Models for discriminating image blur from loss of contrast

Joshua A. Solomon & Michael J. Morgan
Centre for Applied Vision Research, City, University of London, United Kingdom

Abstract

Observers can discriminate between blurry and low-contrast images (Morgan, 2017). Wang and Simoncelli (2004) demonstrated that a code for blur is inherent to the phase relationships between localized pattern detectors of different scale. To test whether human observers actually use local phase coherence when discriminating between image blur and loss of contrast, we compared phase-scrambled chessboards with unscrambled chessboards. Although both stimuli had identical amplitude spectra, local phase coherence was disrupted by phase-scrambling. Human observers were required to concurrently detect and identify (as contrast or blur) image manipulations in the 2x2 forced-choice paradigm (Nachmias & Weber, 1975; Watson & Robson, 1981) traditionally considered to be a litmus test for "labelled lines" (i.e. detection mechanisms that can be distinguished on the basis of their preferred stimuli). Phase scrambling reduced some observers' ability to discriminate between blur and a reduction in contrast. However, none of our observers produced data consistent with Watson & Robson's most stringent test for labelled lines, regardless whether phases were scrambled or not. Models of performance fit significantly better when either a) the blur detector also responded to contrast modulations, b) the contrast detector also responded to blur modulations, or c) noise in the two detectors was anticorrelated.

Keywords

Detection, Modeling, Psychophysics

Introduction

When an image is blurred, its higher spatial frequencies become disproportionately attenuated relative to lower frequencies. The visual system is less sensitive to high than to medium spatial frequencies, so it can be relatively difficult to detect blur. However, as the amount of blur increases, lower and lower spatial frequencies become affected, including those near the peak of the contrast sensitivity function (CSF; Campbell & Robson, 1968), which describes how just-detectable image contrast varies with spatial frequency. Ordinary observers without optical training can easily discriminate between blurry and sharp images. Of course, they can also discriminate between low-contrast images and high-contrast images. Are these two visual tasks really different? Reviewing the literature on blur discrimination, Watson and Ahumada (2011) found that, to a first approximation, just-detectable changes in image blur could be predicted from the CSF. Consequently, they suggested that the visual system might have no mechanism capable of detecting blur *per se*. What it does have is a mechanism capable of discriminating between different levels of image contrast, and it uses that mechanism to discriminate between different levels of image blur.

To avoid any misunderstanding, please note that this paper is concerned with blurry images in normal viewing conditions. Although the best-fitting Gaussian blur kernel has become one of the standard metrics for quantifying all forms of blur (e.g. Levi & Klein, 1990; Watson & Ahumada, 2011), optical blur, such as that caused by retinal defocus, cannot be described as "Gaussian" with 100% accuracy (Cholewiak, Love, & Banks, 2018).

Morgan (2017) found that human observers can not only discriminate between different levels of contrast and blur, they can also discriminate between these two image manipulations, possibly by using a computation of edge blur that makes it independent of contrast (Watt & Morgan, 1983). Wang and Simoncelli (2004) also suggested that blur perception might be influenced by local computations of spatial phase near image contours (such as the edges between the black squares and white squares in Morgan's chessboard-like stimuli). We present a test of this hypothesis below, using phase-scrambled and unscrambled chessboards. Although both types of stimulus have identical amplitude spectra, phase-scrambled chessboards do not have well-defined edges (see Fig. 1).

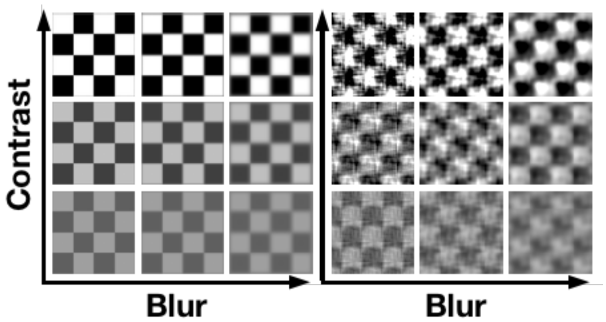


Fig. 1. Example baseline stimuli (i.e. without modulation). Three levels of blur are fully crossed with three levels of contrast in each nine-panel array. Left array: unscrambled chessboards; right array: phase-scrambled chessboards.

In the experiment we report here, observers were required to concurrently detect and identify (as contrast or blur) image manipulations in the two-by-two forced-choice (2 × 2FC) paradigm (Nachmias & Weber, 1975; Watson & Robson, 1981), traditionally considered to be a litmus test for "labelled lines." (i.e. detection mechanisms that can be distinguished on the basis of their preferred stimuli).

According to one review article (Rose, 1999), different philosophers meant different things when they invoked labelled lines, but the reader might imagine tiny signs attached to each neural fiber, describing the stimuli that match its receptive field. Of course, no contemporary scientists actually believe our brains contain homunculi capable of reading tiny signs like that. Instead, information regarding stimulus identity is thought to be inherent in the cerebral positions of active neurons. That is why stimulus preferences vary systematically in the cortex, forming multi-dimensional "maps" of retinal position, spatial orientation, and possibly other stimulus attributes such as spatial frequency, binocular disparity, and chromaticity.

This paper is concerned with selectivity and labelling. Our methodology is psychophysical rather than physiological. Accordingly, we will discuss our findings in terms of channels rather than sensory neurons, but — other than the latter's restriction to (or selectivity for) a relatively small region in the visual field — the two ideas are virtually interchangeable. Like sensory neurons, channels transform sensory information. That is, they both perform a kind of computation. Input to the computation varies with the similarity between the preferred stimulus and the actual stimulus, and output increases monotonically with input.

For non-zero channel input, some aspect of the stimulus must be modulated. Spatial-frequency channels (Campbell & Robson, 1968), for example, obtain non-zero input from modulations in stimulus luminance. Although, by definition, these channels are selective for certain periodicities of luminance modulation, spatial-frequency channels do not have infinitely narrow bandwidth. Thus, if we were to increase the modulation depth (i.e., the contrast) of a sinusoidal luminance grating, we would excite more and more channels whose preferred stimuli are less and less similar. If it were possible to isolate a channel with psychophysics, it would require a stimulus with very little contrast. In the limit, i.e., if the stimulus were just detectable, it is conceivable that it would excite only one channel. Consequently, it wouldn't be unreasonable to describe that channel as a labelled line if the brain could successfully identify a just-detectable stimulus.

At least, that's the logic used by Nachmias and Weber (1975), when they introduced what later became known as the 2 × 2FC paradigm, a variant on the more popular, two-alternative forced-choice (2AFC) paradigm. In addition to deciding whether a small patch of grating was presented within the first or second of two temporal intervals (a "detection" task), Nachmias and Weber's observers had to decide whether the grating contained relatively high or low spatial frequencies. This latter task can be considered "discrimination" or "identification" or "classification" or "categorization." We will use all the latter terms interchangeably.

Rather than present data from Nachmias and Weber's original paper, we shall present data from a follow-up study by Watson and Robson (1981). The task was virtually identical, except Watson and Robson manipulated temporal frequency rather than spatial frequency. Their chief innovation was to establish two quantitative criteria for psychophysical channels to qualify as differently labelled lines. The first criterion is that the identification thresholds must not be significantly higher than the detection thresholds. The second criterion will be discussed below.

Among the channels that satisfied the first of Watson and Robson's criteria were those responsible for discriminating between 0 Hz (or static) Gabor patterns and otherwise identical Gabors flickering at 8 Hz. Blue points in Fig. 2a show the relationship between the contrast (i.e. the modulation depth) of the static Gabor and observer ABW's ability to determine whether it was in the first or second temporal interval. Blue points in Fig. 2b show the analogous relationship for the flickering Gabor. Black points in these two panels show how frequently the Gabors were correctly identified as "static" or "flickering." We have fit these psychometric data with four smooth (Weibull) functions, all of which were constrained to have the same basic shape and upper asymptote. (Pattern detection was a well-studied task, and there was ample empirical support for fixing the Weibull shape parameter at $\kappa = 3.5$; Robson & Graham, 1981. Note also that whereas logic dictates the blue curves must share a lower asymptote of 0.5, the lower asymptotes of the black curves need only sum to 1.) Although the black curve in Fig. 2a has a slight rightward shift with respect to the blue curve, a likelihood-ratio test (Mood, Graybill, & Boes, 1973) reveals this shift to be insignificant [$\chi^2(1) = 0.05, p = 0.825$].

Thus, these data were not inconsistent with Watson and Robson's (1981) first criterion for detection by differently labelled lines.

