Available online at www.sciencedirect.com
ScienceDirect
Vision Research 47 (2007) 2912–2917
Vision Research
www.elsevier.com/locate/visres

Commentary

A local measure for modeling contrast discrimination: Response to Katkov, Tsodyks and Sagi

Stanley A. Klein

School of Optometry, University of California, Berkeley, USA

Received 14 February 2007; received in revised form 14 February 2007

Abstract

Katkov, et al. [Katkov, M., Tsodyks, M., & Sagi, D. (2007). Singularities explained: Response to Klein. *Vision Research,* preceding article] commented on my comments [Klein, S. A. (2006). Separating transducer nonlinearities and multiplicative noise in contrast discrimination. *Vision Research, 46*, 4279–4293] regarding fitting the Kontsevich, et al. [Kontsevich, L. L., Chen, C. C., & Tyler, C. W. (2002). Separating the effects of response nonlinearity and internal noise psychophysically. *Vision Research, 42*, 1771–1784] contrast discrimination data. Klein [Klein, S. A. (2006). Separating transducer nonlinearities and multiplicative noise in contrast discrimination. *Vision Research, 46*, 4279–4293] focused on the question of whether the singularity associated with the constant noise model makes it difficult to reject that model. The present paper acknowledges the presence of a singularity but shows that even in the presence of a strong singularity and using a reasonable number of 2AFC trials, it is possible to not only reject the constant noise hypothesis, but also to place confidence limits on estimates of the magnitude of the multiplicative noise. Our analysis is based on measuring contrast discrimination among triplet of contrast levels.

Keywords: Contrast discrimination; Multiplicative noise; Signal detection theory; Two-alternative forced-choice; Singularity; Gaussian noise; Hypothesis testing

1. Introduction

It is notoriously difficult to determine whether the increasing TvC function in contrast discrimination is caused by a saturating contrast response function (CRF) or by multiplicative noise. Kontsevich, Chen, and Tyler (2002, referred to as KCT) claim that their 2AFC contrast discrimination data could not be accounted for by a constant noise model and thus multiplicative noise was needed. Georgeson and Meese (2006) and Katkov, Tsodyks, and Sagi (2006a, b, referred to as KTS06) raised objections. Klein (2006) responded to those objections by fitting the KCT data using the KTS06 strategy of allowing the contrast response function to take on an arbitrary shape, constrained only by monotonicity. The chi square of my new fits, including one parameter for multiplicative noise, was very good. Table 5 of Klein (2006) showed that for three of the four KCT subjects, when the noise was constrained to be constant, the chi square jumped up to a level that could be rejected. That is, my use of the KTS06 approach was able to reject (each with $p < .05$) the constant noise hypothesis.

A problem with the KTS06 papers was their emphasis on a singularity that occurs for fitting data generated by a constant noise model. The casual reader could be left with the (incorrect) impression that the singularity would disqualify the KCT approach from being useful in testing models for contrast discrimination, contrary to Klein (2006). In the preceding article, Katkov et al. (2007, referred to as KTS07) commented on my comments. My response to KTS07 focuses on a three level dataset that illustrates most of the important points.

In order to clarify the goal of the present paper it is useful to reproduce the KTS07 abstract:

> Klein (2006) *questions the existence of intrinsic singularities in two-alternative force-choice (2AFC) signal detection theory (SDT) models, suggesting that the*

E-mail address: sklein@socrates.berkeley.edu

0042-6989/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.visres.2007.04.006

singularities found in Katkov, Tsodyks, and Sagi (2006a, 2006b) *are due to discarding higher order terms in the Taylor expansion of d′ and/or limited to steep psychometric functions. Here we provide some simple intuitive examples that illustrate the results described in* Katkov et al. (2006a, 2006b). *We show, for the constant noise model, that singularities exist when exact values of d′ are computed and that the singularities are not limited to steep psychometric functions. In these cases the disambiguation of the different models requires millions of trials.*

This paper considers the points raised by KTS07 and makes the following claims:

(1) There is indeed a singularity for the constant noise model if by singularity one means the vanishing of a linear term in a Taylor's series expansion of the quantity that measures the multiplicative noise.
(2) That singularity need not interfere with estimating the amount of multiplicative noise using 2AFC methods.
(3) The ability to place useful confidence limits on the amount of multiplicative noise depends on going away from the Taylor series approximation in the small noise regime.
(4) Rather than requiring millions of trials as suggested by KTS07, estimates of multiplicative noise can be obtained in a practical number of trials, provided that stimuli are strategically placed.
(5) The amount of multiplicative noise in one of the KCT observers was so strong that it deserves special mention. That observer yielded strong evidence of multiplicative noise in less than 1000 trials.

