

Reply

Visual resolution: operational definitions with an eye towards historical precedence

1. Introduction

The perceptual task of “resolving” one from two objects can be operationally defined and measured psychophysically. However, experimental methods aside, terms used to describe the measurement can have historical significance and should be chosen carefully as [Westheimer \(in press\)](#) points out. The term in question in this case is “resolution” which historically involves describing the optical resolving power of a telescope for distinguishing a single from a double star. [Westheimer \(in press\)](#) suggests that a resolution task should involve the detection of a trough (a dark thin dimple) between the two stars, so that one is sure that the image did not come from an extended single object.

In modern astronomy the task of resolving a double star from a single star makes use of prior knowledge (e.g. stars are point-like objects) and does not rely on detecting a trough in the luminance distribution. Deconvolution methods that take into account the telescope point spread function (see [Heydari-Malayeri, Remy, & Magain, 1989](#); [Karovska, 2002](#)) are used to reveal fine detail in modern astronomical images of low contrast objects in noise. The article, “Using a priori information in image restoration: Natural resolution limit” by [Terebizh \(1999\)](#) uses numerical simulation to study the limiting separation between two point-like components of a double star.

[Westheimer \(in press\)](#) raises the question about whether prior knowledge should be used in the psychophysical definition of resolution. Consider the task of discriminating a single line from a pair of lines (for psychophysics we will discuss lines rather than dots). If one had prior knowledge that the lines were thin (a signal detection methodology like 2AFC provides the needed prior knowledge), then the task becomes one of blur discrimination in that the two-line stimulus would produce a broader retinal image than the single line. This is the operational definition of resolution used by [Carney and Klein \(1997\)](#) in their one thin line vs. two thin lines two alternative forced choice task (2AFC). If however, there was no

prior knowledge about whether the line stimuli were thin, by intermixing the two-line stimulus with a random assortment of blurred lines then the size of the retinal image would not be a useful cue. In that case one might want to see a faint dimple between the two lines before saying it could not have been a single line. The presence of a dip in the retinal image is similar to the Rayleigh criterion which defines two points as being resolvable when the center of the image from one point falls on the first diffraction zero of the second point. At the Rayleigh criterion separation a dark dimple is found in the middle of the retinal image of two lines. A problem with basing resolution on the presence of a dimple is that by the time the dimple cue becomes the crucial cue, the line is sufficiently blurred that the task does not feel like a resolution task. Resolution as defined by [Carney and Klein \(1997\)](#) is actually more consistent with modern astronomical use of the term resolution than a Rayleigh criterion based definition such as proposed by [Westheimer \(in press\)](#). Our use of resolution is also consistent with that of [Geisler \(1984\)](#) who applied ideal observer theory to discriminations tasks that he calls intensity discrimination, resolution and separation discrimination. The ideal observer uses all available information in a 2AFC psychophysical paradigm. In a subsequent paper which included human data [Geisler and Davila \(1985\)](#) report that “the 2AFC task permits high degree of scrutiny and the use of any available cue to make the discrimination” to explain the low threshold compared to earlier data. Moreover, they reported a subjective impression of slight blurring or widening of the resolution target.

2. Blur vs. width cue distinction

As pointed out by [Westheimer \(in press\)](#) and described above, multiple cues are generally available when performing a simple two-line resolution task. [Westheimer \(in press\)](#) suggests that in our one-line vs. two-line task a width cue was used. He points out that the difference between the two stimuli corresponds to a

change of 11" in total width at half height. He further notes that this width change is greater than the 6" width discrimination found by Westheimer and McKee (1977) in a bar width discrimination task (threshold defined at $d' = 0.76$). This brings up an important distinction that must be made between blur discrimination (Carney & Klein, 1997) and hyperacuity width discrimination (Westheimer & McKee, 1977). We argue that blur width discrimination is a candidate for being called "resolution" whereas hyperacuity width discrimination is not. The simplest target for discussing hyperacuity width vs. blur involves width discrimination using a thin bar. Fig. 1 shows the threshold width change Δw (ordinate) at which a bar of width $w + \Delta w$ can be discriminable from a bar of width w (abscissa). The bar strength (Weber contrast of the bar times its width) is kept constant so that the bar strength does not provide a discrimination cue. The threshold for $w = 0$ (single pixel line) is what we will call the blur detection threshold and what Carney and Klein (1997), called the resolution threshold. The blur and hyperacuity threshold limits in Fig. 1 are given by equations (1) and (2), respectively.