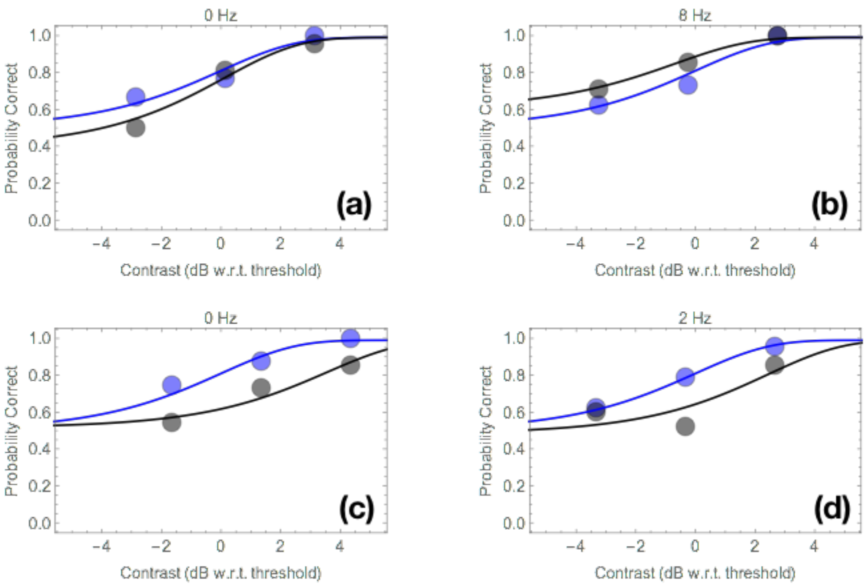


Fig. 2. 2 × 2FC results from Watson and Robson (1981). Panels (a) and (b) illustrate results in which observer ABW had to detect a Gabor pattern and identify its temporal frequency as either 0 Hz or 8 Hz. Panels (c) and (d) illustrate analogous results with Gabor patterns having temporal frequencies of either 0 Hz or 2 Hz. Blue symbols indicate detection performance and black symbols indicate identification. Smooth curves are maximum-likelihood Weibull distributions (all having shape κ = 3.5). All symbols have been shifted laterally by the Weibull scale parameter (λ), which can be considered the observer's 81%-correct detection threshold. Consequently, all blue curves are identical and contain the point (0, 0.81). Note that 0.5 is the minimum probability correct in the detection task. We further assume that the maximum is somewhat less than 1, due to attentional lapses and/or "finger errors." Thus, the blue curves have been scaled to span the interval (0.5, 0.99). There is no corresponding minimum for the discrimination task, thus the black curves in (a) and (b) have been scaled to span the intervals (γ, 0.99) and ($1 - \gamma$, 0.99), respectively; where the guess-rate γ was fit simultaneously with the Weibull scale parameters. Black curves in (c) and (d) were obtained in the analogous fashion.

Data illustrated in Figs. 2c and 2d were collected in an analogous experiment, where the flicker was only 2 Hz. In this case, the black curves have a significant rightward shift with respect to the blue curves, and thus these data do not satisfy Watson & Robson's first criterion for detection by differently labelled lines. One possibility is that both stimuli were (at least sometimes) detected by the same channel. Other possibilities are discussed below.

Whereas Watson & Robson examined selectivity and labelling in channels stimulated by different frequencies of luminance modulation, our goal was to examine selectivity and labelling in channels stimulated by modulations of stimulus contrast and stimulus blur. Both types of modulation are illustrated in Figure 3. Given sufficient time for inspection, all readers should be able to discriminate between the two dimensions of modulation.

General Methods

The methods for this study were reviewed and approved by The School of Health Science (Reference no. ETH1819-1850), City, University of London. The observer's head was placed on a chinrest with an adjustable forehead rest. Viewing was binocular, through the observers' natural pupils. Steady fixation was neither encouraged nor discouraged. An Apple computer controlled stimulus presentations and response collection. The experimental protocol was implemented using the PsychToolbox (Brainard, 1997; Pelli, 1997). (Software will be made available upon request.) Maximum and minimum luminances were 149.8 and 0.277 cd/m^2, respectively. The screen's background luminance was set to the midpoint of these values, and the rest of the room was dark.

All stimuli were based on simple, 4 × 4 chessboards, like those in Fig. 1. Each chessboard had random polarity; the lower right square could be white or black, with equal probability. The amplitude spectrum of each phase-scrambled chessboard was equal to that of an unscrambled chessboard. In all other respects, the methods for phase-scrambled chessboards were identical to those for unscrambled chessboards.

In an attempt to foil "context-coding" (Durlach & Braida, 1969) detection strategies based on a chessboard's (or one of its arbitrarily chosen square's) average or total blur -- or average or total contrast -- we randomly interleaved baseline levels along these stimulus dimensions. On each trial, we exposed one modulated chessboard and one unmodulated chessboard for 1.43 s, with a 1.43-s gap between the two successive exposures. Each chessboard had a one of three randomly and independently selected levels of "baseline" Gaussian blur, and each had one of three randomly and independently selected levels of baseline Michelson contrast. Gaussian blur kernels had spatial extents (σ) equal to 1/16th, 1/8th, or 1/4th the length of one of the chessboard's 16 squares; these spatial extents correspond to 5.6, 11.2, and 22.4 arcmin of visual angle. Baseline contrasts (before blurring and phase-scrambling) were 1, 0.5, and 0.25. Intermediate levels of baseline blur and contrast were comparable to those in Morgan's (2017) "standard" stimuli.

The modulated chessboard was a composite of two chessboards: alternate one-square-wide columns (starting at either the left-hand side or the right-hand side) came from the baseline chessboard, the other columns came from an otherwise-identical chessboard with either more blur or less contrast (see Fig. 3).

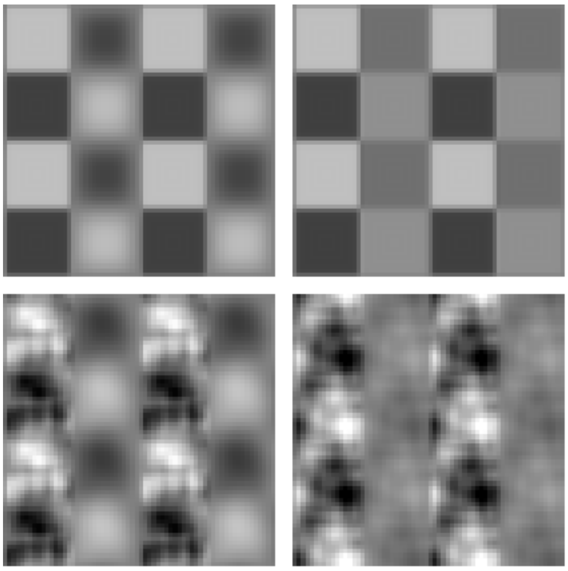


Fig. 3. Unscrambled (top) and scrambled (bottom) chessboards with heavily modulated blur (left) and contrast (right). All panels have intermediate levels of baseline blur and contrast.

Observers indicated which of the two chessboards was modulated by pressing the o key (for "one") or the t key (for "two") on the Apple's keypad. They then indicated whether the modulation was in the dimension of blur (by pressing the b key) or contrast (by pressing the c key). Immediately after this classification, two tones were played in quick succession. The frequency of each tone indicated whether the corresponding response had been correct (low tone) or incorrect (high tone). Feedback of this nature may facilitate perceptual learning and/or help to stabilize response criteria (Tanner, Rauk, & Atkinson, 1970).

For each combination of modulation identity (blur or contrast) and baseline level (low, intermediate, or high) we used two randomly interleaved Quest+ (Watson, 2017) staircases to obtain estimates of the thresholds and psychometric slopes for detection and identification, as well as the guess rate and lapse rate for identification. (Guess rate -- i.e. accuracy in the limit, as the modulation amplitude approaches zero -- is necessarily 0.5 for the detection task. Lapse rates aren't necessarily 0.01, nonetheless, we feel secure in adopting an estimate of 99% correct for the upper asymptote of our very experienced observers' psychometric functions for detection.)

Each of our four observers completed 1728 trials with unscrambled chessboards (JAS completed an extra 22 trials in a session that had to be discontinued, due to a fire alarm) divided into (eighteen) 96-trial sessions. In separate sessions, each observer completed another 1728 trials with scrambled chessboards. "U" sessions with unscrambled chessboards and "S" sessions with scrambled chessboards were run in the following sequence: USSUUSSUUSUUSSUUSS. Quest+ staircases were initialized at the beginning of session 1, and again at the beginning of session 10.

Methods Specific to Experiment 1

Both authors served as observers. Visual stimuli were presented on a gamma-linearized LCD display screen, placed at 0.845 m of viewing distance. There were 21.4 screen pixels per degree of visual angle.

Each chessboard occupied the screen's central 128 × 128 pixels. The phase spectrum of each phase-scrambled chessboard set equal to that of a 64-pixel × 64-pixel "noise image," each pixel of which had a Weber contrast that was selected independently from a zero-mean Gaussian distribution.

Methods Specific to Experiment 2

At a referee’s request, retinal resolution was increased for observers ST and AC, who were naïve to the purposes of this experiment. These 20-year-old university students had no previous experience with psychophysics. They practiced the 2x2 FC task with both scrambled and unscrambled chessboards for one hour before any data were collected. (A third naïve observer practiced for 2 hours but proved incapable of attaining 81% correct performance in the detection task. Her data are not reported here.) For these observers, the display screen was placed at 2.112 m of viewing distance. There were 53.5 screen pixels per degree of visual angle.

Each chessboard occupied the screen's central 320 × 320 pixels. The phase spectrum of each phase-scrambled chessboard set equal to that of a 160-pixel × 160-pixel "noise image," each pixel of which had a Weber contrast that was selected independently from a zero-mean Gaussian distribution.

Results

Detection

As with Watson and Robson's (1981) data (see Fig. 2), we obtained separate, maximum-likelihood fits of the Weibull distribution to each observer's probability of correctly detecting a blur modulation in scrambled and unscrambled chessboards with each level of baseline blur. Similarly, we obtained fits to each observer's probability of correctly detecting a contrast modulation with each level of baseline contrast. Unlike Watson and Robson, who could appeal to a relatively large literature on the detection of luminance modulations, we have decided to make no assumptions regarding the shape parameters of the best-fitting Weibull distributions. Consequently, it was free to vary in all our fits.

With the exception of contrast modulations in phase-scrambled chessboards, 81%-correct detection thresholds (i.e. the scale parameters of the best fitting Weibull distributions) increased disproportionately (i.e. more slowly than would be predicted on the basis of Weber's Law) with baseline levels of blur and contrast. In this paper, we will not offer any firm conclusions regarding why Weber's Law fails for these stimuli. Nonetheless, a variety of potential explanations are offered here.

