A reevaluation of Whittle (1986, 1992) reveals the link between detection thresholds, discrimination thresholds, and brightness perception

David Kane Universitat Pompeu Fabra, Barcelona, Spain

Marcelo Bertalmío Universitat Pompeu Fabra, Barcelona, Spain

In 1986, Paul Whittle investigated the ability to discriminate between the luminance of two small patches viewed upon a uniform background. In 1992, Paul Whittle asked subjects to manipulate the luminance of a number of patches on a uniform background until their brightness appeared to vary from black to white with even steps. The data from the discrimination experiment almost perfectly predicted the gradient of the function obtained in the brightness experiment, indicating that the two experimental methodologies were probing the same underlying mechanism. Whittle introduced a model that was able to capture the pattern of discrimination thresholds and, in turn, the brightness data; however, there were a number of features in the data set that the model couldn't capture. In this paper, we demonstrate that the models of Kane and Bertalmío (2017) and Kingdom and Moulden (1991) may be adapted to predict all the data but only by incorporating an accurate model of detection thresholds. Additionally, we show that a divisive gain model may also capture the data but only by considering polarity-dependent, nonlinear inputs following the underlying pattern of detection thresholds. In summary, we conclude that these models provide a simple link between detection thresholds, discrimination thresholds, and brightness perception.

Introduction

In this paper, we remodel and reevaluate the data from two papers by Paul Whittle. Both papers investigate luminance perception. The first study (Whittle, 1986) was a luminance discrimination experiment. The stimulus was achromatic and consisted of two small uniform patches presented horizontally on either side of a central fixation point. The background was uniform, and a schematic is illustrated in Figure 1. The task was to discriminate between the luminance of the test (I_t) and the reference (I_r) patches. Luminance discrimination thresholds were investigated using a reference luminance that was either above (positive/bright pedestals: $I_r > I_b$) or below (negative/dark pedestals: $I_r < I_b$) the background luminance level I_b. A subset of Whittle's discrimination data is shown in Figure 1, and it can be seen that substantially different results are obtained for the two polarities. For positive/bright pedestals, thresholds increase approximately in proportion to the pedestal luminance, which is defined as the luminance difference between the background and the reference patch ($I_p = |I_b - I_r|$). However, for negative/dark pedestals, a more complex pattern of thresholds is observed with the form of an inverted "U."

To account for the data, Paul Whittle proposed a contrast term W:

$$W = \begin{cases} I_p/I_r, & \text{if } I_r \leq I_b \\ I_p/I_b, & \text{if } I_r \geq I_b \end{cases}. \quad (1)$$

As W is determined by a fixed denominator for positive/bright pedestals and a variable denominator for negative/dark pedestals, a polarity-dependent result is obtained that is broadly consistent with the experimental data. To understand why, it is helpful to reformulate Whittle's model in the form $\Delta I = f(I_p)$, and the calculations for doing so can be found in appendix A of Kingdom and Moulden (1991); for positive pedestals, this gives

$$\Delta I = kI_p, \quad (2)$$

which can be thought of as a simple Weber's law on the contrast dimension, and for negative pedestals, this gives

$$\Delta I = k\frac{I_r I_p}{I_b}. \quad (3)$$

Citation: Kane, D., & Bertalmío, M. (2019). A reevaluation of Whittle (1986, 1992) reveals the link between detection thresholds, discrimination thresholds, and brightness perception. *Journal of Vision*, *19*(1):16, 1–13, https://doi.org/10.1167/19.1.16.

https://doi.org/10.1167/19.1.16 Received January 26, 2018; published January 29, 2019 ISSN 1534-7362

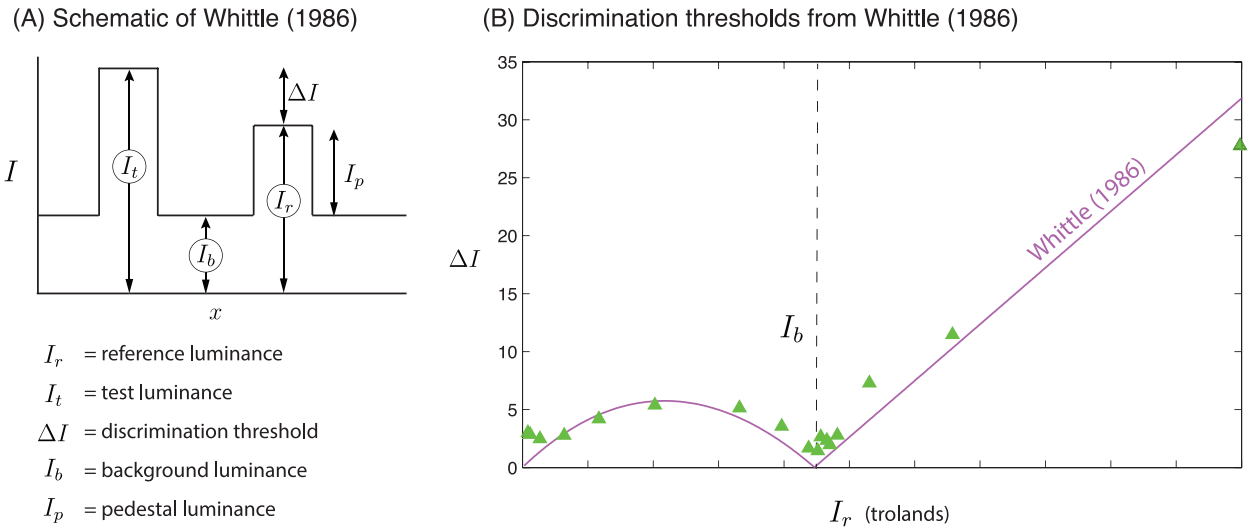

Figure 1. (A) Stimulus schematic and key for the luminance discrimination experiment. (B) Discrimination thresholds plotted over the range of a CRT monitor. The solid line denotes the predictions of Whittle (1986).

The product of the ascending variable I_r and the descending variable I_p produces an inverted "U." The predictions from these two equations are shown by the solid line in Figure 1B.

The second paper by Whittle (1992) investigated the perception of luminance modulations that were greater than a single threshold. This was achieved using the stimulus configuration illustrated in the insets of Figure 2. The stimuli consisted of a series of circular patches arranged in a spiral formulation, all viewed upon a uniform background and on a CRT monitor. The outermost patch was maximally light (I_{max}), and the innermost patch was maximally dark (I_{min}). Subjects were asked to manipulate the luminance of the intermediary patches until each luminance step appeared to be of equal magnitude. In Figure 2, we reproduce the data from two conditions. In the leftmost figure, the test patches were yellow, and the background green and the resulting nonlinearity is compressive with no inflexion. This form is common to most of the nonlinearities reported in the literature (Wyszecki & Stiles, 1982). However, when the test patches and the background were both achromatic, the function steepened around the background luminance level: This is known as the "crispening" effect (Takasaki, 1966) and mirrors the high sensitivity around the background luminance noted in the discrimination threshold data set. To make this point, Whittle plotted the discrimination thresholds alongside the luminance intervals, and we replot this data in Figure 3. The luminance intervals (blue dots) are consistently 4.2 times greater than the discrimination thresholds (green triangles). This strong correlation indicates that the two experimental paradigms were able to probe the same underlying mechanism. This is highly encouraging as it suggests that (under some circumstances) the suprathreshold percept can be reliably measured and, conversely, that the task of discriminating between the luminance of two spatially separated patches is a meaningful way to probe luminance perception. Additionally, the work indicated that the effect known as crispening could be studied using threshold measures, which are generally consid-

