



Published in final edited form as:

*J Vis.*; 6(4): 387–413. doi:10.1167/6.4.8.

## Classification images with uncertainty

**Bosco S. Tjan** and

Department of Psychology and Neuroscience Graduate Program, University of Southern California, Los Angeles, CA, USA

**Anirvan S. Nandy**

Department of Psychology, University of Southern California, Los Angeles, CA, USA

Bosco S. Tjan: btjan@usc.edu; Anirvan S. Nandy: nandy@usc.edu

### Abstract

Classification image and other similar noise-driven linear methods have found increasingly wider applications in revealing psychophysical receptive field structures or perceptual templates. These techniques are relatively easy to deploy, and the results are simple to interpret. However, being a linear technique, the utility of the classification-image method is believed to be limited. Uncertainty about the target stimuli on the part of an observer will result in a classification image that is the superposition of all possible templates for all the possible signals. In the context of a well-established uncertainty model, which pools the outputs of a large set of linear frontends with a max operator, we show analytically, in simulations, and with human experiments that the effect of intrinsic uncertainty can be limited or even eliminated by presenting a signal at a relatively high contrast in a classification-image experiment. We further argue that the subimages from different stimulus-response categories should not be combined, as is conventionally done. We show that when the signal contrast is high, the subimages from the error trials contain a clear high-contrast image that is negatively correlated with the perceptual template associated with the presented signal, relatively unaffected by uncertainty. The subimages also contain a “haze” that is of a much lower contrast and is positively correlated with the superposition of all the templates associated with the erroneous response. In the case of spatial uncertainty, we show that the spatial extent of the uncertainty can be estimated from the classification subimages. We link intrinsic uncertainty to invariance and suggest that this signal-clamped classification-image method will find general applications in uncovering the underlying representations of high-level neural and psychophysical mechanisms.

### Keywords

classification image; reverse correlation; spatial uncertainty; invariance; nonlinearity

### Introduction

If a system responds linearly to its input by correlating it with a single template (by taking the dot product), then this template can be recovered by presenting the system with samples of white noise and averaging those noise samples that led to the same response. Since the 1980s, this simple form of reverse correlation, also known as spike-triggered averaging, has been routinely applied to map the receptive fields of neurons in the early stages of the sensory

© ARVO

Corresponding author: Bosco S. Tjan. Email: btjan@usc.edu Address: Department of Psychology, 3620 South McClintock, SGM 501, University of Southern California, Los Angeles, CA 90089-1061.

Commercial relationships: none

systems (e.g., de Boer & de Jongh, 1978; de Boer & Kuypers, 1968; Jones & Palmer, 1987). When applied to visual psychophysics, where the stimulus noise is in the form of an image (or a movie), the technique is often referred to as the “classification-image method” (Ahumada 2002; Beard & Ahumada, 1999), which owes its roots to the early work of Ahumada and colleagues in auditory psychophysics (Ahumada & Lovell, 1971; Ahumada & Marken, 1975).

In recent years, classification image and similar techniques have been applied to study vernier acuity (Beard & Ahumada, 1999), stereopsis (Neri, Parker, & Blakemore, 1999), illusory-contour perception (Gold, Murray, Bennett, & Sekuler, 2000), identification of facial expression (Adolphs et al., 2005; Gosselin & Schyns, 2003), and surround effect on contrast discrimination (Shimozaki, Eckstein, & Abbey, 2005), to name a few. We can divide these applications into two broad categories. In one category, the primary goal of the investigation was to discern from where in the stimulus an observer extracts information. For example, Gold et al. (2000) found that when an observer was asked to judge whether the shape of an illusory square was “thin” or “fat,” observers often based their decision on the left and right illusory edge, while ignoring the top and bottom ones, which are equally informative. Adolphs et al.’s (2005) work, showing that a patient’s failure to use information in the eye region of faces impaired the perception of fear, is another example of this category. In the second category, the main purpose of using the classification-image method was to infer the “perceptual template” used by an observer to perform a given task. For example, Beard and Ahumada (1999) showed with classification images that vernier discrimination was mediated by an orientation-tuned mechanism, as had been previously suggested.

The method of classification image can recover a mechanism’s template if the mechanism is equivalent to a linear noisy correlator (Ahumada, 2002). Murray, Bennett, and Sekuler (2002) argued that this requirement can be relaxed to include observer models that have an additive noise whose variance is proportional to the contrast energy of the input (as opposed to being a constant) and to models with nonlinear transducer functions when tested over a narrow range (for a more precise description of the requirements regarding nonlinear transducer functions, see Neri, 2004). Even with these generalizations, the range of observer models for which the classification-image method is valid for inferring the perceptual template appears to be restricted. Physiologists have long maintained that spike-triggered averaging (identical to classification image) is of very limited use for uncovering the receptive field structures of higher order visual neurons. Various higher order techniques, such as spike-average covariance (de Ruyter van Steveninck & Bialek, 1988; Rust, Schwartz, Movshon, & Simoncelli, 2004, 2005), are used to augment spike-triggered averaging. Neri and Heeger (2002) recently extended the classification-image method to include the analysis of covariance. Despite its theoretical limitation, the linear version of the classification-image method has been applied to increasingly complex visual tasks, such as face recognition and object categorization, yielding intriguing results.

A simple yet ubiquitous form of nonlinearity generally believed to pose a severe problem to the method of classification image is uncertainty. Murray et al., (2002) described this problem succinctly:

One type of nonlinearity that does pose a problem for the noisy cross-correlator<sup>1</sup> model is stimulus uncertainty. Even when observers are told the exact shape and location of the signals that they are to discriminate between, they sometimes behave as if they are uncertain as to exactly where the stimulus will appear or what shape it will take (e.g., Manjeshwar & Wilson, 2001; Pelli, 1985). We can model spatial uncertainty by assuming that the observer has many identical templates that he applies over a range of spatial locations in the stimulus, but the effects of this operation are complex, and it is not obvious precisely how a classification image is related to the

template of such an observer, or how the SNR of the classification image is related to quantities such as the observer's performance level or internal-to-external noise ratio. If an observer is very uncertain about some stimulus properties, such as the phase of a grating signal, a response classification experiment may produce no classification image at all (Ahumada & Beard, 1999).

This problem is more serious because of the equivalence between feature invariance and intrinsic uncertainty (an uncertainty internal to the observer, as opposed to that in the stimuli, or extrinsic uncertainty), which we shall explain next.

Visual processing entails the extraction of "features" from retinal inputs that are relevant to behavior. In theories of object perception, the degree of invariance that a feature possesses is a central issue. Biederman (1987) and Marr (1982), for example, viewed visual processing as a stage-wise process designed to recover, from retinal images, non-accidental features of increasing complexity and invariance (edges, contours, corners, simple volumes, and structural description of volumes). For example, an edge feature is invariant to local contrast and immune to changes in local illumination; a volumetric feature is invariant not only to local and global illumination but also to the observer's viewpoint. All theories of object recognition involve invariance but differ in the degree of invariance they rely on to make the final determination of object identity (cf. Tjan, 2002; Tjan & Legge, 1998).

Consider a detector that signals the presence of a particular feature (e.g., an edge) while ignoring the specific image properties that the feature was rendered with (e.g., the colors across the edge). It is as if the detector is obligatorily considering all possible versions of the feature (e.g., white-black edge, white-gray edge, red-green edge, etc.). Such a feature detector will exhibit an amount of intrinsic uncertainty, equal to the effective number of orthogonal instances in the equivalent set of the input images that lead to the same response. The notions of "invariance" and "uncertainty," albeit different in their historical and theoretical origins, are therefore the same.

If the method of classification image indeed could not handle uncertainty, it would be of limited use as a tool to reveal the mechanisms of vision, which undoubtedly involve invariance. Limitations imposed by uncertainty have been noted and partially addressed in the past. Using a vernier-offset detection task, Barth, Beard and Ahumada (1999) rejected the linear observer model (one that does not have uncertainty) by showing a significant discrepancy between classification images from human observers and those from a linear model, when the classification images from the offset-present and offset-absent trials were considered separately. They also estimated the amount of positional and orientation uncertainty in the human observers by explicitly modeling uncertainty as a small Gaussian weighting function. Solomon (2002) likewise pointed out that for a yes-no signal detection task, any difference in the shapes of templates estimated from target-present trials as compared with target-absent trials may be due to uncertainty. Abbey and Eckstein (2002) extended this observation to 2AFC tasks and provided a statistical test for using classification images to detect the presence of observer nonlinearities. Although these studies showed how observer nonlinearity, such as uncertainty, could be detected from classification images, they did not provide a general method for template estimation in the face of large uncertainty. Eckstein, Shimozaki, and

<sup>1</sup>In vision, and particularly in the context of signal-detection theory, the term "cross-correlation" has often been used, at least since the 1970s, to refer to a dot product, or in the functional form:  $s(x,y) \cdot t(x,y) = \iint s(x,y)t(x,y)dx dy$ . Hence, the result of a cross-correlation is a scalar. Yet, in mathematics, cross-correlation is defined as a convolution with a flipped and conjugated kernel. For a real function of two dimensions, this is  $s(x,y) \otimes t(x,y) = \iint s(x-u,y-v) t(-u,-v) du dv$ . Hence, the result is not a scalar but a function of  $(x,y)$ . This confusion is particularly unfortunate when we try to describe a mechanism that is shift invariant (i.e., the mechanism is a cross-correlation in the second but not the first sense of the term). Murray et al. (2002) used the term in the first sense (dot product). In this paper, we will avoid the use of the term cross-correlation all together. We will use "correlation" when referring to a dot product and describe cross-correlation in terms of convolution.

Abbey (2002) went one step further to show that, at least for a small amount of positional uncertainty in the task (two possible positions), the classification image computed at each possible position was an unbiased estimate of the underlying templates of a Bayesian ideal observer.

The goal of this paper is to show that with a slight modification to the current practice, the method of classification image *is* generally applicable even when the task, the visual system, or both possess a great deal of uncertainty (and invariance). This is achieved by understanding the role that a signal plays in a classification-image experiment in the context of a well-established uncertainty model (cf. Pelli, 1985), first proposed by Tanner (1961). Specifically, we will demonstrate the theoretical feasibility and empirical practicality of recovering the perceptual templates of an observer for tasks with a high degree of spatial uncertainty. We will also demonstrate how the degree of uncertainty may be estimated from the resulting classification images.

## Overview

The rest of this paper is organized as follows. In the first half of the paper, we will explore the theoretical under-pinnings that allow the use of classification-image methods in conditions with high uncertainty. We will illustrate the various aspects of our proposed method analytically and by means of simulations using an ideal-observer model for which we know the ground truth about the observer's templates. In the second half of the paper, we will demonstrate the practicality of our method via three sets of experiments with human observers. Experiments 1 and 2 will show that we can uncover, within a reasonable number of trials, the perceptual templates for letter identification and detection tasks in conditions with varying degrees of spatial uncertainty. We will also show that the degree of uncertainty can be estimated from the classification images. In Experiment 3, we will demonstrate the potential of our method by using it to measure both the quality of the perceptual templates and the amount of intrinsic spatial uncertainty in human peripheral vision.

## Theory

We first consider an ideal observer for identifying known patterns in additive Gaussian noise (Tjan, Braje, Legge, & Kersten, 1995; Tjan & Legge, 1998). An ideal observer is a theoretically optimal decision mechanism for a given task and its stimuli. Strictly speaking, an ideal observer is not a model of any actual observer. Its formulation is completely determined by the given task and its stimuli. An ideal observer establishes the upper bound of the level of performance achievable by any observer, biological or otherwise, and often provides a good starting point for modeling human observers.

A typical task used in a classification-image experiment is to discriminate between two patterns embedded in additive Gaussian white noise. A detection task is a special case of this, where one of the patterns is a blank (noise-only) display. For each of the two patterns, there may be one or more instances. Consider for example a task to identify if the noisy stimulus contains the letter "O" or "X." A single-instance version of this task is one where there is only one version of "X" and one version of "O." For a single-instance task, the signal for each response is known exactly, and there is no stimulus uncertainty. Stimulus uncertainty is introduced when different image patterns are to be associated with the same response—for example, in a multiple-instance version of the task, the letters may appear in different fonts, sizes, or positions.

Let  $T_{r,j}$  be the  $j$ th version of a noise-free contrast pattern with a response label  $r$ . (Unless the context suggests otherwise, we generally present a 2-D pattern as a column vector by concatenating all columns of an image into a single column.) Let  $N_\sigma$  be a sample of a Gaussian

white noise (a multinomial normal distribution of zero mean and diagonal covariance  $\sigma^2 I$ ). A noisy stimulus with a signal contrast of  $c$  is,

$$I=cS+N_{\sigma}, S \in \{T_{r,j}\}. \quad (1)$$

The general form of the ideal observer for identifying the embedded pattern in  $I$  with maximum accuracy is to select the response label  $r$  that maximizes the posterior probability (Duda & Hart, 1973; Green & Swets, 1974; Peterson, Birdsall, & Fox, 1954). That is,

$$r=\arg \max _r \Pr (r|I)=\arg \max _r \sum _j \Pr (r, j|I). \quad (2)$$

The summation over  $j$  (marginalization) in the second expression follows strictly from probability theory because the occurrence of the different versions of a pattern is mutually exclusive in a single presentation.

Assuming that all patterns are equally likely to occur, by applying the Bayes theorem and the probability density function (p.d.f.) of a normal distribution and by collecting into a constant the terms that do not vary with either  $r$  or  $j$ , we have:

$$\begin{aligned} \Pr (r|I) &= \sum _j \Pr (r, j|I) \\ &= \sum _{j=1}^M \frac{\Pr (I|T_{r,j}) \Pr (T_{r,j})}{\Pr (I)} = k \sum _{j=1}^M \Pr (I|T_{r,j}) \\ &= k' \sum _{j=1}^M \exp \left( -\frac{\|I - cT_{r,j}\|^2}{2\sigma^2} \right) \\ &= k'' \sum _{j=1}^M \exp \left( \frac{2I^T cT_{r,j} - c^2 T_{r,j}^T T_{r,j}}{2\sigma^2} \right), \end{aligned} \quad (3)$$

where  $M$  is the number of distinct patterns with the same response label, the  $k$ s are constants, and the superscript  $T$  denotes matrix transpose. We note that  $I^T I$  does not vary with either  $r$  or  $j$  and has therefore been treated as a constant.

Equation 3 provides us with the optimal decision rule for pattern identification with or without stimulus uncertainty. The optimal decision rule is to choose the response  $r$  that maximizes a univariate decision variable  $\lambda(r)$ :

$$\lambda(r)=\sum _{j=1}^M \exp \left( \frac{2I^T cT_{r,j} - c^2 T_{r,j}^T T_{r,j}}{2\sigma^2} \right). \quad (4)$$

An appendix in Tjan and Legge (1998) provides a computationally efficient way of implementing this decision mechanism when the stimulus uncertainty ( $M$ ) is large (in the tens of thousands).

Two special cases of the optimal decision rule are noteworthy. For a task where all signal patterns have the same contrast energy, the dot product  $T_{r,j}^T T_{r,j}$  is a positive constant and can be removed from the decision rule:

$$\lambda(r) = \sum_{j=1}^M \exp\left(\frac{I^T c T_{r,j}}{\sigma^2}\right). \quad (5)$$

For a task with equal-energy signals and no stimulus uncertainty ( $M = 1$ ), the optimal decision rule can be further reduced to that of a linear correlator by taking advantage of the fact that the exponential function is monotonically increasing and by removing all constant terms:

$$\lambda(r) = I^T T_r. \quad (6)$$

What we have shown is that the popular linear observer model (Equation 6), which makes a decision by linearly correlating the input with a template, is the optimal decision mechanism when there is no stimulus uncertainty and when the stimulus noise is white, a well-known result that is worth reiterating. To maintain optimality under these conditions, it is not necessary to know either the signal contrast ( $c$ ) or the noise variance ( $\sigma^2$ ). The assumption of a linear observer is the singularly most important assumption for the classification-image method (Ahumada, 2002; Murray et al., 2002). Also evident from our derivation is the reason why uncertainty presents a significant challenge to the classification-image method—because there are no apparent means of approximating the optimal decision variable of Equation 5 to something similar to Equation 6.

### The uncertainty model

When stimulus uncertainty is due to the task (multiple input patterns are to be associated to the same response per task requirement), such uncertainty is often referred to as “extrinsic uncertainty” because it is external to an observer. With extrinsic uncertainty, Equation 4 or 5 is the optimal decision rule if the equal contrast energy condition is met. Extrinsic uncertainty is contrasted with “intrinsic uncertainty,” which refers to the uncertainty assumed by the observer. For example, in a letter identification task where there is only one instance of “X” and one instance of “O”, observers may still insist on considering different versions of the letters during a trial either because they lack the precision for encoding certain attributes of the instances (e.g., the exact stimulus size or position) or because they are misinformed about the task. With intrinsic uncertainty, Equation 4 or 5 becomes an ideal-observer *model*<sup>2</sup> of the observer. When  $M$  in Equation 4 or 5 is greater than 1, the decision rule would be suboptimal for the *task*, which has no uncertainty, but it is optimal for the *observer* with the explicit limitation that the observer had assumed that there was uncertainty in the task. Tanner (1961) pointed out that if an observer did not know the signal exactly and had to consider a number of possibilities, the observer, which could be otherwise ideal, would have a steeper psychometric function compared with that of an ideal observer. Early studies in audition (cf. Green, 1964) and vision (e.g., Foley & Legge, 1981; Nachmias & Sansbury, 1974; Stromeyer & Klein, 1974; Tanner & Swets, 1954) found that when a subject was asked to detect a faint but precisely defined signal, the resulting psychometric function had a slope consistent with the presence of a significant intrinsic uncertainty.