$$\text{Blur limit} = \sqrt{w^2 + 12 * \text{Quad}/\text{Bar}} - w \tag{1}$$

$$\text{Hyperacuity limit} = \text{Line}/\text{Michelson Bar Contrast} \tag{2}$$

The mathematics behind Fig. 1 is in Carney and Klein (1997) and Klein (1989).

The only assumptions going into this plot are that the quadrupole detection threshold is $1\% \text{min}^3$ and the line detection threshold is $2\% \text{min}$, values taken from Carney and Klein (1997), for slightly different stimulus conditions. The bar strength chosen for Fig. 1 was taken to

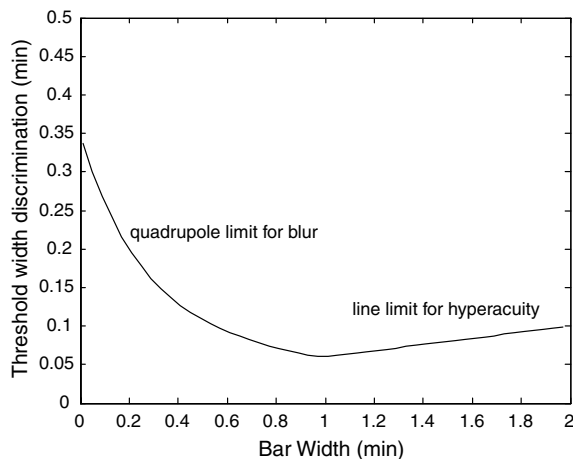


Fig. 1. Plotted are width discrimination thresholds for different bar widths as predicted by formulations provided by Carney and Klein (1997) and Klein (1989). For line widths less than about 1min, threshold is limited by the quadrupole detection threshold and for greater widths performance is limited by the line detection threshold.

be 100%min. Fig. 1 shows that for base widths less than 1 min (the blur regime) the discrimination threshold is determined by the quadrupole detection threshold and for widths greater than 1 min (the hyperacuity regime) it is determined by the line detection threshold. The hyperacuity threshold is easily understood in terms of an edge shift being produced by adding a line to the edge (Klein, 1989; Klein, Casson, & Carney, 1990). The linearly increasing threshold in the hyperacuity regime is due to the edge contrast decreasing as the bar width increases. The point being made by Fig. 1 is that it is misleading to compare the 6" threshold ($d' = 0.76$) found by Westheimer and McKee for bar widths or line separations of greater than one min to the blur discrimination threshold of 20" for a single pixel line.

3. Characterization of the blur and dimple resolution cues

The blur cue is dominated by low spatial frequency differences between the two-line and the comparison stimulus in the 2AFC task, whereas the spatially localized dimple cue is dominated by high spatial frequency differences between the targets. The cue distinctions can be seen in frequency domain plots of potential comparison targets. Fig. 2 shows the Fourier transforms of three stimuli: 2-Line, Qmatch and Zmatch, where 2-Line is a pair of thin lines (solid curve), Qmatch is a bar with matched quadrupole strength (dot-dashed curve) and Zmatch is a bar with matched zero-crossing (dashed

