, A. (2006). Neural correlations, population coding and computation. *Nature Neuroscience Reviews*, *7*(5), 358–366.

Downloaded from http://direct.mit.edu/neco/article-pdf/27/11/2318/952316/neco_a_00778.pdf by guest on 16 May 2021

Baldassi, S., & Burr, D. C. (2000). Feature-based integration of orientation signals in visual search. *Vision Research*, *40*(10), 1293–1300.

Baldassi, S., & Verghese, P. (2002). Comparing integration rules in visual search. *Journal of Vision*, *2*(8), 3.

Beck, J., Bejjanki, V., & Pouget, A. (2011). Insights from a simple expression for linear Fisher information in a recurrently connected population of spiking neurons. *Neural Computation*, *23*, 1–19.

Beck, J. M., Ma, W. J., Pitkow, X., Latham, P. E., & Pouget, A. (2012). Not noisy, just wrong: The role of suboptimal inference in behavioral variability. *Neuron*, *74*(1), 30–39.

Berens, P., Ecker, A., Gerwinn, S., Tolias, A., & Bethge, M. (2011). Reassessing optimal neural population codes with neurometric functions. *Proceedings of the National Academy of Sciences*, *108*, 4423–4428.

Bhardwaj, M., van den Berg, R., Ma, W., & Josić, K. (2015). *Do people take stimulus correlations into account in visual search?* Manuscript submitted for publication.

Brady, T. F., & Alvarez, G. A. (2011). Hierarchical encoding in visual working memory ensemble statistics bias memory for individual items. *Psychological Science*, *22*, 384–392.

Brady, T. F., & Tenenbaum, J. B. (2013). A probabilistic model of visual working memory: Incorporating higher order regularities into working memory capacity estimates. *Psychological Review*, *120*(1), 85.

Chelaru, M., & Dragoi, V. (2008). Efficient coding in heterogeneous neuronal populations. *Proceedings of the National Academy of Sciences*, *105*(42), 16344–16349.

Chen, Y., Geisler, W., & Seidemann, E. (2006). Optimal decoding of correlated neural population responses in the primate visual cortex. *Nature Neuroscience*, *9*(11), 1412–1420.

Cohen, M., & Kohn, A. (2011). Measuring and interpreting neuronal correlations. *Nature Neuroscience*, *14*(7), 811–819.

de la Rocha, J., Doiron, B., Shea-Brown, E., Josić, K., & Reyes, A. (2007). Correlation between neural spike trains increases with firing rate. *Nature*, *448*(7155), 802–806.

Ecker, A., Berens, P., Keliris, G., Bethge, M., Logothetis, N., & Tolias, A. (2010). Decorrelated neuronal firing in cortical microcircuits. *Science*, *327*(5965), 584–587.

Geisler, W. (2008). Visual perception and the statistical properties of natural scenes. *Annual Review of Psychology*, *59*, 167–192.

Geisler, W. S., & Perry, J. S. (2009). Contour statistics in natural images: Grouping across occlusions. *Visual Neuroscience*, *26*(1), 109–121.

Graham, N., Kramer, P., & Yager, D. (1987). Signal-detection models for multidimensional stimuli: Probability distributions and combination rules. *Journal of Mathematical Psychology*, *31*(4), 366–409.

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

Harville, D. A. (1997). *Matrix algebra from a statistician's perspective*. New York: Springer.

Horn, R., & Johnson, C. (2012). *Matrix analysis*. Cambridge: Cambridge University Press.

Downloaded from http://direct.mit.edu/neco/article-pdf/27/11/2318/952316/neco_a_00778.pdf by guest on 16 May 2021

Hu, Y., Zylberberg, J., & Shea-Brown, E. (2014). The sign rule and beyond: Boundary effects, flexibility, and noise correlations in neural population codes. *PLoS Computational Biology*, *10*(2), e1003469.

Josić, K., Shea-Brown, E., Doiron, B., & de la Rocha, J. (2009). Stimulus-dependent correlations and population codes. *Neural Computation*, *21*(10), 2774–2804.

Latham, P., & Nirenberg, S. (2005). Synergy, redundancy, and independence in population codes, revisited. *Journal of Neuroscience*, *25*(21), 5195–5206.

Ma, W., Navalpakkam, V., Beck, J. M., van den Berg, R., & Pouget, A. (2011). Behavior and neural basis of near-optimal visual search. *Nature Neuroscience*, *14*(6), 783–790.

Mathis, A., Herz, A. V., & Stemmler, M. B. (2013). Multiscale codes in the nervous system: The problem of noise correlations and the ambiguity of periodic scales. *Physical Review E*, *88*(2), 022713.

Mazyar, H., van den Berg, R., & Ma, W. (2012). Does precision decrease with set size? *Journal of Vision*, *12*(6), 10.