In a seminal paper, Pelli (1985) made the case that intrinsic uncertainty could account for a large range of psychophysical data related to contrast detection and discrimination. Pelli

---

<sup>2</sup>We make the distinction between an ideal observer and an ideal-observer model. Whereas an ideal observer is defined solely with respect to a given task and its stimuli, an ideal-observer model, in contrast, includes also the assumed limitations of a human observer. That is, an ideal observer is optimal with respect to a task and its stimuli; an ideal-observer model, on the other hand, is optimal with respect to the task, stimuli, and the explicitly assumed limitations of the human observer.



demonstrated that a simple model of intrinsic uncertainty, which was already quite popular at the time of his writing but with properties not well understood, provided an excellent fit to psychophysical data for contrast detection and discrimination in many different conditions. In a nutshell, the uncertainty model makes a decision based on a decision variable of the form:

$$\lambda(r) = \max_j (I^T T_{r,j}). \quad (7)$$

The model essentially says that the observer selects a response associated with the “loudest” channel. With hindsight, it is not difficult to see why Equation 7 is a reasonable approximation to the optimal decision rule (Equation 5):

$$\begin{aligned} \lambda(r) &= \sum_{j=1}^M \exp\left(\frac{I^T c T_{r,j}}{\sigma^2}\right) \\ &\approx \max_{j \in [1, M]} \exp\left(\frac{I^T c T_{r,j}}{\sigma^2}\right) \\ &\leftrightarrow \max_{j \in [1, M]} \left(\frac{I^T c T_{r,j}}{\sigma^2}\right) \leftrightarrow \max_{j \in [1, M]} (I^T T_{r,j}). \end{aligned} \quad (8)$$

We use  $\leftrightarrow$  to indicate that two functions are monotonically related such that replacing one with the other does not affect the rank order of the values of the function. The only approximation in Equation 8 is the replacement of the sum of a set of exponentials by the largest value from the set (Equations 12 and 13 in Nolte & Jaarsma, 1967). This approximation is reasonable if the largest value is very large relative to the other values to be summed, as is often the case with an exponential function.

The uncertainty model (Equation 7) is the key theoretical foundation that led to our proposed method for obtaining a classification image in the face of uncertainty. The results of Pelli (1985) showing the general validity of this model to a large set of empirical data and the rather ubiquitous applications of the model in visual psychophysics justified this starting point. Nevertheless, we note that our approach does not depend on any subtle assumptions of the uncertainty model beyond Equation 7 and is theoretically robust.

### Isolating a channel in the uncertainty model by using a signal

If there is no uncertainty and if the linear observer model (Equation 6) is a good approximation of an actual observer, then it has been established that the classification-image method could uncover the observer templates  $T_r$  (cf. Ahumada, 2002). The same, however, could not be said when there is significant extrinsic or intrinsic uncertainty.

An inherent property of the uncertainty model (Equation 7) offers a way to reduce or eliminate intrinsic uncertainty and thus reduces the uncertainty model to a linear observer model. Because the channel with the highest response drives the net output of the uncertainty model, the presence of a relatively strong signal in the noisy stimulus will bias one channel over the others in terms of its contribution to the observer’s response. When the observer made an incorrect response *while the signal is present*, we know with relative certainty that it was that channel that often responded maximally to the signal that was suppressed. The linear kernel associated with this channel can then be recovered using the conventional classification-image technique.

We can illustrate this logic more precisely by combining Equation 7 with the definition of the stimulus (Equation 1):

$$\begin{aligned}
\lambda(r) &= \max_{j \in [1, M]} (I^T T_{r,j}) \\
&= \max_{j \in [1, M]} (cS^T T_{r,j} + N_\sigma^T T_{r,j}) \\
&\approx (cS^T T_{r,z} + N_\sigma^T T_{r,z}) = I^T T_{r,z},
\end{aligned} \tag{9}$$

where we let  $T_{r,z}$ ,  $z \in [1, M]$  denote the channel that has the highest response for signal  $S$ . The last line of approximation is justified because (1) for equal-energy signals,  $N_\sigma^T T_{r,j}$  is statistically identical for all channels  $j$ , and (2) the term  $S^T T_{r,z}$  leads one particular channel to have the highest response most of the time and thus to single-handedly drive the decision variable  $\lambda(r)$ . What is critical for this approximation is that the response  $S^T T_{r,z}$  must be significantly larger than the responses from the other channels. We refer to this requirement as the “signal-clamping” requirement and the approximation in Equation 9 as the signal-clamping approximation.

In short, we are using a fixed signal to hold on to a specific channel and a varying noise to map the linear kernel of the channel. We refer to this approach as the signal-clamped classification-image method. Our logic is essentially the same as the two-bar method by Movshon, Thompson, and Tolhurst (1978) for mapping the linear component of the receptive field of a complex cell. We can think of a complex cell as an observer with uncertainty in the phase of a grating and is approximately equivalent to a detector that does a max-response pooling from a large set of detectors, each selective to a specific phase.<sup>3</sup> With this perspective, we can think of the two-bar method as using one bar to select a channel of a specific phase, and the other bar, with varying positions relative to the first bar, to map the receptive field of the selected channel.

### Properties of signal-clamped classification images

Signal-clamped classification images have distinct properties that can be exploited to estimate the amount of intrinsic uncertainty (or equivalently, the degree of invariance) of an observer. We will illustrate these properties, first analytically and then by simulation using an ideal-observer model (Equation 4), for which we know the ground truth about the observer’s internal templates and the amount of intrinsic uncertainty. We used an ideal-observer model in the simulation instead of the uncertainty model (Equation 7) to show that the analytical properties of signal clamping derived from the uncertainty model does not depend on the absolute validity of the uncertainty model. Whether these properties are valid for a human observer is an empirical question. The three human experiments in the second half of this paper will confirm that these analytical properties of signal-clamped classification images are indeed valid and robust.

**Contrast of signal-clamped classification images**—For the rest of this paper, we will consider a two-letter identification task (“O” vs. “X”) and a single-letter detection task (detecting “O” against a noisy background). We restrict the form of uncertainty to the uncertainty about the location of the stimulus on the display. The templates (or channels) for a given response are shifted versions of one another but are otherwise identical; that is,

$$T_{r,j} = \text{shift}(T_r, p_j), \tag{10}$$

<sup>3</sup>The more conventional model for a complex cell is to sum the square of the outputs of a quadrature pair (a contrast energy model). Here, we are making a qualitative point and thus are ignoring the quantitative difference between the energy model and the uncertainty model.



where  $T_r$  is the position-normalized template for response  $r$  and  $p_j$  is a position on the display. Our goals are to recover  $T_r$  and the range of  $p_j$ . Possible generalizations of the signal-clamping technique to other types of uncertainty beyond that of shift invariance will be addressed in the General discussions section.

A conventional classification image is a composition of a set of classification subimages. A subimage  $CI_{AB}$  is the average of all the noise patterns  $N_\sigma$  (Equation 1) from trials where the signal in the stimulus was  $A$  and the observer's response was  $B$ . Consider the two-letter identification task ("O" vs. "X"). The subimage  $CI_{OX}$  is the average of the noise patterns  $N_{OX}$  from trials where "O" was in the stimulus but the observer responded "X" (we refer to this as an OX trial). An "X" response implies that the internal decision variable for an "X" response was greater than that for an "O" response; that is,  $\lambda(\text{"X"}) > \lambda(\text{"O"})$ . Appealing to the uncertainty model (Equation 7) and the composition of a stimulus (Equation 1) and letting  $X_j = T_{x,j}$  and  $O_j = T_{o,j}$  to improve readability, we have

$$\begin{aligned} \lambda(\text{"X"}) > \lambda(\text{"O"}) \\ \iff \\ \max_{j \in [1, M]} (cO^T X_j + N_{ox}^T X_j) > \max_{j \in [1, M]} (cO^T O_j + N_{ox}^T O_j), \end{aligned} \quad (11)$$

where  $O$  (without any subscript) is the "O" signal in the noisy stimulus presented to the observer. If there is no uncertainty ( $M = 1$ ), Equation 11 becomes the familiar form that underlies the conventional classification image:

$$\begin{aligned} cO^T X + N_{ox}^T X > cO^T O + N_{ox}^T O \\ \Rightarrow N_{ox}^T (X - O) > (O^T O - O^T X)c. \end{aligned} \quad (12)$$

The right-hand side of the inequality is a positive number because a noiseless "O" stimulus will activate the "O" channel ( $O_1$ ) more than the "X" channel ( $X_1$ ); that is,  $O^T O_1 > O^T X_1$ . For this inequality to hold, the average noise pattern on the left-hand side must have a positive correlation with the X template and a negative correlation with the O template. Ahumada (2002) showed analytically that

$$E[N_{ox}] \propto (X_1 - O_1), \quad (13)$$

where  $E[\cdot]$  denotes a mathematical expectation (see also Abbey & Eckstein, 2002; Murray et al., 2002). The proportional constant is affected by the probability of an OX trial (stimulus "O", response "X"), and the internal-to-external noise ratio (ratio between the variances of the noise internal to an observer and that in the stimuli; e.g., see Equation A3 in Murray et al., 2002).  $CI_{OX}$  approaches  $E[N_{OX}]$  as the number of OX trials ( $N_{OX}$ ) approaches infinity. For a finite number of trials, the variance of  $CI_{OX}$  is rather cumbersome because the probability density of  $CI_{OX}$  is a truncated version of the multidimensional Gaussian ( $N_\sigma$ ) used to form the stimuli. Ahumada (2002) pointed out that the variance of  $CI_{OX}$  is upper bounded by the variance of the nontruncated distribution. Murray et al. (2002, Appendices A and F) further argued that the difference between the upper bound and the actual variance is negligible for a typical classification-image experiment where (1) the amount of the stimulus noise is comparable to the level of the observer's internal noise, (2) the number of independent image pixels (and hence the dimensionality of stimulus) is large, and (3) the accuracy level is above 75%. All of the experiments in the current study met these three conditions. Thus,  $CI_{OX}$  can be approximated as

$$CI_{OX} \approx E[N_{OX}] + N_{\sigma} / \sqrt{n_{OX}}, \quad (14)$$

where  $N_{\sigma}$  is a sample of white noise from the distribution used to form the stimuli (Equation 1).

Equations 13 and 14 show that in a conventional classification-image experiment, where a great deal of effort is directed toward the elimination of uncertainty in the experiment, each classification subimage contains both a positive image of one template and a negative image of the alternative template. In the case of an error trial, the negative image is the template for the presented signal, and the positive image is the template associated with the response.

Now consider a condition where there is no extrinsic uncertainty (i.e., the  $X$ s, and  $O$ s were always presented at the same position on the display) but with a significant amount of intrinsic uncertainty ( $M \gg 1$ ). Applying the signal-clamping approximation (Equation 9) to the right-hand side of Equation 11, we have

$$\begin{aligned} \lambda(\text{"X"}) > \lambda(\text{"O"}) \\ \iff \\ \max_{j \in \{1, M\}} (cO^T X_j + N_{OX}^T X_j) > cO^T O_z + N_{OX}^T O_z. \end{aligned} \quad (15)$$

The signal-clamping approximation applies only to  $\lambda(\text{"O"})$  because the "O" signal in the stimulus consistently biases one particular "O" channel ( $O_z$  in the equation). There is no such trial-to-trial consistency among the "X" channels because none of them are tuned to the "O" signal. Hence, the signal-clamping approximation does not apply to  $\lambda(\text{"X"})$ . Following the logic of Ahumada (2002), we can show that (Appendix)

$$E[N_{OX}] \propto (E[X_j] - O_z). \quad (16)$$

Furthermore, the relationship between the expected value of the noise ( $E[N_{OX}]$ ) and the classification subimage ( $CI_{OX}$ ) remains the same as stated in Equation 14.

Equation 16 shows that the average of the noise patterns of the error trials contains a negative image of exactly one of the many templates for the presented signal. Important for our purpose is that this negative image is not affected by uncertainty and thus provides a good estimate of the unknown template. This is due to the signal-clamping approximation applied to the right-hand side of Equation 11. That is, the presence of a relatively strong "O" signal in the stimulus biased the signal response to precisely one of the many "O" channels; when the observer made an error and responded "X", we are relatively certain that the noise pattern suppressed the particular "O" channel ( $O_z$  in Equations 15 and 16) that would otherwise be responding.

Critically, the signal-clamping approximation is applicable only when the signal contrast in the noisy stimulus is sufficiently strong. The disadvantage of this requirement is that when the signal contrast is high, the number of the error trials, which is more informative than the correct trials, will be low, and the average of the noise patterns from the error trials will have a high variance (Equation 14). Hence, the contrast of the signal must be sufficiently high but not too high. As will be shown in the simulations and with human data, a contrast that achieves an accuracy of 75% correction reaches this balance.

Unlike the negative image, the positive image in the average noise pattern is severely affected by uncertainty. This positive image ( $E[X_j]$  of Equation 16) corresponds to the average of all

the channels associated with the response (“X” in our example). As a result, there will not be any clear positive image in the classification subimages when there is significant intrinsic uncertainty. The clarity of the positive image provides a way to estimate the degree of uncertainty.

**Estimation of spatial uncertainty**—In the case of spatial uncertainty, the channels (or templates) are assumed to be shifted versions of one another (Equation 10). If we represent the spatial distribution of the channels with an image  $S$ , with each pixel corresponding to a location in the image and the pixel value representing the probability of a channel at the location responding erroneously to noise, then

$$E[X_j] = X_z * S, \quad (17)$$

where  $*$  denotes a convolution and  $X_z$  is the position-normalized template for “X”. Combining Equations 16 and 17, we have

$$E[N_{OX}] \propto (X_z * S - O_z). \quad (18)$$

If  $S$  can be parameterized with a small number of parameters (e.g.,  $S$  being a square region with uniform distribution), then Equation 18 provides a way to estimate both the perceptual templates and the amount of spatial uncertainty. We can obtain these estimates in stages. The classification subimage  $CI_{OX}$ , which, in the limit, approaches  $E[N_{OX}]$ , contains a negative image of the “O” template, unaffected by uncertainty. Likewise, the subimage  $CI_{XO}$  provides a direct estimate of the “X” template. Knowing both the “O” and “X” templates, Equation 17 and the corresponding equation for  $E[N_{XO}]$  can be used to estimate the spatial uncertainty  $S$ .

In practice, the estimation of the templates is never precise and methods for removing the noise term in the sub-images tend to introduce various idiosyncratic artifacts. Fortunately, we will show with simulation that estimation of the spatial uncertainty  $S$  appears robust, particularly if it can be parameterized with very few parameters.

**Classification images with extrinsic uncertainty**—So far, we have assumed that there is no spatial uncertainty in the experiment, and the only uncertainty is intrinsic to the observer. In this case, the presentation of an “O” signal at a fixed location will most likely elicit a response from one particular “O” channel. The classification subimages (e.g.,  $CI_{OX}$ ) can be calculated by averaging the noise patterns in the conventional matter:

$$CI_{OX} = \frac{1}{n_{OX}} \sum_i N_{OXi}. \quad (19)$$

We can obtain a clear template despite the spatial uncertainty caused by having a relatively strong signal at a fixed position. No special operation is needed to reconstruct the classification images.

With a small but important modification, Equations 16 and 18 will hold even when the spatial uncertainty is both in the stimuli and the observer. Such a condition arises in experiments when we want to test a shift-invariant observer by using signals whose positions vary from trial to trial. The modification is to simply shift the noise pattern (with wraparound) by an amount that either recenters the signal with respect to the image or otherwise normalizes its spatial position. That is, if a stimulus at trial  $i$  was created by shifting the stimulus  $O_1$  by an amount  $p_i$ :

$$O = \text{shift}(O_1, p_i) + N_{OX_i}, \quad (20)$$

then we will replace  $N_{OX}$  in all of the preceding equations with a shifted version  ${}^S N_{OX}$ , where

$$\begin{aligned} {}^S N_{OX_i} &= \text{shift}(N_{OX_i}, -p_i), \text{ and} \\ CI_{OX} &= \frac{1}{n_{OX}} \sum_i {}^S N_{OX_i} \\ &= \frac{1}{n_{OX}} \sum_i \text{shift}(N_{OX_i}, -p_i). \end{aligned} \quad (21)$$

This modification is valid under the assumption that the templates for a given response are shifted versions of one another (Equation 10).

**Simulations**—To illustrate the various properties of the signal-clamped classification images, we consider an observer model that is otherwise optimal except for two limitations: (1) it uses templates that are slightly different from the presented signal and (2) it may have a high degree of intrinsic spatial uncertainty—spatial uncertainty that is not present in the stimuli but nevertheless assumed by the observer. The decision rule for such an ideal-observer model is given by Equation 5. We assume that there is no internal noise in the ideal-observer model. The presence of internal noise before or after the template comparison stage will lower the contrast of the resulting classification images without qualitatively affecting the critical properties that we are trying to illustrate. In contrast, internal noise during template matching will interact with intrinsic uncertainty and can lead to complex effects on the classification images. Template estimation by the signal-clamping method will remain robust under this type of noise; however, such noise will lead to a biased estimation of the spatial extent of an observer's intrinsic uncertainty when the method that we will be describing in Equations 23a and 23b is used.

We simulated two tasks, a two-letter identification task and a single-letter detection task. For each task, we simulated two levels of intrinsic spatial uncertainty. For each pair of conditions (task and uncertainty level), we estimated the observer templates and the amount of the intrinsic spatial uncertainty from the classification images. We also illustrated the effect of signal clamping by simulating the tasks at two different signal contrast levels, one leading to a 55% correct performance level and another to a 75% correct level.

