European Journal of Plant Pathology (2006) 114:199–213
DOI 10.1007/s10658-005-4732-9

The role of psychophysics in phytopathology: The Weber–Fechner law revisited

Forrest W. Nutter Jr and Paul D. Esker
Department of Plant Pathology, Iowa State University, 351 Bessey Hall, Ames, IA, 50011, USA (Fax: + 1-515-294-9420; E-mail: fwn@iastate.edu)

Accepted 25 October 2005

Key words: accuracy, Fechner's law, phytopathometry, precision, Stevens's power law, Weber's law

Abstract

The accuracy and precision of disease severity assessment data might be improved if there was a better understanding of how the laws of psychophysics actually relate to the theory and practice of phytopathometry. In this regard, we utilized a classical method developed in the field of psychophysics (the method of comparison stimuli) to test Horsfall and Barratt's claim that raters cannot accurately discriminate disease severity levels between 25% and 50% because, according to the Weber–Fechner law, visual acuity is proportional to the logarithm of the intensity of the stimulus. We show for two pathosystems, wheat leaf rust and grapevine downy mildew, that raters can accurately discriminate disease severity levels between 25% and 50%, and that although Weber's law appears to hold true, Fechner's law does not. Furthermore, based upon our results, the relationship between actual (true) disease severity (X) and disease severity estimated by raters (Y) is linear, not logarithmic as proposed by Horsfall and Barratt.

Introduction

Disease assessment: an exercise in stimulus – response relationships

The measurement of disease severity is often based on the visual perceptions of individual raters. Horsfall and Cowling (1978) suggested that the eye is an objective photocell for measuring disease severity, but it is well established in the fields of psychophysics and plant pathology that individuals vary greatly in their ability to discriminate among visual stimuli (Baird and Noma, 1978; Goldstein, 1989; Forbes and Jeger, 1987; Nutter and Schultz, 1995). Psychophysics is the study of the relationship between a physical stimulus (e.g. actual disease severity of a sampling unit) and the observer's perceived (estimated) response to that stimulus (e.g. estimating percentage disease severity of a sampling unit) (Goldstein, 1989; Nutter and Schultz, 1995; Nutter, 2001). Psychologists refer to objects that are not in direct contact with our major sensory systems (e.g. vision and hearing) as distal stimuli, indicating that our sensory systems are stimulated indirectly by reflected light or sound energy. The energy patterns that do reach and affect our sense organs are called proximal stimuli (Goldstein, 1989). Visible light, the band of electromagnetic energy with wavelengths between 360 and 700 nm, is the proximal stimulus for human vision. Because individuals differ in their sensitivity to different wavelengths in the visible spectrum, raters assessing disease severity will differ in their perception of visual stimuli, and thus their ability to discriminate between disease levels (Nutter and Schultz, 1995).

Although numerous disease assessment keys, scales, computer training programmes, and rating systems have been developed to improve visual estimates of disease severity (Kranz, 1988; O'Brien and van Bruggen, 1992; Gaunt, 1995; Nutter, 1997; Nutter et al., in press), the accuracy and precision of these methodologies have rarely been evaluated and compared. Moreover, relatively few studies have been conducted that provide the operational criteria needed to help researchers, select one disease assessment method over another to ensure that the assessment method that is chosen best meets the researchers, needs (O'Brien and van Bruggen, 1992; Nutter et al., 1993; Gaunt, 1995; Guan and Nutter, 2003).

Choosing a scale of measurement and types of scales

At least three factors should be considered when choosing a scale and method of assessment: (i) the type of research information needed; (ii) the statistical analyses or models that will eventually be applied to the data; and (iii) how the information will be used (Nutter and Guan, 2001). For example, assessments that will be used to quantify the relationship between disease intensity (X) and yield loss (Y) will require greater accuracy and precision than assessments used to determine the relative levels of cultivar resistance (low, moderate, high) in a plant breeding programme (Nutter and Guan, 2001; Guan and Nutter, 2004). In more advanced breeding lines, however, more accurate and precise information may be needed to quantify levels of rate-reducing resistance (Nutter and Gaunt, 1996; Nutter and Parker, 1997).