Mazyar, H., van den Berg, R., Seilheimer, R., & Ma, W. (2013). Independence is elusive: Set size effects on encoding precision in visual search. *Journal of Vision*, *13*(5), 1–14.

Meyer, C. (2000). *Matrix analysis and applied linear algebra*. Philadelphia: Society for Industrial and Applied Mathematics.

Montemurro, M. A., & Panzeri, S. (2006). Optimal tuning widths in population coding of periodic variables. *Neural Computation*, *18*(7), 1555–1576.

Moreno-Bote, R., Beck, J., Kanitscheider, I., Pitkow, X., Latham, P., & Pouget, A. (2014). Information-limiting correlations. *Nature Neuroscience*, *17*(10), 1410–1417.

Nolte, L. W., & Jaarsma, D. (1967). More on the detection of one of *M* orthogonal signals. *Journal of the Acoustical Society of America*, *41*(2), 497–505.

Palmer, J. (1990). Attentional limits on the perception and memory of visual information. *Journal of Experimental Psychology: Human Perception and Performance*, *16*(2), 332.

Palmer, J., Ames, C. T., & Lindsey, D. T. (1993). Measuring the effect of attention on simple visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *19*(1), 108.

Palmer, J., Verghese, P., & Pavel, M. (2000). The psychophysics of visual search. *Vision Research*, *40*(10), 1227–1268.

Palmer, S. (1999). *Vision science: Photons to phenomenology*. Cambridge, MA: MIT Press.

Pelli, D. G. (1985). Uncertainty explains many aspects of visual contrast detection and discrimination. *Journal of the Optimal Society of America A*, *2*(9), 1508–1531.

Perkel, D., Gerstein, G., & Moore, G. (1967). Neuronal spike trains and stochastic point processes: II. Simultaneous spike trains. *Biophysical Journal*, *7*(4), 419–440.

Peterson, W. W., Birdsall, T. G., & Fox, W. (1954). The theory of signal detectability. *Transactions of the IRE Professional Group on Information Theory*, *4*(4), 171–212.

Rosenbaum, R., & Josić, K. (2011). Mechanisms that modulate the transfer of spiking correlations. *Neural Computation*, *23*(5), 1261–1305.

Rosenbaum, R., Trousdale, J., & Josić, K. (2010). Pooling and correlated neural activity. *Frontiers in Computational Neuroscience*, *4*, 9.

Rosenholtz, R. (2001). Visual search for orientation among heterogeneous distractors: Experimental results and implications for signal-detection theory models

Downloaded from http://direct.mit.edu/neco/article-pdf/27/11/2318/952316/neco_a_00778.pdf by guest on 16 May 2021

of search. *Journal of Experimental Psychology: Human Perception and Performance*, *27*(4), 985.

Schneidman, E., Bialek, W., & Berry, M. (2003). Synergy, redundancy, and independence in population codes. *Journal of Neuroscience*, *23*(37), 11539–11553.

Shamir, M., & Sompolinsky, H. (2006). Implications of neuronal diversity on population coding. *Neural Computation*, *18*(8), 1951–1986.

Shaw, M. L. (1980). Identifying attentional and decision-making components in information processing. In R. S. Nickerson (Ed.), *Attention and performance VIII* (pp. 277–295). Mahwah, NJ: Erlbaum.

Sompolinsky, H., Yoon, H., Kang, K., & Shamir, M. (2001). Population coding in neuronal systems with correlated noise. *Physical Review E*, *64*(5), 051904.

Townsend, J. T. (1974). Issues and models concerning the processing of a finite number of inputs. In B. H. Kantowitz (Ed.), *Human information processing: Tutorials in performance and cognition* (pp. 133–168). Mahwah, NJ: Erlbaum.

van den Berg, R., Vogel, M., Josić, K., & Ma, W. (2012). Optimal inference of sameness. *Proceedings of the National Academy of Sciences*, *109*(8), 3178–3183.

Verghese, P. (2001). Visual search and attention: A signal detection theory approach. *Neuron*, *31*(4), 523–535.

Vincent, B. T., Baddeley, R. J., Troscianko, T., & Gilchrist, I. D. (2009). Optimal feature integration in visual search. *Journal of Vision*, *9*(5), 15.

Wolfe, J. M. (1998). Visual search. In H. Pashler (Ed.), *Attention*. London: University College London Press.

Zohary, E., Shadlen, M., & Newsome, W. (1994). Correlated neuronal discharge rate and its implications for psychophysical performance. *Nature*, *370*(6485), 140–143.

Received February 26, 2015; accepted June 19, 2015.

Downloaded from http://direct.mit.edu/neco/article-pdf/27/11/2318/952316/neco_a_00778.pdf by guest on 16 May 2021