For the letter identification task, the signals were lowercase “o” and “x” in Times New Roman font with an x-height of 21 pixels. The signals were always presented at the center of a  $128 \times 128$  pixel image. The ideal-observer model used lowercase “p” and “k” from the same font and size as its templates for “o” and “x”, respectively. In the case of no uncertainty ( $M = 1$ ), the templates were positioned to have the maximum overlap with the signal. In case of high spatial uncertainty, the center position of a template is uniformly distributed within the center  $64 \times 64$  pixels of the image. There were 1,000 spatially shifted templates for each response ( $M = 1,000$ ). The relative positions of the signals and the templates are shown in Figure 1a. For each trial, the observer model made a decision according to Equation 5. The external noise had a variance of  $1/16$  ( $\sigma = 0.25$ ), identical to that used in the human experiments. The signal contrast was set to a level to obtain an accuracy of 55% correct (low contrast) or 75% correct (high contrast). The observer model was assumed to know the signal contrast (parameter  $c$  in Equation 5).

Figure 1b shows the four sets of classification sub-images from these four simulated conditions. Consider the high signal contrast conditions (middle column). When there was no spatial

uncertainty (first row), the subimages contain an equal portion of both a positive and a negative image of the two templates (“p” and “k”) used by the observer model. As predicted by Equation 13, these are the templates of the observer and not the presented stimuli. Compare these subimages to the ones obtained with high degree of spatial uncertainty (second row, middle column). As predicted by Equation 16, only one clear template is visible in each subimage. Specifically, for the trials where the signal was “o” and the response was “x”, the classification subimage  $CI_{OX}$  contains a clear negative image of the template for the “o” response, which in this case was the letter “p”—the template we built into the ideal-observer model. Remarkably, this image is sharp and unaffected by the high degree of intrinsic uncertainty. This is the main result of the signal-clamping technique. Also, as predicted by Equation 16, there is no clear positive template in  $CI_{OX}$ , which is the single most important difference between the two uncertainty levels ( $M = 1$  vs.  $M = 1,000$ ). It is important to reiterate the point that the negative image in  $CI_{OX}$  resembles the observer’s template “p” and not the signal “o” that was presented. The “o” signal biased a “p” template at a particular location, allowing the effect of noise on that particular template to accumulate over all the error trials when the presented signal was “o”. The effect of the noise was on the nonzero regions of the biased template, although these regions may not overlap with the signal (e.g., the descender of the lowercase “p”).

The signal-clamping approximation that led to Equation 16 relies on the fact that there is sufficient signal contrast in the stimulus to select a particular channel for imaging. When the signal contrast was reduced, the image quality of the signal-clamped classification images is markedly degraded (left column of Figure 1b). This is in stark contrast to the conventional classification-image method (or reverse correlation) without uncertainty. When there is no uncertainty, the overall image quality of the classification images improved with decrease in signal contrast, as is commonly observed. The improvements are due to a decrease in noise for the error-trial subimages (because the number of error trials increases) and an increase in signal for the correct-trial subimages (because with a weak signal, correct responses are often aided by coincidence with noise). With uncertainty, however, these improvements were overridden by a failure of the signal-clamping approximation, allowing the uncertainty to affect the accumulated template images and rendering the templates invisible. This effect is clearly shown in the left column, second row of Figure 1b, where signal contrast was set to a low value to achieve an accuracy of 55%.

We next turn to the estimation of the extent of the spatial uncertainty intrinsic to the observer using Equation 18. We assumed  $S$  to be a uniform square region centered in the image with  $d$  pixels on a side. Thus,

$$S_d(x,y)=\begin{cases} 1/d^2 & \text{if } |x| \leq d/2 \text{ and } |y| \leq d/2 \\ 0 & \text{otherwise.} \end{cases} \quad (22)$$

From Equations 14, 18, and 22, we have

$$\begin{aligned} CI_{OX} &\approx E[N_{OX}] + N_{\sigma} / \sqrt{n_{OX}} \\ &= k(X_z * S_d - O_z) + N_{\sigma} / \sqrt{n_{OX}}. \end{aligned} \quad (23a)$$

Likewise,

$$CI_{XO} \approx k(O_z * S_d - X_z) + N_{\sigma} / \sqrt{n_{XO}}. \quad (23b)$$

The noise terms in Equations 23a and 23b are white and can be made to have the same variance if we multiply both sides of Equations 23a and 23b by  $\sqrt{n_{OX}}$  and  $\sqrt{n_{XO}}$  respectively. If we knew

the observer's signal-clamped templates ( $O_z$  and  $X_z$ ), then  $k$ , and most importantly the extent of the spatial uncertainty  $d$ , can be estimated from the classification subimages for the error trials by minimizing the least-squared error. The right-most column of Figure 1b plots the residual sum-of-squares error for different values of  $d$  (with the value of  $k$  chosen to minimize the residual at each level of  $d$ ). The solid green curves were obtained using the veridical observer templates (lowercase "p" and "k"). The value of  $d$  at which a global minimum is achieved provides the estimate of the extent of the spatial uncertainty. The estimated values for the two levels of uncertainty are 1 and 35 pixels, respectively, and are indicated by the first character of the template label "pk". For the high-uncertainty condition, the residual landscape suggests that although the lower bound of  $d$  is well defined, the upper bound is not. In the context of this limitation, the estimated values are in good agreement with the veridical values (1 for the nouncertainty condition and 64 for the high-uncertainty condition).

The black curves and the one red curve represent the residual landscape of  $d$  computed using incorrect observer templates. Each of the three black curves was obtained with a pair of lowercase letters (except "p" and "k") that resembled the classification subimages as the presumed observer templates. The red curve was obtained with the presented signals ("o" and "x") as the presumed observer templates. Note that the values of  $d$  at the global minimum of each of these residual curves are very similar. This result demonstrates the robustness of the estimate of the spatial extent  $d$  of the underlying uncertainty, even when the observer template is not precisely known. In practice, this means that we can obtain a reasonable estimate of the spatial extent by assuming that the observer templates were identical to the presented signals.

Figure 2 shows the results of the single-letter detection task. The signal in this task is the lowercase letter "o" from the two-letter identification task. The ideal-observer model used a lowercase "e" as the template to detect the signal (Figure 2a). Two levels of intrinsic uncertainty were simulated: spatial extents with a uniform distribution of 32 (medium uncertainty) and 64 (high uncertainty) pixels on a side of a square centered on the image. For the condition with the smaller spatial extent, two types of spatial uncertainty were considered: one with a constant  $M$  for both levels of spatial extents ( $M = 1,000$ ) and another with a constant density ( $M = 1,000$  for high uncertainty,  $M = 250$  for medium uncertainty).<sup>4</sup> The telltale sign of uncertainty is evident in the classification subimages for all conditions (Figure 2b). In particular, the classification subimage from the miss trials ( $CI_{\text{miss}}$ ) shows a negative image of the observer's template (a lowercase letter "e"), whereas the subimage of the false-alarm trials ( $CI_{\text{FA}}$ ) shows only a positive haze (if there were no uncertainty, it would be a positive image of observer's template).

Performance of the ideal-observer model in the two medium-uncertainty conditions was essentially the same in terms of threshold contrast ( $C_{250}/C_{1,000} = 1.1$ ) and classification images (Figure 2b, first row, left and middle columns). This is consistent with the finding of Tjan and Legge, (1998) that there exists a task-dependent upper bound of the *effective* level of uncertainty, which can be substantially less than the highest possible level of *physical* uncertainty. With respect to our current letter detection task, this means that increasing  $M$  beyond a density of 250 possible positions per  $32 \times 32$  pixels has no consequence in performance.

For the signal-clamping approximation (Equation 9) to be exact, an observer's internal templates should be orthogonal, the signal should be strong, or both. Orthogonality is effectively reduced when the spatial extent of the templates are confined to a smaller space.

<sup>4</sup>In the simulations, the possible positions were randomly drawn with replacement. The number of unique positions for the different conditions were as follows: 870 for high uncertainty, 611 for medium uncertainty with  $M = 1,000$ , and 223 for medium uncertainty with  $M = 250$ .



That is, a randomly selected channel will tend to be in closer proximity to the channel at the stimulus position. A reduction in the spatial extent also reduced the threshold contrast for detection (by a factor of about 1.5 for the ideal-observer model). The combined effect of reduced orthogonality and reduced signal contrast was incomplete signal clamping, which resulted in the noticeable dark haze around the negative image of the observer template in  $CI_{\text{miss}}$  in both of the medium-uncertainty conditions. This dark haze was absent in the high-uncertainty condition.

The white haze in  $CI_{\text{FA}}$  is noticeably broader and fainter in the high-uncertainty condition compared with the medium-uncertainty condition.

Equations 23a and 23b were used to estimate the spatial extent ( $d$ ) of the uncertainty. Note that for a detection task, one of the templates ( $X$  in this case) is an image of zeros; that is,

$$\begin{aligned} CI_{\text{miss}} &\approx k(-O_z) + N_\sigma / \sqrt{n_{\text{miss}}} \\ CI_{\text{FA}} &\approx k(O_z * S_d) + N_\sigma / \sqrt{n_{\text{FA}}}. \end{aligned} \quad (24)$$

The residual landscape for estimating  $d$  is plotted in the second row of Figure 2b. As with the case of the letter identification simulation, the green curve represents using the veridical observer template (“e”) to perform the estimation, the red curve represents using the signal in the stimuli as the template, and the three black curves were obtained using other lowercase letters that resembled the classification subimages. Again, the values of  $d$  that minimize these residual functions are relatively independent of the assumed observer templates. The averaged estimated value of  $d$  was 14.6 pixels for the medium-uncertainty condition and 37.4 pixels for the high-uncertainty condition. Although showing the same ratio of difference as the veridical values (32 vs. 64 pixels, respectively), the estimated values are admittedly a factor of 2 less. This is probably because the simulation used only 1,000 positions within  $S_d$ , as opposed to a true uniform distribution of positions.

### Summary of the method of signal-clamped classification image

Our main finding here is that by presenting a relatively strong signal in the stimulus, the observer template for the presented signal can be imaged using the conventional classification-image method in the face of a high degree of intrinsic spatial uncertainty. We called this type of classification image obtained with a relatively strong signal embedded in the stimulus the “signal-clamped” classification image. If spatial uncertainty is extrinsic (i.e., in the stimulus), then the only minor change to the calculation of classification images is to shift the noise pattern (with wraparound) to recenter the presented signal in the image (Equation 21). How this finding may be generalized to other types of uncertainties will be addressed in the General discussions section.

We have shown analytically and with simulations the following properties of signal-clamped classification images obtained with a high degree of spatial uncertainty:

1. Each of the classification subimages from the error trials contains a clear negative image of the observer’s template for the presented signal, unaffected by spatial uncertainty intrinsic or extrinsic to the observer. However, in the presence of uncertainty, the clarity of the template image markedly deteriorates if the contrast of the presented signal is not sufficiently high. The need for a high-contrast signal goes opposite to the conventional practice of using a low-contrast signal to increase the effect of noise on the observer’s response.
2. Any positive image of the alternative template in a classification subimage for the error trials is blurred by spatial uncertainty, often rendering it indiscernible.

3. The extent to which these positive template images are blurred provides an estimate of the spatial extent of the uncertainty.
4. Because of the presence of a relatively strong signal in the stimulus, the classification subimages from the correct trials contain very little contrast and are relatively uninformative. As a result, we do not advocate combining the subimages to form a single classification image as in the conventional approach.

Our discussions have been and will continue to be focusing on the subimages from the error trials, although the general properties of signal-clamped classification images derived in this section also apply to the subimages from correct trials with merely a sign change. We ignore the correct-trial subimages for the sake of simplicity. We do not lose much because with a relatively strong signal in the stimulus, the signal-to-noise ratio (SNR) of correct-trial subimages are often quite low.

## Experiments

Three sets of human experiments were conducted to determine the practicality and utility of the signal-clamped classification-image method. Experiments 1 and 2 paralleled the simulation studies and aimed to demonstrate the feasibility of the proposed method and to empirically validate the various properties of signal-clamped classification images. Experiment 1 used the two-letter identification task, whereas Experiment 2 used the single-letter detection task. To compare the effects of spatial uncertainty, both experiments were performed in the fovea where spatial uncertainty of a human observer can be effectively manipulated with the stimulus. We introduced spatial uncertainty into the task by randomizing the signal position within a given region in the stimulus display. Knowing the actual spatial extent of the stimulus-level spatial uncertainty provides a reference for evaluating the estimated spatial extent obtained from the signal-clamped classification-image method.

Experiment 3 tested letter identification in the periphery. The visual periphery is known to have a considerable amount of intrinsic spatial uncertainty (Hess & Field, 1993; Hess & McCarthy, 1994; Levi & Klein, 1996; Levi, Klein, & Yap, 1987). No spatial uncertainty was added to the stimulus. The objective of this experiment was to demonstrate that the method of signal-clamped classification images can be used to uncover the perceptual template in the presence of spatial uncertainty and to estimate the spatial extent of the uncertainty.

## General methods

### Procedure

In the identification experiments (Experiments 1 and 3), the task was to indicate which of the two lowercase letters “o” or “x” was presented. In the detection experiments (Experiment 2), the task was to indicate whether the lowercase letter “o” was presented.

Each experiment consisted of 10 blocks with 1,050 trials per block. In each trial, a white-on-black letter was presented in a field of Gaussian white noise. The noisy stimuli (letter + noise) were presented at the fovea for Experiments 1 and 2 and at 10 deg in the inferior visual field for Experiment 3. The first 50 trials in each block were calibration trials in which the letter contrast was dynamically adjusted using the QUEST procedure (Watson & Pelli, 1983) as implemented in the Psychophysics Toolbox extension in MATLAB (Brainard, 1997; Pelli, 1997) to obtain a “calibrated” threshold letter contrast for reaching an accuracy level of 75%. The remaining 1,000 trials were divided into five subblocks of 200 trials each, and QUEST was reinitialized to the calibrated value at the beginning of each subblock. During the initial 50 calibration trials, the standard deviation of the prior distribution of the threshold value was set to 5 log units (a practically flat prior), but for each subblock, the prior was narrowed to a

standard deviation of 1 log unit. This restricted the variability of the test contrast but still allowed adequate flexibility for the procedure to adapt to the observers' continuously improving threshold levels.

For the foveal experiments, the letter size was fixed at 48 pt in Times New Roman font (x-height = 22 pixels). For the peripheral experiments, an acuity measurement was first performed for each subject, in which the subject was instructed to identify any of the 26 letters presented at a 10-deg retinal eccentricity in the inferior field. The size of the presented letter was varied using the QUEST procedure to achieve an identification accuracy of 79%. Twice the acuity size so determined was used in the main experiment.

## Stimuli

The stimulus for each trial consisted of a white-on-black letter added to a Gaussian, spectrally white noise field of  $128 \times 128$  pixels. Before being presented to the observers, each pixel of this noisy stimulus was duplicated by a factor of 2, such that four screen pixels were used to render a single pixel in the stimulus. This was done to increase the spectral density of the noise. The noise contrast was fixed at 25% rms. At a viewing distance of 105 cm, the noisy stimulus was of size 4.7 deg, and the noise has a two-sided spectral density of  $85.5 \mu\text{deg}^2$ . The mean luminance of the noisy background was  $19.8 \text{ cd/m}^2$ .

For the fovea experiments (Experiments 1 and 2), the letters were of size 0.81 deg (x-height) in visual angle. For the periphery experiment (Experiment 3), the letters were of size 0.85 deg for one subject and 1.15 deg for the other subject. The periphery letter size was 0.3 log units above the subject's letter acuity at 10 deg eccentricity. The contrast of the target letter was adjusted with a QUEST procedure as described in the Procedure section.

For the experiments with spatial uncertainty, 1,000 uniformly distributed random positions, representing the center of a presented letter, were preselected with replacement from an imaginary square centered in the noise field. The spatial extent of the spatial uncertainty was manipulated by changing the size of the imaginary square: 32 stimulus pixels on a side (i.e., 64 screen pixels because of the factor-of-2 blocking to increase noise spectral density, 1.18 deg of visual angle) for the "medium" level of uncertainty and 64 stimulus pixels (128 screen pixels, 2.37 deg of visual angle) for the "high" level of uncertainty. For the experiments without spatial uncertainty, the letter was always presented at the center of the noise field, marked by a fixation cross before and after stimulus presentation. Figure 3a depicts a noisy stimulus used in the experiment.

The stimuli were displayed in the center of a 19-in. CRT monitor (Sony Trinitron CPD-G400), and the monitor was placed at a distance of 105 cm from a subject. The monitor has 11 bits (2,048 levels) of linearly spaced contrast level. All 11 bits of the contrast levels were addressable to render the noisy stimulus for each trial. This was achieved by using a passive video attenuator (Pelli & Zhang, 1991) and a custom-built contrast calibration and control software implemented in MATLAB. Only the green channel of the monitor was used to present the stimuli.

The stimuli were presented according to the following temporal design: (1) a fixation beep immediately followed by a fixation screen for 500 ms, (2) a stimulus presentation for 250 ms, (3) a subject response period (variable) with positive feedback beep for correct trials, and (4) a 500-ms delay before onset of the next trial (see Figure 3b).

At the end of each trial, the following data were collected for the subsequent classification-image reconstruction: the center position of the target letter, the state of the pseudorandom

number generator used to produce the noise field, the identity and contrast of the presented letter, and the response of the subject.

## Subjects

Five subjects (one of the authors and four paid students at the University of Southern California who were unaware of the purpose of the study) with normal or corrected-to-normal vision participated in the experiments. All had (corrected) acuity of 20/20 in both eyes. Subjects viewed the stimuli binocularly in a dark room. Written informed consent was obtained from each subject before the commencement of data collection. Because of the monotonous nature and long duration of each experiment (approximately 8–10 hr), subjects were allowed (and encouraged) to take breaks whenever they so desired. All the subjects completed their respective experiments in three to five sessions.