The scale of measurement can refer to the dimensions (units of measure) used to measure a variable (meters, seconds, kg ha^{-1}, %, etc.) or to the type of scale represented in a given set of units (Bordens and Abbott, 1996). Four basic types of scales listed in order of the least to the most quantitative information provided are: nominal, ordinal, interval, and ratio scales (DeVellis, 1991). Variables whose values differ by category are said to fall along a nominal scale and such values have different categorical names (fallow vs. soybean, male plants versus female plants, irrigated versus non-irrigated), and no ordering of these values is implied. The next level of measurement is that in which variables are measured along an ordinal scale. The different values of a variable in an ordinal scale not only have different names (just as in the nominal scale), but these values can also be ranked according to some descriptive measure of quantity. Ordinal scales are commonly applied by plant breeders to score plant genotypes for disease resistance (highly resistant, moderately resistant, moderately susceptible, highly susceptible). For example, a plant breeder may score genotype A as being moderately resistant to a plant pathogen (based on disease severity), whereas genotype B is scored as being highly susceptible. It cannot be said that genotype A is five times more resistant than genotype B; all that can be ascertained is that genotype A is more resistant compared to genotype B.

If the spacing between values along a scale is known, then the scale is either an interval scale or a ratio scale (DeVellis, 1991; Bordens and Abbott, 1996). With interval scales, the position of the zero point is established on the basis of convenience (e.g. using a 1–9 scale to assess disease resistance, where 1 = no disease and 9 = maximum disease). Rating scales that describe disease severity in semi-quantitative terms that are broadly and subjectively defined (e.g. 1 = no disease, 2 = few lesions on the upper leaves, ... 7 = 25–50% defoliation, ... 9 = dead plants) are useful when there is a large number of treatments and sampling units to be evaluated (e.g. cultivars, genotypes, or screening new fungicides). One such scale is the ICRISAT 1–9 scale, which is used to evaluate breeding lines of legumes for disease resistance (Subrahmanyam et al., 1982). Thus, interval scales provide an assessment method to obtain relative rankings of treatments with respect to disease severity, as well as some (limited) quantitative information. Unfortunately, such scales are sometimes inappropriately used to generate disease progress curves or to quantify temporal disease progress.

In contrast to interval scales, ratio scales have a zero point that indicates the absence of a quantity (i.e. disease incidence and/or severity in a population of sampling units could be zero for a particular disease and assessment date). Assessing disease prevalence, disease incidence, disease severity, and/or percentage defoliation are examples of ratio scales. Ratio scales with small intervals (i.e. measuring disease severity using a 0–100% scale in 1% intervals) will provide data with greater resolution than 1–9, 0–4, or 1–10 interval scales, provided they have adequate accuracy and precision. Not all scales fall precisely into interval or ratio scale types, such as some descriptive, semi-quantitative rating scales. The Horsfall–Barratt scale (Horsfall and Barratt, 1945) is an example of an interval scale, but the intervals represent unequal (log) steps in percentage disease severity. For practical purposes, an important difference between interval and ratio scales concerns the types of mathematical operations and statistical analyses that can legitimately be applied to these data. For example, dividing a 1–9 descriptive rating scale by 9 (or the highest response in the experiment) and then applying transformations to the resulting disease proportions in order to quantify temporal or spatial disease dynamics is inappropriate. Instead various non-parametric methods should be employed to analyze such data. This is why the intervals (grades) for the Horsfall–Barratt scale are converted back to percentages of disease severity, before attempting to quantitatively model temporal or spatial spread (Redman et al., 1969).

The role of psychophysics in phytopathometry

The difference threshold or 'Just Noticeable Difference' (JND)

When quantifying a difference threshold, the reason for doing so is to determine the minimum difference between two stimuli that can be detected. As plant pathologists, we would like to know the minimum difference required for a rater to perceive that two disease severity levels are actually different. Researchers in the field of psychophysics would ask: what is the 'just noticeable difference (JND)' required to perceive that a comparison stimulus is different from a standard (or reference) stimulus? For example, if the standard or 'reference' stimulus represents a disease severity level of 25%, what level of disease severity is needed in a comparison stimulus for a rater to perceive that there is a difference between the two disease severity levels? Is it a disease severity level of 30% that is required (i.e. the JND = 30% − 25% = 5%)? Is it 40% (JND = 40% − 25% = 15%)?

Horsfall and Barratt (1945) have argued that raters cannot accurately discriminate disease severity levels between 25% and 50% because visual acuity in the human eye is proportional to the logarithm of the stimulus. Horsfall and Barratt assumed this to be true by invoking what they referred to as the 'Weber–Fechner law' (Horsfall and Cowling, 1978). Actually, the Weber–Fechner law is not one, but two separate laws: Weber's and Fechner's (Goldstein, 1989). Furthermore, not all human senses (including vision) perceive stimuli in logarithms (Baird and Noma, 1978; Goldstein, 1989). According to Horsfall and Cowling (1978), the Weber–Fechner law was accidentally 'discovered' and applied by Horsfall and Barratt (1945) to the visual estimation of disease severity (e.g. a distal stimulus) without fully realizing that not all human senses respond to distal stimuli according to this law (in fact many do not) (Stevens, 1961). Moreover, Stevens (1957) showed that sound 20 JND units above a threshold is perceived as being much more than twice as loud as a sound 10 JNDs above threshold. The following sections will review and define the laws of classical psychophysics in relation to percentage disease severity as a stimulus.