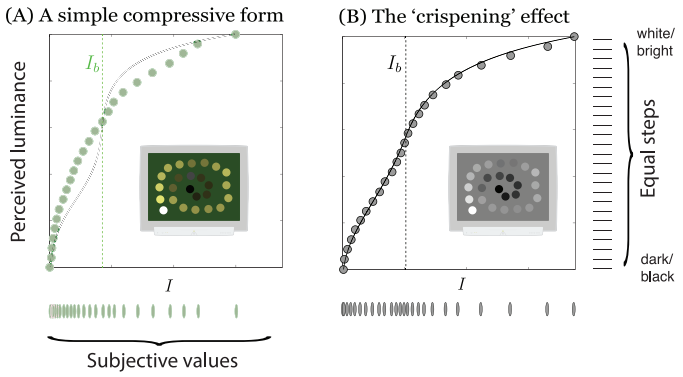

Figure 2. The insets illustrate the two experimental conditions from Whittle (1992). The task was to set the luminance levels such that the luminance steps between adjacent patches appeared equal. The luminance levels set by subjects are then plotted on abscissa against equal steps on the ordinate. The left plot shows the data when the test patches are yellow and the background green. The function is compressive with no inflexion. The right plot shows the data when the stimuli is achromatic. This function is much steeper around the background luminance level, an effect known as crispening (Takasaki, 1966). Whittle's model (blue line) can predict the data in the achromatic condition but not the data in the yellow–green condition—in effect, crispening is hard-coded in this model.

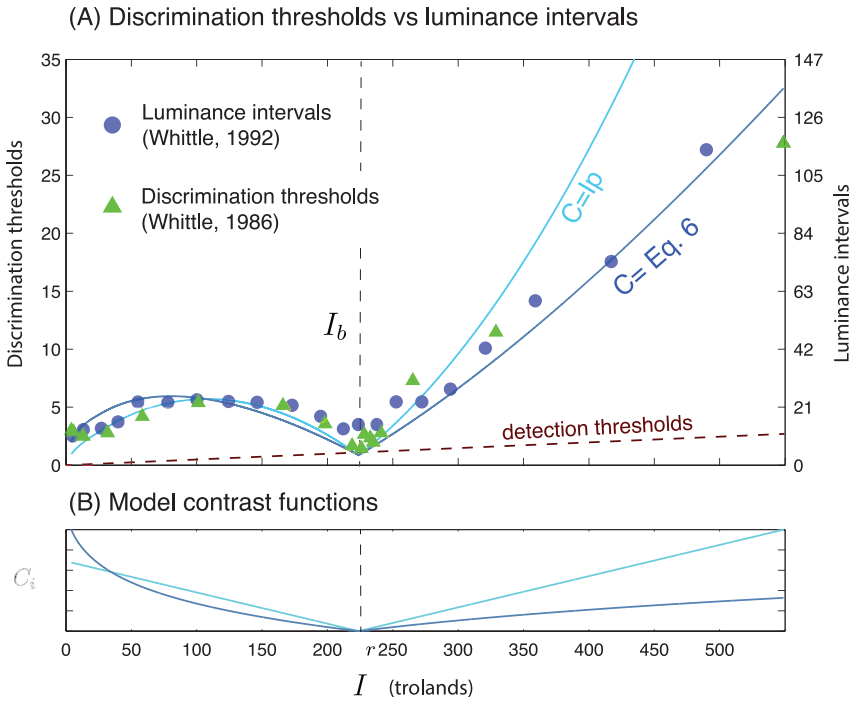

Figure 3. (A) The discrimination thresholds from Whittle (1986) are plotted with green triangles, and the luminance intervals set by subjects in Whittle (1992) are denoted by blue circles. The discrimination threshold values are denoted by the left ordinate, and the luminance intervals are denoted by the right ordinate, which is scaled by a factor of 4.2. A detection threshold function obtained from Blackwell (1946) is shown by the dashed line. We note that the detection threshold is valid only when ($I_b = I_p$). The predictions from Kane and Bertalmío (2017) by the solid lines (see text) use two models of contrast gain: one in which contrast C is simply $C = I_p$ and a nonlinear formulation of contrast as defined by Equation 6. The respective linear and nonlinear contrast functions are plotted in panel B. Note how the amplitude of the function is considerably less for positive pedestals for the contrast function of Equation 6; this is critical to allowing the accurate prediction of discrimination thresholds, which are otherwise predicted as too high.

ered to be more reliable than absolute or relative judgments of suprathreshold luminance.

Whittle (1992) referred to his data using the term "brightness," but we wish to note that, for stimuli in which there is no clear change in the overall illuminance across a scene, judgments of *brightness*, *lightness*, or local *brightness contrast* (all suprathreshold judgments) cannot be distinguished (Arend & Goldstein, 1987; Arend & Spehar, 1993; McCourt & Blakeslee, 2008). We shall use the term "brightness" to be consistent with Whittle, but we note that our interpretation is that, for simple stimuli such as that used by Whittle, suprathreshold judgments probe the same mechanism as threshold judgments and that the use of the term "brightness" could be interchanged with the term "lightness."

The close relationship between the discrimination thresholds (Whittle, 1986) and the luminance intervals set by subjects (Whittle, 1992) meant that Whittle could model the brightness functions to a high degree of accuracy by integrating over the inverse of Equations 2 and 3. An example curve is shown in Figure 2B and provides an excellent fit although we note (as did Whittle) that an additional constant was required to avoid an integration over one/zero. However, no parameter combination could allow the prediction of the data in the yellow–green condition as shown in Figure 2A. In short, the crispening effect is hard-coded into the model of Whittle (1986).

Overview

The motivation for this study was to find and evaluate models that could capture the full set of effects noted in Whittle (1986, 1992). As discussed above, Whittle was able to accurately model both the discrimination threshold data and the brightness data for achromatic stimuli for which strong crispening was observed but was unable to model the brightness data for functions that did not exhibit crispening. As such, we identify three other models that have the potential to overcome this difficultly and can also capture two other features of Whittle's data sets that we have not yet discussed. These features are the "dipper" effect, which describes a decrease in sensitivity at low pedestal luminance levels (Solomon, 2009), and the heterogeneity of Whittle's discrimination thresholds at different

background luminance levels that is observed for negative/dark pedestals.

The three additional models we shall test are an extension of the model by Kane and Bertalmío (2017), which posits a link between detection thresholds and discrimination thresholds; a model introduced by Kingdom and Moulden (1991), which was designed to overcome the requirement of Whittle's model to specify different contrast functions for positive/bright and negative/dark luminance pedestals; and finally, a divisive gain model that has previously been used to model contrast discrimination data sets and the dipper effect (Foley, 1994; Watson & Solomon, 1997).

In this paper, we extract and evaluate the models on the twin data sets of Whittle (1986, 1992) reviewed in the introduction. In both cases, the data was extracted from the figures using WebPlotDigitizer (Rohatgi, 2011).