## Classification-image reconstruction

For the purpose of reconstructing the classification images, the calibration trials in each experiment block were ignored. For the rest of the 10,000 trials, the noise field was first regenerated using the stored random number state. Next, the noise field was shifted with wraparound based on the stored target position information as if to recenter the presented letter (Equation 21). This shifting procedure was obviously unnecessary when there was no spatial uncertainty at the stimulus level (Experiment 3, and one condition in Experiment 1). For each trial, the recentered noise field was then classified into one of four bins based on the presented stimulus and the subjects' response. The noise fields in each bin were then averaged pixel-wise to form the corresponding classification subimages.

## Relative SNR of classifications subimages

The most practical concern in the signal-clamped classification method is whether the method would require an unreasonably large number of trials to make up for the loss in the number of error trials due to the need to use a relatively strong signal. For our experiments, as it will transpire, 10,000 trials were sufficient to obtain classification images of good quality. We sought to estimate from our data the minimum number of trials that would be needed when uncertainty is high. We did so by computing the relative SNR (rSNR; Murray et al, 2002) as a function of the number of trials; we then compared this function across different uncertainty levels.

Murray et al. (2002) defined rSNR of a classification image  $C$  as:

$$\text{rSNR} = \frac{(T'^T C)^2}{\sigma_C^2} - 1, \|T'\| = 1, \quad (25)$$

where  $T'$  is an assumed template and  $\sigma_C$  is the pixel-wise standard deviation of the image  $C$ . Murray et al. showed that the discrepancy between  $T'$  and the observer's actual template only leads to a reduction in the amplitude of rSNR by a constant factor relative to the inherent variability of a classification image, thereby making the measurement less reliable. We modified this approach to measure only the classification subimages of the error trials (e.g.,  $CI_{OX}$  and  $CI_{XO}$  for the letter identification experiment) and only the negative template images in these subimages.

For the two-letter identification task, we define rSNR as follows:

$$r\text{SNR} = \frac{(O^T \text{CI}_{OX})^2}{(O^T O)\sigma_{OX}^2} + \frac{(X^T \text{CI}_{XO})^2}{(X^T X)\sigma_{XO}^2} - 2. \quad (26)$$

Here,  $X$  and  $O$  are the presented letter stimuli. Equation 26 is applicable to the letter detection task by setting  $X$  to zero. In essence, Equation 26 measured the SNR of the pixels that overlap the negative  $O$  template in the subimage  $\text{CI}_{OX}$  and the negative  $X$  template in the subimage  $\text{CI}_{XO}$ .

## Experiment 1: Letter identification with and without spatial uncertainty

Experiment 1 was conducted in two different conditions, as was the case for the simulation study: The first condition (no uncertainty) was intended to replicate past findings without spatial uncertainty; the second condition was intended to verify our signal-clamped classification-image method and the associated theoretical claim that perceptual templates can be uncovered under conditions of spatial uncertainty.

Two subjects (A.O. and B.B.) participated in the nouncertainty condition. In this condition, the letters (“o” and “x”) were presented at fixation without any spatial uncertainty. The task was to indicate which of the two letters (“o” or “x”) was presented at each trial. The subjects were explicitly told that the letters were always centered at fixation.

Subject A.O., and a third subject, A.S.N., participated in the high-uncertainty condition in which the letter stimuli (“x” or “o”) were presented at any one of 1,000 different random positions (see General methods section). The set of random positions was chosen from within a square of  $64 \times 64$  stimulus pixel (128 screen pixels or 2.37 deg on a side) centered in the stimulus area. The extent of the spatial uncertainty was not known explicitly to the subjects (except the author A.S.N.).

## Results and discussions

Introducing a high degree of uncertainty into the stimulus elevated the contrast threshold by a factor of 1.74 on average across subjects, although the effect of uncertainty on contrast threshold is not of interest here. The left column of Figure 4 shows the classification subimages for both levels of spatial uncertainty. The results of Experiment 1 bear out the theoretical predictions described earlier, that a clear classification image showing what could be an observer’s perceptual templates can be obtained under high spatial uncertainty within a reasonable number of trials (10,000 in this case). Consider only the classification subimages from the error trials (top right— $\text{CI}_{OX}$  and bottom left— $\text{CI}_{XO}$ ). The most crucial finding is that across uncertainty conditions, there was little or no difference between the negative components of the error-trial classification subimages. This was true both within and between subjects, confirming the general validity of the signal-clamping approximation (Equations 9 and 15).

There was a subtle difference in the estimated “x” templates from the two uncertainty conditions. One stroke appeared missing in the high-uncertainty condition. With the Times New Roman font used in the experiment, the missing stroke was about one third the width of the other stroke. As a result, the lowercase “x” is not isotropic in its ability to limit spatial uncertainty. It is less able to “clamp” spatial shift of an observer’s internal template along the thicker stroke than across the thinner stroke. Shift or spatial uncertainty, along the thicker stroke blurred the image of the thinner stroke, rendered it invisible for subject A.S.N, and only partly visible for subject A.O. In other words, we do not think that the observer template for “x” changed as a function of spatial uncertainty; rather, the difference in the observed templates

was a result of imperfect signal clamping, which is not always avoidable. We will return to this issue when we consider the perceptual templates in the visual periphery in Experiment 3.

The most noticeable difference between the classification subimages obtained from the two uncertainty conditions is that for the condition without uncertainty in the stimuli (top two rows), the subimages from the error trials showed both a negative and a positive component; for the condition with a high degree of uncertainty (bottom two rows), only the negative component was apparent, with the positive component being smeared out due to the spatial uncertainty. This is predicted by Equations 18, 23a, and 23b and consistent with our simulation results.

To aid visual inspection, particularly regarding the absence of the positive components in the high-uncertainty condition, we blurred the classification subimages with a Gaussian kernel with a space constant of 1.4 stimulus pixels (right column of Figure 4). In the condition where there was no spatial uncertainty in the stimulus, the positive component in the error-trial subimages appeared considerably weaker and less defined than the negative component. Because the positive component is susceptible to uncertainty (Equations 23a and 23b), it stands to reason that there existed measurable amounts of spatial uncertainty internal to the observers. Having no uncertainty in the stimuli does not guarantee the absence of uncertainty intrinsic to an observer.

To estimate the spatial extent of the uncertainty (extrinsic and intrinsic) from the classification images, we fitted Equations 23a and 23b to the two error-trial subimages to obtain a numerical estimation of  $d$  in terms of stimulus pixels, using the lowercase stimuli as the presumed templates. As demonstrated in the simulation, the choice of the presumed templates, which may be different from the actual observer templates, does not significantly affect the estimated value of  $d$ . The residual landscapes are plotted in the middle column of Figure 4. The standard error of the estimate was determined by bootstrapping (Efron & Tibshirani, 1994). The results are summarized in Table 1. As expected, the estimated spatial extent ( $d$ ) of the combined uncertainty (extrinsic and intrinsic) was significantly higher in the high-uncertainty condition than the no-uncertainty condition. Moreover, these values are in reasonable agreement with the veridical values (1 for the no-uncertainty condition and 64 for the high-uncertainty condition).

Lastly, we sought to determine the minimum number of trials that would be required to obtain classification subimages of sufficient quality. Figure 5 plots the rSNR (Equation 26) of the error-trial classification subimages for subject A.O. as a function of the number of trials for both the no-uncertainty and high-uncertainty conditions. rSNR linearly increased as a function of the number of trials. This is expected because the pixel-wise variance of a classification subimage linearly decreases with the number of trials. What is noteworthy is that the rSNR for the high-uncertainty condition was higher than that for the no-uncertainty condition, which is opposite to the results of the ideal-observer simulation (Figure 1). This remains to be the case even when we changed Equation 26 to include both the negative and positive template images in the calculation. We will address the relationship between rSNR and uncertainty in the General discussions section.

Subjectively speaking, with 10,000 trials, both of the error-trial subimages for the no-uncertainty condition were of sufficient quality. If we use this as a standard, then we only need about 8,000 trials in the high-uncertainty condition to reach the same level of rSNR. We note with interest that although uncertainty leads to an increase in threshold, the increase in threshold, in turn, keeps in check the number of trials required for a signal-clamped classification-image experiment.



## Experiment 2: Letter detection with medium and high degree of uncertainty

The ideal-observer simulations described earlier (see Figure 2) predict that the extent of the spatial uncertainty ( $d$  as opposed to  $M$ ) can be estimated from the classification images. This prediction was tested in Experiment 2.

The task was to detect a lowercase letter “o” in noise. In each trial, the target was either presented at any one of 1,000 different random positions (see General methods section) or not presented with equal probability. Subjects were asked to indicate whether the letter was presented or not. In the medium-uncertainty condition, the total set of random positions was chosen from within a central square of  $32 \times 32$  stimulus pixels (1.18 deg). The extent of the spatial uncertainty was indicated to the subjects by means of a white rectangular bounding box that was displayed during the fixation period immediately prior to the stimulus onset. Subjects J.H. and M.J. participated in this condition.

In the high-uncertainty condition, the 1,000 different random positions were chosen from a central square of  $64 \times 64$  stimulus pixels (2.37 deg), and the extent of this uncertainty range was not explicitly indicated to the subjects. In all other respects, the high-uncertainty condition was identical to the medium-uncertainty condition. Two subjects, J.H. (who also participated in the medium-uncertainty condition) and B.B., participated in this condition.

### Results and discussions

The resulting classification images and the estimation of the spatial extent of the uncertainty are shown in Figure 6. As predicted by Equation 24 and consistent with the simulation result, a clear negative signal was visible in  $CI_{\text{miss}}$  (the subimage from the miss trials) in the high-uncertainty condition. Also, as predicted, there was no clear image of the target in  $CI_{\text{FA}}$  (the subimage from the false-alarm trials). The positive haze in  $CI_{\text{FA}}$  is not as pronounced as that in the simulation, probably due to the presence of internal noise and intrinsic spatial uncertainty. The presence of a significant amount of intrinsic uncertainty in the observers may also explain the absence of any blurring of the negative template image in  $CI_{\text{miss}}$  in the medium-uncertainty condition, which was observed in the simulation.

The positive haze in  $CI_{\text{FA}}$  is more visible for the medium-uncertainty condition if we blur the subimages (using a Gaussian kernel with a space constant of 14.1 stimulus pixel, right column of Figure 6). Such a positive haze around the center of the image appears to be absent from  $CI_{\text{FA}}$  in the high-uncertainty condition.

The quantitative results for the estimation of spatial extent are depicted as plots of residual versus  $d$  (middle column of Figure 6) and summarized in Table 2. These results were obtained by fitting Equation 24 to the classification subimages, using the target letter “o” as the presumed observer template. The spatial extent of the uncertainty ( $d$ ) was significantly higher in the high-uncertainty condition as compared with the medium-uncertainty condition, both within and between subjects. The standard errors were estimated with bootstrap.

For the subject who participated in both of the uncertainty conditions (J.H.), we plotted rSNR versus number of trials in Figure 7. Consistent with the result of Experiment 1, we found that the rSNR was higher for the condition with a large extent in spatial uncertainty (and a higher detection threshold). However, unlike Experiment 1, this result is consistent with the result of the corresponding ideal-observer model (Figure 2), which also exhibited a higher rSNR in the high-uncertainty condition.

### Experiment 3: Letter identification in the periphery

We explicitly manipulated spatial uncertainty in Experiments 1 and 2 to test if the various properties of signal clamping derived from analysis and simulation were empirically relevant. The results from the two preceding experiments suggest that these properties are indeed valid. In Experiment 3, we used the method of signal-clamped classification image to estimate the letter templates in the visual periphery and to determine the level of intrinsic spatial uncertainty in the periphery (10 deg in the inferior field). It has been suggested that one reason for an impoverished form vision in the periphery was because of a high degree of intrinsic spatial uncertainty in the human periphery. The cause of the intrinsic spatial uncertainty can be due to undersampling of the visual space (Levi & Klein, 1996; Levi et al., 1987) or to an uncalibrated disarray in spatial sampling (Hess & Field, 1993; Hess & McCarthy, 1994). The theory of uncalibrated disarray would predict a distorted perceptual template, whereas that of undersampling would not. These predictions are contingent on the possibility of recovering the observer's template despite the high intrinsic spatial uncertainty in the periphery.

Prior to the main experiment, an acuity measurement was first performed on each subject (see General methods section for details). In the main experiment, letter stimuli (lowercase "x" and "o") were presented at a fixed retinal eccentricity of 10 deg. There was no stimulus-level spatial uncertainty, and the letter was always presented at the center of the noise field. The subjects were apprised of this fact before the commencement of data collection. Subjects maintained fixation at a green LED and were asked to identify which letter was presented in each trial.

The experiment was conducted on two subjects who had previously participated in one of the earlier experiments. Subject A.S.N, (who had participated in Experiment 1 in the high-uncertainty condition) had a peripheral acuity measurement of 0.42 deg in x-height. A letter of 50 pt Times New Roman (x-height = 0.85 deg in visual angle) was used for A.S.N. Subject B.B. (who had participated in Experiment 1 in the no-uncertainty condition) had an acuity of 0.57 deg in x-height. A letter size of 66 pt (1.15 deg in x-height) was used for B.B.

### Results and discussions

The classification images for the two subjects are shown in the left column of Figure 8. Qualitatively, the classification images in the periphery are very similar to the one in the fovea with high stimulus-level spatial uncertainty (Rows 3 and 4 of Figure 4) and differ noticeably from the fovea results without stimulus-level uncertainty (Rows 1 and 2 of Figure 4). The recovered templates are not distorted in shape and almost identical to those obtained in the foveal conditions. As was the case for the fovea condition with high extrinsic uncertainty, the observers' "x" templates obtained in the periphery, without any extrinsic uncertainty, appear to involve only one stroke. We have attributed this effect to the possibility that spatial uncertainty was not equally reduced in all directions with the Times New Roman "x" stimulus because the two strokes of "x" differ in width by a factor of 3.

A very weak or nonexistent positive image in the error-trial subimages implies that there was a significant amount of intrinsic spatial uncertainty in the periphery. Unlike the fovea experiment (Experiment 1), the uncertainty in this experiment was entirely intrinsic to the observers. We estimated the spatial extent of this intrinsic uncertainty using Equations 23a and 23b. The residual functions for the estimation are plotted in the middle column of Figure 8, and the estimated spatial extents, in units of stimulus pixels, are summarized in Table 3. Table 3 also restates the results from the fovea experiment (Experiment 1) for comparison. Comparing the fovea (from Experiment 1) and periphery results obtained without any spatial uncertainty in the stimuli, it is clear and not surprising that intrinsic spatial uncertainty in the visual periphery is much higher than that in the fovea. Averaging across the two subjects (B.B. and A.S.N.), the intrinsic spatial uncertainty as measured with an isolated letter target in noise at

10 deg eccentricity was 48 pixels or 1.78 deg, compared with 0.25 deg in the fovea. Figure 9 plots the estimated extent of spatial uncertainty in units of visual angles for the fovea and the periphery conditions across subjects.

As to the debate of whether the primary source of spatial uncertainty in the periphery is uncalibrated disarray (Hess & Field, 1993; Hess & McCarthy, 1994) or (calibrated) undersampling (Levi & Klein, 1996; Levi et al., 1987), our results side with the latter. This is because the negative templates for the letter “o” (and for the visible stroke of the letter “x”) are sharp and undistorted, despite the sizable amount of intrinsic spatial uncertainty revealed by the lack of positive template images and the estimated value of  $d$ .

## General discussions

We showed that by presenting a signal of sufficient contrast in noise, we could uncover the linear kernel (template) of a shift-invariant mechanism, using an otherwise conventional classification-image method (or reverse correlation). In the context of a well-established uncertainty model (cf. Pelli, 1985), spatial uncertainty or shift invariance can be modeled with a set of linear front-end channels of identical kernels at different spatial positions. The responses from these channels are pooled by a max operator. A signal of sufficient strength can positively bias one of these channels, making it the most likely one to drive the system’s response. Noise samples that are negatively correlated with the kernel of the selected channel will suppress its response, occasionally leading to an error. Hence, by averaging the noise sample from the error trials associated with a particular signal, we can obtain a negative image of the linear kernel of the channel that normally responded to this signal. We demonstrated the validity of this theory with simulations and in three human experiments. We also showed how the spatial extent of the uncertainty could be estimated from the classification images.

The key to this method is to present the signal at a sufficient strength such that one particular channel often generates the highest response. In the simulations, we showed that the resulting classification images revealed the observer’s internal template and not the presented signal.

Another important departure from the conventional classification-image method is that we do not combine the classification subimages. Keeping the subimages separate allows us to preserve the blurry positive template images such that we can numerically estimate the spatial extent of the uncertainty.

Although this paper has focused on spatial uncertainty (or shift invariance), the signal-clamped classification-image method can be generalized to the other types of uncertainties. This is because the signal-clamping approximation (Equation 9) depends solely on the validity of the uncertainty model, which is not specific to spatial uncertainty.

The following discussions consider the potentials and limitations of how this signal-clamped classification-image method may in general be applied to uncover internal representations.

### Feature specificity and invariance

The spatial structure of the receptive field of a neuron in a higher cortical area (e.g., V4, IT) is hard to characterize because the cell’s responses are both specific and invariant. Specificity means that the cell may respond to a face but not to the lower half of a face. Invariance means that the cell may respond equally well to either a frontal view of a face or a quarter view, although these two images are very different.

Increases in both specificity and invariance are the hallmarks of visual processing. However, both are forms of nonlinearities that render the conventional reverse correlation or

classification-image method inapplicable. Specificity implies that it is statistically unlikely to come across a noise pattern that happens to activate a mechanism (a neuron) because a partially composed target may not elicit any response. Invariance causes distinct image patterns to be sorted into the same response bin. Averaging such patterns often results in a blur and a pattern to which the mechanism does not respond at all.