Weber's law

E. H. Weber was a pioneer in the field of psychophysics and it was Weber who developed the concept of the difference threshold or JND (Weber, 1834; Goldstein, 1989). Weber published the results of experiments in which he asked observers to lift a standard weight and a comparison weight and then judge which was heavier. By having observers compare a large number of different standard and comparison weights, Weber was able to determine the difference threshold, i.e. the smallest difference between two weights that could be reliably detected. He found that the difference threshold (or JND) was dependent upon the weight of the standard (reference) stimulus. For example, an observer can just notice the difference between a 100 g standard weight and a 103 g comparison weight, so the JND in this example is 3 g. Weber found, however, that if the weight of the standard was increased to 1000 g, the JND was no longer 3 g but increased to 30 g (i.e. the comparison weight must be heavier than 1030 g to perceive a difference). Weber investigated further and found that the size of the JND for most human senses (e.g. sight, sound, taste, touch, etc.) is a constant fraction of the size of the standard stimulus (Goldstein, 1989). Expressed mathematically, this is known as Weber's law: $\mathrm{JND} = kS$, where, k is a constant called the Weber fraction and S is the value of the standard stimulus. This equation is usually expressed in the form: $k = \mathrm{JND}/S$. It should be obvious to the reader at this point how the JND concept could be applied to the assessment of disease severity. To date, the JND for assessing percentage disease severity has not been quantified, nor has Weber's law been tested and verified in plant pathology.

Fechner's law

Fechner (1860) derived a relationship between stimulus intensity and perceived magnitude (Baird and Noma, 1978; Goldstein, 1989) by making two important assumptions: (i) Weber's law holds that the JND is a constant fraction of the stimulus, and (ii) the JND is the basic unit of perceived magnitude, so that one JND is perceptually equal to another JND. Fechner accepted these assumptions and then hypothesized that the magnitude of a stimulus is determined by starting at the detection threshold and adding JNDs. Thus, a light that is 10 JNDs above the detection threshold should be perceived as being twice as bright as a light with an intensity of 5 JNDs above the detection threshold. Based upon Fechner's two assumptions, he derived the following mathematical relationship between perceived magnitude (P) and stimulus intensity (I): $P = k \log I$ where, k is a constant. Using Fechner's law, it can be determined whether doubling the intensity of a light makes it appear twice as bright. If we set $k = 1$ and $I = 10$, then $P = 1.0$ since the log of $10 = 1.0$; however, if the intensity of light is doubled to 20, $P = 1.3$. Thus, doubling the light's intensity does not double brightness (perceived magnitude). Both assumptions made by Fechner have since been questioned by those working in the field of psychophysics (Goldstein, 1989). Furthermore, Fechner's assumptions have never been tested with regards to the estimation of percentage disease severity as a distal stimulus, yet the so-called Weber–Fechner law is often invoked to legitimize logarithm-based disease assessment scales. It should be questioned whether or not percentage disease severity as a stimulus evokes the same response as brightness. Since Horsfall and Barratt (1945) published their abstract, it has been assumed for 60 years (but not yet proven) that a log increase in the stimulus (disease severity) is required to perceive an increase in the estimated magnitude of the stimulus (estimated disease severity) (Horsfall and Barratt, 1945; Horsfall and Cowling, 1978).