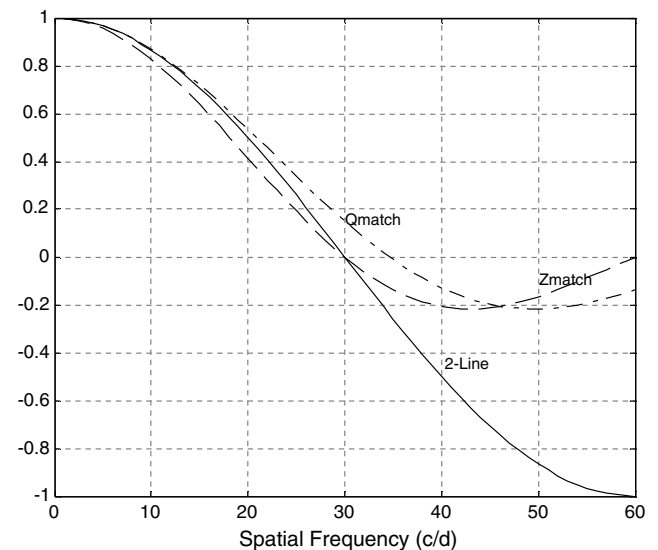


Fig. 2. Blur and resolution stimulus cues in the frequency domain. The solid line plot is a comparison 2-Line stimulus with 1 min separation. The dashed line plot is for a bar stimulus with matched zero crossing (30 c/d) that deviates significantly from the 2-Line comparison at low spatial frequencies, like the stimulus used by Westheimer and Beard (1998). The dash-dot plot depicts a bar stimulus with matched standard deviation (1/2min) that matches more closely the 2-Line comparison stimulus at low frequencies and thereby reduces the blur cue but deviates more at high frequencies.

Table 1
Stimulus categories, discrimination thresholds and stimulus characterization

	Observer# 1 (min)	Observer# 2 (min)	Profile	Width (min)	SD (min)	Fourier	ZC (c/d)
2-Line	Reference for 1 line		[1, 1]/2	$s = 1/3$ min	$s/2$	$\cos(1.5fs)$	90
1-Line	0.31	0.38	1	0	0	1	
2-Line	Reference for 3 line and bar		[1, 0, 0, 1]	$s = 1$ min	$s/2$	$\cos(1.5fs)$	30
3-Line, SD	1.04	0.91	[9, 0, 14, 0, 9]/16	$4s/3$	$s/2$	$(7 + 9\cos(2sf))/16$	35.3
3-Line, ZC	0.97	1.04	[10, 0, 10, 0, 10]/15	$4s/3$	$16s/27$	$(5 + 10\cos(2sf))/15$	30
3-Line, blend	1.0	1.0	[10, 0, 12, 0, 10]/16	$4s/3$	$15s/27$	$(6 + 10\cos(2sf))/16$	31.7
Bar, Zmatch	0.9	0.93	[1, 1, 1, 1, 1]/3	2s	$s/\sqrt{3}$	$\sin(3fs)/3fs$	30

curve). For the conditions of Table 1, Zmatch has a width of 2 min (Table 1, row 7) and Qmatch has a width of $\sqrt{3}$ min. The matched quadrupole strength means the two-line stimulus and the bar stimulus have the same standard deviation, corresponding to having the same Fourier amplitude in the limit of low frequency. In Fig. 2 we chose a two line stimulus to have a separation of 1 min (close to the resolution threshold presented in Table 1) so the spatial standard deviation is $1/2$ min and the Fourier zero crossing is 30 c/d.

The Zmatch bar stimulus is very similar to the resolution stimulus used by Westheimer and Beard (1998) to limit the blur cue, but note that substantial low frequency differences remain. The purpose of comparing the 2-Line and Qmatch stimuli is to remove not only the strength cue but also the quadrupole (blur) cue. Two cues are available for discriminating both bar stimuli from the two-line stimuli: (a) for spatial frequencies below about 30 c/d the Fourier amplitudes of both bar stimuli deviate from the two-line stimulus, introducing a blur cues. (b) For spatial frequencies above 30 c/d the 2-Line stimulus has much stronger negative Fourier amplitudes than both bar stimuli, corresponding to a dark dimple. Visual resolutions as measured by standard eye charts likely involve both classes of cue.