2. Mathematical background

The KCT experiment involves measuring the discriminability of stimuli at several contrast levels. The result of the experiments are d_{mn}, the signal detection d' values between levels m and n. The 2AFC d_{mn} values are obtained by converting the 2AFC % correct to a z-score and then multiplying by sqrt(2) (see Klein, 2001 for details). In order to discriminate among various hypotheses for contrast discrimination a model prediction, $\hat{d}_{mn}$ is needed to fit the data d_{mn}. I will follow the KCT, Klein (2006) and KTS signal detection theory model based on unequal variance Gaussians

$$\hat{d}_{mn} = (R_n - R_m)/((S_m^2 + S_n^2)/2)^{1/2} \quad (1)$$

where R_m and R_n are the contrast response functions (CRF) at the two contrast levels, m and n. We are free to define $R_1 = 0$ and $S_1 = 1$ so that

$$\hat{d}_{1n} = R_n/((1 + S_n^2)/2)^{1/2} \quad (2)$$

Signal detection theory experts will realize that R_n is d' defined at the horizontal intercept of the ROC curve (a common definition of d' in yes/no tasks (Nachmias & Kocher, 1970)) and $\hat{d}_{1n}$ is the 2AFC definition that is the distance from the origin to the ROC curve (on z-score axes) times sqrt(2).

A good way to think about Eq. (1) for optimally separated levels is that the CRF parameters, R_m, are used to fit the d' experimental 2AFC values between adjacent levels $d_{m\ m+1}$, and the noise strength parameters, S_m, are used to fit the non-adjacent data $d_{m-1\ m+1}$. For good testing efficiency it is useful to have neighboring levels separated by about $d_{m\ m+1} = 1.5$, in which case the d' for separations greater than two levels have too large a variance to be useful (see Section 3).

For simplicity the present paper will deal with just three levels (an important, brief, discussion of four levels is in Section 3), so the total experimental data consists of only three d' values: d_{12}, d_{23} and d_{13}. We are interested in the combination

$$D = d_{12} + d_{23} - d_{13} \quad (3)$$

where D can be considered as a measure of multiplicative noise, since as will be seen, $D = 0$ for the constant noise case.

The variance of D is given by the sum of the variances, σ_{mn}^2, of the three terms measuring the discriminability of levels m and n. Binomial statistics (Klein, 2001) gives

$$\sigma_{mn}^2 = 4\pi \exp(z_{mn}^2) p_{mn}(1 - p_{mn})/N_{mn} \quad (4)$$

where z_{mn}, the z-score between the two levels is: d_{mn}/sqrt(2), p_{mn} is the corresponding probability, and N_{mn} is the number of trials for the runs. The standard error of the statistic D is therefore

$$SE_D = (\sigma_{12}^2 + \sigma_{23}^2 + \sigma_{13}^2)^{1/2} \quad (5)$$

Nachmias and Kocher (1970) found that near detection threshold the ratio $k_m = S_{m+1}/S_m$ was an increasing function of the d' between levels. In order to obtain useful scaling properties across a large range of levels (for future use), I will parameterize the dependence of noise variance on d' as

$$k_m = S_{m+1}/S_m = \kappa^{d'} \quad (6)$$

where κ was found by Nachmias and Kocher (1970) to be close to 1.25 for the yes–no task (using the ROC horizontal intercept definition of d') corresponding to $\kappa = 1.3$ for the 2AFC task (see comment following Eq. (2)). I chose to put d' in the exponent rather than as a linear term in order to localize the scaling property when there are more than three levels. Note that k_m is the inverse of the ROC slope where the blank and signal distributions are levels m and m + 1, respectively.

For the three-level case to be considered in this paper, there are four free parameters R_2, R_3, S_2, and S_3 since $R_1 = 0$ and $S_1 = 1$. This is an underdetermined situation. In order to constrain the fit we will put a smoothness constraint on the k_m ratio of adjacent values of S_m. Without that constraint one gets wildly changing values of S as was shown in Table 1 of KTS07 even for multilevel data when the number of parameters is less than the number of data. In the simulations for the present paper I will take $d_{12} = d_{23} = d'$ (the experimenter has the freedom to vary contrast to vary $d_{\mathrm{m\ m+1}}$) and the noise ratios $k_1 = k_2 = k$ will be fixed, our smoothness constraint. Section 3.4 will show that the assumption that k_{m} is constant can be relaxed and the rate of change of that ratio is experimentally determinable by adding a fourth test level.

