Available online at www.sciencedirect.com
SCIENCE DIRECT®
Vision Research 44 (2004) 287–296
Vision Research
www.elsevier.com/locate/visres

In search of an elusive hard threshold: a test of observer's ability to order sub-threshold stimuli

A.D. Logvinenko [a,*], W. Byth [a], E.E. Vityaev [b]

[a] School of Psychology, Queen's University of Belfast, David Keir Bldg, Belfast, Northern Ireland BT9 5BP, UK
[b] Institute of Mathematics, Russian Academy of Science, Novosibirsk, Russia

Received 10 October 2002; received in revised form 1 September 2003

Abstract

The contrast transducer function (d' vs. contrast) for sine gratings was claimed to come up from some non-zero contrast value rather than from the origin. This implies that there is a point (a hard threshold) on the grating contrast axis below which observers could not distinguish between presentations containing gratings and those containing a homogeneous field. We studied the ability to order sub-threshold square wave gratings and found, to the contrary, that observers were able to do this no matter how low the contrasts. At the same time, the observers failed to order the sub-threshold gratings when they were of the same contrast. The latter is inconsistent with signal detection theory which predicts that an observer's judgements are based on the same ordered set of sensory states irrespective of whether the stimuli differ or are the same. On the other hand, these data can be reconciled with the notion of a threshold if the latter is thought of as a fuzzy rather than a sharp margin on the contrast axis.

1. Introduction

It is no exaggeration to say that the history of experimental psychology began with the concept of a threshold (e.g. Boring, 1950; Woodworth & Schlosberg, 1955). However, few other ideas have been subjected to such criticism, and even its very existence has been questioned. Indeed, signal detection theory (SDT), which refutes the idea of a threshold as such, has dominated psychophysics for the past four decades.

SDT (Green & Swets, 1966) suggests that under a threshold task observer responses to a stimulus (e.g. "yes" or "no") are mediated by a sort of random variable (sensory state) which may even, in the absence of any stimulus, come into existence due to noise. An observer rests a decision on the likelihood of that particular sensory state being induced by the stimulus plus noise or by noise alone. According to the theory, an index of detectability, d', is not zero for any stimulus however small its intensity. In particular, a transducer function, relating d' to stimulus intensity is generally believed to be monotone, accelerating, and positive everywhere on the stimulus intensity continuum (see e.g. Nachmias & Kocher, 1970; Nachmias & Sansbury, 1974).

However, the question of the existence of a threshold still provokes debate. In particular, evidence has recently appeared that d' can be zero for some non-zero stimulus magnitude. Specifically, the transducer function for sine wave grating contrast was reported to exhibit two distinct segments, the first being a horizontal interval coinciding with the contrast axis (Simpson, Falkenberg, & Manahilov, 2003). This and previous similar observations (Beard, Klein, & Tyler, 1995; Kontsevich & Tyler, 1997, 1999) lead to the suggestion that the observers in these experiments might have not distinguished between gratings and a homogeneous field when the gratings' contrast was below a certain low value. If so, such a critical point on the contrast axis may be interpreted as a sort of threshold (hard threshold to distinguish it from a classical threshold which is defined as a grating contrast which yields the conventional probability of correct responses), thus restoring life to one of the oldest concepts in psychology.

It should be noted, however, that establishing a hard threshold by finding a region on the contrast axis where $d' = 0$, is a very difficult task from the statistical point of view. As a matter of fact, there is no satisfactory

* Corresponding author. Tel.: +44-2890-274542; fax: +44-2890-664144.
E-mail address: a.logvinenko@qub.ac.uk (A.D. Logvinenko).

0042-6989/$ - see front matter © 2003 Elsevier Ltd. All rights reserved.
doi:10.1016/j.visres.2003.09.015

statistical procedure to test the hypothesis that the index of detectability d' derived from the empirical data obtained in the standard two-alternative forced choice procedure is equal to zero. Here we deal with the same problem as that of differentiating two close probabilities. In particular, an adequate solution implies a number of presentations so large that it could not in principle be implemented in a psychophysical experiment—an observer's sensitivity will unavoidably change due to visual fatigue, habituation, ageing etc.

For this reason we decided to test the hypothesis of the existence of a hard threshold by using a rather different approach with the following rationale. If there is a hard threshold on the contrast axis then there should be a zone below the hard threshold where $d' = 0$. If so, observers will not be able to order stimuli from this zone relative to their certainty that what they observe results from presentation of a pattern rather than a homogeneous field.

We have attempted to find such a zone in an experiment designed as a direct test of the observer's ability to order low contrast square wave gratings whose contrast may be presumed to fall within the hypothetical zone where $d' = 0$. Specifically, forced choice comparisons of perceived contrast were made between gratings whose detection probability for the observer had been shown not to differ significantly from chance level (50%) in a two alternative forced choice paradigm.

2. Theory

An essential feature of any order is its transitivity (e.g. Krantz, Luce, Suppes, & Tversky, 1971). An order $\succ$ on a set S is called transitive if for every triad of stimuli $s_1, s_2, s_3 \in S$, the following axiom holds true:

$$s_1 \succ s_2 \cap s_2 \succ s_3 \Rightarrow s_1 \succ s_3. \tag{1}$$

This means that, presented with a triad of stimuli s_1, s_2, s_3, if an observer prefers s_1 over s_2, and s_2 over s_3, then transitivity implies that the observer also prefers s_1 over s_3.

An experimental test of transitivity of an order on low contrast gratings (or any other stimuli around threshold) is complicated by the fact that an observer's judgements are subject to random variability. In other words, the same stimulus brings about different comparative judgements for the same observer on different occasions. Therefore, we run into another problem, namely, how to test transitivity under stochastic uncertainty.

A traditional approach to this problem is to replace the order (1) with one defined in terms of probability of preferences and then to test transitivity of this new, stochastic, order (e.g. Falmagne, 1985; Suppes, Krantz, Luce, & Tversky, 1989). There are several definitions of such stochastic transitivity (e.g. Luce & Suppes, 1965, p. 340; Suppes et al., 1989, p. 389). For instance, one of these (so-called strong stochastic transitivity) is

$$\boldsymbol{P}\{s_1 \succ s_3\} \geqslant \max\{\boldsymbol{P}\{s_1 \succ s_2\}, \boldsymbol{P}\{s_1 \succ s_3\}\}$$
$$\text{for each triad } s_1, s_2 \text{ and } s_3, \tag{2}$$

which says that the probability of an observer preferring stimulus s_1 over s_3 is not less than either of the probabilities of preferring stimulus s_1 over s_2, and s_2 over s_3.