4. Operational definition of resolution

To use modern objective psychophysical testing methods an operational definition of resolution is needed. Can we devise a stimulus that emphasizes the historical Rayleigh based trough cue? Westheimer and Beard (1998) have already devised a stimulus (see Table 1, row 7) that attempts to minimize the blur and luminance cues with the intention of limiting the available cue to the trough in the retinal light distribution. This was a step in the direction of operationally defining a particular resolution task that emphasizes features based on the classical Rayleigh criteria. Proceeding along the same lines, we propose two operational definitions of visual resolution based on the comparison of a two-line stimulus to: (1) a single line (Res_{2vs1}) and (2) three closely spaced lines (Res_{2vs3}). We prefer the 3-Line comparison target over the bar target because of its smaller

width, as shown in the fifth column of Table 1. We will also discuss the significance of the presence of a trough in the luminance distribution ($\text{Res}_{\text{dimple}}$) advocated by Westheimer. For direct comparisons, psychophysical discrimination thresholds for the two different categories for two observers are provided.

5. General methods

A two alternative spatial forced choice method was used for each discrimination task. Each stimulus was presented 250 times (in blocks of 50) at each of three viewing distances, 4.5, 5 and 5.5 m (0.372, 0.335 and 0.305 min/pixel at the respective distances). The stimulus lines were 104 min long with comparison patterns separated by 20 min (at 5 m) in a simultaneous 2AFC paradigm. Stimuli were presented for 1.5 s, enabling the observers to scan the two patterns being compared. The simultaneous presence of the two patterns, together with scanning, minimized the effect of accommodative fluctuations across time. The display background luminance was 13.0 cd/m² and the maximum luminance was 84 cd/m². The maximum line contrast is $(84 - 13)/13 = 546\%$. The strength of a one-pixel, peak luminance line at the 5.0 m viewing distance was $546\% \times 0.33 \text{ min} = 182\% \text{ min}$, which is about 100 times the typical line detection threshold. The WinVis psychophysical testing system (www.neurometrics.com) was used to generate and present the stimuli. Table 1 identifies the different stimulus types (column 1) along with the normalized adjacent pixel intensities, orthogonal to the line orientation (column 4, profile). The total luminance flux between patterns within a stimulus category is the same. Details about the individual stimulus patterns are presented in the corresponding operational definition sections below.

The probability of correct pattern discrimination (p) at each of the three viewing distances was converted to d' , where $d' = 2 \times \text{erfinv}(2 \times p - 1)$ and erfinv is the inverse error function. To estimate “resolution” threshold, the d' function was fit by a power function $d' = 1.5(s/\text{th})^p$, where s is the 2-Line stimulus separation, th is the 2-Line “resolution” threshold, and p is the exponent of the d' function. We have chosen to define threshold at

$d' = 1.5$ corresponding to 85.6% correct rather than a lower d' value, to improve the accuracy of the threshold estimate (Klein, 2001; Green, 1990).

5.1. Two lines vs. one line resolution (Res_{2vs1})

The resolution task used by Carney and Klein (1997) was Res_{2vs1} . This 1-line vs. 2-Line task is a blur discrimination task limited by the observer's quadrupole threshold. This differs from spatial localization tasks which are limited by an observer's line or dipole detection threshold, depending the particular pattern (Carney & Klein, 1997, 1999; Klein et al., 1990). Knowledge of the quadrupole threshold is able to predict the Res_{2vs1} threshold quite well. The optimal line "resolution" thresholds reported by Carney and Klein (1997) for a Res_{2vs1} stimulus were between 0.3 and 0.4 min, better than traditional grating resolution of 1 min but not in the hyperacuity range associated with a width discrimination task.

We tested two observers on this Res_{2vs1} stimulus. The one-line target was one pixel wide with luminance, 84 cd/m² and the comparison two-line target consisted of two adjacent pixel lines (42 cd/m²), corresponding to a one pixel separation. The total luminance flux was matched to avoid a luminance cue. The probability of correct discrimination for these two patterns was measured at viewing distances of 4.5, 5.0 and 5.5 m. Further details and results are provided in Table 1 (rows 1 and 2). The calculated two-line "resolution" thresholds for the two subjects were 0.31 and 0.38 min, respectively. These results are similar to previously reported findings (Carney & Klein, 1997) even though the display conditions were significantly different. Clearly, a trough in the luminance profile at the retina was not a cue in this task since, as we will discuss, the trough only becomes visible for two lines separated by more than 1 min. Both targets appeared as a very narrow line, with the two line target appearing more blurred than the one line target.