An extension of Kane and Bertalmío (2017)

An influential approach to modeling the perception of luminance is via the assumption that the integral of ΔI^{-1} can provide a meaningful estimate of the perceived luminance for intervals greater than a single threshold (Fechner, 1966). ΔI^{-1} may refer to the detection threshold or the discrimination threshold. It is well established that integrating over detection thresholds produces a compressive nonlinearity (with no inflexion) and provides a reasonable estimate of the global perception of luminance under many circumstances (Fechner, 1966; Bartleson & Breneman, 1967). The theory has had considerable influence upon the design of response nonlinearities used in industry for the storing and transmission of visual content as it forms the basis of the CIE color space (Fairchild, 2013) and the recently developed PQ curve (Miller, Nezamabadi, & Daly, 2013) for high dynamic range content. Clearly then, detection thresholds have proven useful to predict luminance perception under some circumstances, but as described above, luminance discrimination thresholds are more appropriate for the stimulus configurations investigated by Whittle (1986, 1992). As such, a model that can account for both types of phenomena would be able to unify the two sets of results, and this was the idea behind the work in Kane and Bertalmío (2017), which proposes a relationship between detection thresholds and discrimination thresholds as follows:

$$\Delta I(I_r) = \Delta I_{det}(I_r)(e + cC(I_r)), \quad (4)$$

where ΔI is the discrimination threshold, e and c are constants to be fit to the data, and C some function of the contrast between I_b and I_r.

The basic idea behind the model is that when $I_r = I_b$, discrimination thresholds reduce to detection thresholds, but when $I_r \neq I_b$, contrast gain decreases sensitivity and gives rise to the crispening effect. In the original model $C = I_p$; however, we can see in Figure 3 that, although it captures the basic effect, it does not produce a particularly accurate fit. The alternative is to consider nonlinear processing of luminance and, more specifically, that there is good reason to expect contrast to be computed differently depending on the polarity of the reference patch. To estimate the initial contrast function, we begin with a detection threshold function ΔI_{det} derived from the psychophysical data by Blackwell (1946; see details in Appendix 2), which is used to estimate the initial model contrast as the integral of one over thresholds, as follows:

$$Z(I_r) = \frac{1}{s}\sum_{I_{min}}^{I_r}\frac{1}{\Delta I_{det}(I_i)}, \quad (5)$$

where s is a normalization constant (the number of samples per troland required to ensure a result that is independent of the sampling rate). The contrast C in the modified model is then

$$C(I_r) = |Z(I_r) - Z(I_b)|. \quad (6)$$

This nonlinear computation of contrast is shown in Figure 3B. Critically, contrast is now an expansive function when the reference is negative/dark and a compressive function when the reference is positive/bright, and this contrast term allows accurate prediction of thresholds.

Finally, a second modification is made to allow the capture of the dipper function. The reason for doing so is explored fully later in the paper, but for now, we note that this model reduces to the simpler form when $c_2 = 0$ and $e_2 = 1$, and we use this model from this point forth.

$$\Delta I(I_r) = \Delta I_{det}(I_r)\left(\frac{e_1}{e_2 + c_2 C(I_r)} + c_1 C(I_r)\right). \quad (7)$$

Kingdom and Moulden (1991)

Kingdom and Moulden (1991) criticized the model of Whittle because it required a different formulation for incremental and decremental pedestals. As such, they introduced an alternative model in which thresholds could be predicted from a single equation. This was based upon a logarithmic formulation for contrast G:

$$G = ln(I/I_b) \quad (8)$$

and the observation that (like Whittle's formation of W), G could also predict the pattern of discrimination thresholds.

As shown in appendix B of Kingdom and Moulden (1991), their model predicts thresholds as follows:

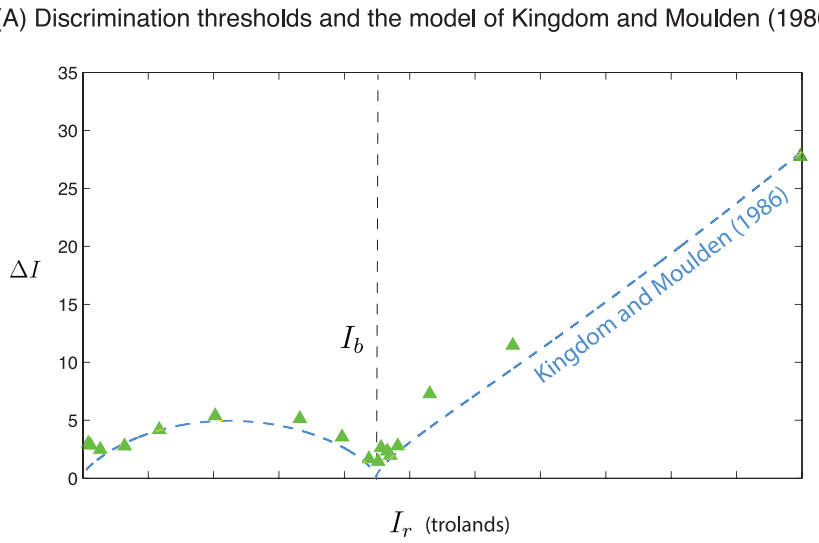

Figure 4. (A) Discrimination thresholds plotted over the range of a CRT monitor. The solid line denotes the predictions of Kingdom and Moulden (1991).

$$\Delta I = kI_r\left|log\left(\frac{I_r}{I_b}\right)\right|^n, \quad (9)$$

and this model produces the dashed blue predictions shown in Figure 4. Interestingly, when expressed in the manner of $\Delta I = f(I_r)$, the model can be thought of as a Weber's term for detection thresholds multiplied by a gain term that models the contrast between I_b and I_r.

We note that, by relaxing the Weber's law constraint to include the square root behavior of detection thresholds at low luminance levels, Kingdom and Moulden (1991) obtained the following:

$$\Delta I = k(I_r - h)\left|ln\left(\frac{I_r + h}{I_b + h}\right)\right|^n. \quad (10)$$

The additional parameter h allows the prediction of the non-Weber behavior of detection thresholds. We use this latter formulation throughout the paper, and we later note that it is critical to the modeling of data we present in the section "The background luminance."

Divisive gain

In this section, we demonstrate how one can modify a divisive gain model to account for Whittle's (1986) threshold data set. A large number of divisive gain formulations have been proposed in the literature (Carandini & Heeger, 2012), and divisive gain models have previously been used to capture the dipper effect in contrast discrimination experiments (Foley, 1994; Watson & Solomon, 1997). Probably the simplest divisive gain formulation is as follows:

$$R(x) = \frac{x^n}{s^n + x^n}. \quad (11)$$

This formulation has been used to model the V1 cell response R to a single grating, where x is the grating contrast (Albrecht & Hamilton, 1982) and s the semisaturation constant. Because the numerator and the denominator are both driven by the same variable x, this is a model of self-divisive gain, and the response R will saturate with increasing contrast even without the presence of a masking stimulus. To predict Whittle's data, we substitute x for C, which represents some measure of contrast between the background and the reference stimulus; we also use different exponents in the denominator and the numerator, and add the constant β to the denominator as follows:

$$R = \frac{C^m}{\alpha + \beta C^n}. \quad (12)$$

To derive model predictions, we make the simplifying assumption that thresholds ΔI are proportional to the inverse of the derivative of the response:

$$\Delta I = \frac{1}{R'} = \Delta I_{det}(\alpha + \beta C^n)^2 \times \frac{1}{C^{m-1}(m\alpha + m\beta C^n - \beta n C^{m+n-1})}. \quad (13)$$