However, such an approach is not relevant to our purpose since we want to test the original deterministic order (1) rather than a new one defined by Eq. (2). In other words, we want to test statistically a logical statement (1) the same way as we usually test statistically quantitative statements, that is, we want to find out whether or not the observers' responses are in line with the statement (1).

To test the logical hypothesis (1) one has to present an observer with various triads of stimuli (of the set S) $\{(s_1, s_2), (s_2, s_3), (s_1, s_3)\}$ and to record the observer's responses to these stimuli. Let us write $R_1(s_1, s_2)$ if the observer prefers the first stimulus, s_1, to the second one, s_2; and $R_2(s_1, s_2)$ otherwise, that is, when the stimulus s_2 is preferred to s_1. There are exactly eight possible triplets of responses associated with a stimulus triad $\{(s_1, s_2), (s_2, s_3), (s_1, s_3)\}$:

$$\begin{aligned}
&\{R_1(s_1,s_2), R_1(s_2,s_3), R_1(s_1,s_3)\},\\
&\{R_1(s_1,s_2), R_1(s_2,s_3), R_2(s_1,s_3)\},\\
&\{R_1(s_1,s_2), R_2(s_2,s_3), R_1(s_1,s_3)\},\\
&\{R_1(s_1,s_2), R_2(s_2,s_3), R_2(s_1,s_3)\},\\
&\{R_2(s_1,s_2), R_1(s_2,s_3), R_1(s_1,s_3)\},\\
&\{R_2(s_1,s_2), R_1(s_2,s_3), R_2(s_1,s_3)\},\\
&\{R_2(s_1,s_2), R_2(s_2,s_3), R_1(s_1,s_3)\},\\
&\{R_2(s_1,s_2), R_2(s_2,s_3), R_2(s_1,s_3)\}.
\end{aligned} \tag{3}$$

Only one of the eight outcomes, namely, $\{R_1(s_1, s_2), R_1(s_2, s_3), R_2(s_1, s_3)\}$ is not consistent with the transitivity hypothesis (1). From deterministic point of view, once at least one such a triplet is observed in the experiment the transitivity hypothesis (1) should be rejected. However, from stochastic point of view, for the transitivity hypothesis (1) to be rejected one has to compare the number of times when the outcome $\{R_1(s_1, s_2), R_1(s_2, s_3), R_2(s_1, s_3)\}$ was observed with that for other outcomes in order to decide whether the responses which contradict transitivity can be accounted for by random fluctuations.

A statistical test of transitivity employed in the present study, rests upon the following fact established recently (Logvinenko & Vityaev, 2003). Given the transitivity hypothesis (1), the occurrence of any of the two antecedents in (1) increases the conditional probability of occurrence of the inference. To be more exact, it is proved that the logical hypothesis of transitivity (1)

entails that for every triad $s_1, s_2, s_3 \in S$ the following strict inequality for conditional probabilities holds:

$$\boldsymbol{P}\{(s_1 \succ s_3)|(s_1 \succ s_2) \cap (s_2 \succ s_3)\} > \max\{\boldsymbol{P}\{(s_1 \succ s_3)|(s_1 \succ s_2)\}, \boldsymbol{P}\{(s_1 \succ s_3)|(s_2 \succ s_3)\}\}. \quad (4)$$

Inequality (4) can be reduced to the following two inequalities (Logvinenko & Vityaev, 2003):

$$\boldsymbol{P}\{AB\} > \boldsymbol{P}\{A\}\boldsymbol{P}\{B\}, \quad (5)$$

where A is the event which occurs if an observer prefers s_2 to s_3 on condition that s_1 was preferred to s_2, and B is the event which occurs if an observer prefers s_1 to s_3 on the same condition; and

$$\boldsymbol{P}\{A'B'\} > \boldsymbol{P}\{A'\}\boldsymbol{P}\{B'\}, \quad (5')$$

where A' is the event which occurs if an observer prefers s_1 to s_2, and B' is the event which occurs if an observer prefers s_1 to s_3, both on condition that s_2 was preferred to s_3.

Hence, we observe that statistically testing transitivity (1) can be reduced to testing statistical interdependence between the antecedents and inference. The following experiment has been undertaken to determine such interdependence for sub-threshold grating contrast comparisons. The null (no transitivity) hypotheses to be tested have that the inference judgement $s_1 \succ s_3$ is probabilistically independent from both antecedent judgements $s_1 \succ s_2$ and $s_2 \succ s_3$. More specifically, the following two null hypotheses:

$$H_0 : \boldsymbol{P}\{AB\} = \boldsymbol{P}\{A\}\boldsymbol{P}\{B\}, \quad (6)$$

$$H_0' : \boldsymbol{P}\{A'B'\} = \boldsymbol{P}\{A'\}\boldsymbol{P}\{B'\}; \quad (6')$$

have been tested against the alternatives (5) and (5′). The hypothesis of transitivity (1) has been accepted when both the null hypotheses (6) and (6′) are rejected in favour of the alternatives (5) and (5′) respectively.

3. Experiment

3.1. Method

Three different experiments were carried out with the same two male observers. Both observers were experienced in psychophysical observations. One (BB—one of the authors) was partially aware of the purpose of the experiment and the other (RM) was completely unaware.

The *Preliminary* experiment was designed to determine individual contrast thresholds for each observer which thereafter were used to set the contrast of the gratings for each observer in the subsequent experiments. The *Main* experiment examined the observers' ability to order different low (sub-threshold) contrast gratings. As the task to compare the visibility of subthreshold gratings is rather hard (especially for naive observers), the main experiment was conducted in two steps. In the main experiment 1 gratings of a little below threshold contrast were used. After the observers were used to this task, the Main experiment 2 was run with gratings well below threshold, at contrasts as low as could reliably be displayed on the computer monitor screen. The *Control* experiment was the same as the Main, except that the gratings to be compared were of equal contrast, and was designed to reveal any bias in the observers' responses when there were no differences between stimuli.