Should the Res_{2vs1} task that was the focus of the Carney and Klein (1997) paper be called resolution? To avoid confusion with the Raleigh definition of resolution, we suggest blur discrimination may be a better task descriptor for this type of "resolution" target. Similarly, blur discrimination may also be the appropriate descriptor for the task of detecting a dipole added to an edge (Carney & Klein, 1997).

5.2. Two lines vs. three lines (Res_{2vs3})

Should the definition of resolution involve the visibility of a depression in the image (Res_{dimple})? The vision research community may want to reserve the word resolution for an operational definition that offers similarity to the classical Rayleigh definition without the great difficulty of specifying the actual cue used. Westheimer and Beard (1998) tried to emphasize the trough cue by redu-

cing the blur cue using a bar comparison stimulus of six adjacent lines. The two line stimulus was the second and fifth of these lines, each three times the intensity (note that Westheimer and Beard describe their stimulus differently). They called this a pure resolution task since the average flux per unit retinal area was about the same. Unfortunately, while the bar stimulus has the same first zero crossing in the Fourier domain (see Table 1, row 7), it still differs significantly from the two-line stimulus at the low spatial frequencies (the blur cue) as shown in Fig. 3. Fig. 3 is a frequency domain plot of the two-line target, the bar target and our new 3-Line targets, all at viewing distance equivalent to 1/3 min pixels (about 5 m in our apparatus). While the trough cue is dominated by high frequency content, the differences at low spatial frequencies might still be utilized as a cue for this task. As expected, thresholds (Table 1) using the bar stimulus are significantly higher as compared to Res_{2vs1} thresholds. Observers report using the trough cue at the 4.5 m distance but reverting to other stimulus cues at the 5 and 5.5 m viewing distance.

To reduce the low frequency blur cue we compared three, three-line targets for comparison with the two-line target. Their profiles are [10, 0, 10, 0, 10]/15, [9, 0, 14, 0, 9]/16 and [10, 0, 12, 0, 10]/16 (see Table 1). The first of these stimuli has the same Fourier domain zero crossing as the two-line stimulus [1001]. The second stimulus has the same standard deviation as the two-line stimulus. The third, 3-Line stimulus is a compromise blend of the first two. Fig. 3 (left panel) shows the frequency plot of the standard 2-Line resolution target (solid line) along with three 3-Line comparison targets and the 6-Line bar target (Zmatch). The equations for the five frequency plots are presented in Table 1, column 7 for the second category of stimuli. Fig. 3 (right panel) is the same plot except the 2-Line plot has been subtracted from the other plots so that spatial frequency differences from the 2-Line stimulus are more easily seen. The 3-Line target, with equal strength lines (Table 1, row 5) is the dash-dotted line. This target, like the bar target, Zmatch (bold dotted line), has the same zero crossing as the 2-Line target with significant remaining low spatial frequency differences. It is closer to the 2-Line target at all spatial frequencies below the 60 c/d limit of vision. This closer match to the 2-Line target is not surprising given the tighter overall width of the 3-Line target relative to the bar target as seen in the fifth column of Table 1. The dashed line in the figure is for a 3-Line target with normalized line intensities of (9, 0, 14, 0, and 9)/16. This stimulus has its first zero crossing at 35.3 c/d and more closely matches the 2-Line stimulus at low frequencies as would be expected since the stimulus has the same standard deviation (quadrupole moment) as the 2-Line stimulus. However, it deviates even more from the 2-Line stimulus at high frequencies. The dotted line in the figure is for a 3-Line target with normalized line intensi-