For one thing, our task requires the detection of modulation away from a baseline, rather than discrimination between increments of different magnitude. And whereas the latter task can reliably produce thresholds consistent with Weber's Law (e.g. when the dimension is luminance), the former task does not (Cornsweet & Pinsker, 1965). Furthermore, not even the discrimination between different contrast increments will reliably produce thresholds consistent with Weber's Law (Nachmias & Sansbury, 1974). Finally, it must be noted that, whereas detection with the intermediate baselines almost certainly requires a visual mechanism that responds to the modulation, context-coding strategies may be used with the other baselines. For example, an observer who selected the chessboard with the greatest average blur would be relatively successful when the baseline blur was high. Consequently, with high baseline blur, the observer's 81%-correct threshold for blur modulation would be relatively low, even though the observer never really detected that modulation *per se*.

As we were particularly keen to determine whether the visual system contained labelled lines for modulations of contrast and modulations of blur, we focused the remainder of our analyses on performance with modulations away from the intermediate baselines (a Gaussian blur kernel with $\sigma = 11.2$ arcmin and a contrast of 0.5), where context-coding strategies were unlikely to facilitate performance.

Just-noticeable Weber fractions (JNWFs; Solomon, 2010) are shown in Fig. 4. Each JNWF is the ratio between the 81%-correct detection threshold and the baseline (a.k.a. “pedestal”) level of blur or contrast. The younger, naïve observers were significantlty more sensitive (they had smaller JNWFs) to blur modulations on scrambled chessboards than the authors. This may be related to their use of relatively high-resolution stimuli (see Methods Specific to Experiment 2, above).

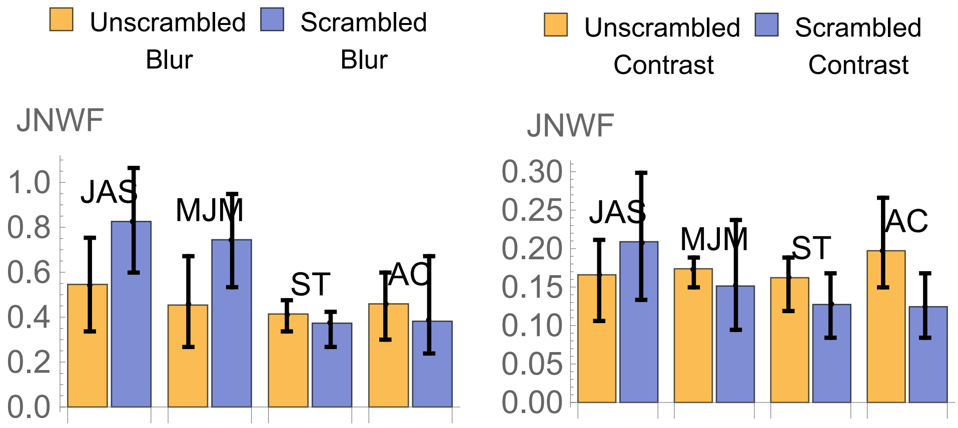


Fig. 4. Just-noticeable Weber fractions for detecting contrast and blur modulations away from the intermediate baselines illustrated in Fig. 3. Error bars contain 95% credible intervals.

For the purposes of illustration, we have provided detailed results from one observer in Fig. 5. Results for the other observers appear in Appendix A. The format of Fig. 5 is analogous to that of Fig. 2. Specifically, the blue points in Fig. 5a show the relationship between the modulation depth of blur in an unscrambled chessboard and MJM's ability to detect whether it was in the first or second temporal interval. Blue points in Fig. 5b show the relationship between detection and the modulation depth of contrast. Smooth curves show the maximum-likelihood Weibull fits. Figs. 5c and 5d illustrate corresponding results that were collected using phase-scrambled chessboards.

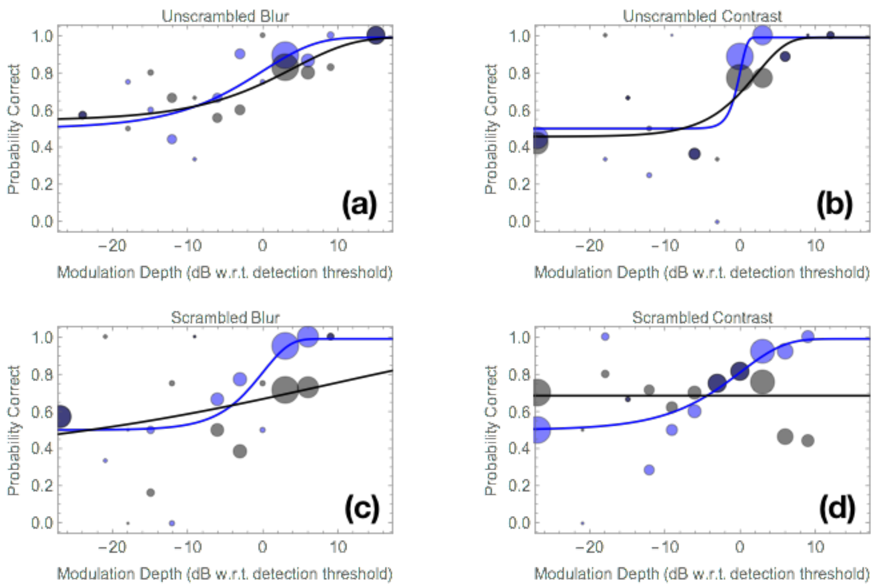


Fig. 5. 2 × 2FC detection (blue) and identification (black) of modulations away from the intermediate baseline levels. Panels (a) and (b) illustrate results in which observer MJM had to detect the

modulation in an unscrambled chessboard and identify its dimension either blur or contrast. Panels (c) and (d) illustrate analogous results with phase-scrambled chessboards. Symbol diameter is proportional to the number of trials. Smooth curves are maximum-likelihood Weibull distributions with unconstrained shape parameters. All other formatting conventions identical to those in Fig. 2.

Identification

In some cases (MJM scrambled contrast, ST unscrambled blur) it proved impossible to measure a threshold modulation depth for identification: the psychometric functions were flat (see Figs. 5d and A1e). In 12 of the remaining 14 cases, threshold for identification was greater than threshold for detection (exceptions were JAS unscrambled contrast and AC unscrambled contrast). Likelihood-ratio tests indicate a significant $[\chi^2(1) > 3.84, p < 0.05]$ difference between thresholds in 9 of the aforementioned 12 cases. It is noteworthy that all three exceptions occurred with unscrambled chessboards (MJM contrast, MJM blur, JAS blur). Consequently, it seems safe to conclude that the removal of edge information (via phase scrambling) decreased our observers' ability to identify the dimension of modulation as "blur" or "contrast." In other words, this rather superficial summary of our results is broadly consistent with the hypothesis that edges are important for the visual discrimination between blur and loss of contrast. Observers were capable of *detecting* a modulation in stimulus contrast or blur, but their ability to *identify* that modulation as such seems to have been compromised, even when that modulation was several decibels[i] above the threshold for detection.

Models

(1) High Threshold Theory

The model

Figs. 5a and 5b reveal that, when edges were present, MJM wasn't significantly worse at identifying the dimension of modulation (i.e. blur or contrast) than he was at determining whether that modulation occurred in the first or second temporal interval. What Figs. 5a and 5b do not reveal is whether or not MJM got the dimension and the interval correct on the same trials. Of course, there is no reason that an error in one task must accompany an error in the other task, but to quantify the conditional probabilities we need a model. One such model was offered by Watson and Robson (1981). Its basis is High Threshold Theory, which can be stated quite succinctly: a stimulus modulation might or might not excite any channel, but channels are never excited in the absence of stimulus modulation.

Within the framework of High Threshold Theory, a channel can be considered a labelled line if its excitation ensures correct identification. Obviously, this cannot be possible if the same channel can be excited by different types of modulation. Accordingly, when establishing their second and more stringent criterion for detection by differently labelled lines, Watson and Robson (1981) assumed "no overlap" between channel sensitivities. Given this assumption, only two parameters are required to calculate the likelihoods of all four possible outcomes in any trial:

O_1–Correct interval, correct identity.

O_2–Correct interval, incorrect identity.

O_3–Incorrect interval, correct identity.

O_4–Inorrect interval, incorrect identity.

If, on the other hand, the joint likelihood of trial outcomes is significantly better fit by a more saturated model (i.e., with three free parameters per modulation depth), then we must reject the idea that excitation ensures correct identification. Accordingly, Watson and Robson's second criterion for detection by channels with labelled lines is that the saturated model does not provide a significantly better fit. Note that if we are to maintain the assumption of no overlap between channel sensitivities, then the saturated model's third free parameter can be considered a "fudge factor," allowing observers to mis-identify an arbitrary proportion of stimuli that nonetheless do succeed in exciting a channel. Instead, we prefer to relax the assumption of no overlap.