With these choices for d' and k and a bit of algebra, Eqs. (2) and (6) give the following expression for D in Eq. (3):

$$D = d'(2 - \mathrm{sqrt}((1 + k^2)/(1 + k^4))(1 + k)) \quad (7)$$

where we made use of $S_3 = k^2$, $R_3 = R_2\,(1 + k)$ and $R_2 = d'$ $\mathrm{sqrt}((1 + k^2)/2)$. A good choice for d' is $d' = 1.5$. This value is a balance between improved efficiency (higher d') and avoidance of excessive finger errors of parameter estimation (lower d'). With this choice, the function D is plotted as the solid line in the left panel of Fig. 1. The standard errors of D are given by Eqs. (4) and (5), and are plotted as dotted lines on Fig. 1. To calculate the SE we assumed that total number of trials was 8000, close to the value used for each of the KCT observers. We used 2000 trials for measuring d_{12} and d_{23} and 4000 trials for d_{13} since the standard error for d_{13} is more than double that for the adjacent trials, $d_{\mathrm{m\ m+1}}$.

The vertical line shows that if the measured value of D was 0.24, it would be more than 2.5 standard errors from $D = 0$, which is a solid rejection of the constant noise null hypothesis. The horizontal line indicates that $D = 0.24$ would give 1 SE error bars of about $\kappa = 1.300 \pm 0.075$.

An important feature of Fig. 1 is that the curve is parabolic at $\kappa = 1$, the constant noise point, corresponding to the KTS singularity. It would be impossible to detect small values of $|\kappa - 1|$. However, one can put confidence limits on larger values of $|\kappa - 1|$, as shown in Fig. 1.

The right panel of Fig. 1 is for the case where $d' = 2$ between adjacent levels, instead of $d' = 1.5$. The larger d' means the noise ratio k_{m} is larger so that D will be larger. The penalty is that the value of D is much more sensitive to finger errors. However, since KTS07 did not consider finger error I felt it was okay to include the second panel, showing smaller error bars. However, in practice one would want to choose the experimental conditions that are less sensitive to finger errors.

3. Discussion

3.1. Model parameters

I followed the approach of KCT, Klein (2006) and KTS in using a signal detection theory, Gaussian noise approach. The model parameters are specified by the contrast response function, R_{m}, and the noise standard deviation, S_{m}, for the level m stimulus. There is freedom to set $R_1 = 0$ and $S_1 = 1$. It is worth noting that R_{m} can be understood as equaling the d' between the mth level and the first level, as long as d' is measured at the horizontal intercept of the ROC curve (the z-score of the false alarm rate for a 50% hit rate). In the presence of multiplicative noise the ROC slope would not be unity, so one needs to know the slope in order to estimate the horizontal intercept (see Eq. (2)).

3.2. What happened to contrast?

Nowhere in any equation was contrast mentioned. What's going on? The answer is that our entire analysis is unchanged by any arbitrary initial nonlinearity of the system before the dominant stage of noise. These early nonlinearities are revealed when the contrast response function, $R(c)$ is visualized as a function of contrast rather than as a function of test levels, R_{m}.

3.3. Do the Ns justify the means?

In order to get reasonable confidence limits on the parameter estimates we suggested using $N = 8000$, the number of trials used by KCT. We also suggested using half of those trials to measure d' between levels 1–3