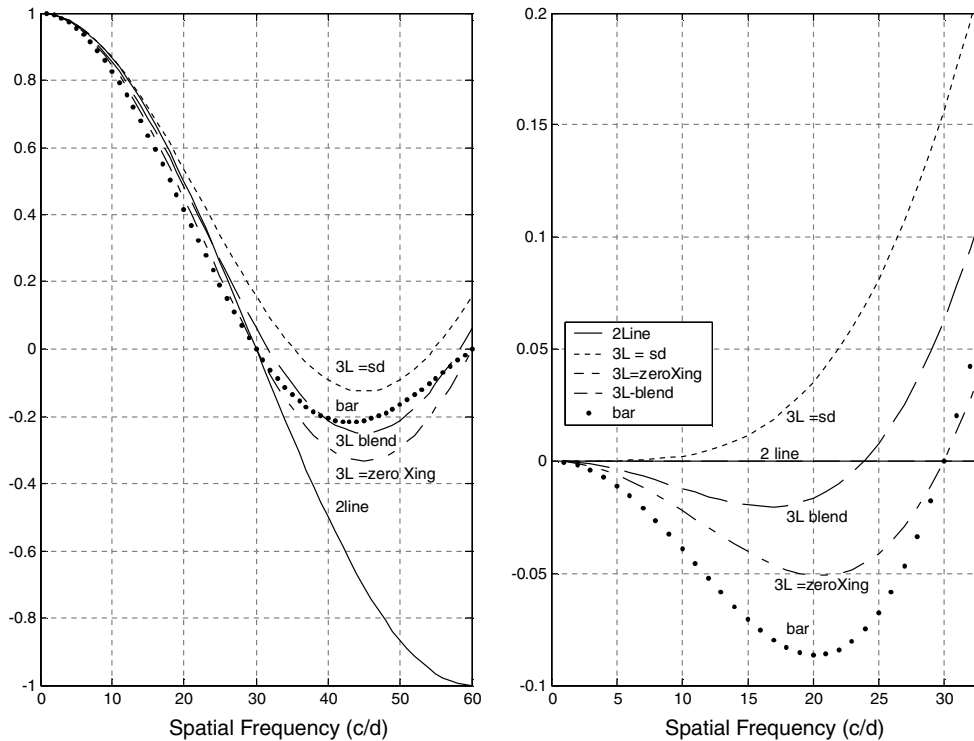


Fig. 3. Frequency domain plots of the five stimuli used in this study. In the left panel the two-line reference stimulus is plotted along with the three 3-Line stimuli plus the bar stimulus, Zmatch, labeled 'bar', used by Westheimer and Beard (1998). Details of the five targets are presented in the second stimulus category of Table 1. To view the differences between the five targets the two-line spectrum was subtracted from the each bar target and plotted in the right hand panel. For frequencies below about 24c/d the bar comparison stimulus used by Westheimer and Beard (1998) deviates more from the two-line comparison, and contribute more to the blur cue than any of the three line targets.

ties of (10, 0, 12, 0, and 10)/16. This stimulus, a blend of the other two 3-Line stimuli, is an excellent match to the 2-Line stimulus at very low frequencies, which should minimize blur cues while the larger deviations at high frequencies should emphasize the trough cue.

The $Res_{(2vs3)}$ thresholds for these 3-Line targets are shown in Table 1. For observer 1 the results are as expected based on the low frequency information, the better the match at the lowest spatial frequencies the larger the resolution thresholds. For observer 2 the ordering of the 3-Line results is compatible with using the high frequency information. The effect of different line strengths in the 3-Line stimulus is subtle and will require more detailed methods to tease them apart. The thresholds obtained probably reflect a differential weighting of available cues. When performing the task it did not seem like a single cue was involved but rather the observer had to pay more or less attention to various cues depending on the viewing distance and accommodative state. In view of these findings, we recommend any of the 3-Line stimuli as excellent comparison targets with the 2-Line stimulus. One approach to minimizing the low frequency blur cue might be to randomly use the different 3-Line stimuli on successive trials to further confuse use of the blur cue.