The full high-threshold model can be described as follows. Let p_{ijk} denote the probability that stimulus *i* excites channel *k* when it has a modulation amplitude of *j*. Channels are "labelled," such that $i,k \in \{1,2\}$. On each trial there are four mutually exclusive possibilities: Channel *k* is excited, Channel l is excited ($l = 3 - k$), both are excited, and neither is excited. The corresponding probabilities are:

$$q_1 = p_{ijk}\left(1 - p_{ijl}\right), \tag{1}$$

$$q_2 = p_{ijl}\left(1 - p_{ijk}\right), \tag{2}$$

$$q_3 = p_{ijk} p_{ijl}, \tag{3}$$

and

$$q_4 = \left(1 - p_{ijk}\right)\left(1 - p_{ijl}\right). \tag{4}$$

Let r_1 and r_2 denote the probabilities of selecting interval 1 and interval 2, respectively, in the absence of any excitation, such that $r_2 = 1 - r_1$. For stimuli in interval *m*, the outcome probabilities are:

$$P\left(O_1\right) = q_1 + b_i q_3 + n_i r_m q_4, \tag{5}$$

$$P(O_2) = q_2 + (1-b_i)q_3 + (1-n_i)r_m q_4 , \tag{6}$$

$$P(O_3) = n_i r_{3-m} q_4 , \tag{7}$$

and

$$P(O_4) = (1-n_i) r_{3-m} q_4 , \tag{8}$$

where b_i and n_i are the probabilites that stimulus i is selected when both channels are excited and neither channel is excited, respectively. (NB: $b_{3-i} = 1 - b_i$ and $n_{3-i} = 1 - n_i$.) An observer can be considered unbiased when $r_1 = r_2 = b_1 = b_2 = n_1 = n_2 = 1/2$.

The results were fit assuming that the probability of channel excitation increased as a Weibull function of the stimulus modulation, i.e.

$$p_{ijk} = (1-\delta)\left(1 - \exp\left[-\left(j/\lambda_{ik}\right)^{\kappa_k}\right]\right). \tag{9}$$

Note that there are three free parameters in Eq. (9). The Weibull function's scale parameter λ_{ik} can be considered channel *k*'s sensitivity to modulations in stimulus dimension *i*. The Weibull function's shape parameter κ_k, on the other hand, is independent of stimulus dimension *i*. It describes the relationship between input and output within channel *k*. Attentional lapses and finger errors can be accommodated by allowing the remaining parameter to exceed zero (i.e. $\delta > 0$). This parameter was not allowed to vary across the dimension of modulation, as different dimensions were randomly interleaved in our procedure.

Model fits

Although it is conceivable that observers used the same computations (i.e., the same channels) for scrambled and unscrambled chessboards, nothing in our methods encouraged them to do so. Consequently, we decided that the data collected with scrambled chessboards should be fit separately from the data collected with unscrambled chessboards.

Two fits of the high-threshold model to MJM's data with unscrambled chessboards are shown in Figs. 6a and 6b. (Analogous fits to the other observers' data appear in Appendix A.) Solid curves illustrate fits of the most general version of the model, without the restriction on overlapping sensitivities. Dashed curves illustrate fits of a nested model, in which overlap was prohibited by setting $\lambda_{12} = \lambda_{21} = \infty$. Examination of the right-hand sides of the dashed curves reveals that, on trials in which observers selected the correct temporal interval, the nested model's predictions for the probability of a correct identification [i.e., *P*(Identification|Detection)] tend to be a little too high.

Nonetheless, overall, this version of High Threshold Theory seems to fit the data obtained with unscrambled chessboards fairly well.

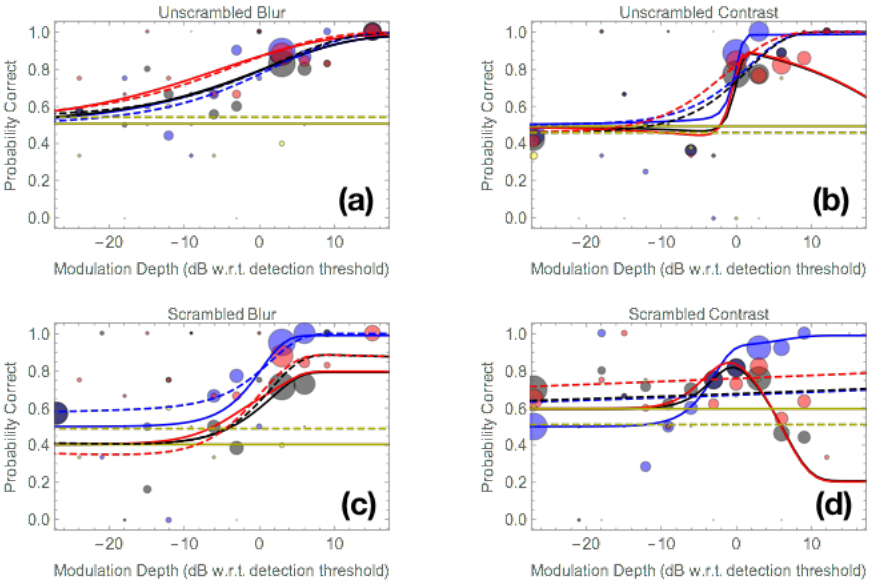


Fig. 6. Conditional probabilities fit with High Threshold Theory. As in Fig. 5, here the blue and black symbols indicate MJM's detection and identification performances, respectively. Red and amber symbols indicate the conditional probabilities *P*(Identification|Detection) and *P*(Identification|~Detection), respectively. The relative paucity of amber symbols is due to the small number of trials in which identification was successful, even though detection was not. Solid curves illustrate maximum-likelihood fits, allowing for overlap in the two channels' sensitivities (see text). Dashed curves illustrate maximum-likelihood fits without overlap.

The nested model cannot achieve anywhere near as good a fit to results obtained with phase-scrambled chessboards. It radically underestimates the difference between (unconditional) probabilities of detection and identification (note the similarity between dashed blue and black curves in Fig. 6d, they're virtually identical and almost flat; compare with the blue and black curves in Fig. 5d). It should be apparent that the model fits significantly better when channels are allowed overlapping sensitivities. Indeed, a generalized likelihood-ratio test indicated a significant improvement [$\chi^2(2) > 6, p < 0.05$] for each observer with each type of chessboards (i.e. even the unscrambled ones). Thus, *none* of our results satisfy Watson and Robson's (1981) second criterion for detection of blur and contrast modulations by differently labeled lines.

(2) Signal Detection Theory

The model

Signal Detection Theory (Green & Swets, 1966) was developed as an alternative to High Threshold Theory, which proved to be inconsistent with several empirical results (e.g. better-than-chance second responses in mAFC detection experiments, when $m > 2$, Swets, Tanner, & Birdsall, 1961; Solomon, 2007). In this section, we use Signal Detection Theory to describe the detection of modulations along any arbitrary stimulus dimensions A and B. Output from channels in this model can be used for both detection and identification within the 2 × 2FC paradigm.

Although the stimulus dimensions A and B are arbitrary, in this paper they can be understood as blur and contrast, respectively. Consider a sinusoidal modulation along dimension A. Its amplitude and phase are a and θ_{A}, respectively. A general formula for the expected output of a linear mechanism is $a\alpha\cos\left(\theta_{\mathrm{A}} - \theta_0\right)$, where α is the mechanism's sensitivity (or "gain") and θ_0 is its preferred phase.

Phase-independence (and square-law transduction) can be achieved using a non-linear transformation of the output from a quadrature pair of linear mechanisms:

$$\begin{aligned}&\left[a\alpha\cos\left(\theta_{\mathrm{A}} - \theta_0\right)\right]^2 + \left[a\alpha\cos\left(\theta_{\mathrm{A}} - \theta_0 - \frac{\pi}{2}\right)\right]^2 \\ &= a^2\alpha^2\left[\cos^2\left(\theta_0 - \theta_{\mathrm{A}}\right) + \sin^2\left(\theta_0 - \theta_{\mathrm{A}}\right)\right] \; . \\ &= a^2\alpha^2\end{aligned} \qquad (10)$$

Arbitrary power-law transduction can be achieved without sacrificing phase-independence by raising this expression to the arbitrary power $p/2$ [ii].

Now consider two sinusoidal modulations having the same frequency, one along dimension A and one along dimension B. Amplitudes and phases are a and b and θ_{A} and θ_{B}, respectively. A general formula for the expected output of a linear mechanism is $a\alpha\cos\left(\theta_{\mathrm{A}} - \theta_0\right) + b\beta\cos\left(\theta_{\mathrm{B}} - \theta_0\right)$ where α and β are the mechanism's sensitivities and θ_0 is its preferred phase. Again, phase independence (and square-law transduction) with respect to θ_0 can be achieved using a quadrature pair:

$$\begin{aligned}&\left[a\alpha\cos\left(\theta_{\mathrm{A}} - \theta_0\right) + b\beta\cos\left(\theta_{\mathrm{B}} - \theta_0\right)\right]^2 \\ &\quad + \left[a\alpha\sin\left(\theta_{\mathrm{A}} - \theta_0\right) + b\beta\sin\left(\theta_{\mathrm{B}} - \theta_0\right)\right]^2 \; , \\ &= a^2\alpha^2 + b^2\beta^2 + 2a\alpha b\beta\cos\Delta\theta\end{aligned} \qquad (11)$$

where $\Delta\theta = \theta_{\mathrm{A}} - \theta_{\mathrm{B}}$. This too can be raised to the arbitrary power $p/2$, if necessary.

Putting it all together, we can write

$$\mu_X = \left(a^2\alpha^2 + b^2\beta^2 + 2a\alpha b\beta\cos\Delta\theta\right)^{\frac{p}{2}} \tag{12}$$

for the expected output from a quadrature pair, given two sinusoidal inputs with amplitudes *a* and *b* and phase angle $\Delta\theta$.

Detection in the 2 × 2FC and 2AFC paradigms is determined on the basis of the difference between outputs to the first and second interval. In this paper, we use the random variable *X* to represent this differential output. Without loss of generality, we may assume that the variance is $\sigma_X^2 = 1$.