3.1.1. Preliminary experiment

In the preliminary experiment psychometric functions (probability of correct response vs. contrast) for a square wave grating were obtained using the two-alternative spatial forced choice technique for both observers. Gratings were displayed on an 8-bit Silicon Graphics monitor, which had a resolution of 1152×1080 pixels, using standard X-Window library routines. The monitor was gamma corrected using standard IRIX gamma correction tool "gamma". The 8-bit display allowed the use of 256 shades of grey. Spatial frequency of the grating was 33 c/deg. Nine gratings of Michelson contrast from 3% to 13% were used to evaluate the psychometric function for the observer BB and eight, from 2% to 14%, for RM.

On every trial, the observers were presented with a set of three pairs of continuously visible circular windows, each 43 min arc in diameter (from a distance of 4.36 m) on a grey background, arranged in a rectangle (Fig. 1(a)). The space-average luminance of the windows was greater (66 cd/m^2) and the colour slightly bluer (CIE chromaticity coordinates $x = 0.278$, and $y = 0.313$) than that of the background ($x = 0.266$, $y = 0.307$, and luminance 45 cd/m^2).

In each trial observers were required to indicate the window of each pair in which the grating appeared by pressing the appropriate button of a mouse. The order in which the observers had to make their responses on each individual trial was random and indicated by an arrow appearing on the screen (Fig. 1(a)).

All the gratings in all three pairs were presented simultaneously and remained present until the last (third) response in the current trial was made. After the third response was registered all the windows reverted to blank fields until the next trial. There were no time constraints on responses and observers could cancel a response which they felt was mistaken. Observers launched each presentation themselves with a key press when they were ready.

There were 50 trials in each daily test session. The test contrasts for each trial were chosen randomly with the constraint that all the three gratings presented at a trial

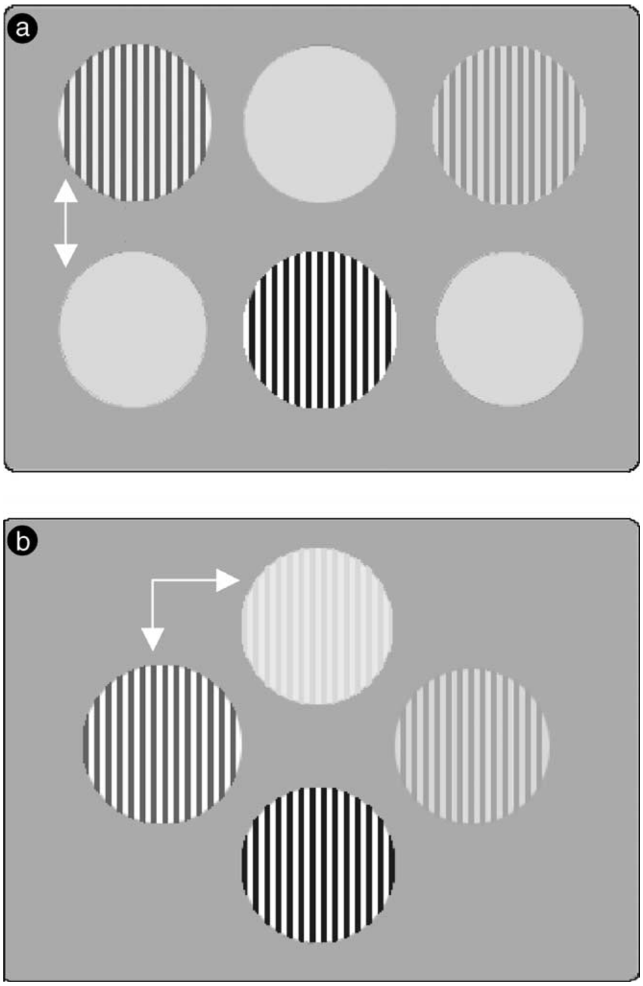

Fig. 1. Stimulus configurations for the preliminary (a) and main (b) experiments (see explanation in text).

have different contrast. Thus, observers made 150 two-alternative spatial forced choice decisions each day. One experimental session lasted approximately 15–20 min. Experimental conditions were kept as constant as possible throughout this period. Each test session was conducted at the same time of day in the same semi-darkened laboratory room with the same equipment. Viewing was binocular.

3.1.2. Main experiment

In the main experiment the observers were presented with pairs of sub-threshold gratings and they were asked to estimate which of the gratings in each pair was more visible, that is, to prefer one of the two on the basis of their visibility. Observers were instructed to base their judgements on sensory factors alone, such as apparent contrast or other apparent spatial structure or inhomogeneity. They were told to evaluate neither their confidence that the grating was presented in this particular window nor anything like subjective or a posteriori probabilities. They were also instructed not to feel obliged by logic to chose $s_1 \succ s_3$ if they had chosen $s_1 \succ s_2$ and $s_2 \succ s_3$. When both gratings were completely invisible and observers felt completely uncertain they were allowed to guess which grating was closer to threshold.

Four square wave gratings of the same 33 c/deg spatial frequency were employed in both parts of the main experiment, so that in all there were six possible pairs of gratings to compare. Gratings were presented on the same computer screen and under the same conditions as in the preliminary experiment. The luminance and colour of the windows, the grey background and the average luminance of the gratings was the same as that in the preliminary experiment. The gratings were presented in four circular windows (43 min. diam) arranged in a rhombus (Fig. 1(b)). The window in which each grating would appear was varied randomly from trial to trial. In each trial all four gratings appeared simultaneously and remained present until all six judgements were completed at which point the four windows reverted to blank fields until the observer launched the next presentation. Each of the six possible pairs of gratings was singled out in random order by a pointing arrow appearing on the grey background which pointed out the next pair to compare as soon as the response was made (Fig. 1(b)). The observers were required to indicate their preferences (as to which of two gratings was "more visible") by pressing a mouse button.

In the first part, main experiment 1, contrast of gratings was selected so as to make the task of comparison comfortable for each observer. For the naive observer RM the contrasts were 4.5%, 8.5%, 11%, and 12.5% which corresponds to the 50–81% correct response probability band on the psychometric function. The contrast range for the observer BB was narrower (3%, 6%, 7.5% and 8%) and farther from the threshold point, namely, 50–58%.