As we did above for Kane and Bertalmío (2017), we evaluate the divisive gain model using two variants for C in Figure 5. If $C = I_p$, the model produces the black curves and cannot accurately model thresholds. However, if we use the nonlinear formulation of contrast described previously in Equation 6, the model produces

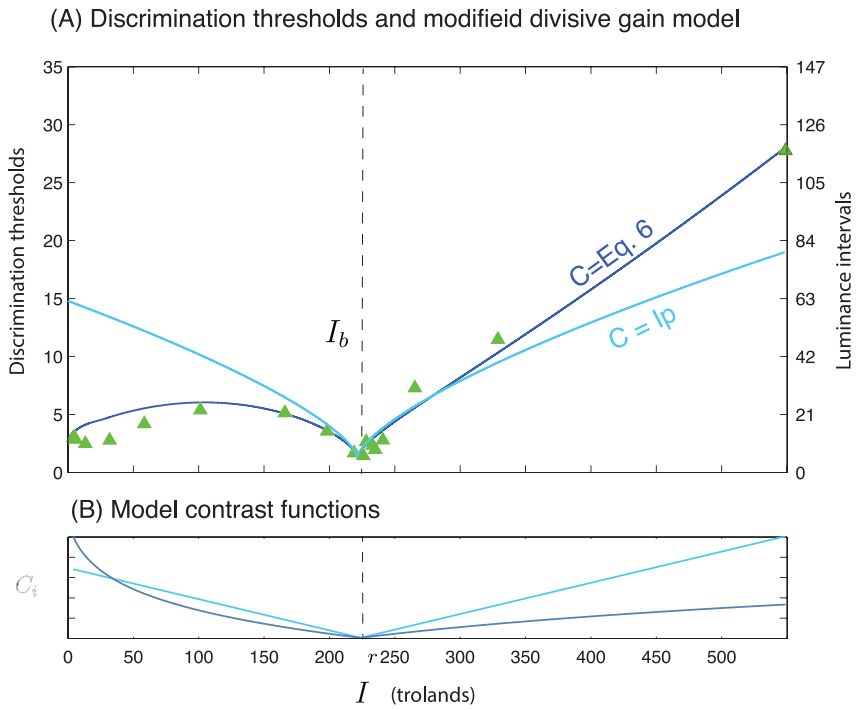

Figure 5. The prediction of a divisive gain model using two models of *C*.

accurate predictions. Thus, as with the model of Kane and Bertalmío, an accurate modeling of discrimination thresholds requires a polarity-dependent input; for negative/dark pedestals, the function must be expansive, and for positive/bright pedestals, it must be compressive.

Brightness and crispening

To derive predictions of brightness from the four models discussed thus far, one can integrate over the inverse of the model thresholds as follows:

$$B(I_r) = \sum_{I_{min}}^{I_r} \frac{1}{\Delta I(I_i)}. \quad (14)$$

In Figure 2, we plot the data from Whittle's yellow test, green background condition, in the left-hand plot, and we plot the data from the three achromatic conditions in the right-hand plot. Unlike the model of Whittle in which crispening is hard-coded, the other three models can capture both classes of function, and the best-fitting parameters are shown in Table 1. Interestingly, the three models can be considered to act in a broadly analogous manner to allow the moderation of the crispening effect (Figure 6). This is most obviously demonstrated in the model of Kane and Bertalmío (2017) in which the parameter c_1 controls the degree of contrast gain. For the achromatic condition, the best-fitting parameter is $c_1 = 0.08$, and for the equivalent luminance yellow–green condition, the parameter is much lower at $c_1 = 0.02$. Recall from Equation 7 that at

Condition	Whittle		Kane and Bertalmio				Kingdom and Moulden			Divisive gain			
Category													
I_b	*k*	*n*	e_1	e_2	c_1	c_2	*h*	*k*	*n*	*m*	*n*	α	β
Achromatic													
0.131	*0.085*	*0.010*	1.00	1.00	*0.00*	0.00	0.20	0.03	*0.004*	1.00	1.00	0.33	*0.00001*
20.40	*0.015*	*0.035*	1.00	1.00	*0.08*	0.00	0.20	0.03	*0.48*	1.00	1.00	0.33	*0.00075*
85.10	*0.010*	*0.075*	1.00	1.00	*0.14*	0.00	0.20	0.03	*0.64*	1.00	1.00	0.33	*0.00093*
Yellow–green													
19.05	*0.100*	*0.095*	1.00	1.00	*0.02*	0.00	0.20	0.03	*0.17*	*1.00*	1.00	0.33	*0.00040*

Table 1. Best-fitting parameters for fitting Whittle's brightness data set. *Notes*: All parameters were fit to the data, but parameters in regular text were fixed for all conditions, and the italic text indicates that parameters were allowed to vary with the underlying condition.

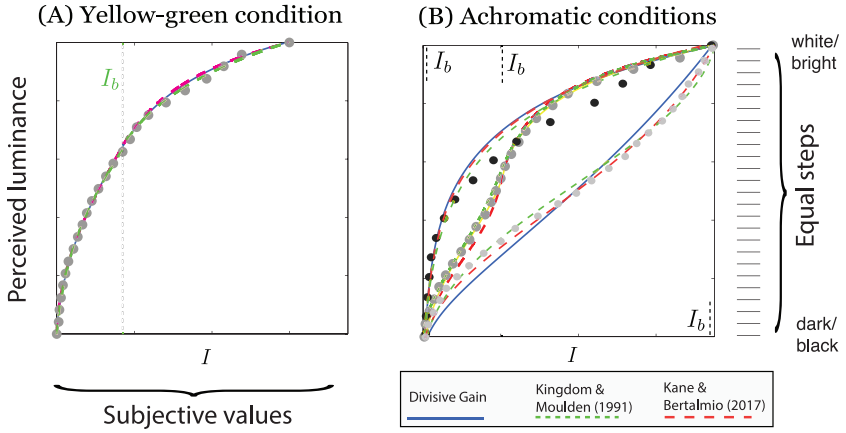

Figure 6. (A) Perceived luminance magnitude functions from the yellow–green condition and (B) the three achromatic luminance conditions for which the background luminance levels were set to either the lowest, middle, or highest luminance of the CRT monitor. All three models indicated by the legend can capture both the yellow–green function, which only exhibits weak crispening, and the achromatic stimuli for which strong crispening is observed.

$c_1 = c_2 = 0$ the discrimination thresholds reduce to the detection thresholds. An analogous pattern of behavior is noted in the model of Kingdom and Moulden (1991). If we return to Equation 9, we can see that, as the exponent n decreases, the impact of the contrast term is reduced, leaving only a Weber function of luminance. In keeping with this, the best-fitting values are $n = 0.48$ for the achromatic condition and $n = 0.17$ for the yellow–green condition.

Finally, a similar pattern of behavior is shown in the divisive gain model. To capture the brightness functions, we set $m = n = 1$. Doing so considerably reduces the complexity of the resulting threshold model, as follows:

$$\Delta I = \Delta I_{det}(\alpha + \beta C^n)^2. \quad (15)$$

In this model, the parameter β is found in the denominator; when this is set to zero, then the denominator reduces to a divisive constant. The best-fitting values are $\beta = 0.00075$ for the achromatic condition and $\beta = 0.00040$ for the yellow–green condition.