Now consider another mechanism, with expected output $\mu_Y = \left[a^2\alpha'^2 + b^2\beta'^2 + 2a\alpha' b\beta'\cos\Delta\theta\right]^{\frac{p'}{2}}$, variance $\sigma_Y^2 = 1$, and covariance $\mathrm{cov}\left(X,Y\right) = \rho$. This mechanism is identical to the first, except for different gains and a possibly different power-function transducer.

Both mechanisms may be used for the task of detection. The simplest decision rule is linear. Imagine the plane of all possible outputs $\left(X,Y\right)$ and divide it into two regions with the line $y = m_\theta x + b_\theta$. The observer should select interval 1 if and only if output $\left(x,y\right)$ lies in the region below the line. Detection will be unbiased only if $m_\theta = -1$ and $b_\theta = 0$.

These same two mechanisms can be used for discrimination. Again, the simplest decision rule is a line $y = m_\phi x + b_\phi$ separating each of the aforementioned two regions into quadrants (see Fig. 7 for an illustration). For the unbiased observer, $m_\phi = -1$ and $b_\phi = 0$.

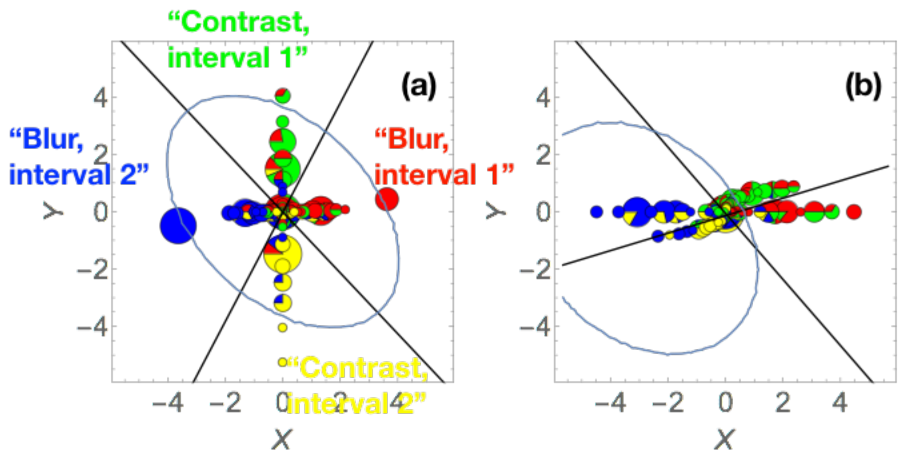


Fig. 7. Graphical interpretation of the signal-detection model's fit to one observer's results with unscrambled (a) and phase-scrambled (b) chessboards. Each piechart represents one combination of

modulation interval (1 or 2), dimension of modulation (blur or contrast), and modulation depth. Larger piecharts indicate more trials. Red, green, blue, and yellow sectors illustrate the frequencies with which observer MJM selected each of the four possible responses, as indicated in panel (a). The horizontal position of each piechart shows the X channel's expected output, and the vertical position shows the Y channel's expected output. Not shown are Gaussian blobs centred on each one of these piecharts. Each blob describes the density of the joint likelihood for the two channels' responses. That likelihood has unitary standard deviation in each dimension (X and Y) but its covariance was left as a free parameter. MJM's data were best fit with a negative covariance; perhaps there was some competition between channels. Covariance is illustrated by the ellipses, which describe four standard deviations in every direction around the origin in panel (a) and the point (–2.91, –0.93) in panel (b), whose coordinates correspond to the expected channel outputs for a first-interval contrast modulation having a depth that is 10 dB greater than MJM's detection threshold. We have assumed that observers divide the space of all possible channel outputs into the four types of response. The simplest possible decision rule uses two linear discriminants. These are represented by the lines in each panel. Sample outputs in the right quadrant are classified as "Blur, interval 1," sample outputs in the top quadrant are classified as "Contrast, interval 1," and so on (as indicated in panel a).

As illustrated in Fig. 7, each trial can be considered one sample from a joint density function on the plane of all possible channel outputs. If there were no attentional lapses or finger errors, the probability of any specific response (e.g. "blur, interval 1") would correspond to the fraction of that density function that lies within the quadrant associated with that specific response. However, when fitting the model, we allowed for the possibility of a non-zero lapse rate, i.e. a proportion of trials (denoted δ') on which the observer selects one of the four possible responses at random (regardless of the modulation depth, with probability $1/4$).

Model fits

We used Mathematica's implementation of Brent's (2002) principal-axis method to find maxima (with 2 digits of accuracy) in the function mapping parameter values to log likelihood. The full signal-detection model has 12 free parameters: four (m_θ, b_θ, m_ϕ, and b_ϕ) for the discriminant lines, plus one (δ') for the lapse rate, plus one (ρ) for the channel covariance, plus two (p and p') for the power-function transducers, plus four channel gains (α, β, α', and β'). In addition to this full model, we fit a version constrained to exclude overlap between channel sensitivities (called "leakage" by Raphael & Morgan, 2016; and Morgan, 2017). Specifically, both channels were prohibited from responding to more than one dimension of modulation, i.e. $\beta = 0$ and $\alpha' = 0$. This constraint significantly reduced the model's maximum likelihood [$\chi^2(2) > 6, p < 0.05$] only for JAS's data with the unscrambled chessboards (see Fig. 9). We also fit a version constrained to exclude any correlation between channel outputs (by forcing $\rho = 0$). This constraint did not significantly reduce the model's maximum likelihood for any of the data sets [in all cases, $\chi^2(1) < 2.2, p > 0.18$; see Fig. 9]. Finally, we fit a version constrained to exclude both overlap and correlation. This constraint did significantly reduce the model's maximum likelihood for each of the data sets [in all cases, $\chi^2(3) > 8, p < 0.05$; see Fig. 9].

Psychometric functions illustrating fits of the full model appear in Fig. 8. Perhaps the most salient feature of this figure is the downward trend of some amber curves, illustrating P(Identification|~Detection). Whereas High Threshold Theory predicts that this conditional probability should be independent of modulation depth; in the absence of attentional lapses and finger errors (i.e. when $\delta' = 0$), Signal Detection Theory predicts that this conditional probability should mirror P(Identification|Detection), as modulation depth increases. Some of the amber curves have a kink on the right side, where the curve suddenly shoots back up toward a probability of 0.5. This is due to non-zero lapse rates, which are the only explanation for the failure to detect massively suprathreshold modulations.

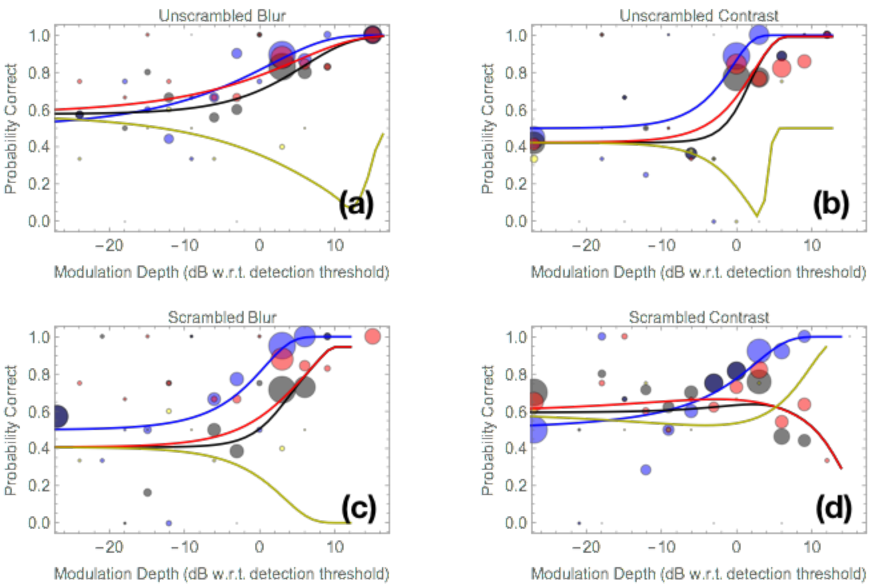


Fig. 8. Conditional probabilities fit with Signal Detection Theory. As in Fig. 5, here the blue, black, red, and amber symbols indicate MJM's P(Detection), P(Identification), P(Identification|Detection), and P(Identification|~Detection), respectively. Curves illustrate maximum-likelihood fits of the full, 12-parameter model.

A visual comparison of the amber curves with the amber points suggests little compelling evidence for P(Identification|~Detection) dropping to zero. With few exceptions, the amber symbols tend to congregate around 0.5, consistent with High Threshold Theory. However, we cannot form any firm conclusions in this regard. For each of the conditions summarised by one panel in Fig. 8, the adaptive staircases produced just 16 (out of a total 189) trials above threshold, on which MJM failed to detect the modulation. One fairly strong conclusion that can be drawn from these results is this: despite their potential value towards selecting between Signal Detection and High Threshold Theories, suprathreshold detection errors are too rare pursue with any vigor.

Perhaps surprisingly, the full signal-detection model has no trouble accounting for MJM's decline in *P*(Identification|Detection) with increasingly large modulations of stimulus contrast in scrambled chessboards (as illustrated by the red curve in Fig. 8d)[iii]. Examine Fig. 7b to see how this arises. Notice that the "X" channel has non-zero gain to both blur modulations and contrast modulations. (Contrast signals "leak" into the channel that responds to blur modulations.) Consequently, piecharts aren't confined to the vertical axis. Unlike the ellipse in Fig. 7a, which was centered on the origin, the ellipse in Fig. 7b is centered on the coordinates (–2.91, –0.93), which correspond to the expected channel outputs for a first-interval contrast modulation having a depth that is 10 dB greater than MJM's detection threshold. On trials such as these, *P*(Identification|Detection) can be visualized as ratio between two areas: the intersection between the ellipse and the bottom quadrant and the intersection between the ellipse and the union of bottom and left quadrants. This ratio is 0.47. *P*(Identification|~Detection) varies with the ratio between two different areas: the intersection between the ellipse and the top quadrant and the intersection between the ellipse and the union of top and right quadrants. This ratio is 0.87. Fig. 9 summarizes how well the various models fit each set of data.