Stevens's power law

Twelve years after Horsfall and Barratt (1945) published their abstract on an improved grading system based upon the Weber–Fechner law, Stevens proposed that the perceived magnitude (P) equals a constant (k) times the stimulus intensity (S), raised to a power (n). Now known as Stevens's power law (1957), this law was demonstrated by plotting the logarithm of the intensity of the stimulus (X) versus the logarithm of the perceived (estimated) magnitude of the stimulus (Y). Stevens's law was found to better quantify the stimulus-response curves for a number of sensory phenomena, such as loudness, brightness, smell, taste, vibration, line length, and electric shock (Goldstein, 1989). Stevens described three types of stimulus intensity-estimated response curves (Figure 1) (Stevens, 1957, 1961). The first is called a response compression curve (curve A for brightness), because the stimulus-response curve bends downwards. Doubling light intensity (X) resulted in only a small change in perceived brightness (Y), because the exponent (n) is < 1.0. Thus, as light intensity increases, the response also increases, but not as rapidly as the intensity. A second type of stimulus-response curves includes those that bend upward, thereby exhibiting response expansion. For electric shock applied to the finger (Figure 1, curve B), the exponent (n) is 3.5 (much higher than 1.0). This indicates that a doubling of the intensity of the shock results in more than a doubling of the sensation of being shocked (pain). Finally, the third type of curve is one that is linear or approaches linearity (i.e. the power of the exponent n, is 1.0 or close to 1.0). Observers estimating the lengths of straight lines (Y) in response to being shown different line lengths (X) were found to produce linear curves (Figure 1, curve C). Stevens stated that for all senses, in general, the relationship between any stimulus intensity (X) and estimated response magnitude (Y) is best described by

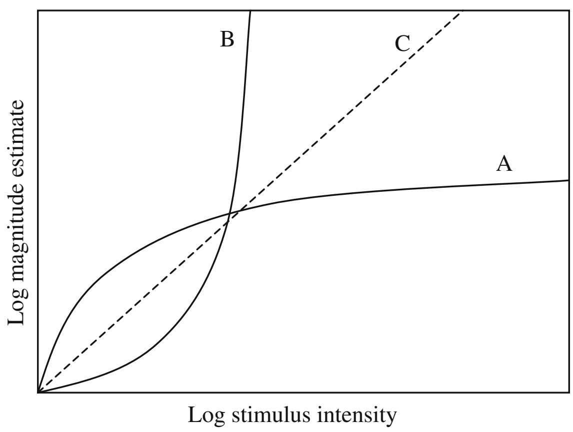

Figure 1. Curves showing the relationship between perceived magnitude (Y) and stimulus intensity (X) for curve A (brightness), curve B (electric shock), and curve C (line length), adapted from Goldstein (1989).

a power law and the exponent of the power law indicates whether doubling of the stimulus results in doubling in the perceived stimulus (i.e. linear) or whether a doubling of the stimulus causes more or less than a doubling of the response. Horsfall and Barratt (1945) assumed that doubling the intensity of the stimulus (disease severity) results in less than a doubling of the response (estimated severity), i.e. response compression, and they did not consider that response expansion or linearity were also potential stimulus-response curves that might best describe the relationship between actual disease severity (X) and perceived (estimated) disease severity (Y). To date, the applicability of Fechner's law (Log X – linear Y), Stevens's power law (log X – log Y), or a simple linear model (actual severity X – estimated severity Y) has yet to be elucidated.

The laws of psychophysics in relation to phytopathometry

The purpose of this article is to: (i) employ the method of comparison stimuli to test Weber's law by quantifying the JND threshold for percentage disease severity in the leaf rust of wheat and downy mildew of grape pathosystems, and (ii) determine the JND, Weber's fraction, and crossover points for three reference (standard) levels of disease severity (25%, 37%, and 50%) in the leaf rust of wheat and grapevine downy mildew pathosystems.

Materials and methods

Method of comparison stimuli

The method of comparison stimuli (Baird and Noma, 1978; Goldstein, 1989) was used to determine the JND for two plant pathosystems: leaf rust of wheat and downy mildew of grape. These two pathosystems were chosen in the event that lesion size or the colour of diseased areas might affect a rater's response when estimating disease severity. For leaf rust of wheat, small orange-brown pustules appear on a green (wheat leaf image) background, whereas, for grapevine downy mildew, large, medium, and small lesion sizes appear randomly as white lesions against a green (grape leaf image) background. A computer programme 'Comparison Stimuli' was written (Nutter et al., 2001), and all disease severity images for this programme were generated using the computer programme 'Severity.Pro' (Nutter and Litwiller, 1998).

Using the method of comparison stimuli, a reference stimulus was presented alongside a 'comparison stimulus' to each rater as shown in Figure 2. Each rater was then asked to determine if the comparison stimulus (on the left) was less than, greater than, or equal to the reference stimulus displayed to the right. Once each response ($<$, $>$, $=$) was recorded, a new pair of comparison and reference stimuli would be displayed on the computer screen. This process was repeated for a wide range of comparison stimuli, in random order. In this study, three reference stimuli (25%, 37%, and 50%) were selected for study (Figure 3), and comparison stimuli varied plus/minus 25% in 1% increments. Each comparison stimulus was randomly paired with one of the reference stimuli (either 25%, 37%, or 50%) on a computer screen as shown in Figure 2. Therefore, for each reference stimulus, there were 50 comparison stimuli, plus one comparison stimulus that was equal to the reference stimulus in percentage disease severity (i.e. 51 comparison images $\times$ 3 reference stimuli $=$ 153 comparison images per disease pathosystem). Ten raters participated in each pathosystem study and each study was repeated 3 times. To avoid rater fatigue, each rater was provided with two 5 min rest periods, one after the 51st comparison and the second after the 102nd comparison.