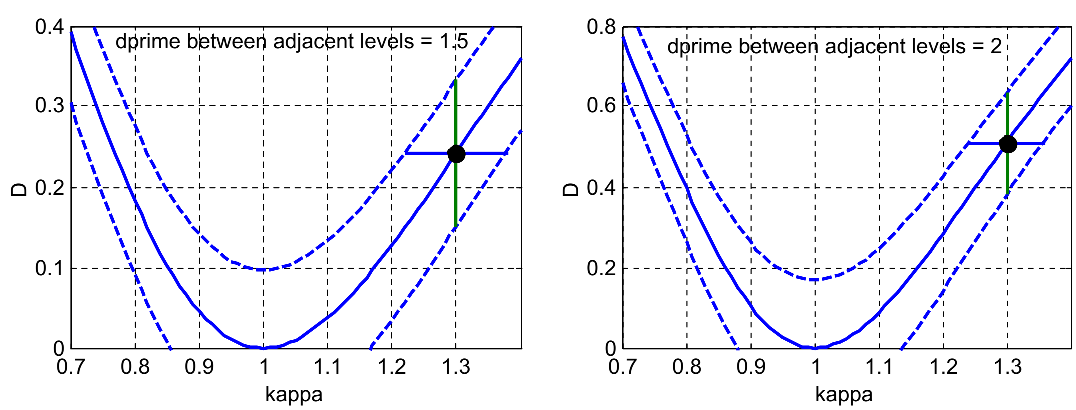

Fig. 1. The deviation from d' additivity as specified by D in Eqs. (3) and (7) is plotted on the ordinate as solid lines. The abscissa is κ, specified in Eq. (6) as a measure of the growth of the Gaussian standard deviation. The two panels correspond to a d' separation between adjacent levels of $d' = 1.5$ (left) and 2.0 (right). The black dot corresponds to a value of noise growth that is discussed in the text. The standard error of D (Eq. (5)) is shown by the dashed lines and by the vertical bar. The standard error of κ is shown as the horizontal line. The parabolic behavior near the point of perfect additivity, $\kappa = 1$, corresponds to the KTS singularity.

(d_{13}) because that discrimination had the largest standard error by Eq. (4). The simulations in Fig. 1 showed that N needs to be fairly large in order to place reasonably tight error bars on estimates of κ. It is instructive to see what $N = 4000$ means in terms of actual data collection. For the conditions of the left panel of Fig. 1, $d_{13} = 3$ corresponds to $p = .9831$ with a standard error $SE = \text{sqrt}(p * (1 - p)/N) = 0.002$. For a 2 SE confidence limit that would be ± 0.004. As long as finger errors are fewer than one part in 250, very small but significant differences in probability correct will alter the value of D, our measure of multiplicative noise.

3.4. Is the constancy of κ a reasonable assumption?

Fig. 1 was based on data at three contrast levels, providing three data points (d_{12}, d_{23}, d_{13}) to be fit by four parameters, R_2, R_3, k_2 and k_3, where $k_m = S_{m+1}/S_m$ was defined in Eq. (6). Having three equations and four unknowns is under constrained. However, inspection of the KTS Table 1 shows that additional constraints are needed, even if the equations were over constrained. In one of their fits (the third one) the percent of noise change in going from level 1 to level 2 was $k_1 - 1 = 36.1\%$ and in going from level 2 to 3 it changed to $k_2 - 1 = 18.8\%$. This nearly twofold decrease in the growth of the noise is quite surprising given that the levels were spaced by a mere $d' = 1$. Another calculation in KTS Table 1 (the 2nd one) had a noise change ($k - 1$) by more than a factor of 10,000 across a $d' = 1$ spacing. For a well behaved system, as long as one is away from threshold, the change in the multiplicative noise ($k - 1$) should be a slowly varying function. For calculating the plot in Fig. 1, I reduced the number of free parameters from 4 to 3 by assuming that $k_3 = k_2$. This assumption is able to be empirically validated by carrying out another experiment at a higher contrast. In the spirit of the KCT data one experiment could be a triplet of levels around 30% and a second triplet of levels could be done at contrasts near 60%. If the calculated value of k is relatively unchanged across this doubling of contrast then the constancy of k assumption is fine. If there is a measurable change in k, then that change could be used to constrain the fit by putting in a slow dependence of κ on contrast (see Eq. (6)).

Rather than using two groups of three levels one could assess the dependence of k (or κ) on level by testing a single group of four levels. There would be two extra data points, d_{34} and d_{24}. The possible datum, d_{14}, would have too large an error bar to be useful (too few errors in performance). The extra pair of data allows an extra pair of parameters to be estimated. One parameter, R_4, would be used to fit d_{34}. The last parameter would be used to specify how κ depends on level using a linear relationship. We expect to find the slope term to be small since we are doing the measurements well above threshold where κ should be fairly constant.

3.5. The singularity

The quadratic behavior near $\kappa = 1$ in Fig. 1 shows that there is indeed a KTS-type singularity at the constant noise condition, $\kappa = 1$. Near $\kappa = 1$, D, the measure of multiplicative noise is insensitive to κ. What was not emphasized by KTS is that if the multiplicative noise were $\kappa = 1.3$ and if κ varies slowly across levels (discussed in preceding section), then Fig. 1 showed confidence limits can be placed on the amount of multiplicative noise using a reasonable number of trials. The next section considers whether the $\kappa = 1.3$ value we looked at in Fig. 1 was excessively high.