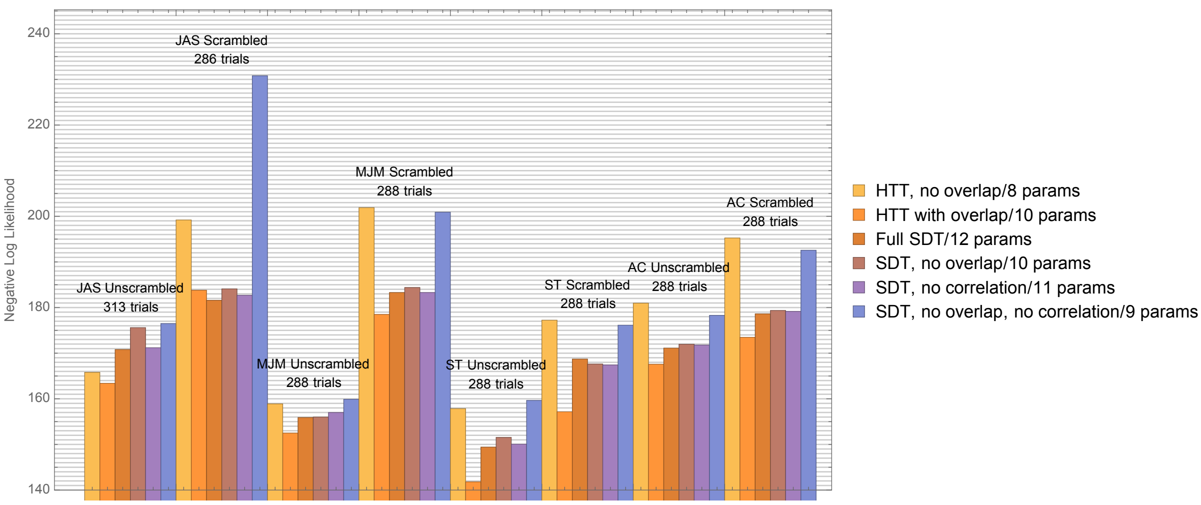


Fig. 9. Negative log likelihoods for the fit of six models to four separate data sets; all data collected with modulations away from the intermediate baseline levels.

Discussion

Some observers (e.g. the authors JAS and MJM) appear to be capable of discriminating between a reduction in contrast that is limited to the high spatial frequencies (i.e. blur) and a reduction in contrast that is uniform across the spatial frequency spectrum. However, none of our observers were capable of consistently identifying the dimension of modulation when edges were removed via phase-scrambling. When asked to do so, they adopted idiosyncratic and ineffective strategies. For example, MJM's data suggest a slight preference for labelling large modulations as "blur," but his ability to report contrast modulations as "contrast" never rose beyond a baseline frequency of about 68%, regardless of modulation depth (see Figs. 5c and 5d).

Our methodology, with its interleaved, adaptive staircases, effectively decorrelated modulation depth from modulation identity (i.e., blur vs contrast). Consequently, decisions based on the output of a single channel could not attain an identification accuracy better than 50% correct overall (i.e. when blur and contrast trials are combined). Some observers may not have attained 81% correct with modulations in stimulus blur or stimulus contrast, but all observers' identification accuracies were well in excess of 50% correct overall. Accordingly, we can reject the idea that there is just one channel. Better-than-chance identifications imply at least two.

Given the logical necessity of two channels, we must turn to theory for why identification performance with scrambled chessboards is so bad. One potential explanation is overlap between the two channels' sensitivities: at least one channel responds both to blur modulations and contrast modulations. Thus, a single modulation can excite both channels. The high-threshold model of Watson and Robson (1981) does not allow for this possibility. In our elaboration of that model, observers make an arbitrary (but possibly biased) decision regarding stimulus identity, when both channels are excited.

Sensitivity overlap can produce identity confusions within the context of Signal Detection Theory as well. When the expected response of both channels to a contrast modulation isn't very different from their expected response to a blur modulation, observers will often err in their attempt to identify the modulation. Moreover, since Signal Detection Theory's channels are never quiescent, identity confusions can arise when the channels' noises are negatively correlated. Random activity in the "blur channel" favoring interval 1 could increase the probability of random activity in the "contrast channel" favoring interval 2, and vice versa.

Our modelling addresses the relationships between modulation amplitude and decision. We have intentionally remained agnostic regarding how the visual system represents the quantities that serve as input to the blur and contrast channels. Nonetheless, it seems reasonable to assume those quantities are computed from the output of visual pattern analyzers (Graham, 1989) conjointly selective for retinal position and spatial frequency. Analyzer outputs could be weighted (or unweighted), forming an input to the contrast channel that correlates with spatial modulations in stimulus visibility. This idea is similar to the Visible Contrast Energy (ViCE) model of Watson and Ahumada (2011). Alternatively, observers may adopt a bespoke weighting of analyzers (ignoring those with preferred frequencies that are far from our chessboards' 2 cycles/image, say). We are even less certain how the visual system represents image blur. Although blur can be computed from an arithmetic combination of analyzer outputs (e.g. the difference between outputs from low-frequency and high-frequency analyzers, perhaps divided by their sum), it can also be computed from the spatial separation between maximally stimulated analyzers (Watt & Morgan, 1983; Georgeson, May, Freeman, & Hesse, 2007) or the coherence of spatial phase across different scales of analyzer, as demonstrated by Wang and Simoncelli (2004).

Although unequivocally successful identification at the detection threshold can be considered evidence in favour of labelled lines, identification errors need not imply the absence of labelled lines. Indeed, these sorts of errors are sometimes taken as evidence *for* labelled lines (e.g. Ramachandran & Hubbard, 2001; Periera & Alves, 2011). Consequently, we conclude that it would be best to compare detection and identification with regards to their implications for interactions between channels. Specifically, we can assert that channel-based models of detection are unable to satisfactorily fit our results without sensitivity overlap or anticorrelated noise. Morgan (2017) arrived at a similar conclusion (i.e., in support of signal leakage between channels) using unscrambled, black-and-white chessboards.

Whereas black-and-white chessboards can be considered relatively naturalistic stimuli, our phase-scrambled chessboards cannot; they lack well-defined edges. There is no reason to think that observers perform the same computations when making decisions about these two classes of stimulus. Indeed, the larger JNWFs unambiguously indicate lower sensitivity in the channels responsible for detecting blur in the phase-scrambled stimuli. However, for both phase-scrambled and unscrambled stimuli, the conditional probabilities indicate that successful detection does not imply successful identification. Within the context of Signal Detection Theory, these probabilities demand either sensitivity overlap or anticorrelation between the outputs from the channels responsible for detecting modulations of blur and those responsible for detecting modulations of contrast.