The background luminance

In this section and the following, we now evaluate the four models upon the complete discrimination threshold data set obtained by Paul Whittle. The data set was very extensive and investigated a very wide range of reference luminance levels and four background luminance levels, which spanned four orders of magnitude ($I_b = [10^{1.35}, 10^{2.35}, 10^{3.35}, 10^{4.35}]$). For each condition, Whittle probed a very wide range of I_r. We note that the lowest level of I_r tested was constrained by optical scatter, which, in effect, limits the lowest luminance level that can be seen to some fraction of the background luminance level. We take this into account in all modeling, and the rationale for doing so is described in Appendix 2.

The results of all four models are shown in Figure 7, and the best-fitting parameters are shown in Appendix 1, Table A1, and in A, C, E, and G, we plot ΔI against I_r for the four models. In B, D, F, and H, we plot $\Delta I/I_b$ against I_r/I_b. The latter plots normalize the functions such that they superimpose. The plots emphasize that the functions are more or less identical for positive pedestals but substantially heterogeneous for negative pedestals. This heterogeneity for negative pedestals cannot be captured by the (unmodified) model of Whittle. The other three models can capture the heterogeneity. In the case of our model and the divisive gain model, this is captured because the function C is dependent on the underlying detection functions, which are heterogeneous as a function of I_b. In the case of the Kingdom and Moulden (1991) model, the parameter h can play the same role if it is allowed to vary with the background condition. The best-fitting parameters for each model are shown in Table A1. Both the models of Whittle (1986) and Kingdom and Moulden report zero thresholds at zero pedestals; this feature can easily be addressed by the addition of a constant as Whittle (1992) proposes. In our model, this feature is addressed by the incorporation of the detection function.

The dipper effect

When Whittle plotted his data as a function of I_r, he excluded a number of data points near the background luminance level; however, these data points were included when he plotted his data as a function of I_p. This reveals the dipper effect, which is an

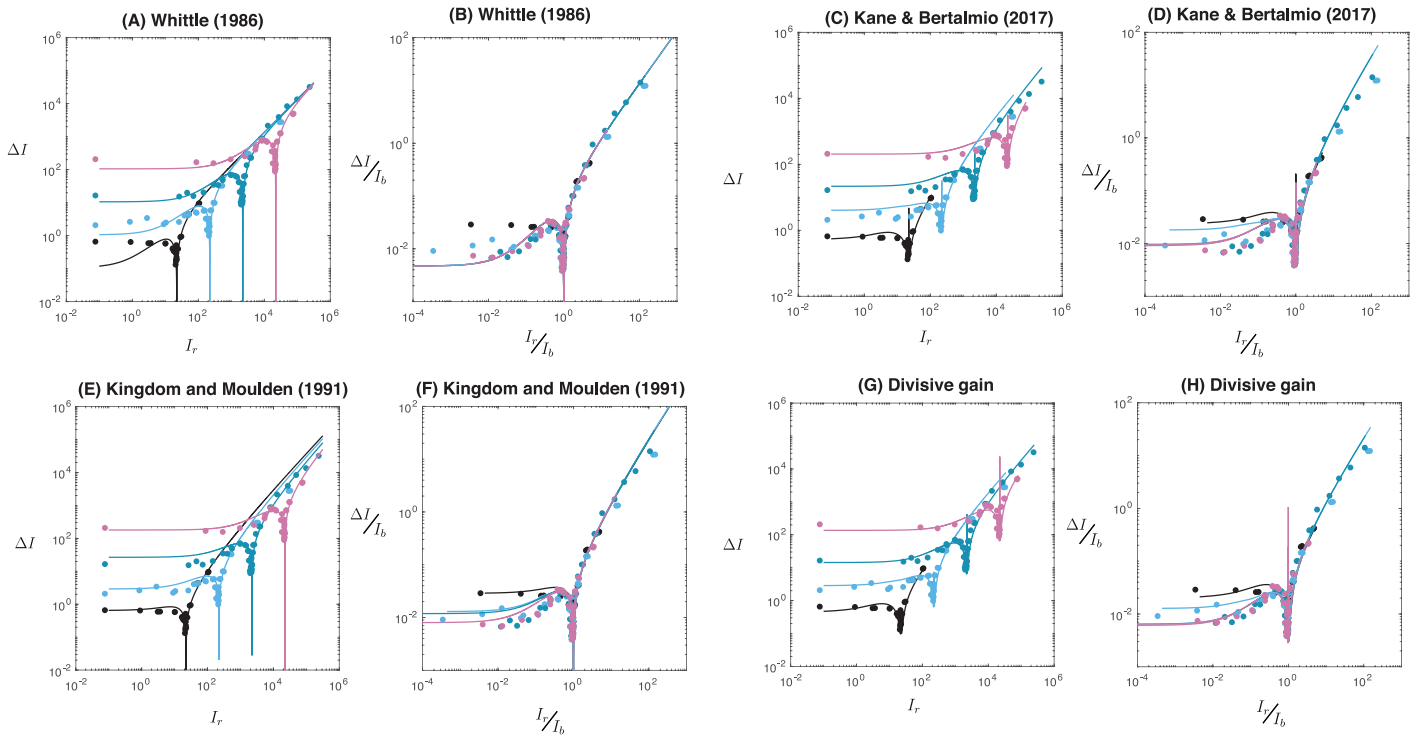

Figure 7. Subplots A, C, E, and G plot ΔI as a function of I_r, and subplots B, D, F, and H plot $\Delta I/I_b$ as a function of I_r/I_b. The functions are approximately homogeneous above the background luminance level but heterogeneous below the background luminance level. The feature is not captured by the model of Whittle but is by the other models.

increase in thresholds as the pedestal approaches zero, and this data is replotted in Figure 8. We plot the data for both positive/bright and negative/dark pedestals together.

Neither of the models Whittle (1986) or Kingdom and Moulden (1991) can predict the dipper effect in their current form, but the divisive gain model can do so as long as $m > n$. The requirement for unequal exponents means the more complex derivative must be used (Equation 13 rather than Equation 15). The finding that m must be greater than n is consistent with previous research that probed the dipper effect using contrast discrimination thresholds (Watson & Solomon, 1997).

The model of Kane and Bertalmío (2017) needed to be modified to capture the dipper effect. To do so, we added the term $1/(e_2 + c_2 C)$, which can be thought of as a gain term that decreases thresholds near the background.