In the main experiment 2 grating contrast was chosen from the region of the psychometric function where the probability of detection was not significantly different from 50%, or chance level. For both observers, these contrasts were 1%, 2%, 3% and 4%, which was still reliably above the contrast resolution of the display (0.5%). Although these gratings were never actually visible to the observers (contrary to main experiment 1 where gratings were occasionally, though barely, experienced as visible), they accepted the task of comparing such low contrast gratings as an extension of that required in the first experiment.

Fifty trials were carried out in each daily test session so that observers made 300 forced choice decisions each day. One experimental session lasted approximately 20–25 min.

3.1.3. Control experiment

To ensure that observers followed the instructions to take into account only sensory factors, a control ex-

periment was undertaken. Its design was analogous to that of the main experiment except that three of the four gratings (chosen at random for each stimulus set) contained the same contrast. The contrast of the three identical gratings was selected randomly from the four available contrasts.

The control experiment was run simultaneously with the main experiment 1, i.e. "catch" presentations for the control experiment were interleaved randomly with presentations designed for the main experiment. Observers were not informed that sometimes three of the gratings had the same contrast, that is, they were not actually aware of the control experiment.

Preliminary tests showed that if all four windows contained identical gratings observers became aware that sometimes all the gratings were the same. The inclusion of a single different grating rendered the control presentations indistinguishable from the main experiment presentations. At least, observers did not report that they noticed any difference between the main experiment 1 containing "catch" presentations and the main experiment 2 where there were no control presentations. The probability of control presentations was set at 20% and data comparisons including the fourth (different) grating were not of course included in the transitivity analysis of the control data.

The rationale of the control experiment was that should the observers be influenced by a non-sensory (logical) response strategy or factors such as window position, then such a tendency would also appear in the control experiment data.

3.1.4. Learning and training

As would be expected, it was found that the observers' ability to detect and discriminate the gratings improved from session to session in the preliminary and main experiments but not in the control one. For this reason there was a training period during which the data obtained in the paradigms of the preliminary and main experiments were recorded but not included in the final analysis presented below. [1] The training period was continuing until the observers' performance stopped improving. The formal criteria to decide that the learning was over was as follows. During the learning period matrices of pair comparisons obtained in two successive sessions (50 trials each) were tested on homogeneity. We believed that it was safe to start the preliminary experiment when at least three matrices of pair comparison in succession were homogeneous.

After the main and control experiments were completed we measured the psychometric function again to make sure that it was stable, that is, there was no improvement in detection of the gratings. A test for homogeneity showed that the function was indeed stable for both observers.

[1] However, the data from all the control presentations through the whole experimental period including training were combined to give a total of 255 presentations for BB and 143 for RM.

4. Results

The individual data on detection collected in the preliminary experiment amalgamated with the data obtained after the main and control experiments are presented in Tables 1 and 2, and shown in Fig. 2.

Smooth lines are the approximations based on the Weibull distribution:

$$P(c) = 1 - 0.5\exp\left[-\left(\frac{c}{c_0}\right)^{\beta}\right], \tag{7}$$

where the parameter β specifies the slope of the psychometric curve and the shift parameter c_0 is taken as an estimate of the threshold. Threshold contrast gives a probability of success of $P(c_0) = 0.816$. These parameters have been estimated by the maximum likelihood method for each observer separately. The estimates are as follows: for BB $c_0 = 0.126$ and $\beta = 5.3$; for RM $c_0 = 0.125$ and $\beta = 5.9$. There is no significant difference either between individual thresholds or between the slope parameters.

The binomial test shows that not only for all the stimuli used in the main experiment 2 but also for two stimuli in the main experiment 1 detection was also at chance level ($p > 0.05$).

Table 1
Detection performance for observer BB

Contrast (%)	Total number of presentations	Number of correct responses	Proportion of correct responses
3	343	171	0.4985
4	338	150	0.4438
5	346	187	0.5405
6	348	177	0.5086
7	338	171	0.5059
8	330	190	0.5758
9	351	206	0.5869
10	383	231	0.6031
13	223	189	0.8475

Table 2
Detection performance for observer RM

Contrast (%)	Total number of presentations	Number of correct responses	Proportion of correct responses
2	187	88	0.4706
4.5	364	178	0.4890
6.5	183	92	0.5027
8.5	315	171	0.5429
10	179	109	0.6089
11	310	220	0.7097
12.5	319	258	0.8088
14	179	167	0.9330