Appendix A: Illustrating the performances of observers JAS, ST, and AC

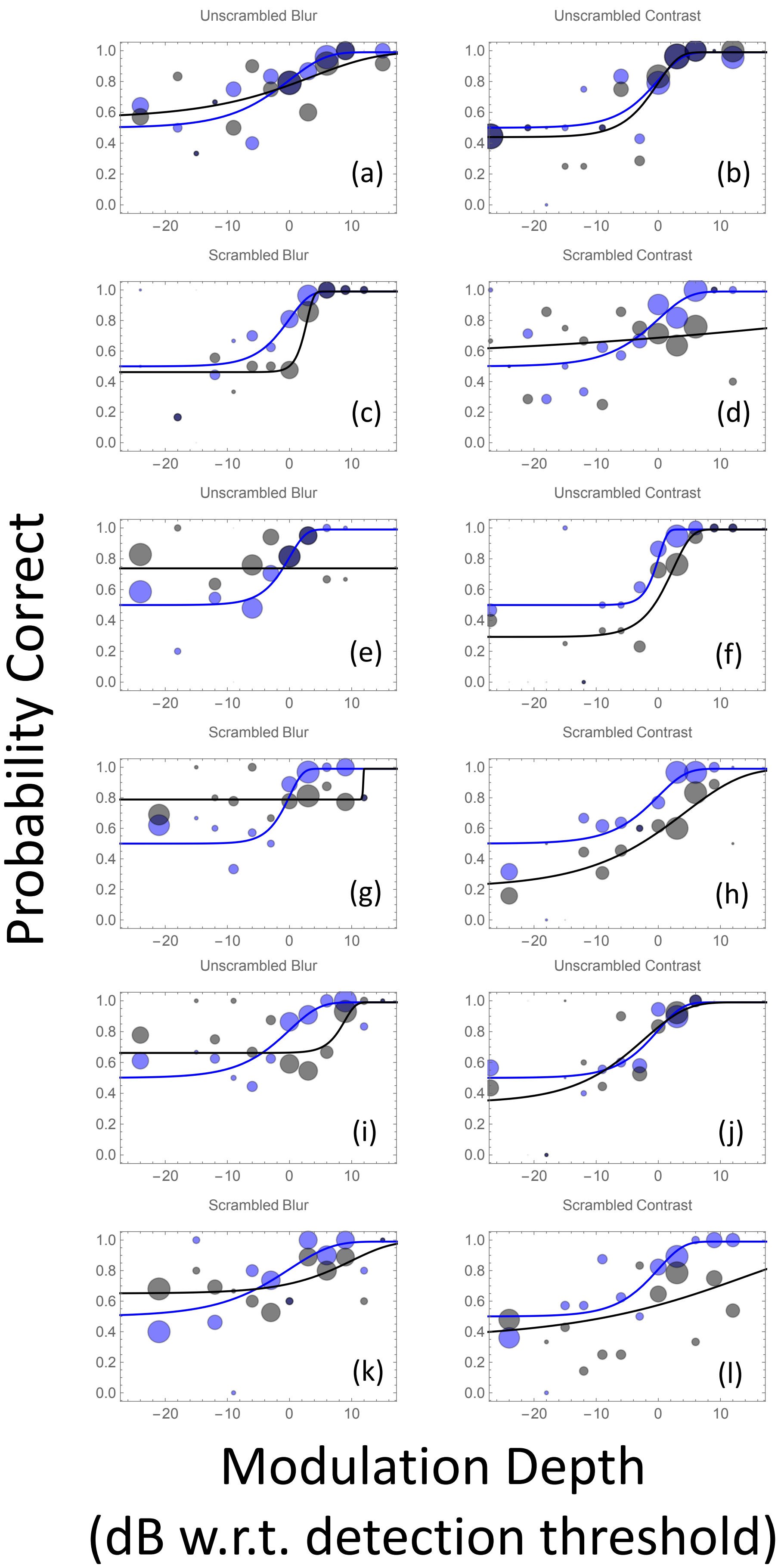
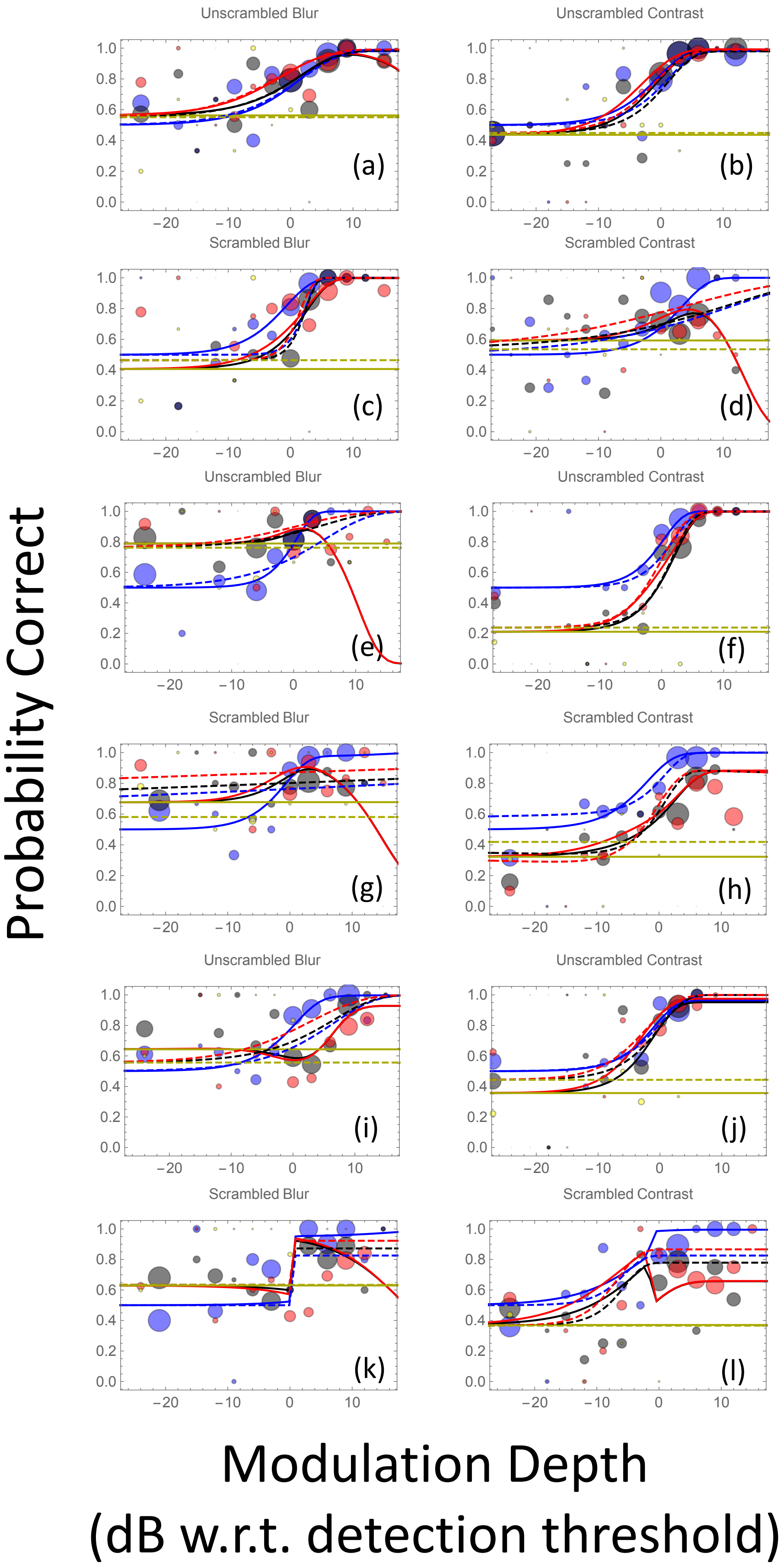

Probability Correct
Unscrambled Blur
Unscrambled Contrast
Scrambled Blur
Scrambled Contrast
(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
(i)
(j)
(k)
(l)
Modulation Depth
(dB w.r.t. detection threshold)

Fig. A1. 2×2FC detection (blue) and identification (black) of modulations away from the intermediate baseline levels. Panels (a) – (d): observer JAS; panels (e) – (h): observer ST; panels (i) – (l): observer AC. All formatting conventions identical to those in Fig. 5.



Unscrambled Blur
Unscrambled Contrast
Scrambled Blur
Scrambled Contrast
Unscrambled Blur
Unscrambled Contrast
Scrambled Blur
Scrambled Contrast
Unscrambled Blur
Unscrambled Contrast
Scrambled Blur
Scrambled Contrast
(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
(i)
(j)
(k)
(l)
1.0
0.8
0.6
0.4
0.2
0.0
−20
−10
0
10
Probability Correct
Modulation Depth
(dB w.r.t. detection threshold)

Fig. A2. Conditional probabilities fit with High Threshold Theory. As in Fig. A1, here the blue and black symbols indicate detection and identification performances, respectively. Panels (a) – (d): observer JAS; panels (e) – (h): observer ST; panels (i) – (l): observer AC. All formatting conventions identical to those in Fig. 6.

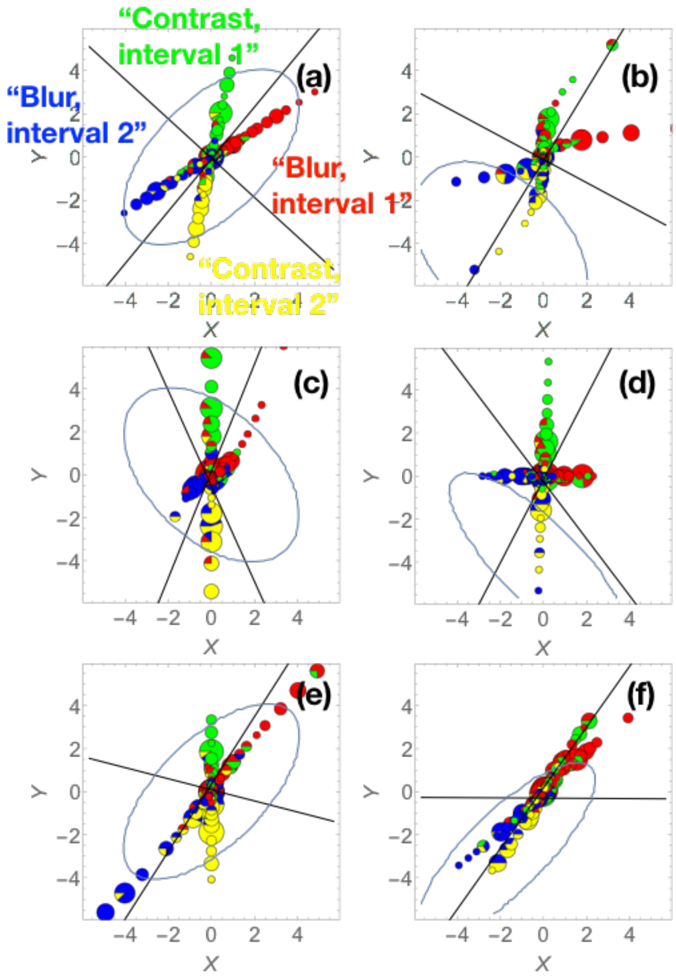
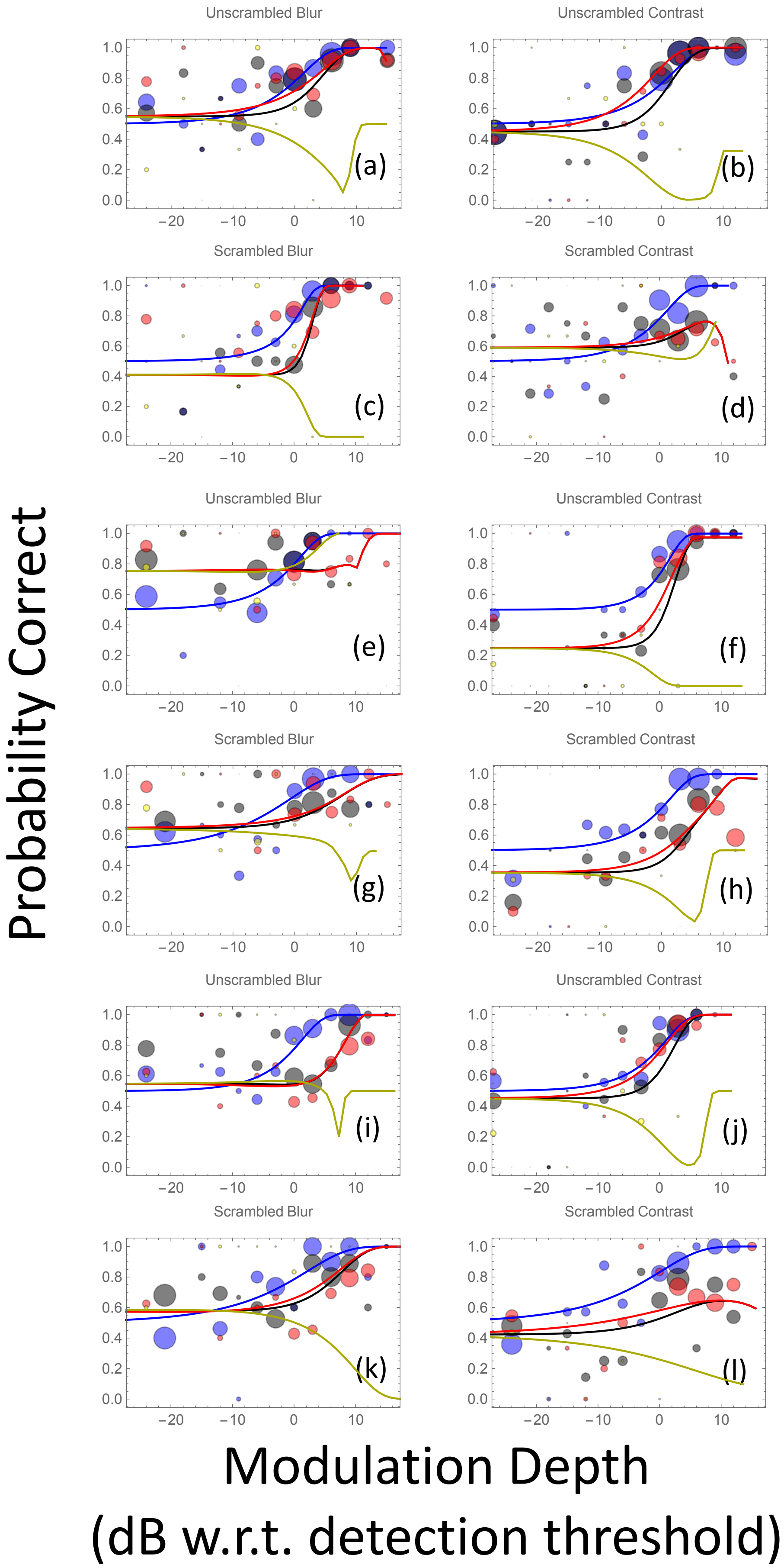