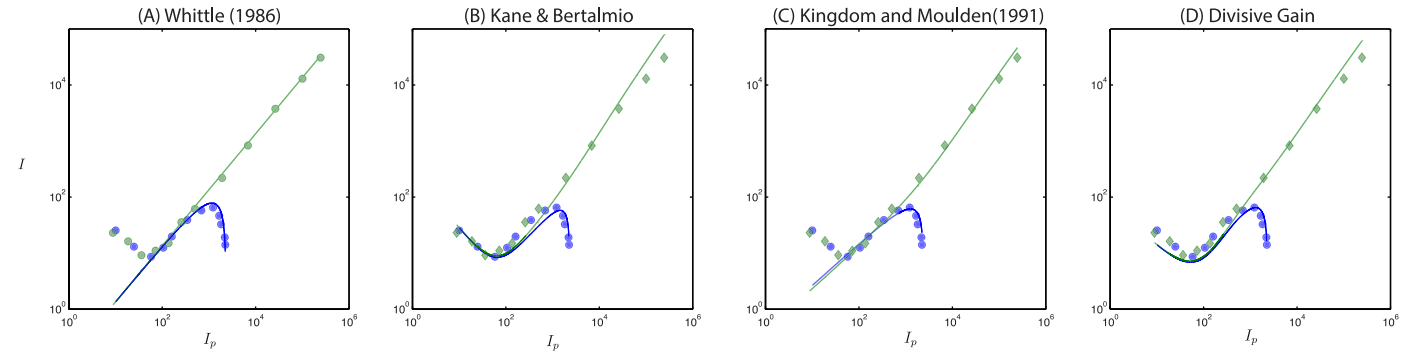

Figure 8. The dipper effect describes a counterintuitive increase in thresholds close to a pedestal of zero. Please note that the data points at low pedestals were excluded from all other plots in this paper. We plot thresholds for positive/bright pedestals using green diamonds and use blue circles for negative/dark pedestals. The models of Whittle (1986) and Kingdom and Moulden (1991) cannot capture the dipper effect, but the models of Kane and Bertalmío (2017) and the divisive gain model do provide a reasonable approximation to the data.

	Crispening	Heterogeneity (-ve)	Dipper
Original paper	Whittle (1992)	Whittle (1986)	Whittle (1986)
Figure number	2	7	8
Kane and Bertalmío (2017, modified)	soft-coded	✓	✓
Whittle (1986)	hard-coded	×	×
Kingdom and Moulden (1991)	soft-coded	✓	×
Divisive gain (modified)	soft-coded	✓	✓

Table 2. Table highlights the success of the four models tested in this paper at capturing the various features in Whittle twin data sets.

Discussion and model comparison

The performance of each of the four models tested is highlighted in Table 2. To understand how the four models operate, it is helpful to deconstruct the equations into their constituent terms. This is done in Table 3 and illustrated in Figure 9. This analysis reveals some remarkable similarities between the models. As can be seen, the first component of each model has a monotonically increasing form except for the Whittle model, which uses different equations depending on the polarity. (As this model is simply proportional to I_p for positive/bright pedestals, we illustrate this first term with a plateau at one when $I_r > I_b$.) The second component of each model can be considered a contrast

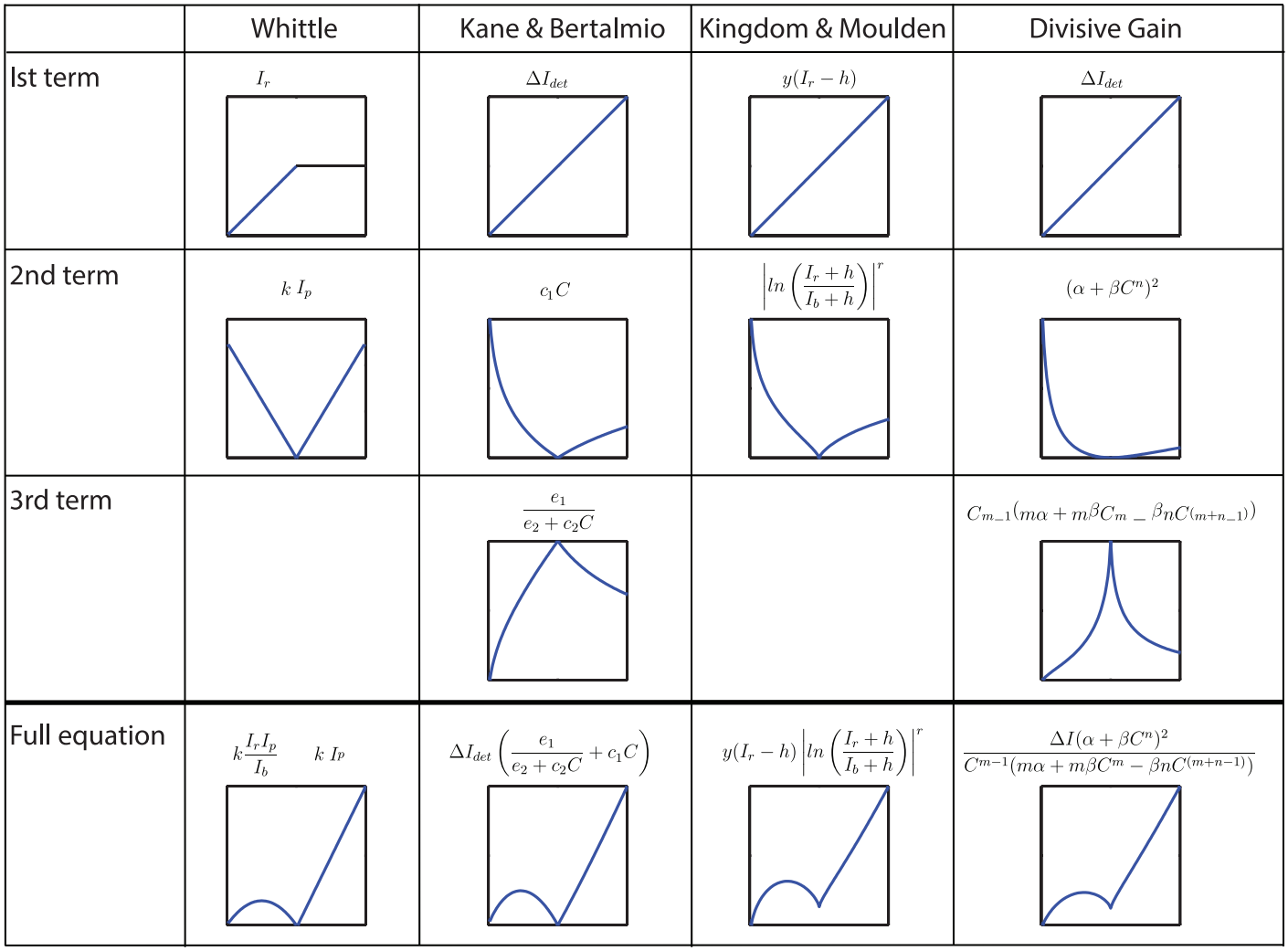

Figure 9. Each column denotes a different model. Rows 1, 2, and 3 denote the mathematical terms of each model as described in Table 3, and the bottom row shows the threshold predictions of each model. Note the remarkably similar forms that all models have. This is particularly surprising for the divisive gain model, which, unlike the other models, was not originally designed to be able to capture the threshold data of Whittle.

Model	Term A	Term B	Term C	Full equation
Kane and Bertalmío (2017, modified)	ΔI_{det}	cC	$\frac{e_1}{(e_2+cC)}$	$A \times (B + C)$
Whittle (1986)	$\frac{I_r}{I_b}$ if $I_r < I_b$ else 1	kI_r	N/A	$A \times B$
Kingdom and Moulden (1991)	$k(I_{r-h})$	$log\left(\frac{I-h}{I_b h}\right)$	N/A	$A \times B$
Divisive gain	ΔI_{det}	$(\alpha + \beta C^n)^2$	$\frac{1}{(C^{m-1}(m\alpha+m\beta C^n-\beta)}$	$A \times B \times C$

Table 3. In this table, we split the various models into their constituent terms. *Notes*: Each model has either two or three terms, and the right-hand column describes how these terms must be combined to return to the original model.

term that increases monotonically with I_p. Those models that can capture the dipper effect include a third term that monotonically decreases with I_p. It is worth recalling at this stage that this third term is only required for the divisive gain model if $m \neq n$.