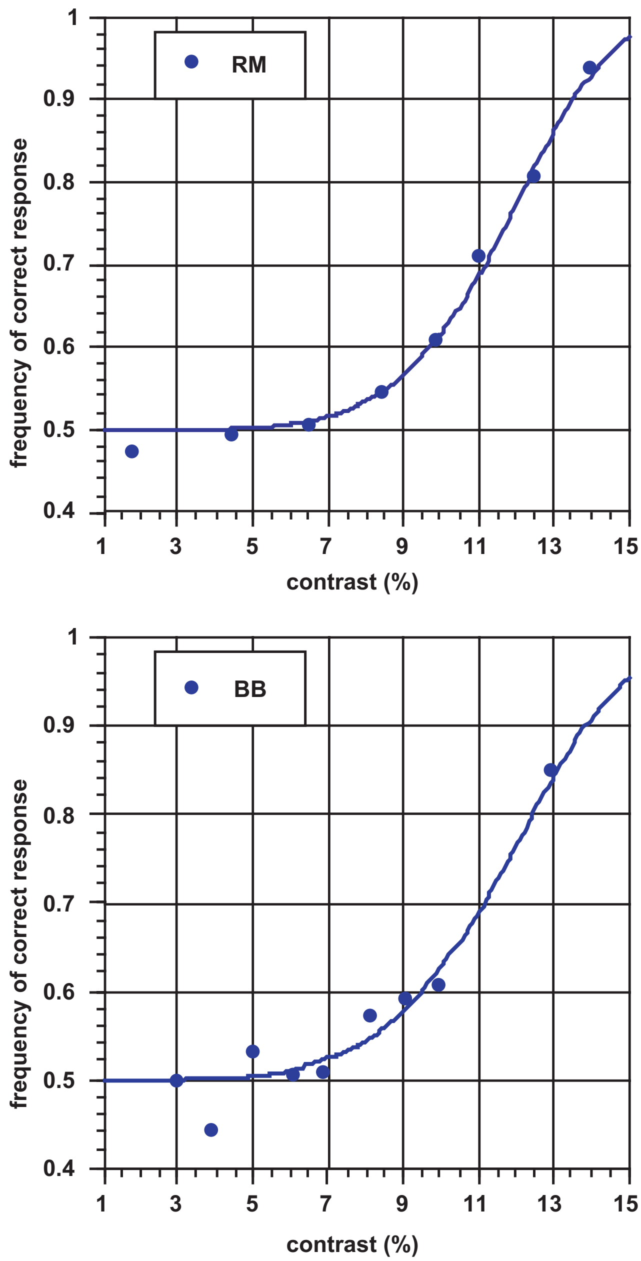

Fig. 2. Individual psychometric functions measured in the preliminary experiment. Continuous lines are Weibull approximations (see Eq. (6) in text).

Matrices of paired comparisons for the gratings used for each observer in both main experiments are given in Table 3. These data are based on 360 trials for the observer RM and 333 trials for BB in the main experiment 1 and on 300 trials for both observers in the main experiment 2. No proportion of correct discriminations in the main experiment 2 was significantly greater than chance level (50%) for RM, and only one pair was successfully discriminated above chance level for BB. So for the set of gratings used in the second experiment not only contrast detection but also contrast discrimination were no better than chance level.

Table 3
Proportion of preferences for each observer for stimulus s_i over stimulus s_j where the contrast of s_i are given horizontally above the tables and the contrast of s_j are vertically along the side

RM	2%	3%	4%	RM	8.5%	11%	12.5%
1%	0.50	0.51	0.52	4.5%	0.50	0.49	0.56*
2%		0.45	0.50	8.5%		0.56*	0.62*
3%			0.51	11%			0.58*
BB	2%	3%	4%	BB	6%	7.5%	8%
1%	0.46	0.52	0.51	3%	0.56*	0.61*	0.67*
2%		0.54	0.57*	6%		0.56*	0.59*
3%			0.50	7.5%			0.59*

The proportion denotes the number of times the preference was expressed for s_i over s_j divided by the total number of times a choice was made between them (333 choices for RM and 360 for BB). The asterisk (*) denotes proportions significantly greater than 0.50 ($p < 0.05$).

5. Statistical analysis of transitivity

Testing the null hypotheses (6) and (6′) against the alternatives (5) and (5′) can be reduced to a classical problem of testing independence in a 2×2 contingency table (see e.g. Kendall & Stuart, 1979). Indeed, with each of the two null hypotheses a corresponding 2×2 table can be associated. Particularly, Table 4 is associated with the hypotheses (6). An analogous 2×2 table can also be produced for the null hypothesis (6′).

Given that the number of trials in our experiment was large enough, we used the large-sample normal test (Kendall & Stuart, 1979, p. 582) to test independence in such 2×2 tables. The data for both the main experiments and for the control experiment were tested separately for each observer. Both the null hypotheses, (6) and (6′), were rejected for both observers in both the main experiments but not in the control experiment (corresponding p-values are presented in Table 5).

It should be noted, however, that if we redo such an analysis for the main experiments the set of p-values in Table 5 will, generally, change because the set of the sample triads will be different (as only one of the four triads from each presentation is randomly picked up during a particular run). In other words, if we repeat the analysis of the main experiment data a few times we will obtain a whole set of different p-values for each null

Table 4
A 2×2 table associated with the null hypotheses H_0 (Eq. (6))

	B	$\overline{B}$	Totals
A	n_{11}	n_{12}	n_{1*}
$\overline{A}$	n_{21}	n_{22}	n_{2*}
Totals	n_{*1}	n_{*2}	n_{**}

Here n_{11}, n_{12}, n_{21}, and n_{22} are the number of times when the events $AB = \{R_1(s_1,s_2), R_1(s_2,s_3), R_1(s_1,s_3)\}$, $\overline{A}B = \{R_1(s_1,s_2), R_2(s_2,s_3), R_1(s_1,s_3)\}$, $A\overline{B} = \{R_1(s_1,s_2), R_1(s_2,s_3), R_2(s_1,s_3)\}$, and $\overline{AB} = \{R_1(s_1,s_2), R_2(s_2,s_3), R_2(s_1,s_3)\}$ were observed in the experiment respectively. The totals $n_{1*} = n_{11} + n_{12}$; $n_{2*} = n_{21} + n_{22}$; $n_{*1} = n_{11} + n_{21}$; $n_{*2} = n_{12} + n_{22}$; $n_{**} = n_{1*} + n_{2*}$.

Table 5
The significance p-values for the null hypotheses H_0 and H_0'

Observer	RM		BB	
Hypothesis	H_0	H_0'	H_0	H_0'
Main expt 1	0.034	0.032	0.000	0.000
Main expt 2	0.002	0.001	0.031	0.004
Control expt	0.097	0.732	0.678	0.437

hypothesis. A distribution of these p-values will, however, be different depending on whether the null hypothesis or the alternative holds true. More specifically, under the null hypothesis p-values are to be distributed approximately uniformly [2] whereas under the alternative they will gravitate towards zero.

So it seemed quite natural to repeat the statistical analysis as described above a few times for the main experiments. In this case, first, we could use all the information obtained (not only that contained in the three sampled triads), and second, a more sophisticated statistical analysis (based on the type of distribution of p-values) could be applied. Namely, we could base our decision on the whole distribution of the p-values rather than on their individual values. Particularly, a shift in the distribution towards 0 would provide a basis for rejecting the null hypotheses (6) and (6′) in favour of the alternatives (5) and (5′).

As a matter of fact, we repeated the analysis 100 times for both the main experiments making a resampling at every run. To be more exact, the resampling procedure was as follows. For every presentation made in the main experiment a sample triad, (s_{i_1}, s_{i_2}), (s_{i_2}, s_{i_3}) and (s_{i_1}, s_{i_3}), where s_{i_1}, s_{i_2}, and s_{i_3} are three of the four gratings used in the main experiment, was constructed randomly. If in reality (i.e. during the experimental presentation) the observer preferred the first stimulus in the pair, the response R_1 was assigned to this pair; otherwise the complementary response, R_2, was assigned. Then p-values were evaluated for every resampled triad for each observer.