We have shown that in the case of shift invariance, the problem associated with invariance can be resolved by signal clamping. This method can be generalized to other types of invariance or uncertainty because the uncertainty model (and as a consequence, the signal-clamping approximation of Equation 9) is not specific to spatial uncertainty. If a fixed signal is used to probe a mechanism, it will remain the case that the signal will bias precisely one channel from among the many to respond. Noise patterns with pixels that are negatively correlated with this channel will likely lead to an error in the response. Hence, a classification subimage obtained from the error trials will contain a clear negative image of the template of that one channel of the invariant mechanism. Furthermore, the absence of any clear positive image in an error-trial subimage will indicate that the mechanism is indeed invariant to some aspect of the stimulus, although the precise nature of the invariance is not known. However, unlike shift invariance, there is no general method to normalize the equivalent templates of an arbitrary invariant mechanism. The template images revealed in a signal-clamped classification-image experiment correspond only to those channels that responded to the presented signals. For example, if the side view of a face was used as the signal to probe a face-selective neuron, then only the template responding to the side view of the face will be revealed by the experiment, although the mechanism may respond equally to all views of a face (i.e., the mechanism is viewpoint invariant).

It may come as a surprise that signal clamping can also be useful to overcome the difficulties associated with feature specificity of a high-level mechanism. Recall that feature specificity means that a mechanism is highly nonlinear such that a partial signal often leads to no response. The mechanism requires a conjunction of features to be present before it generates a response. Random noise patterns are therefore unlikely to elicit any response. Signal clamping gets around this problem by using the noise to disrupt, as opposed to activate, a mechanism. For example, if a mechanism is tuned to the conjunction of two features (a AND b), and such a mechanism is activated by a stimulus, then a noise sample that masks either feature “a” or feature “b” will lead to an error response, and averaging such noise samples will reveal both features “a” and “b”.

It is often possible to find a pseudominimal stimulus that sufficiently activates a mechanism. A classic example is the reduction method that Saleem, Tanaka, and Rockland (1993) used to investigate shape tuning of IT neurons. The reduction method, which starts with a stimulus that the neuron is known to respond to and successively reduces the feature and complexity of the stimulus until the cell’s response drops significantly, is an effective way of obtaining a seemingly minimal stimulus that the neuron is tuned to. However, the process of reduction is subjective, and a choice of reduction made several steps ago may lead to an end pattern that is neither minimal nor optimal, and there is no way to know which way it is.

We observed that such a pseudominimal or suboptimal stimulus could be used as the signal in a signal-clamped classification-image experiment to select a channel of an invariant mechanism. If this initial signal contains a part that is superfluous, the noise component that happens to mask that part will not have any effect on the mechanism’s response. If a noise component masks a part of the signal that is crucial to the mechanism, then the mechanism’s response will be suppressed, leading to an error (miss). This is particularly true if the mechanism has a high feature specificity and does not respond to a partial target. Critically, the noise patterns that masked the different crucial parts of the signal during different trials can

be “ORed” together by averaging [ $\text{NOT}(a \text{ AND } b) = (\text{NOT}(a) \text{ OR } \text{NOT}(b))$ ], revealing the complete signal that the mechanism is tuned to as a negative image in the classification subimage from the error trials.

Hence, regarding both invariance and specificity, the “trick” is to present a signal that can effectively elicit a response from the mechanism of interest and collect the noise patterns that suppress the response. In some sense, what we propose here is the opposite of spike-triggered averaging. Rather than adding up the noise patterns that led to a spike, we propose to add up the noise patterns that *suppressed* a spike.

### Detecting invariance in a mechanism

Consider the letter identification experiment in the periphery (Experiment 3). Had we summed the classification subimages as is conventionally done, we would have obtained a dual-template image similar to what was obtained in the fovea condition without spatial uncertainty (assuming good signal clamping). There would not be any indication from the classification image alone that the periphery had a high degree of intrinsic spatial uncertainty. However, by examining the individual subimages separately, particularly those from the error trials, it is very clear that the fovea and the periphery results differ qualitatively—the positive component is largely absent from the error-trial subimages in the periphery condition.

The signal-clamped classification-image method provides a qualitative means to detect the presence of intrinsic uncertainty in a mechanism, as well as a quantitative method to estimate the uncertainty. In principle, the method is generally applicable to all types of intrinsic uncertainty and is not restricted to spatial uncertainty. In short, when the signal-clamping approximation (Equation 9) is valid, the negative component in the error-trial subimages will correspond to the template of a single channel in a possibly invariant mechanism that responded to the presented signal, and the positive component will correspond to the average of all the templates for all the equivalent signals associated with the erroneous response of the mechanism (Equation 16). For a two-way discrimination task (e.g., our “o” vs. “x” task), the negative component from one type of error trials (say, XO—signal was “x”, response was “o”) can be compared to the positive component from the complementary error trials (OX). The discrepancy between the two, aside from a sign difference, is indicative of intrinsic uncertainty. This line of reasoning is similar to those in previous works on detecting observer nonlinearity based on the differences between classification subimages (Abbey & Eckstein, 2002; Barth et al., 1999; Solomon, 2002).

Whether the discrepancy between the negative and positive components is easily discernable depends on the type of uncertainty and the stimuli used to probe it. For example, consider a mechanism that has an uncertainty in the size but not in the position of a signal. If we tested this system with the “x” versus “o” task, then the positive component from the OX trials would be an average of x’s of all sizes, centered on one another. The result would still resemble an “x” but with a bright and well-defined center and graded strokes extending outward. That is, the average of x’s of all sizes may not be sufficiently different from a single, medium-sized “x”. In contrast, the average of o’s of all sizes, which would look like a haze, will be quite different from an “o” of any particular size. Thus, it would be easy to detect the presence of size uncertainty with an “o” rather than with an “x” as a signal. We note with interest that the classification subimages from the letter identification experiment in the fovea with no extrinsic spatial uncertainty (Figure 4, Rows 1 and 2) appears to show this kind of size uncertainty.

### Task requirements and invariance

A mechanism of sufficient flexibility (e.g., a human observer) may adjust its degree of invariance to suit the task. For example, when there is no positional uncertainty in the stimuli,

it would be suboptimal to use a mechanism that has a high degree of positional invariance. The form-vision mechanism in the fovea, for example, seems to be capable of limiting its degree of positional invariance, and hence the amount of intrinsic spatial uncertainty, when the target position is precisely known (Experiment 1, nouncertainty condition). In contrast, the form-vision mechanism in the periphery appears to be unable to make the same adjustment (Experiment 3).

Likewise, a flexible mechanism must increase its degree of invariance along the relevant stimulus dimension when the task requires it to do so. The letter detection experiment (Experiment 2) showed that the foveal mechanism appears to make the appropriate adjustment when the extent of the spatial uncertainty of the stimulus changed across conditions.

A mechanism that flexibly adapts to the task, such as a human observer, poses problems to the signal-clamping method. Given that relatively strong signals must be used in an experiment, the presence of these signals can influence how the mechanism will otherwise perform the task. For example, a mechanism that normally has a high degree of positional invariance may limit its processing to a particular region on the display if the signal is always presented at the same location. In the case of spatial uncertainty, we may assume that the mechanism is shift invariant, which allowed us to present a single signal at different positions on the display, and then shift the noise patterns to normalize the position of the presented signal before averaging. However, as we noted earlier, there does not exist a general method for normalizing other types of uncertainty or invariance. Without such methods, an unpractically large number of trials may be needed to obtain the classification images and to maintain a task requirement of high invariance.

### The components of a representation

Consider the “o” versus “x” experiment. What if a mechanism for this task represents “x” as “either a left slash ( $\backslash$ ) or a right slash ( $/$ )”? The psychometric function ( $d'$  vs. signal contrast) of such a mechanism is nonlinear, as opposed to the linear psychometric function of a mechanism that represents “x” as a single template. Unfortunately, linearity of a psychometric function is nondiagnostic in practice because other factors, such as other types of uncertainty, can also lead to a nonlinear psychometric function. In fact, the psychometric function of a human observer is rarely linear for just about any task tested.

For the two-component mechanism, the classification image for the “x” template will look exactly like “x”. There will be no indication of the two distinct components in the representation. Signal clamping will not help to resolve this problem without any a priori assumptions about the possible components. In fact, this is a general problem in all classification-image methods that involve averaging noise samples across trials. For each type of trials, although the signal and the response were the same, the cause of the response might vary from trial to trial. Averaging assumes that the mechanism does not distinguish the higher order structures within a trial from those across trials, which is clearly incorrect. Accumulating higher order statistics across trials seem necessary. Techniques that involve obtaining the covariance (de Ruyter van Steveninck & Bialek, 1988) in addition to the mean appear promising, but their applications remain restricted to relatively simple systems or stimuli (e.g., the spatiotemporal receptive field of macaque VI neurons, Rust et al., 2005; complex cells in cats, Touryan, Lau, & Dan, 2002; bar detection by human observers, Neri & Heeger, 2002). The major challenges to these high-order techniques include the determination of what high-order statistics to collect and whether the number of trials required to obtain such statistics is practical.

To decipher a complex mechanism, intuitions about the underlying representation are not just helpful but essential. Returning to our toy example, if we have an a priori reason to suspect



that the “x” might be represented by a disjunction of two slashes (i.e., “x” = “/” or “\”), we may test this hypothesis with signal clamping by presenting randomly either the left or the right slash in a trial when “x” is supposed to be the target. Assuming that adequate measures have been taken to ensure that nature of the task is not changed by the presented partial signal, we can then left-right reverse the noise patterns from the error trials when “/” was presented and average them with the noise patterns from the error trials when “\” was presented. If the resulting negative image is not that of a single slash, then the hypothesis of “x” being represented as either the left or the right slash can be rejected.

This toy example stresses a general nature of the signal-clamping method—it is as much a hypothesis-driven method as a hypothesis-free exploration tool that the conventional classification-image method is.

### Other methods for measuring uncertainty

Signal-clamped classification images provide one method of estimating the amount of intrinsic uncertainty (Equations 23a, 23b and 24) that is considerably different from the traditional approach. The traditional method for quantifying intrinsic uncertainty is to estimate  $M$ , the number of orthogonal channels possessed by the observer, by measuring the extent by which the psychometric function ( $d'$  vs. signal contrast) of the observer deviates from linearity or, equivalently, its log–log slope deviates from unity (e.g., Foley & Legge, 1981; Green, 1964; Nachmias & Sansbury, 1974; Stromeyer & Klein, 1974; Tanner & Swets, 1954). Pelli (1985) used a Weibull approximation to the psychometric function and established via numerical simulations the relationship between the parameters of the Weibull function and  $M$ . Later work (e.g., Eckstein, Ahumada, & Watson, 1997; Tyler & Chen, 2000; Verghese & McKee, 2002) departed from the Weibull approximation and/or derived analytically the relationship between  $M$  and the parameters of a psychometric function. All these approaches assumed the Max-rule model of uncertainty (observer’s response is determined by the maximally responding channel) and that the  $M$  channels are orthogonal. Most critically, these approaches treat uncertainty in a generic sense and make no distinction regarding the feature dimension of the uncertainty. For example, uncertainty about a signal’s position is not distinguished in these formulations from uncertainty about its orientation. All types of uncertainty are characterized in terms of  $M$ —the equivalent number of orthogonal channels that the observer possesses.

An alternative approach is to use an image-based decision model and measure the intrinsic uncertainty of an observer by matching the model’s performance to that of the observer by varying the amount of uncertainty in the model. With this method, uncertainty must be introduced along one or more specific dimensions of the stimuli. For example, Tjan and Legge (1998) studied the effect of viewpoint uncertainty on 3-D object-recognition tasks, whereas Manjeshwar and Wilson (2001) measured positional uncertainty in a line-detection task. Both studies assumed a sum-of-likelihood decision model. These image-based methods can characterize uncertainty in units specific to the feature dimension of the uncertainty (e.g., visual angle for positional uncertainty or angle of rotation for viewpoint). Moreover, these methods do not require the templates considered by the observer to be orthogonal.

Our method of using signal-clamped classification images to estimate intrinsic certainty is similar to the image-based approach except that it is less model specific. The method works as long as signal clamping is reasonably effective (i.e., Equation 9 is a reasonable approximation). Our method measures intrinsic uncertainty in terms of the spread of the noise patterns along a feature dimension of interest (e.g., spatial positions of the perceived signal) that led to false alarms. Such noise patterns cannot be clamped or normalized by the signal and can therefore be separated from the noise patterns that led to misses, which are clamped by the

signal. Unlike the image-based methods used in earlier studies, our method does not require defining a specific image-based model of the observer.

### rSNR and spatial uncertainty

We noted with interest that the quality of the signal-clamped classification images from human observers, when measured in terms of rSNR (Equation 26), increased as spatial uncertainty in the stimuli increased. In contrast, we found with ideal-observer simulations that the relationship between uncertainty (intrinsic or extrinsic) and rSNR was actually quite complex. Consider the letter detection task (Figure 2). When spatial uncertainty increased from a spatial extent of  $32 \times 32$  stimulus pixels ( $M = 250$ ) to  $64 \times 64$  pixels ( $M = 1,000$ ), the rSNR of the model's classification images increased from 627 to 751, whereas the model's log threshold contrast increased from  $-1.29$  to  $-1.14$  (a factor of 1.4 in contrast). When there was no uncertainty, the model rSNR was 2,020 (not shown in Figure 2) at a log threshold contrast of  $-1.48$ . This U-shape function of rSNR in terms of uncertainty was also evident for the letter discrimination task (Figure 1). The rSNRs of the model's classification images were 1,180, 856, and 973 for spatial extents of  $1 \times 1$ ,  $32 \times 32$  ( $M = 250$ , not shown in Figure 1), and  $64 \times 64$ , respectively.

We do not fully understand why rSNR versus spatial extent is a U-shape function because we do not yet have a close-form expression relating rSNR to uncertainty. We suspect that the U-shape function was a result of the interplay between uncertainty and contrast threshold. Given that the signal-clamping approximation (Equation 9) is never perfect, we expect rSNR to decrease as uncertainty increases. However, as uncertainty increases, so does the threshold contrast. Because masking a signal of higher contrast requires larger instantaneous amplitude in the noise, the negative correlation between the presented signal and the noise pixels in the error-trial classification subimages must therefore be stronger, resulting in a higher rSNR. In short, an increase in uncertainty can cause either a decrease or an increase in rSNR, depending on the amount of uncertainty and the amount of threshold elevation caused by the uncertainty.

In the experiments report here, human rSNR always increased with extrinsic spatial uncertainty, which ranged from none to  $64 \times 64$  pixels (at  $M = 1,000$ ). This pattern of results can be reconciled with data from the ideal-observer models by noting that intrinsic spatial uncertainty was always present in the human observers, even when there was no uncertainty in the stimulus (Table 1). Such intrinsic uncertainty might place human data on the increasing portion of the U-shape function. In addition, internal noise in human observers may also play a role. A more thorough analysis of rSNR versus intrinsic uncertainty in human observers awaits future studies.

### Conclusion

Most human experiments using the classification-image methods present a signal in each trial primarily to keep the observers engaged. Here, we showed that such a signal, if of sufficient strength, could limit or even eliminate the effect of uncertainty on the resulting classification images. As examples, we successfully obtained clear images of human observers' perceptual templates in the face of a high degree of spatial uncertainty.

A hallmark of visual processing is the progressive increase of invariance. Because invariance is a form of uncertainty, our method offers a new tool for uncovering the underlying representations in a visual processing system.

## Appendix

### Appendix

We want to show that that noise sample  $N_{OX}$  from the OX error trials, where the signal was “O” but the response was “X”, has the mathematical expectation as described in Equation 16, where  $O_z$  is the channel that is tuned to the presented “O” signal,  $X_j, j \in [1, M]$  are the channels that are tuned to the possible signals for the “X” response, and  $E[X_j]$  denotes the average across all  $X_j$ s. Our starting points are (1) the result from Ahumada (2002) for  $M = 1$  (Equation 13) and (2) the internal decision variable of the observer during these trials, with the signal-clamping approximation applied (Equation 15).

We shall prove Equation 16 with mathematical induction on  $M$ . The case of  $M = 1$  is true from Ahumada (2002) (i.e., Equation 13). Assuming  $M = k$  is true, we consider the case of  $M = k + 1$ . Let  $v$  be the number of trials where  $X_j, j \leq k$ , were the maximum-responding X channels on the left-hand side of Equation 15. For these trials only, it was as if  $M = k$ , and Equation 16 is true by assumption. The sum of the noise samples from these trials is

$$vE[N_{OX}] \propto (vE[X_j] - vO_z)j \in [1, k]. \quad (\text{A1})$$

Let  $w$  be the number of trials where  $X_{k+1}$  is the maximum-responding X channel. For these trials, it was as if  $M = 1$ , for which the result of Ahumada (2002) (Equation 13) applies. The sum of the noise samples from these trials is

$$wE[N_{OX}] \propto (wX_{k+1} - wO_z). \quad (\text{A2})$$

Adding Equation A1 to Equation A2 and dividing the sum by the total number of trials ( $v + w$ ), we have

$$E[N_{OX}] \propto (E[X_j] - O_z)j \in [1, k+1]. \quad (\text{A3})$$

Thus, Equation 16 will be true for  $M = k + 1$ , if it is true for  $M = k$ . Because it is true for  $M = 1$ , by mathematical induction, Equation 16 is true for all  $M \geq 1$ .

## Acknowledgments

We like to thank Susana Chung and Miguel Eckstein for their valuable comments and critiques. This research was supported by National Institutes of Health Grant EY016391 to BST.