In summary, the use of the word resolution has a long history in optics and astronomy. Today's astronomers

use all available prior knowledge when confronted with resolving one from two stars. In a perceptual two-line resolution task, use of prior knowledge is akin to using all available cues when determining resolution threshold as in the Res_{2vs1} stimulus category. Westheimer (in press) seeks to restrict use of the term resolution to tasks that involve detection of a luminance trough in the light distribution. As it turns out, it is difficult to generate stimuli that only present that particular cue on the observer's retina. We propose two operationally defined 'resolution' target categories. To accommodate the historical usage of resolution thresholds being about 1 min and avoid further confusion we propose assigning the term blur discrimination to the 1-line vs. 2-Line task which is equivalent to detection of a quadrupole added to line as described by Carney and Klein (1997, 1999). The term 'resolution' could be reserved for the 2-Line vs. 3-Line task that, as operationally defined, may involve the detection of luminance trough similar to the Rayleigh criteria. Both of the observers plus two further observers consistently saw the trough in the 2-Line stimulus at the 4.5m viewing distance. This task also may bear the closest relation to resolution as applied to grating detection. It must be pointed out that when the trough is visible, the stimulus looks decidedly blurred.

Westheimer (in press) has correctly noted the confusion over the term resolution and offers a semantic

and historical argument for when the term should be used. While we disagree over the details of the argument, we agree clarification is needed and propose two operationally defined “resolution” tasks, to one of which we apply the label two-line resolution.

Acknowledgement

This research was supported by NEI grant EY04776.

References

- Carney, T., & Klein, S. A. (1997). Resolution acuity is better than vernier acuity. *Vision Research*, 37, 525–539.
- Carney, T., & Klein, S. A. (1999). Optimal spatial localization is limited by contrast sensitivity. *Vision Research*, 39, 503–511.
- Geisler, W. S. (1984). Physical limits of acuity and hyperacuity. *Journal of the Optical Society of America A*, 1, 775–782.
- Geisler, W. S., & Davila, K. D. (1985). Ideal discriminators in spatial vision: two-point stimuli. *Journal of the Optical Society of America A*, 2, 1483–1497.
- Green, D. M. (1990). Stimulus selection in adaptive psychophysical procedures. *Journal of the Acoustical Society A*, 87, 2662–2674.
- Heydari-Malayeri, M., Remy, M., & Magain, P. (1989). Two more massive stars resolved. *Astronomy and Astrophysics*, 222, 41–44.
- Karovska, M. (2002). Deconvolution methods in astronomy, American Astronomical Society, 201st AAS Meeting. *Bulletin of the American Astronomical Society*, 34, 1214.
- Klein, S. A. (1989). Visual multipoles and the assessment of visual sensitivity to displayed images. *SPIE: Human Vision, Visual Processing and Digital Display*, 1077, 83–92.
- Klein, S. A., Casson, E., & Carney, T. (1990). Vernier acuity as line and dipole detection. *Vision Research*, 30, 1703–1719.
- Klein, S. A. (2001). Measuring, estimating, and understanding the psychometric function: a commentary. *Perception and Psychophysics*, 63, 1421–1455.
- Terebizh, V. Y. (1999). Using a priori information in image restoration: natural resolution limit. *Astronomy Reports*, 43, 42–58.
- Westheimer, G., & Beard, B. L. (1998). Orientation dependency for foveal line stimuli: detection and intensity discrimination, resolution, orientation discrimination and vernier acuity. *Vision Research*, 38, 1097–1103.
- Westheimer, G. (in press). The resolving power of the eye. *Vision Research*. doi:10.1016/j.visres.2004.01.019.
- Westheimer, G., & McKee, S. P. (1977). Spatial configurations for visual hyperacuity. *Vision Research*, 17, 941–947.

Thom Carney
Stanley A. Klein
Vision Science
University of California
Berkeley, CA 94720-2020, USA
Neurometrics Institute
Oakland, CA 94619, USA