A computational model was also constructed which simulated an observer responding randomly, irrespective of the contrast of stimulus patterns. Using this model we generated the same number of responses as in

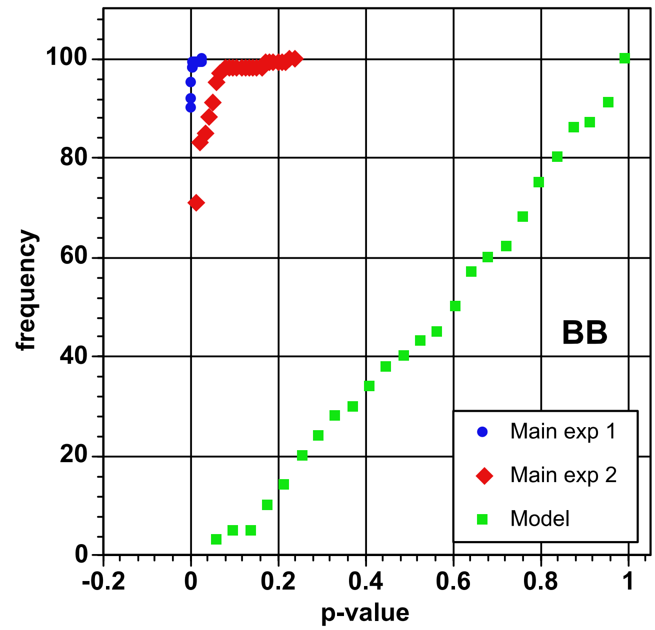

Fig. 3. The cumulative distribution function of responses from observer BB for the main experiments and for modelling. The abscissa gives the p-value (see text). The ordinate shows for every particular x how many samples produced p-values not exceeding x.

the main experiment 1 (i.e. 333 for BB, and 360 for RM). Then the same statistical analysis of transitivity as for the main experiments was carried out. As expected, the model produced an almost uniform distribution of p-values which, when plotted, gives a nearly linear cumulative distribution function (see squares in Figs. 3 and 4).

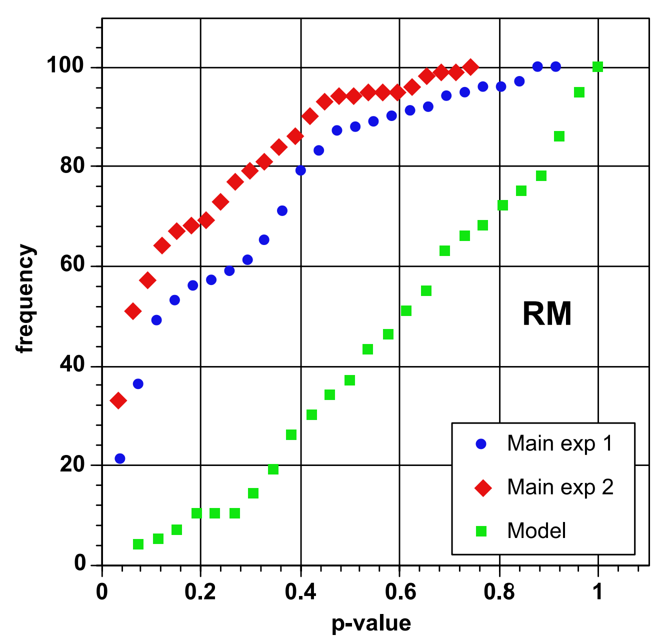

Fig. 4. The cumulative distribution function of responses from observer RM for the main experiments and for modelling (the designations as in Fig. 3).

[2] This can be easily proved when the statistic is a continuous variable. More properly, if a distribution function $F(x) = P\{\xi \leqslant x\}$ for a random variable ξ is continuous, then a random variable $F(\xi)$ is uniformly distributed over the interval [0; 1]. Indeed, given $0 \leqslant z \leqslant 1$ we have the following sequence of obvious equations: $P\{F(\xi) \leqslant z\} = P\{\xi \leqslant F^{-1}(z)\} = F(F^{-1}(z)) = z$. Therefore, the graph of the cumulative distribution function for $F(\xi)$ is linear on the interval [0; 1]. To be more exact, it is a diagonal of the unit square. If ξ is a discrete random variable, then the graph of its cumulative distribution function has a staircase form. The greater the number of points in the distribution, the narrower the width of the staircase and hence the closer the distribution to a straight line.

In contrast, the distributions of p-values from the main experiments for both observers were very far from uniform. For observer BB, for instance, all the p-values from both the main experiments were mainly concentrated in close vicinity to zero, indicating an entire inconsistency with the null hypothesis. More specifically, 99% of the p-values derived from the results of observer BB in the first main experiment and 71% in the second did not exceed 0.01 (Fig. 3). The results for observer RM were also distinctively different from uniformity (Fig. 4) although the inclination from the diagonal was not so pronounced as for BB.

6. Discussion

We have attacked a problem of threshold by using a new methodological approach. More specifically, we have tested a hypothesis that there is a region on the contrast axis where the stimuli are not distinguishable from the background. This theoretical hypothesis can be operationalised in at least three ways: (i) the stimuli from this region produce 50% correct responses in a two-alternative forced choice detection experiment (a classical operational conception of threshold); (ii) the index of detectability d' for the stimuli in this region is equal to zero (a hard threshold); (iii) the stimuli from this region cannot be ordered on the dimension of visibility. The first two predictions (50% correct probability and $d' = 0$) are not easy to test experimentally for the reasons mentioned in Section 1. This is why we have made an alternative prediction from the concept of threshold, namely, ordering of sub-threshold stimuli, and have used an alternative (transitivity) test.

It was found that while the traditional test (i) did not allow us to reject the threshold hypothesis, the new transitivity test (iii) reliably testified against it. Such a difference in results is by no means surprising, and, probably, it can be accounted for by the difference in power of the statistical tests we have employed in the preliminary and main experiments. [3] Namely, we believe that the failure of the grating stimuli to be statistically differentiated from the blank background stimulus in the preliminary experiment was due to insufficient power of the traditional test we used in the preliminary experiment. In other words, we would not exclude that the true detection probability in the preliminary experiment for those stimuli which were employed in the main experiment, was more than 0.5, but the power of the classical test to differentiate between two close probabilities which we used in the preliminary experiment was not enough to reveal it.