## References

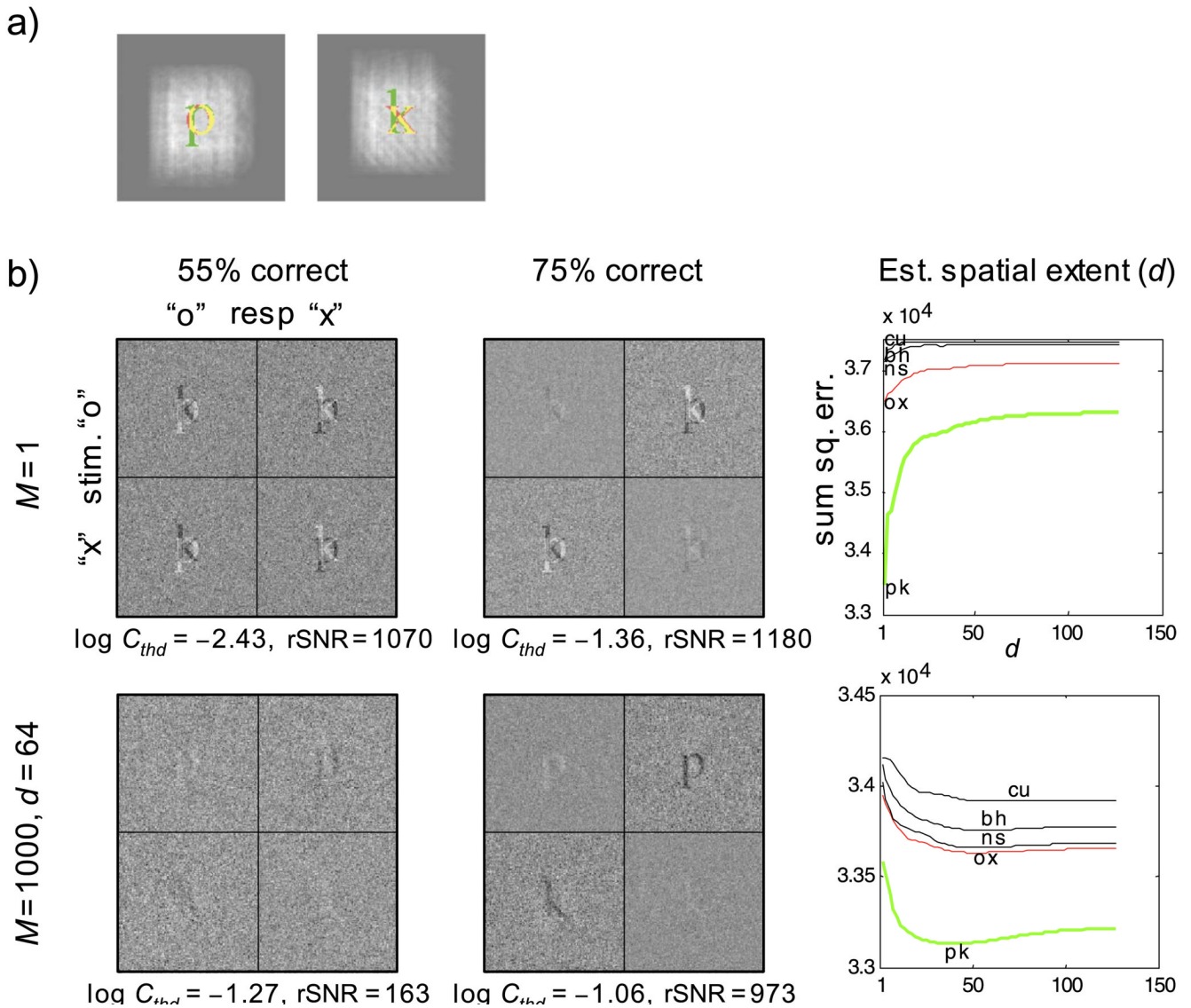
- Abbey CK, Eckstein MP. Classification image analysis: Estimation and statistical inference for two-alternative forced-choice experiments. *Journal of Vision* 2002;2(1):66–78. [PubMed: 12678597] <http://journalofvision.org/2/1/5/>, doi:10.1167/2.1.5. [PubMed] [ Article ]
- Adolphs R, Gosselin F, Buchanan TW, Tranel D, Schyns P, Damasio AR. A mechanism for impaired fear recognition after amygdala damage. *Nature* 2005;433:68–72. [PubMed: 15635411] [PubMed]
- Ahumada AJ Jr. Classification image weights and internal noise level estimation. *Journal of Vision* 2002;2(1):121–131. [PubMed: 12678600] <http://journalofvision.org/2/1/8/>, doi:10.1167/2.1.8. [PubMed] [ Article ]
- Ahumada AJ Jr, Beard BL. Classification images for detection. *Investigative Ophthalmology and Visual Science* 1999;40:S572.

- Ahumada AJ Jr, Lovell J. Stimulus features in signal detection. *The Journal of the Acoustical Society of America* 1971;49:1751–1756.
- Ahumada A Jr, Marken R. Time and frequency analyses of auditory signal detection. *The Journal of the Acoustical Society of America* 1975;57:385–390. [PubMed: 1117090] [PubMed]
- Barth, E.; Beard, BL.; Ahumada, AJ, Jr. Nonlinear features in vernier acuity. In: Rogowitz, BE.; Pappas, TN., editors. *Human Vision and Electronic Imaging III*. SPIE Proceedings. San Jose, CA: SPIE; 1999. p. 88-96. Vol. 3644
- Beard BL, Ahumada AJ Jr. Detection in fixed and random noise in foveal and parafoveal vision explained by template learning. *Journal of the Optical Society of America A, Optics, Image Science, and Vision* 1999;16:755–763. [PubMed]
- Biederman I. Recognition-by-components: A theory of human image understanding. *Psychological Review* 1987;94:115–147. [PubMed: 3575582] [PubMed]
- Brainard DH. The psychophysics toolbox. *Spatial Vision* 1997;10:433–436. [PubMed: 9176952] [PubMed]
- de Boer E, de Jongh HR. On cochlear encoding: Potentialities and limitations of the reverse-correlation technique. *Journal of the Acoustical Society of America* 1978;63:115–135. [PubMed: 632404] [PubMed]
- de Boer E, Kuypers P. Triggered correlation. *IEEE Transactions on Biomedical Engineering* 1968;15:169–179. [PubMed: 5667803] [PubMed]
- de Ruyter van Steveninck R, Bialek W. Real-time performance of a movement-sensitive neuron in the blowfly visual system. *Proceedings of the Royal Society of London B* 1988;234:269–276.
- Duda, RO.; Hart, PE. *Pattern classification and scene analysis*. New York: Wiley; 1973.
- Eckstein MP, Ahumada AJ Jr, Watson AB. Visual signal detection in structured backgrounds: II. Effects of contrast gain control, background variations, and white noise. *Journal of the Optical Society of America A, Optics, Image Science, and Vision* 1997;14:2406–2419. [PubMed]
- Eckstein MP, Shimozaki SS, Abbey CK. The footprints of visual attention in the Posner cueing paradigm revealed by classification images. *Journal of Vision* 2002;2(1):25–45. [PubMed: 12678595] <http://journalofvision.org/2/1/3/>, doi:10.1167/2.1.3. [PubMed] [Article]
- Efron, B.; Tibshirani, RJ. *An introduction to the bootstrap*. New York: Chapman & Hall; 1994.
- Foley JM, Legge GE. Contrast detection and near-threshold discrimination in human vision. *Vision Research* 1981;21:1041–1053. [PubMed: 7314485] [PubMed]
- Gold JM, Murray RF, Bennett PJ, Sekuler AB. Deriving behavioural receptive fields for visually completed contours. *Current Biology* 2000;10:663–666. [PubMed: 10837252] [PubMed] [Article]
- Gosselin F, Schyns PG. Superstitious perceptions reveal properties of internal representations. *Psychological Science* 2003;14:505–509. [PubMed: 12930484] [PubMed]
- Green, DM. Psychoacoustics and detection theory. In: Swets, JA., editor. *Signal detection and recognition by human observers—Contemporary readings*. New York: John Wiley & Sons; 1964. p. 58-91.
- Green, DM.; Swets, JA. *Signal detection theory and psychophysics*. Huntington, New York: Robert E. Krieger Publishing Company; 1974.
- Hess RF, Field D. Is the increased spatial uncertainty in the normal periphery due to spatial undersampling or uncalibrated disparity? *Vision Research* 1993;33:2663–2670. [PubMed: 8296463] [PubMed]
- Hess RF, McCarthy J. Topological disorder in peripheral vision. *Visual Neuroscience* 1994;11:1033–1036. [PubMed: 7947397] [PubMed]
- Jones JP, Palmer LA. The two-dimensional spatial structure of simple receptive fields in cat striate cortex. *Journal of Neurophysiology* 1987;58:1187–1211. [PubMed: 3437330] [PubMed]
- Levi DM, Klein SA. Limitations on position coding imposed by undersampling and univariance. *Vision Research* 1996;36:2111–2120. [PubMed: 8776477] [PubMed]
- Levi DM, Klein SA, Yap YL. Positional uncertainty in peripheral and amblyopic vision. *Vision Research* 1987;27:581–597. [PubMed: 3660620] [PubMed]
- Manjeshwar RM, Wilson DL. Hyperefficient detection of targets in noisy images. *Journal of the Optical Society of America A, Optics, Image Science, and Vision* 2001;18:507–513. [PubMed]

- Marr, D. Vision: A computational investigation into the human representation and processing of visual information. New York: W. H. Freeman; 1982.
- Movshon JA, Thompson ID, Tolhurst DJ. Receptive field organization of complex cells in the cat's striate cortex. *Journal of Physiology* 1978;283:79–99. [PubMed: 722592] [PubMed]
- Murray RF, Bennett PJ, Sekuler AB. Optimal methods for calculating classification images: Weighted sums. *Journal of Vision* 2002;2(1):79–104. [PubMed: 12678598]http://journalofvision.org/2/1/6/, doi:10.1167/2.1.6. [PubMed] [Article]
- Nachmias J, Sansbury RV. Letter: Grating contrast: Discrimination may be better than detection. *Vision Research* 1974;14:1039–1042. [PubMed: 4432385] [PubMed]
- Neri P. Estimation of nonlinear psychophysical kernels. *Journal of Vision* 2004;4(2):82–91. [PubMed: 15005649]http://journalofvision.org/4/2/2/, doi: 10.1167/4.2.2. [PubMed] [Article]
- Neri P, Heeger DJ. Spatiotemporal mechanisms for detecting and identifying image features in human vision. *Nature Neuroscience* 2002;5:812–816. [PubMed] [Article]
- Neri P, Parker AJ, Blakemore C. Probing the human stereoscopic system with reverse correlation. *Nature* 1999;401:695–698. [PubMed: 10537107] [PubMed]
- Nolte LW, Jaarsma D. More on the detection of one of M orthogonal signals. *Journal of the Acoustical Society of America* 1967;41:497–505.
- Pelli DG. Uncertainty explains many aspects of visual contrast detection and discrimination. *Journal of the Optical Society of America A, Optics and Image Science* 1985;2:1508–1532. [PubMed]
- Pelli DG. The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision* 1997;10:437–442. [PubMed: 9176953] [PubMed]
- Pelli DG, Zhang L. Accurate control of contrast on microcomputer displays. *Vision Research* 1991;31:1337–1350. [PubMed: 1891822] [PubMed]
- Peterson WW, Birdsall TG, Fox WC. The theory of signal detectability. *Transactions IRE Professional Group on Information Theory* 1954:171–212.PGIT-4
- Rust NC, Schwartz O, Movshon JA, Simoncelli EP. Spike-triggered characterization of excitatory and suppressive stimulus dimensions in monkey V1. *Neurocomputing* 2004:793–799.58–60
- Rust NC, Schwartz O, Movshon JA, Simoncelli EP. Spatiotemporal elements of macaque v1 receptive fields. *Neuron* 2005;46:945–956. [PubMed: 15953422] [PubMed]
- Saleem KS, Tanaka K, Rockland KS. Specific and columnar projection from area TEO to TE in the macaque inferotemporal cortex. *Cerebral Cortex* 1993;3:454–464. [PubMed: 8260813] [PubMed]
- Shimozaki SS, Eckstein MP, Abbey CK. Spatial profiles of local and nonlocal effects upon contrast detection/discrimination from classification images. *Journal of Vision* 2005;5(1):45–57. [PubMed: 15831066]http://journalofvision.org/5/1/5/, doi:10.1167/5.1.5. [PubMed] [Article]
- Solomon JA. Noise reveals visual mechanisms of detection and discrimination. *Journal of Vision* 2002;2(1):105–120. [PubMed: 12678599]http://journalofvision.org/2/1/7/, doi:10.1167/2.1.7. [PubMed] [Article]
- Stromeyer CF III, Klein S. Spatial frequency channels in human vision as asymmetric (edge) mechanisms. *Vision Research* 1974;14:1409–1420. [PubMed: 4446371] [PubMed]
- Tanner WP Jr. Physiological implications of psychophysical data. *Annals of the New York Academy of Sciences* 1961;89:752–765. [PubMed: 13775211] [PubMed]
- Tanner WP Jr, Swets JA. The human use of information. I. Signal detection for the case of the signal known exactly. *IRE Transactions on Information Theory* 1954:213–221.PGIT-4
- Tjan, BS. Object recognition. In: Arbib, M., editor. *The handbook of brain theory and neural networks*. Vol. 2nd. Cambridge, MA: MIT Press; 2002.
- Tjan BS, Braje WL, Legge GE, Kersten D. Human efficiency for recognizing 3-D objects in luminance noise. *Vision Research* 1995;35(21):3053–3069. [PubMed: 8533342] [PubMed]
- Tjan BS, Legge GE. The viewpoint complexity of an object-recognition task. *Vision Research* 1998;38:2335–2350. [PubMed: 9798003] [PubMed]
- Touryan J, Lau B, Dan Y. Isolation of relevant visual features from random stimuli for cortical complex cells. *The Journal of Neuroscience* 2002;22:10811–10818. [PubMed: 12486174] [PubMed] [Article]

- Tyler CW, Chen CC. Signal detection theory in the 2AFC paradigm: Attention, channel uncertainty and probability summation. *Vision Research* 2000;40:3121–3144. [PubMed: 10996616] [PubMed]
- Vergheze P, McKee SP. Predicting future motion. *Journal of Vision* 2002;2(5):413–423. [PubMed: 12678655] <http://journalofvision.org/2/5/5/>, doi: 10.1167/2.5.5. [PubMed] [Article]
- Watson AB, Pelli DG. QUEST: A Bayesian adaptive psychometric method. *Perception & Psychophysics* 1983;33:113–120. [PubMed: 6844102] [PubMed]

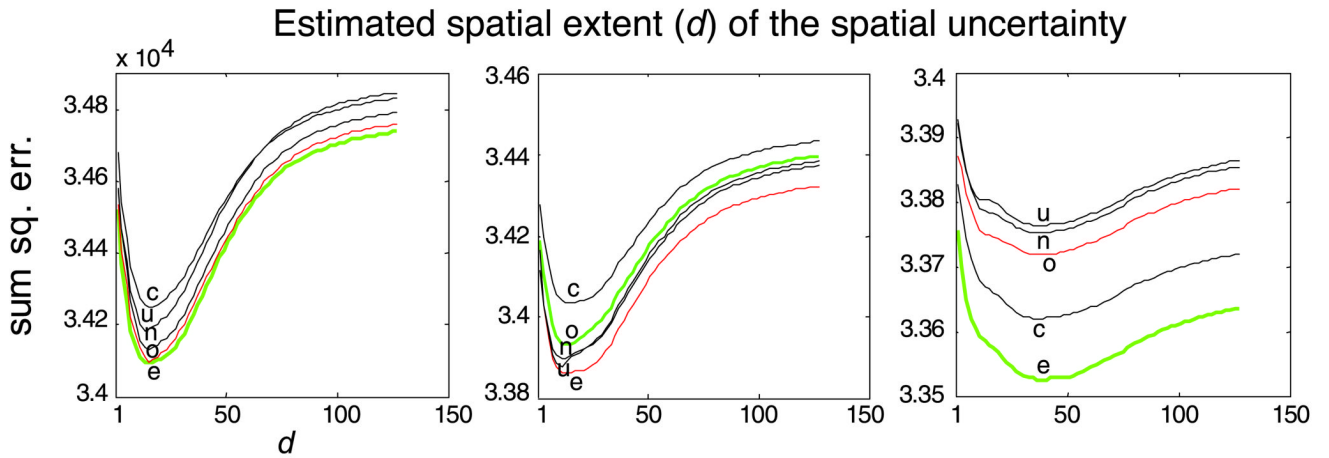
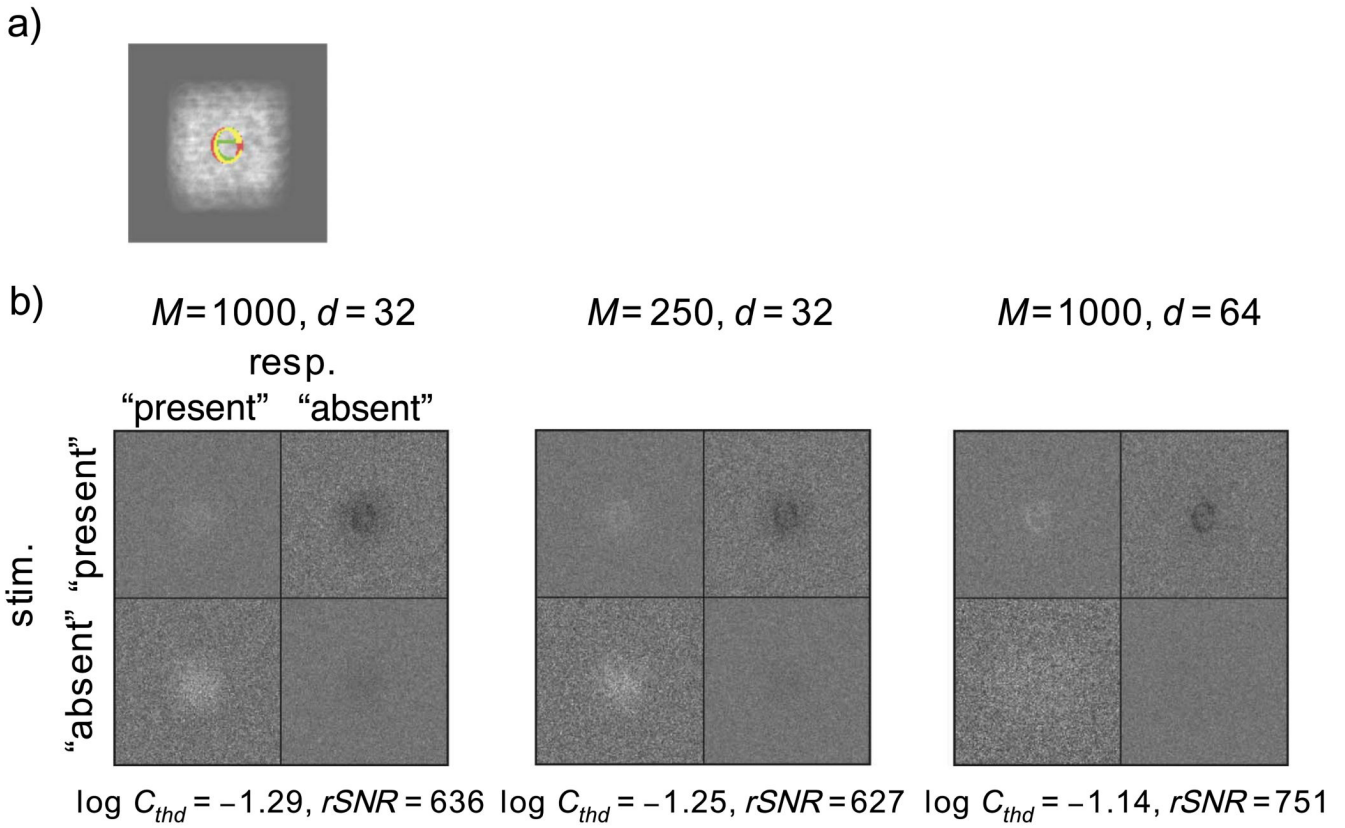