Conclusion

This paper evaluates four models upon the broad range of effects noted in the twin data sets of Paul Whittle (1986, 1992) into the crispening effect. We note that no other paper has attempted to compare these models on the full data set from the two papers. This includes the paper by Kingdom and Moulden (1991), which only evaluated their model on a subset of the discrimination threshold data. As such, our finding that the model of Kingdom and Moulden can capture the luminance heterogeneity via the parameter h is novel as is the finding that this model can predict brightness functions without obvious crispening by manipulating the parameter n.

In regards to the divisive gain model, we are the first to demonstrate that it can account for Whittle's luminance discrimination data via a modification that substitutes a linear input (I_p) for a nonlinear, polarity-dependent input based on the underlying pattern of detection thresholds. The requirement for this polarity-dependent input has been obscured from many previous studies of the dipper effect because the majority of studies use a contrast pedestal rather than a luminance pedestal. When a contrast pedestal is used, the resulting contrast must be the average of the two polarities (McIlhagga & Peterson, 2006), thus the impact of polarity is hidden.

Finally, we note that the models of Kane and Bertalmío (2017) and Kingdom and Moulden (1991) were both developed with Whittle's data in mind. In the former case, the model was designed to incorporate the hypothesis that there was a relationship between detection and discrimination thresholds. In contrast, the model of Kingdom and Moulden was developed by postulating a specific contrast response function. However, the two models can be compared under the simplifying assumption that thresholds are proportional to the inverse of the derivative of the response function. Thus, taking the inverse of the derivative of the Kingdom and Moulden model produces an estimate of thresholds, and integration of the discrimination threshold model can give an estimate of the shape of the contrast response function although doing so does not produce a simple equation. The result of this analysis reveals that the two models result in a very similar prediction of thresholds and that the equations have a very similar form as shown in Table 3 and Figure 9.

Keywords: crispening, *brightness*, *perceived luminance magnitude*, *detection thresholds*, *discrimination thresholds*, *Fechner*, *Weber*, *contrast gain*, *divisive gain*, *the dipper effect*

Acknowledgments

Paul Whittle has a reputation for collecting data sets that go beyond the immediate research question to fully document the phenomena in question (Kingdom, 2010). The twin papers reported here (Whittle, 1986, 1992) are no exception to this and have provided us with a very thorough and very complete data set from which to evaluate various models. We hope the late Paul Whittle would have appreciated this work.

This work has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement number 761544 (project HDR4EU) and under grant agreement number 780470 (project SAUCE), and by the Spanish government and FEDER Fund, grant reference TIN2015-71537-P (MINECO/FEDER,UE).

Commercial relationships: none.
Corresponding author: David Kane.
Email: david.kane@upf.edu.
Address: Universitat Pompeu Fabra, Barcelona, Spain.

References

Albrecht, D. G., & Hamilton, D. B. (1982). Striate cortex of monkey and cat: Contrast response

function. *Journal of Neurophysiology*, *48*(1), 217–237.

Arend, L. E., & Goldstein, R. (1987). Simultaneous constancy, lightness, and brightness. *Journal of the Optical Society of America. A, Optics and Image Science*, *4*(12), 2281–2285.

Arend, L. E., & Spehar, B. (1993). Lightness, brightness, and brightness contrast: 2. Reflectance variation. *Perception & Psychophysics*, *54*(4), 457–468.

Bartleson, C., & Breneman, E. (1967). Brightness perception in complex fields. *Journal of the Optical Society of America*, *57*(7), 953–957.

Blackwell, H. R. (1946). Contrast thresholds of the human eye. *Journal of the Optical Society of America*, *36*(11), 624–643.

Carandini, M., & Heeger, D. J. (2012). Normalization as a canonical neural computation. *Nature Reviews Neuroscience*, *13*(1), 51–62.

Fairchild, M. D. (2013). *Color appearance models*. Hoboken, NJ: John Wiley & Sons.

Fechner, G. (1966). Elements of psychophysics. Vol. 1. New York, NY: Holt, Rinehart and Winston.

Foley, J. M. (1994). Human luminance pattern-vision mechanisms: Masking experiments require a new model. *Journal of the Optical Society of America. A, Optics and Image Science*, *11*(6), 1710–1719.

Kane, D., & Bertalmío, M. (2017). Can "crispening" be explained by contrast gain? *Electronic Imaging*, *2017*(14), 182–187.

Kingdom, F. A. A. (2010). Obituary: Paul Whittle (1938–2009). *Perception*, *39*, 577–580.

Kingdom, F., & Moulden, B. (1991). A model for contrast discrimination with incremental and decremental test patches. *Vision Research*, *31*(5), 851–858.

McCourt, M., & Blakeslee, B. (2008). Coming to terms with lightness and brightness: Effects of stimulus configuration and instructions on brightness and lightness judgments. *Journal of Vision*, *8*(6):291, https://doi.org/10.1167/8.6.291. [Abstract]

McIlhagga, W., & Peterson, R. (2006). Sinusoid = light bar + dark bar? *Vision Research*, *46*(12), 1934–1945.

Miller, S., Nezamabadi, M., & Daly, S. (2013). Perceptual signal coding for more efficient usage of bit codes. *Society of Motion Picture and Television Engineers Motion Imaging Journal*, *122*(4), 52–59.

Moon, P., & Spencer, D. E. (1944). On the Stiles-Crawford effect. *Journal of the Optical Society of America*, *34*(6), 319–329.

Obituary, Paul Whittle. (2010). *Perception*, *39*, 577–580.

Rohatgi, A. (2011). *Webplotdigitizer*. https://automeris.io/WebPlotDigitizer/

Solomon, J. A. (2009). The history of dipper functions. *Attention, Perception, & Psychophysics*, *71*(3), 435–443.

Takasaki, H. (1966). Lightness change of grays induced by change in reflectance of gray background. *Journal of the Optical Society of America*, *56*(4), 504–509.

Watson, A. B., & Solomon, J. A. (1997). Model of visual contrast gain control and pattern masking. *Journal of the Optical Society of America. A, Optics and Image Science*, *14*(9), 2379–2391.

Whittle, P. (1986). Increments and decrements: Luminance discrimination. *Vision Research*, *26*(10), 1677–1691.

Whittle, P. (1992). Brightness, discriminability and the crispening effect. *Vision Research*, *32*(8), 1493–1507.

Wyszecki, G., & Stiles, W. S. (1982). *Color science* (vol. 8). New York, NY: Wiley.

Appendix 1: Best-fitting parameters

Condition	Whittle		Kane and Bertalmio			Kingdom and Moulden			Divisive gain			
Category												
l_b	*k*	*n*	*e_1*	*e_2*	$c_1 = c_2$	*h*	*k*	*n*	*m*	*n*	α	β
Achromatic												
1.35	0.014	N/A	0.005	1.00	0.03	*5*	0.75	0.03	1.75	1.50	0.008	510
2.35	0.014	N/A	0.005	1.00	0.03	*10*	0.75	0.03	1.75	1.50	0.008	510
3.35	0.014	N/A	0.005	1.00	0.03	*80*	0.75	0.03	1.75	1.50	0.008	510
4.35	0.014	N/A	0.005	1.00	0.03	*160*	0.75	0.03	1.75	1.50	0.008	510

Table A1. Best-fitting parameters for fitting Whittle's discrimination threshold data set. *Notes*: All parameters were fit to the data, but parameters in regular text were fixed for all conditions, and the italic text indicates that parameters were allowed to vary with the underlying condition.