It is an unavoidable shortcoming of the classical threshold theories that they make predictions which are hard to test with standard statistical procedures. Sometimes it can even lead to a sort of paradox. For instance, if we adopt the operational definition of threshold based on a 50% correct criterion we would run into the following problem. A set of low contrast gratings which produced not more than 50% probability of detection on a forced choice task in the preliminary experiment nonetheless provided sufficient information for the observer to order these stimuli on a visibility dimension in the main experiment. [4] However, this paradox disappears if the theoretical framework is changed so that the threshold hypothesis within the new framework becomes reliably testable from the statistical point of view.

So we have derived another prediction from the threshold theory, namely, that stimuli with contrast lower than some critical value cannot be ordered. This prediction was not confirmed in our experiment. Contrary to this prediction it was found that gratings with very low (i.e. sub-threshold) contrast could be ordered relative to their visibility, the order satisfying the axiom of transitivity. This finding contradicts the assumption of a threshold on the contrast axis as a margin dividing a stimulus continuum into two subregions, in which stimuli are either visible or not. On the contrary we have found that there is a whole transition zone where visibility of the stimuli gradually changes from one extreme ("completely invisible") to another ("clearly visible"). So we conclude that we have found no evidence in favour of the threshold hypothesis (at least within the contrast range we have explored).

At the first glance our results support SDT as an alternative to the classical threshold concept. Indeed, it is a general assumption in SDT that sensory states are stochastically related to stimuli, and that they mediate the decision making process in threshold tasks. So, if as

[3] Although it is not easy to compare the powers of the two tests, the transitivity test seems to be more powerful one. Furthermore, it is based on a new approach which differs from the traditional one as SDT does from the traditional measuring thresholds. Recall that although we have conducted one experiment, when analysing the data we have undertaken resampling. It allowed us to evaluate not only one p-value, as is usually done, but the whole set. So the analysis was done in terms of distributions of the p-values rather than of an individual p-value. It is equivalent of having ROC curves instead of individual probabilities.

[4] A similar paradox was established by those who worked in the area called "subliminal perception" (e.g. Dixon, 1971; Holender, 1986). They claimed that information from stimuli, which were below threshold, could be used to recognise, discriminate or categorise them. We do not suggest that the stimuli in our experiments were "subliminal" (i.e. sub-threshold) since we do not accept the notion of threshold. In other words, we do not consider "subliminal perception" data as paradoxical because they can be accounted for within the context of some detection theories (see e.g. Macmillan, 1986; Macmillan & Creelman, 1991). However, we do consider that these data show to what logical puzzles the concept of threshold may lead.

is generally assumed, sensory states are ordered [5] such an order on sensory states will naturally induce a corresponding subjective order in the observers' judgements, as in our main experiments.

Nevertheless, SDT cannot account for why transitivity is not satisfied in the control experiment judgements. Indeed, SDT suggests that an observer bases his/her judgements on the same ordered set of sensory states in both cases when the stimuli are different and when there is no difference between the stimuli at all. Therefore, the order on the sensory states should have induced a corresponding order on the pair comparison judgements in both (main and control) experiments. However, the induced order has not been revealed in the control experiment. Hence, we would consider the observer's inability to order stimuli of the same contrast (relative to their visibility) revealed in the control experiment, as evidence against SDT.

One could argue that there are other lines of evidence in psychophysics in favour of SDT, particularly the many experimental results which were found to be in line with SDT. However, note firstly that the evidence in favour of SDT's application to psychophysics actually comes from indirect sources such as the second choice and rating experiments (Green & Swets, 1966; Swets, 1961; Swets, Tanner, & Birdsall, 1961). Secondly, the experimental data supposedly corroborating SDT are also in line with some threshold theories (e.g. Krantz, 1969; Luce, 1963).

Furthermore, recall that psychophysics has resorted to SDT because there is an indeterminacy caused by an observer's inability to divide stimuli into exactly two classes (e.g. "visible" and "invisible"). However, the statistical decision theory (on which SDT is based on) is by no means the only formal framework which enables us to deal with such an indeterminacy. Perhaps even more appropriate would be fuzzy sets theory which has been designed to describe in quantitative terms vague notions and fuzzy categories (Zadeh, 1965). [6]

Indeed, the main problem with a classical notion of threshold is that it cannot be described by a single number, as was the original intention. Our experiments show that an observer cannot make an exact decision as to whether a particular stimulus belongs to the class of "visible" (or "invisible") stimuli, not only because the words of ordinary language used to specify these classes are vague, but also because there is the whole variety of grades of membership in the class of invisible (as well as visible) stimuli. In terms of fuzzy sets theory it means that a subset of "invisible" stimuli is fuzzy. Fuzzy set theory suggests a measure from some scale for an element to be in a class. This measure is called a membership function for a given class. That the order in question satisfies the axiom of transitivity means that the grade of membership in the class of "invisible" stimuli can be expressed in terms of the order scale.

Therefore, an order scaling technique (e.g. Krantz et al., 1971) would be appropriate when measuring a membership function for the class of "invisible" (subthreshold) stimuli. After the membership function is measured we could approximate it by a step function if we want to have a single (not fuzzy) number to characterise an observer's at-threshold-performance. Alternatively, we can use fuzzy numbers following the rules of fuzzy arithmetic (e.g. Rodabaugh, 1983) to represent the border between sub-threshold and super-threshold subsets on the stimulus axis. We believe that the latter would seem to be more relevant to our subjective experience under a threshold task than classical threshold theories and SDT.