**Figure 1.**

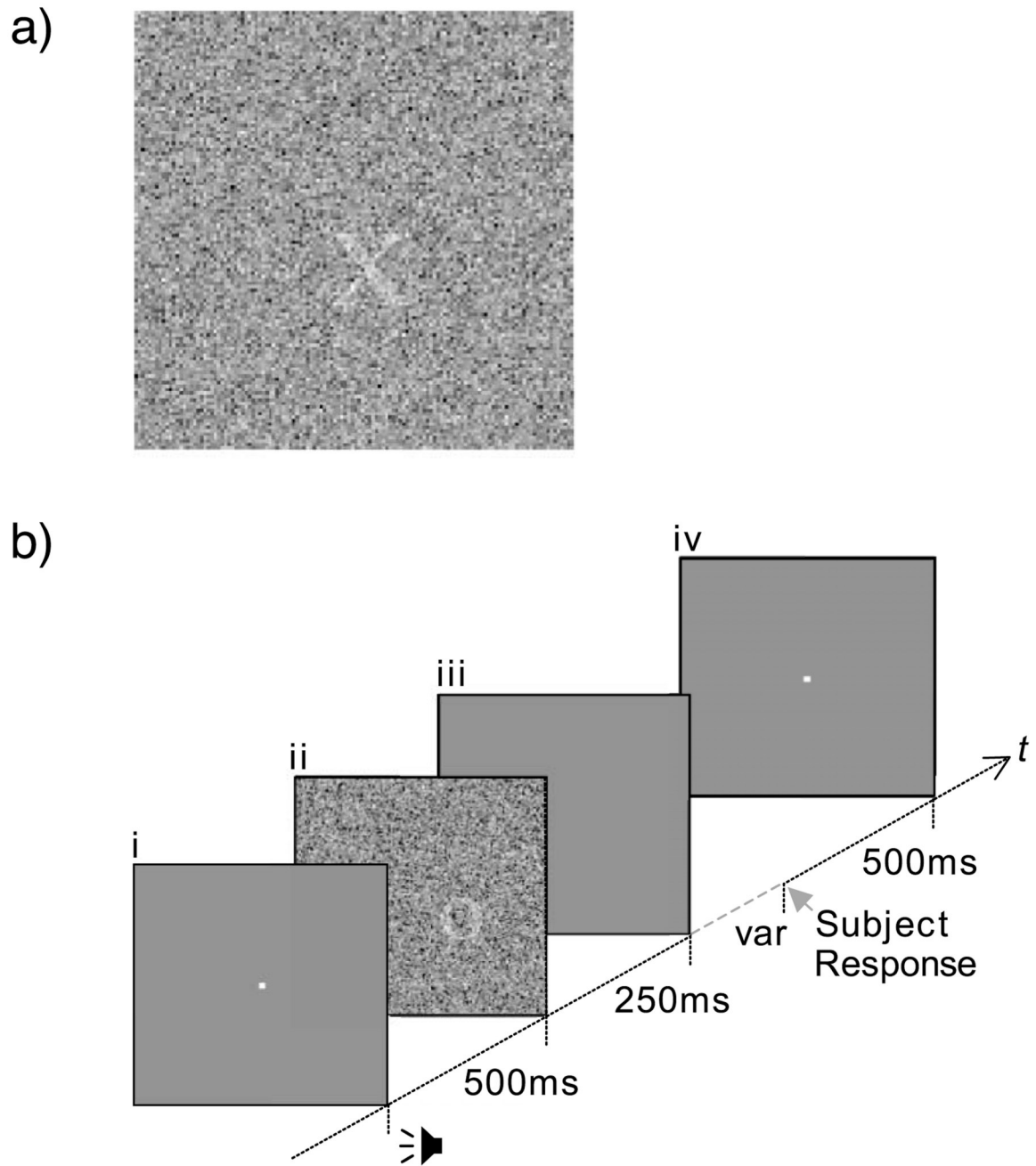
(a) Signals and templates used for simulating the letter identification task using an ideal-observer model. The white haze shows the spatial extent of the intrinsic spatial uncertainty of the model for  $M = 1,000$  and spatial extent ( $d$ ) equal to 64 pixels. The templates used by the model are shown in green. The letter stimuli are shown in red and overlapping regions in yellow, (b) Classification images from the ideal-observer model for the letter identification task: first row, simulations with no spatial uncertainty ( $M = 1$ ); second row, simulations with high spatial uncertainty ( $M = 1,000, d = 64$ ); left column, low signal-contrast simulations at an accuracy criterion of 55% correct; middle column, high signal-contrast simulations at an accuracy criterion of 75% correct; right column, estimations of the spatial extent ( $d$ ) of the uncertainty for the high signal-contrast condition (middle column)—each curve is an error function labeled by the templates used to obtain the estimate. The value of  $d$  at the minimum of each error function represents the estimated spatial extent of the uncertainty. The minimum of each curve is marked by the position of the first character of the corresponding label. The green curves were obtained using the actual observer templates from the model, the red curves were obtained using the stimuli letters as templates, and the black curves were obtained using pairs of letters

that closely resembled (in terms of rms distance) the true templates. The high degree of similarity in the estimated values of  $d$  using different putative templates shows the robustness of the method. The stimulus noise had a pixel-wise standard deviation of 0.25. rSNR was computed using only the error trials, as described in Equation 26.



**Figure 2.**  
 (a) Signal and template used for simulating the letter detection task with an ideal-observer model. The white haze shows the extent of the intrinsic spatial uncertainty of the model observer for  $M = 1,000$  and spatial extent ( $d$ ) equal to 64. The template used by the model is shown in green. The letter stimulus is shown in red. The overlapping regions are shown in yellow, (b) Classification images from the ideal-observer model performing the letter detection task at an accuracy level of 75% correct: first column, classification-image and spatial-extent estimations for the medium spatial uncertainty condition ( $M = 1,000, d = 32$ ); second column, classification-image and spatial-extent estimations for a medium spatial uncertainty condition ( $M = 250, d = 32$ ), which has the same spatial density of templates as the high-uncertainty

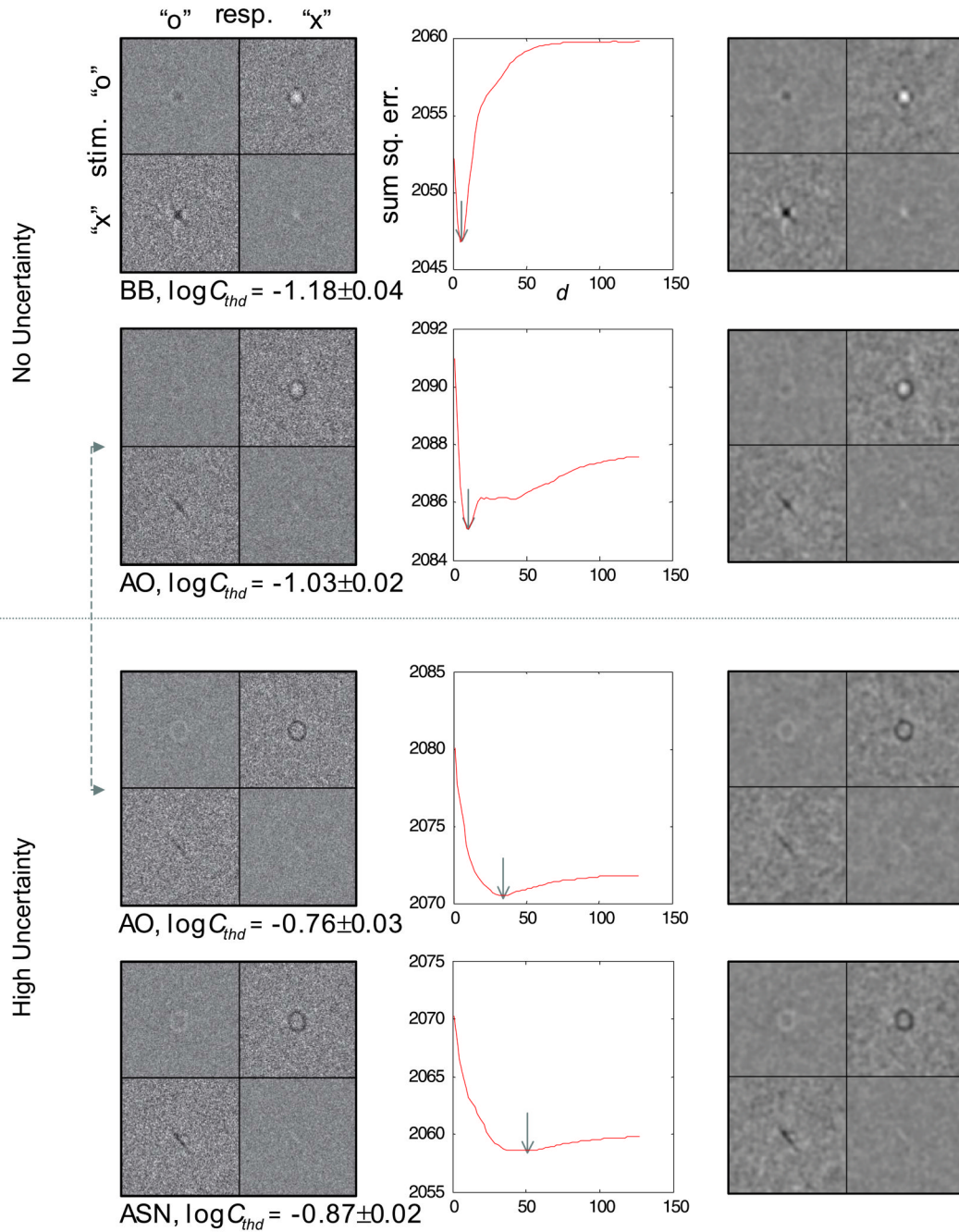
condition; third column, classification-image and spatial-extent estimations for the high spatial uncertainty condition ( $M = 1,000$ ,  $d = 64$ ). The error functions of spatial-extent estimations are labeled by the putative template used for the estimation. The value of  $d$  at the minimum of each curve represents the estimated spatial extent and is marked by the position of the corresponding label. The green curves were obtained using the model's template, the red curves were obtained using the stimulus letter as the template, and the black curves were obtained using letters that resembled (in terms of rms distance) the model template.



**Figure 3.**

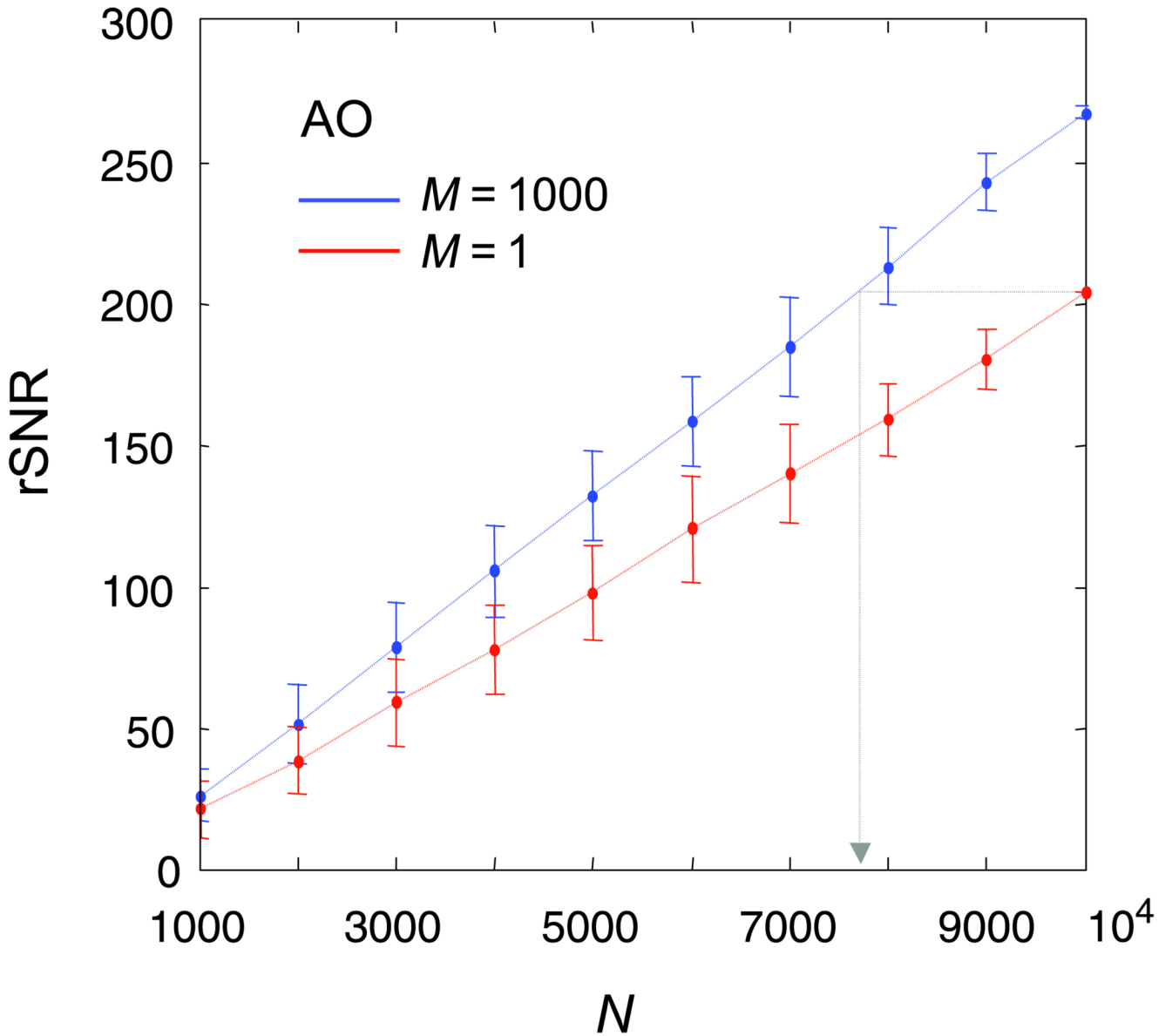
(a) A sample of the noisy stimulus, (b) Timing of stimuli presentation: (1) fixation beep immediately followed by a fixation screen for 500 ms, (2) stimulus presentation for 250 ms, (3) subject response period (variable) with positive feedback beep for correct trials, and (4) 500 ms delay before onset of next trial.