Appendix 2: Detection thresholds and the impact of optical scatter

Whittle (1986) investigated discrimination thresholds using a 2-mm artificial pupil. The stated illuminance levels of I_r were from 0 to 100,000 trolands without considering optical scatter. Whittle modeled optical scatter as follows:

$$I_{r(retina)} = (s-1)I_{r(on-screen)} + sI_b, \quad (16)$$

where $s = 0.035$. This equation means that the lowest illuminance levels presented to the retina depends on the background luminance level and are 0.88, 7.93, 78.53, and 783.65. Without considering the impact of optical scatter, none of the models can capture how thresholds plateau at very low reference luminance levels. This is shown in Figure A1 for Whittle's model. As such, in all the models tested in this article, we use the scatter-corrected values $I_r = I_{r(retina)}$.

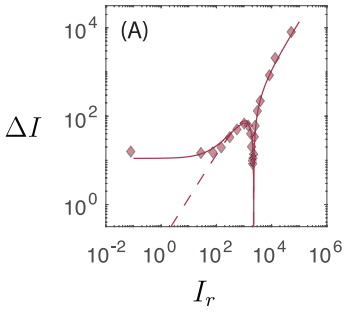

Figure A1. The impact of scatter on the discrimination thresholds of Whittle. Without considering scatter, thresholds keep decreasing (dashed line) at low levels of I_r; in contrast, once scatter has been considered, thresholds plateau (solid line).

The detection threshold data set of Blackwell (1946) was specified in terms of display luminance and was between 0 and 3,462 cd/m^{-2}. To convert to retinal illuminance (trolands), one needs a model of the pupil size. We use the model of Moon and Spencer (1944):

$$P = 4.9 - 3. \times tanh(0.4. \times log(I_b)). \quad (17)$$

Conversion to trolands is then

$$T = I \times p. \quad (18)$$

After conversion to trolands, the minimum and maximum investigated by Blackwell are 0.0017 and 9,810 trolands. To obtain estimates of thresholds for illuminance levels greater than 9,810, we assumed the Weber's law held true, and thus, a doubling of illuminance led to a doubling of thresholds. To generate an estimate of detection thresholds over the desired luminance range and a high resolution, we interpolate the Blackwell data set using MATLAB's `interp1` function using the setting `pchip`. Interpolation is needed because to model brightness we integrate over the inverse of thresholds, and this required a sufficiently high resolution to produce realistic curves.

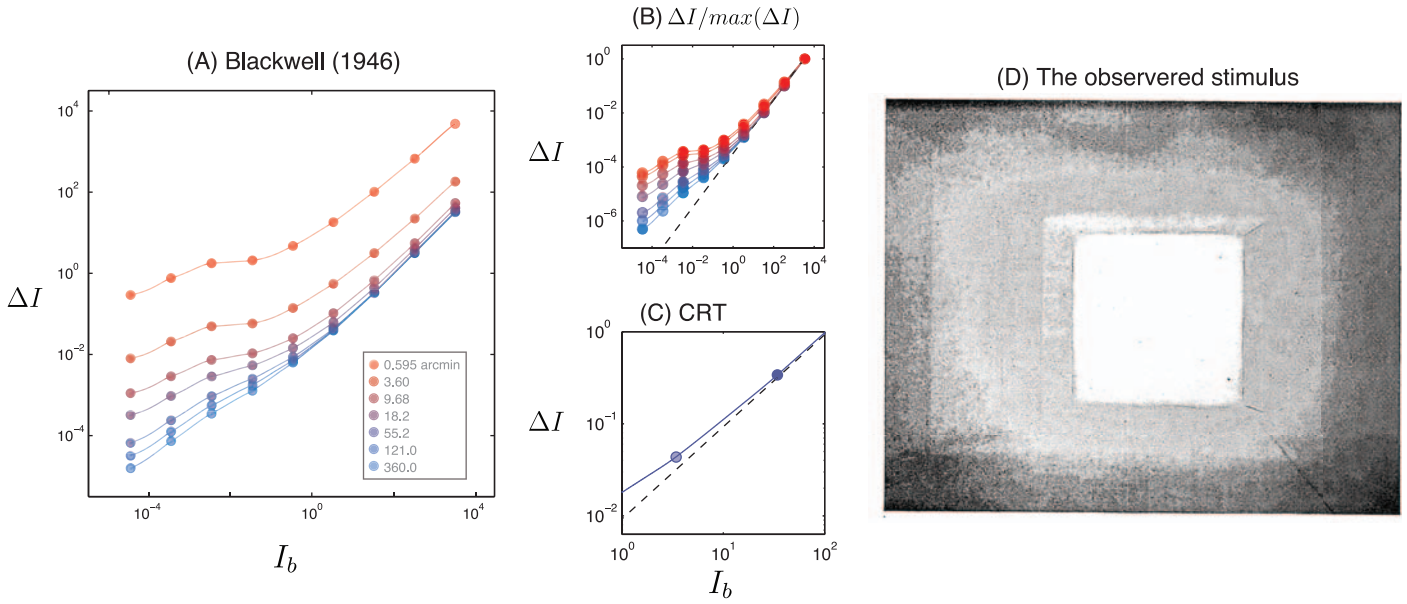

Figure A2. (A) Detection thresholds as a function of background luminance. (B) Detection thresholds normalized by the maximum thresholds. (C) Detection thresholds over the range of a CRT monitor. (D) A screen grab of the stimulus from the original paper (Blackwell, 1946).

Notes on Blackwell (1946)

Note that the Blackwell (1946) model of detection thresholds remains one of the most comprehensive studies of detection thresholds to date, collecting "more than two million responses" from nine observers. The study uses a full field stimulus as shown in Figure A2D by lining the room with troffers whose luminance could be varied from 0 to 3,462 cd/m^{-2}. The central field was designed to be maximally bright, and the brightness in the surround fell off slowly. A number of features of the data set make it appropriate for our use. First, the stimuli used were broadband as in Whittle's study. Second, various stimulus sizes were used (0.59, 3.60, 9.68, 18.20, 55.20, 121, and 360 arcmin). We use the patch size of 59 arcmin to model the data in Whittle's discrimination study, which used a stimulus size of 55.8 arcmin, and we use the 121 arcmin data to model the data from Whittle's "brightness" study, which used a stimulus size of either 126 or 150 arcmin. Third, the study used an effectively infinite response time, and Whittle used a stimulus duration of 200 ms for the discrimination study because he found that the pattern of thresholds did not change greatly for durations greater than this (see figure 6 of the original Whittle, 1986 publication), and the perceived luminance magnitude function used an infinite stimulus duration. Each observer was trained and was fully adapted to each background luminance level prior to testing. On each run, the luminance values were checked with a photometer.

The results from Blackwell (1946) are replotted in Figure A2. Detection thresholds are plotted as a function of the background luminance level and for various test stimulus sizes. At high luminance levels, the functions all follow Weber's law ($k = \Delta I/I_b$), but at low luminance levels, the Weber fraction rises as thresholds enter the square root region, and this transition occurs earlier when small test patches are used.