Fig. A3. Graphical interpretation of the signal-detection model’s fit to JAS’s results (panels a and b), ST’s results (panels c and d), and AC’s results (panels e and f) with unscrambled (a, c, e) and phase-scrambled (b, d, f) chessboards. The ellipses describe four standard deviations in every direction around the origin in panels (a, c, and e), In panels (b), (d), and (f), the ellipses are centred around the points (–1.99, –4.26), (–0.19, –4.02), and (–1.26, –2.80), respectively. These coordinates correspond to the expected channel outputs for a first-interval contrast modulation having a depth that is 10 dB greater than the observer’s detection threshold. All formatting conventions identical to those in Fig. 7.



Unscrambled Blur
Unscrambled Contrast
Scrambled Blur
Scrambled Contrast
Unscrambled Blur
Unscrambled Contrast
Scrambled Blur
Scrambled Contrast
Unscrambled Blur
Unscrambled Contrast
Scrambled Blur
Scrambled Contrast
(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
(i)
(j)
(k)
(l)
1.0
0.8
0.6
0.4
0.2
0.0
−20
−10
0
10
Probability Correct
Modulation Depth
(dB w.r.t. detection threshold)

Fig. A4. Conditional probabilities fit with Signal Detection Theory. As in Fig. A3, here the blue, black, red, and amber symbols indicate *P*(Detection), *P*(Identification), *P*(Identification|Detection), and *P*(Identification|~Detection), respectively. All formatting conventions identical to those in Fig. 8.

---

[i] We have adopted the decibel scale for comparing arbitrary modulation depths with the detection threshold. Thus, if λ represents the detection threshold, then the depth of any arbitrary modulation m can be described as $20 \log_{10}(m/\lambda)$ dB.

[ii] Non-linear transduction is a component common to most psychophysical models within the framework of Signal Detection Theory. Power-law transducers are particularly popular, because psychometric slope is directly proportional to the exponent. Whereas the shape of sinusoidal signal will change following non-linear transduction, the shape of a square-wave signal will not. In our experiment, we utilized square-wave modulations (see Fig. 3) to ensure observers could not use the apparent shape of the modulation as a cue to its identity (i.e. blur vs. contrast). When fitting the signal-detection model to our data, we used the square-wave amplitudes in place of the sinusoidal amplitudes a and b.

[iii] *P*(Identification) also declines. This isn't immediately apparent from Fig. 7d because the black curve is identical to (and hidden by) the red curve. Regardless of their parameters' values, both Signal Detection Theory and High Threshold Theory predict *P*(Identification|Detection) ≥ *P*(Identification).

References

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, 10, 433.

Brent, R. P. (2002). *Algorithms for Minimization without Derivatives*. Dover, Original edition 1973.

Cambell, F. W., & Robson, J. G. (1968). Application of fourier analysis to the visibility of gratings. *Journal of Physiology*, 197, 551.

Cholewiak, S. A., Love, G. D., & Banks, M. S. (2018). Creating correct blur and its effect on accommodation. *Journal of Vision*, 18(9):1, 1–29, https://doi.org/10.1167/18.9.1.

Cornsweet, T. N., & Pinsker, H. M. (1965). Luminance discrimination of brief flashes under various conditions of adaptation. *Journal of Physiology*, 176, 294.

Durlach, N. I., & Braida, L. D. (1969). Intensity Perception. I. Preliminary Theory of Intensity Resolution. *Journal of the Acoustical Society of America*, 46, 372.

Georgeson, M. A., May, K. A., Freeman, T. C. A., & Hesse, G. S. (2007). From filters to features: Scale–space analysis of edge and blur coding in human vision. Journal of Vision, 7(13):7, 1–21, http://journalofvision.org/7/13/7, doi:10.1167/7.13.7.

Green, D. M., & Swets, J. A. (1966). *Signal Detection Theory and Psychophysics*. New York, NY: Wiley.

Mood, A., Graybill, F., & Boes, D. (1973). *Introduction to the theory of statistics* 3rd edn. New York, NY: McGraw-Hill. pp. 440–441.

Morgan, M. J. (2017). Labeled lines for image blur and contrast. *Journal of Vision*, 17 (6): 16.

Nachmias, J., & Weber, A. (1975). Discrimination of simple and complex gratings. *Vision Research*, 13, 217.

Nachmias J., & Sansbury, R. V. (1974). Grating contrast: discrimination may be better than detection. *Vision Research*, 14, 1039.

Pelli, D. G. (1997). The VideoToolbox software for psychophysics: transforming numbers into movies. *Spatial Vision*, 10, 437.

Pereira, J. C. Jr., & Alves, R. C. (2011). The labelled-lines principle of the somatosensory physiology might explain the phantom-limb phenomenon. *Medical Hypotheses*, 77, 853.

Ramachandran, V. S, & Hubbard, E. M. (2001). Synaesthesia–a window into perception, thought and language. *Journal of Consciousness Studies*, 12, 3.

Raphael, S., & Morgan, M. J. (2016). The computation of relative numerosity, size and density. *Vision Research,* 124, 15.

Robson, J. G., & Graham, N. (1981). Probability summation and regional variation in contrast sensitivity across the visual field. *Vision Research*, 21, 409.

Rose, D. (1999). The historical roots of the theories of local signs and labelled lines. *Perception*, 28, 675.

Solomon, J. A. (2007). Contrast discrimination: second responses reveal the relationship between the mean and variance of visual signals. *Vision Research*, 47, 3247.

Swets, J., Tanner, W. P. Jr., & Birdsall, T. G. (1961). Decision processes in perception. *Psychological Review*, 68, 301.

Tanner, T. A., Rauk, J. A., & Atkinson, R. C. (1970). Signal recognition as influenced by information feedback. *Journal of Mathematical Psychology*, 7, 259.

Wang, Z., & Simoncelli, E. P. (2004). Local phase coherence and the perception of blur. In *Advances in Neural Information Processing Systems*. S. Thrun, L. Saul, & B. Schölkopf, Eds. Cambridge, MA: MIT Press, vol. 16.

Watson, A. B. (2017). QUEST+: A general multidimensional Bayesian adaptive psychometric method. *Journal of Vision*, 17 (3): 10.

Watson, A. B., & Ahumada, A. J. Jr. (2011). Blur clarified: A review and synthesis of blur discrimination. *Journal of Vision*, 11 (5): 10.

Watson, A. B., & Robson, J. G. (1981). Labeled lines for image blur and contrast. *Vision Research*, 21, 1115.

Watt, R. J., & Morgan, M. J. (1983). The recognition and representation of edge blur: evidence for spatial primitives in vision. *Vision Research*, 23, 1465-1477.