My quibble with KTS can be clarified by looking at all their model fits reported in KTS07 Tables 1 and 2. All of those combinations of R_m and S_m correspond to $D = 0$. But suppose the 2AFC experiments show that $D > 0$ by a significant amount. Section 3.7 looks at the KCT data and shows that subject AK-T had a value of D that was significantly greater than zero using a relatively small number of trials.

3.6. Was κ = 1.3 a good place to plot the confidence limits in Fig. 1?

The value of $\kappa = 1.3$ is not unreasonable for three reasons:

(1) KTS07 use $\kappa = 1.36$ for their first pair of levels, a value larger than ours. Our assumption of a uniform κ seems at least as reasonable as the rapidly decreasing value used by KTS07.
(2) The ROC slope, given in Eq. (6), is $1/k = \kappa^{-d'}$. Previous studies measuring ROC slope for detection reported finding $\kappa = 1.25$ (Nachmias & Kocher, 1970). They used yes–no methods and defined d' in terms of the ROC horizontal intercept, d_{horiz}. However, we use 2AFC where d' is defined in terms of the ROC area. The connection between the two definitions of d' is $d_{\text{horiz}} = d_{2AFC}\ ((1+k^2)/2)^{1/2}$. For d' in the range of 1.5–2.0 this difference in definition comes very close to the relationship that $\kappa_{\text{yes–no}} = 1.25$ corresponds to the same slope as $\kappa_{2AFC} = 1.30$.
(3) As discussed next, the KCT dataset has a triplet of data at 15%, 21.5% and 30% contrasts that give $D = 0.533 \pm 0.214$. This value is able to reject the constant noise hypothesis and corresponds to $\kappa = 1.44$.

3.7. A KCT triplet for estimating D

Table 1 presents a subset of the KCT data for observer AK-T. In order to do our style of analysis we need to locate a triplet of data. Unfortunately the KCT dataset was not planned with triplets in mind. The only adequate triplet, is for levels 15%, 21.5% and 30%. The first row gives the reference contrast that was fixed throughout a run. The second row gives test contrasts that were intermixed in a

Table 1
Data from observer AK-T of Kontsevich et al. (2002)

Reference contrast	15	15	30	30
Test contrast	21.5	30	15	21.5
Probability correct	0.815	0.915	0.874	0.735
Number of trials	200	200	286	486
d' for 2AFC	1.268	1.941	1.621	0.884
SE of d'	0.146	0.179	0.134	0.086

Reference contrast was one of the alternatives in every trial. The d' values were used to make the D statistic of Eq. (3). The SEs were used to calculated the significance of the D statistic in Eq. (5).

given run and randomized in presentation order for the 2AFC task. The next two rows are the percent correct and the number of trials. The final pair of rows are the d' and its SE for that discrimination, given by Eq. (4).

Although two triplets can be formed since two estimates of the 15–30% pairing are available, the d' for the 30% contrast reference is the most important for reasons discussed by Klein (2006) in connection with the large finger error asymmetry (many more finger errors with the low contrast tests than the high contrast tests). Using the value with the 30% reference gives $D = (d_{12} + d_{23}) - d_{13} = 2.153 - 1.621 = 0.533 \pm 0.214$. This value is able to reject the constant noise hypothesis by a z test with $p < .01$. The average $d' = 1.1$ between adjacent levels leads to a remarkably large value of $\kappa = 2.1$ from Eq. (6). 886 trials were able to reject the constant noise hypothesis with $p < 0.01$ That's why Klein (2006) questioned the KTS emphasis on the constant noise domain. If instead of just using the 30% reference data, we combine the two estimates for the 15–30% pairing one has 433 correct out of 486 trials leading to $d_{13} = 1.742 \pm 0.107$. This value for d_{13} gives $D = 2.152 - 1.742 = 0.410 \pm 0.200$, a value able to reject the constant noise hypothesis ($p < .05$) with 1172 trials. From Fig. 1 (with an expanded scale) $D = 0.41$ corresponds to $\kappa = 1.44$, a multiplicative noise greater than the $\kappa = 1.3$ value used in our earlier discussions.