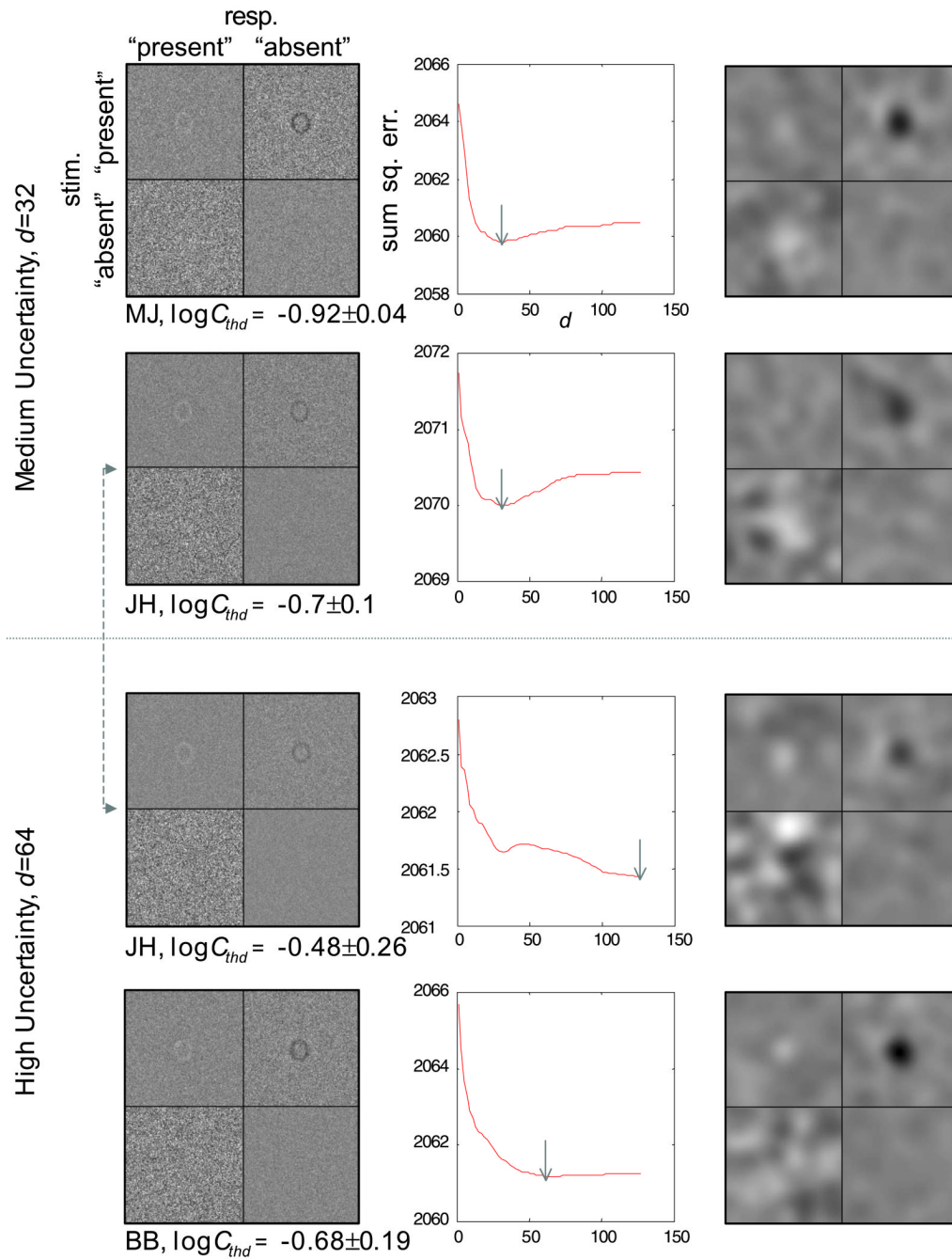


**Figure 4.** Classification images for the human observers in the letter identification task (Experiment 1): top two rows, no spatial uncertainty ( $M = 1$ ); bottom two rows, high spatial uncertainty ( $M = 1,000$ ,  $d = 64$  stimulus pixels); left column, classification images at a signal contrast corresponding to 75% correct; middle column, the spatial extent of the uncertainty estimated from the classification images in the left column; the value of  $d$  at the minimum of each curve (marked by the gray arrow) represents the mean spatial extent of the uncertainty; right column, blurred versions of the classifications images from the left column using a Gaussian kernel with space constant of 1.4 stimulus pixels for visualization purposes only. Image intensities in each column are identically scaled to facilitate across-condition comparisons.



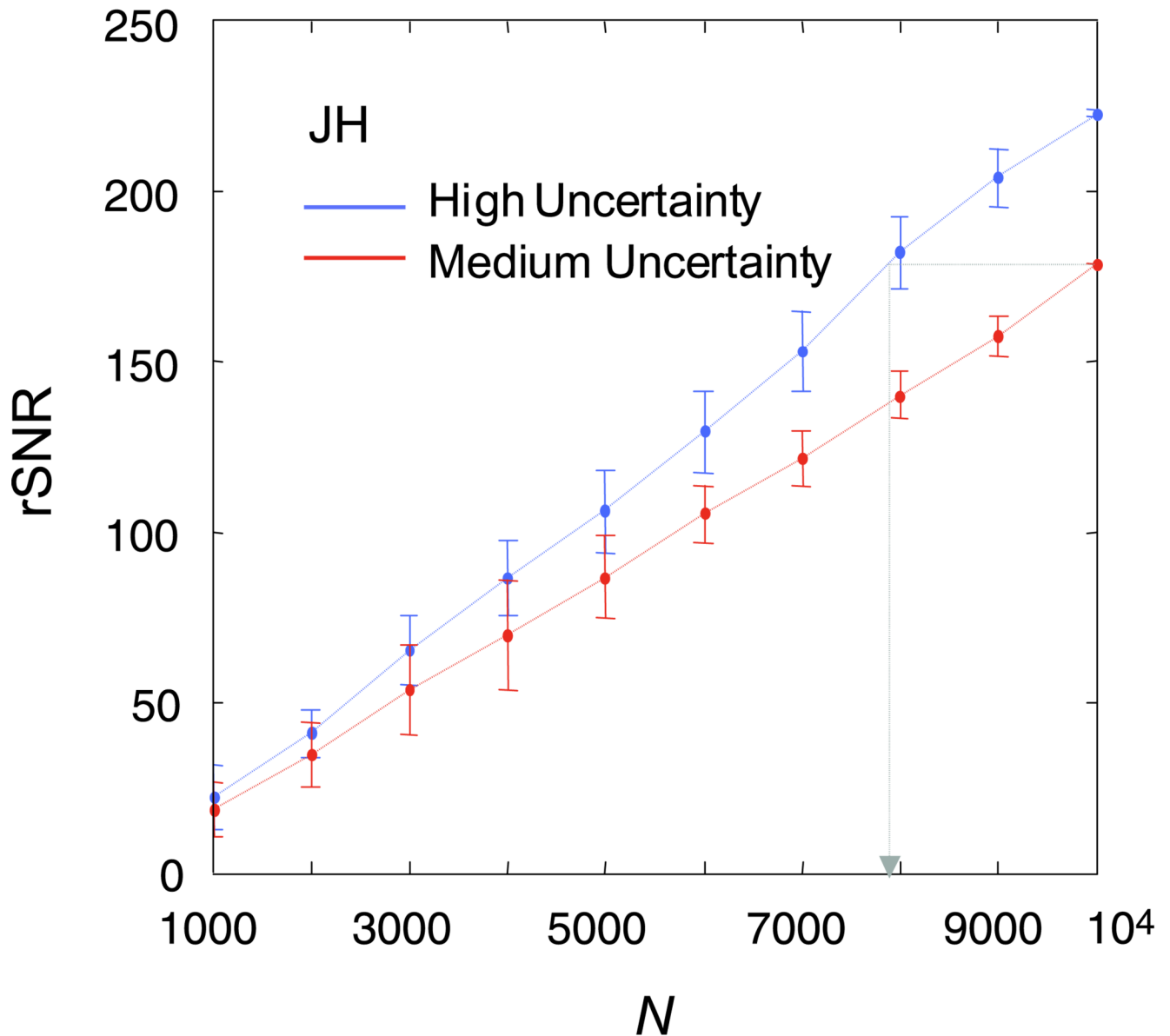


**Figure 5.** rSNR versus number of trials for subject A.O. who participated in both conditions of Experiment 1. The gray arrow marks the approximate number of trials that would be needed in the high-uncertainty condition to achieve the same classification-image quality as the no-uncertainty condition. Error bars are boot-strap standard errors of the mean.

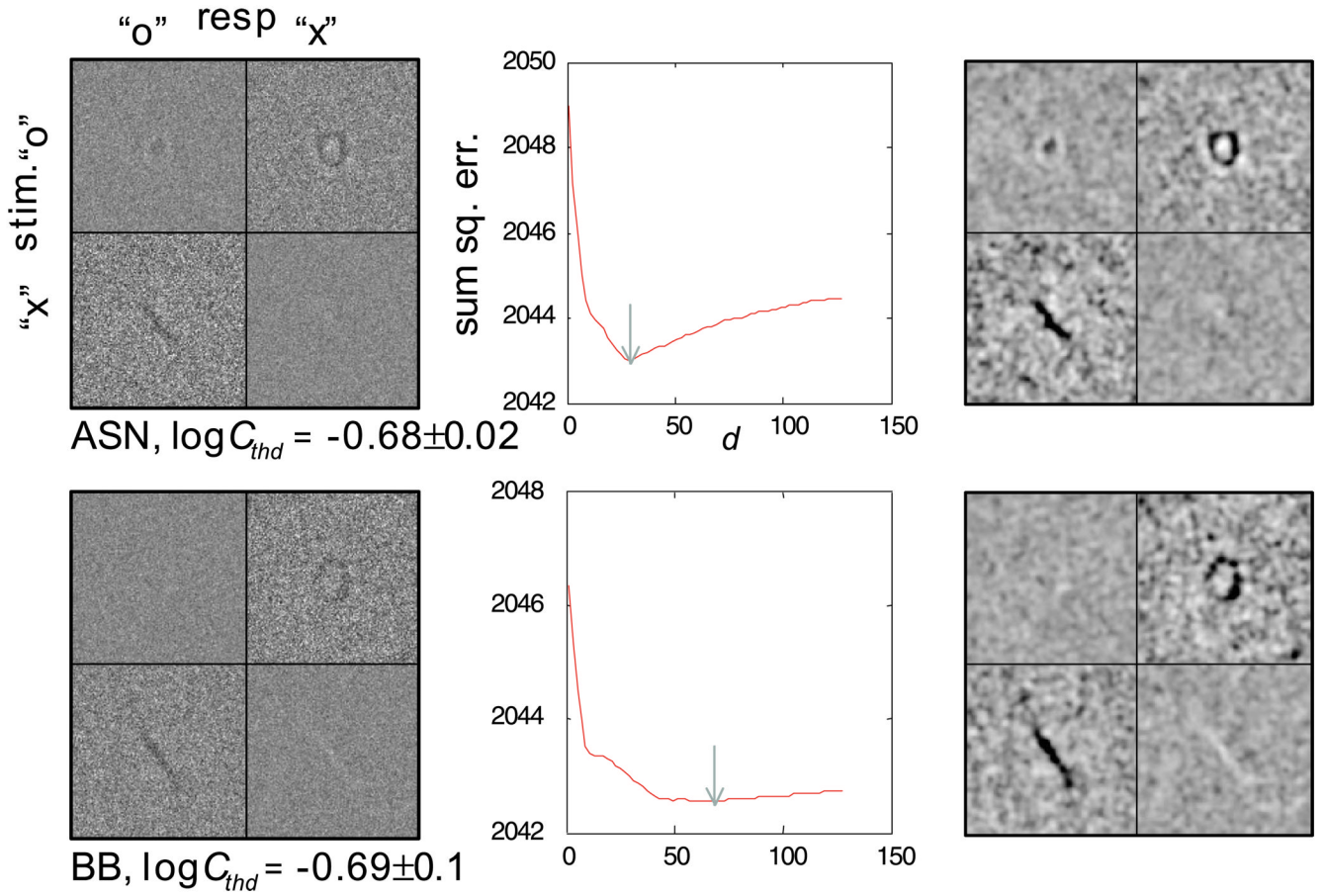


**Figure 6.**

Classification images for the human observers in the letter detection task (Experiment 2): top two rows, medium spatial uncertainty; bottom two rows, high spatial uncertainty; left column, classification images at a signal contrast corresponding to 75% correct; middle column, estimation of spatial extent of the uncertainty from the classifications images in the left column; the estimated value of  $d$  with the minimum residual error is marked by the gray arrows; right column, blurred versions of the classifications images in the left column using a Gaussian kernel of space constant equal to 14.1 stimulus pixels to visualize the positive haze in the false-alarm trials.

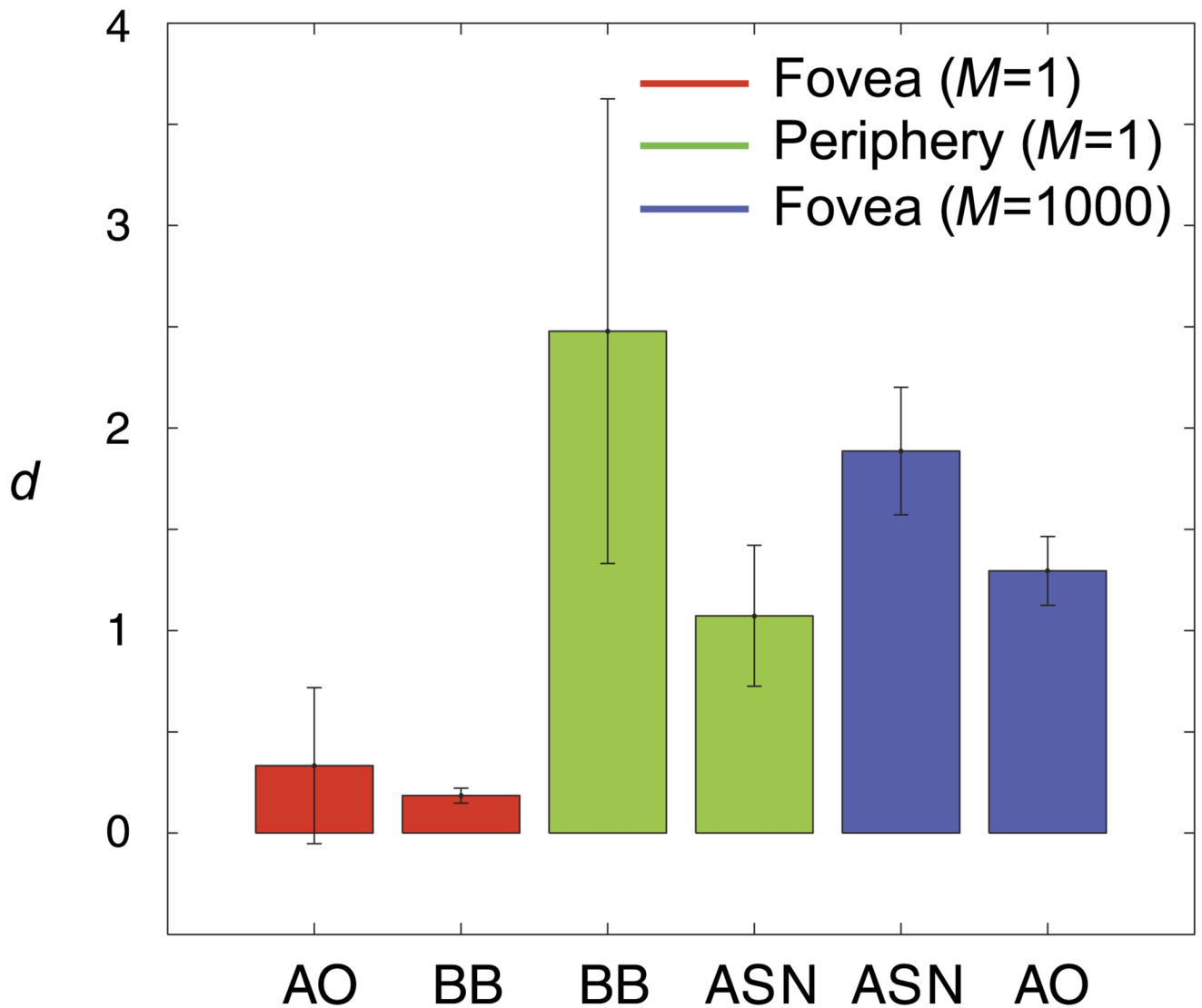


**Figure 7.** Plot of rSNR versus number of trials for subject J.H. who participated in both conditions of Experiment 2. The gray arrow marks the approximate number of trials needed in the high-uncertainty condition to achieve the same classification-image quality as the medium-uncertainty condition.



**Figure 8.**

Classification images for the human observers performing a letter identification task in the periphery (Experiment 3) with no stimulus-level (extrinsic) spatial uncertainty: left column, classification images at a letter contrast sufficient to obtain 75% correct; middle column, estimation of the spatial extent of the intrinsic uncertainty from the classification images in the left column; the value of  $d$  with the minimum error is marked by the gray arrows; right column, blurred versions of the classifications images in the left column using a Gaussian kernel of space constant equal to 1.4 stimulus pixels for visualization.



**Figure 9.**

The spatial extent of uncertainty,  $d$ , in degrees of visual angle, for the subjects who participated in the letter identification task in the periphery (10 deg inferior field, Experiment 3) without stimulus-level spatial uncertainty (green). For comparison, the results for the same task in the fovea with (blue) and without (red) stimulus-level spatial uncertainty are also shown.

**Table 1**

The estimated extents of spatial uncertainty for conditions in Experiment 1 in units of stimulus pixels.

| Condition        | Subject | $d \pm SE$   |
|------------------|---------|--------------|
| No uncertainty   | B.B.    | $5 \pm 1.0$  |
|                  | A.O.    | $9 \pm 10.4$ |
| High uncertainty | A.O.    | $35 \pm 4.6$ |
|                  | A.S.N.  | $51 \pm 8.5$ |



**Table 2**

The estimated extents of spatial uncertainty for conditions in Experiment 2 in units of stimulus pixels.

| Condition          | Subject | $d \pm SE$     |
|--------------------|---------|----------------|
| Medium uncertainty | M.J.    | $31 \pm 22$    |
|                    | J.H.    | $31 \pm 13$    |
| High uncertainty   | J.H.    | $127 \pm 45.3$ |
|                    | B.B.    | $65 \pm 28$    |

**Table 3**

The estimated extents of spatial uncertainty for the periphery condition in Experiment 3 in units of stimulus pixels, as compared with those for the fovea conditions in Experiment 1.

| Condition                  | Subject | $d \pm SE$   |
|----------------------------|---------|--------------|
| Fovea, no stimulus         | A.O.    | $9 \pm 10.4$ |
| uncertainty (Experiment 1) | B.B.    | $5 \pm 1.0$  |
| Periphery, no              | B.B.    | $67 \pm 31$  |
| stimulus uncertainty       | A.S.N.  | $29 \pm 9.4$ |
| Fovea, high stimulus       | A.S.N.  | $51 \pm 8.5$ |
| uncertainty (Experiment 1) | A.O.    | $35 \pm 4.6$ |