7. Conclusion

When measuring only probabilities of correct responses in the threshold task one loses some information essential for understanding human at-threshold-performance. Our data show that even when the probability of correct responses is at chance level an observer can distinguish various grades of visibility of stimuli. More specifically, an observer's judgements of visibility are ordered, and they satisfy the transitivity axiom. The latter means that if visibility of grating a was judged to be better than visibility of grating b and the latter in turn was judged as better than visibility of grating c, then

[5] Strangely enough, despite the fact that the application of SDT in psychophysics is based upon this assumption, the ability to order sensory states has never been put to a direct experimental test. At the same time, some widely used psychophysical techniques are based on this assumption. One example is the rating procedure, where an observer is asked to assign a category to his/her confidence that s/he really did see or hear a stimulus (Green & Swets, 1966). That observers could give a consistent category response in such a task has been known for a long time. In particular, the psychophysical method of "single stimuli" was based on the observer's capacity to confidently classify stimuli within a threshold zone under certain categories: for example high, medium, and low (Woodworth & Schlosberg, 1955, p. 217). Other examples are the multi-alternative forced choice methods to measure threshold (Green & Swets, 1966). Under these methods an observer is urged to choose one of many presentations in which a stimulus might be present, despite the fact that the observer cannot be sure that s/he experienced a stimulus in any of them. In fact during such experiments observers always respond correctly with a probability significantly higher than chance. That is usually explained by assuming an observer's ability to order the "invisible" (or "inaudible").

[6] It should be noted that it is not a novel idea to use the fuzzy mathematics in psychology of perception. Particularly, there are a few perceptual models involving the fuzzy logic (e.g. Crowther, Batchelder, & Hu, 1995; Massaro, 1987; Oden, 1979; Oden & Massaro, 1978). However, none of them has used that notion in a way which might be applied to threshold.

visibility of a was estimated to be higher than visibility of c.

However, this was the case only when all the stimuli, a, b, and c, were different. When they were the same an observer's judgements were not transitive. This finding contradicts the prediction from SDT.

An alternative account of the data obtained is proposed in terms of fuzzy sets theory. Two complementary subsets of stimuli—sub-threshold and superthreshold—and therefore a threshold as a margin dividing them, are fuzzy rather than sharp. This means that there are many grades of belongingness to both the sub-threshold and superthreshold subsets. Hence, an appropriate representation for a threshold is a fuzzy rather than an ordinary (classical) number. This may have strong implications for measuring thresholds since techniques for identification of fuzzy numbers might be quite different from current measurement techniques.

Acknowledgements

We would like to thank Professor Yuri Tyurin of Moscow State University for his advice concerning the statistical analysis in this paper, and Dr. Alexander Sostak of the University of Latvia for discussing the problem of applying fuzzy set theory to psychophysics. We also thank Dr. Martin Sawey without whose programming assistance this study could not have been completed.

References

Beard, B. L., Klein, S. A., & Tyler, C. W. (1995). Hard threshold vs. accelerating transducer in contrast detection. *Investigative Ophthalmology & Visual Science, 36*(Suppl. 4), S904.

Boring, E. G. (1950). *A history of experimental psychology*. New York: Appleton-Century-Crofts.

Crowther, C. S., Batchelder, W. H., & Hu, X. (1995). A measurement-theoretic analysis of the fuzzy logic model of perception. *Psychological Review, 102*, 396–408.

Dixon, N. F. (1971). *Subliminal perception*. McGraw-Hill.

Falmagne, J.-C. (1985). *Elements of psychophysical theory*. New York: Oxford University Press.

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

Holender, D. (1986). Semantic activation without conscious identification in dichotic listening, parafoveal vision, and visual masking: a survey and appraisal. *The Behavioral and Brain Sciences, 9*, 1–66.

Kendall, M., & Stuart, A. (1979). *The advanced theory of statistics, Vol. 2* (4th ed.). London & High Wycombe: Charles Griffin & Co Ltd.

Kontsevich, L. L., & Tyler, C. W. (1997). Nonlinearities of near-threshold contrast transduction. *Perception, 26*(Suppl.), 2.

Kontsevich, L. L., & Tyler, C. W. (1999). Nonlinearities of near-threshold contrast transduction. *Vision Research, 39*, 1869–1880.

Krantz, D. H. (1969). Threshold theories of signal detection. *Psychological Review, 76*, 308–324.

Krantz, D. H., Luce, R. D., Suppes, P., & Tversky, A. (1971). *Foundations of measurement, Vol. I*. San Diego: Academic Press.

Logvinenko, A. D. & Vityaev, E. E. (2003). Testing statistically transitivity. *Journal of Mathematical Psychology*, submitted, Available at: <www.psych.qub.ac.uk/homepages/logvinenko/stochastictransitivity.pdf>.

Luce, R. D. (1963). A threshold theory for simple detection experiments. *Psychological Review, 70*, 61–79.

Luce, R. D., & Suppes, P. (1965). Preference utility and subjective probability. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. III, pp. 249–410).

Macmillan, N. A. (1986). The psychophysics of subliminal perception. *Behavioral and Brain Sciences, 9*, 38–39.

Macmillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide*. Cambridge e.a: Cambridge University Press.

Massaro, D. W. (1987). *Speach perception by ear and eye: A paradigm for psychological inquiry*. Hilsdale, NJ: Erlbaum.

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

Nachmias, J., & Sansbury, R. V. (1974). Grating contrast: discrimination may be better than detection. *Vision Research, 13*, 1039–1042.

Oden, G. C. (1979). A fuzzy logical model of letter identification. *Journal of Experimental Psychology: Human Perception and Performance, 5*, 336–352.

Oden, G. C., & Massaro, D. W. (1978). Integration of featural information in speech perception. *Psychological Review, 85*, 172–191.

Rodabaugh, S. E. (1983). Separation axioms and the fuzzy real line. *Fuzzy Sets and Systems, 11*, 163–183.

Simpson, W. A., Falkenberg, H. K., & Manahilov, V. (2003). A "hard threshold" in detection, summation, and direction discrimination? In *The 3rd Annual Meeting of the Vision Sciences Society*, Sarasota, Florida, May 9–14, 2003, p. 153.

Suppes, P., Krantz, D. M., Luce, R. D., & Tversky, A. (1989). *Foundations of measurement, Vol. II*. San Diego e.a: Academic Press.

Swets, J. A. (1961). Is there a sensory threshold? *Science, 134*, 168–177.

Swets, J. A., Tanner, W. P., Jr., & Birdsall, T. G. (1961). Decision processes in perception. *Psychological Review, 68*, 301–340.

Woodworth, R. S., & Schlosberg, H. (1955). *Experimental Psychology* (3rd ed.). London: Methuen & Co. Ltd.

Zadeh, L. A. (1965). Fuzzy sets. *Information and Control, 8*, 338–353.