Given the importance of the data in the 15–30% range it is useful to examine KCT data at nearby contrasts to double-check that d' depends on which level is the reference. For the 15% reference runs there is a KCT test level at 24% contrast with 200 trials and with 91.5% correct, just like for the 30% test but with substantially lower contrast! And for the 30% reference runs there is a test level as low as 6% contrast with 486 trials and as much as 8.64% correct! In both cases the probability correct values are in the right direction for adding significance to the asymmetry for the two 15–30% pairings. These values make it clear that there is a dramatic, highly significant difference in performance across the two references, as discussed by Klein (2006) in connection with the finger error asymmetry found in two of the KCT datasets.

The last paragraph argues that the difference in estimates of D from the two 15–30% pairings is significant. What might have caused this dependence of the d' on the context of the run (a violation of independence)? Klein (2006) mentioned a number of reasons why the runs with a 30% reference had so many errors in discriminating 30% contrast from low contrast (the finger error asymmetry I discussed in Klein (2006)), including adaptation effects from the intermixed very high contrast test stimuli. Another possibility is that there could be a context effect whereby the runs with the 30% reference had a more heterogeneous set of test contrasts. That heterogeneity could have affected the observer's ability to attend to the relevant contrast range. It is relevant to the present topic that the type of reduction in performance found with the 30% reference might well be entitled to be called a performance reduction due to multiplicative noise. Lu and Dosher (1999), for example, define multiplicative noise as threshold elevations due to previous stimulus presentations, not simply elevation due to the stimuli of the present trial.

3.8. Final thoughts

This paper shows that 2AFC can be used to reject the constant Gaussian noise hypothesis and to put constraints on parameter estimates of the amount of multiplicative noise. In the process of thinking about these questions several new questions came up for future consideration:

(1) What happens if we relax the assumption that the noise is Gaussian, and one is given the ROC curve between adjacent levels. I found that surprisingly one can still reject a constant noise hypothesis in a reasonable number of trials. One of the fascinating aspects is how to go from the ROC curve to the shape of the non-Gaussian distribution that satisfies the constant noise criterion.
(2) How do the KCT power law parameters (q specifies the amount of multiplicative noise) map onto the local parameters such as D and κ? I consider the local nature of the D statistic and the κ parameter to be an important improvement over previous global measures of multiplicative noise such as the KCT q parameter.
(3) My proposed measure of local multiplicative noise, D, depends on the size of the d' interval between levels (see Fig. 1). I would like to come up with a density measure (might it be $D/(d_{12}\ d_{23})$?) that is independent of step size for small steps. The advantage of such a statistic is that it would be a robust measure of deviations from Gaussian additive noise.

One final point should be emphasized. In this paper I have defended the 2AFC method by showing that it can in principle place constraints on the underlying model with a feasible number of trials. However, there are many problems with the 2AFC method, some previously spelled out by Klein (2001). An additional problem is that for the task considered in this paper, of estimating the contrast response function and separately identifying the nature of the noise (is it multiplicative and is it Gaussian) the rating scale method of constant stimuli that Dennis Levi and I have been using for 25 years is far superior to 2AFC.

Acknowledgment

This project was supported by NEI Grant EY04776. The "Do the Ns justify the means" sub-heading came from Patrick Cavanagh by way of Olga Favreau.

References

Georgeson, M., & Meese, T. M. (2006). Fixed or variable noise in contrast discrimination? The jury's still out. *Vision Research*, 46.

Katkov, M., Tsodyks, M., & Sagi, D. (2006a). Singularities in the inverse modeling of 2AFC contrast discrimination data. *Vision Research, 46*, 259–266.

Katkov, M., Tsodyks, M., & Sagi, D. (2006b). Analysis of two-alternative forced-choice Signal Detection Model. *Journal of Mathematical Psychology, 50*, 411–420.

Katkov, M., Tsodyks, M., & Sagi, D. (2007). Singularities explained: Response to Klein. *Vision Research,* preceding article.

Klein, S. A. (2001). Measuring, estimating and understanding the psychometric function: A commentary. *Perception & Psychophysics, 63*, 1421–1455.

Klein, S. A. (2006). Separating transducer nonlinearities and multiplicative noise in contrast discrimination. *Vision Research, 46*, 4279–4293.

Kontsevich, L. L., Chen, C. C., & Tyler, C. W. (2002). Separating the effects of response nonlinearity and internal noise psychophysically. *Vision Research, 42*, 1771–1784.

Lu, Z.-L., & Dosher, B. (1999). Characterizing human perceptual inefficiencies with equivalent internal noise. *Journal of the Optical Society of America, A, 16*, 764–778.

Nachmias, J., & Kocher, E. C. (1970). Visual detection and discrimination of luminance increments. *Journal of the Optical Society of America, 60*, 382–389.