Calculating the JND, Weber's fraction, and crossover points

To facilitate calculating the JND, Weber's fraction, and crossover points for each rater, each reference (standard) stimulus, and each disease, all rater responses were converted into a more visual form as shown in Figures 4–6. A solid circle indicated a correct response for a specific comparison stimulus/reference stimulus pair, and an open circle represented an incorrect response.

The JND and Weber's fraction

The JND was operationally defined as the absolute difference between the disease severity percentage values beyond which all rater responses were

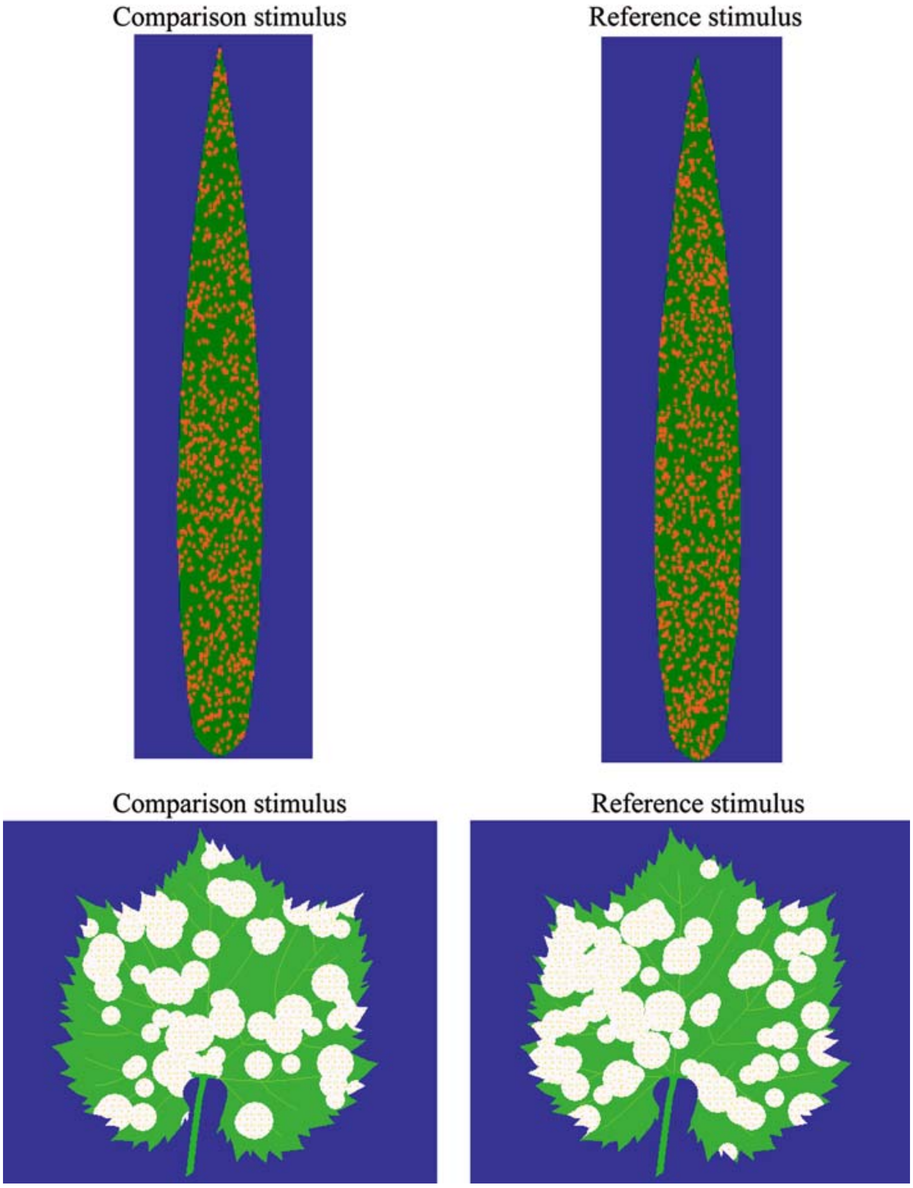

Figure 2. Example of the method of comparison stimuli for grapevine downy mildew in which raters are asked to indicate if the comparison image of disease severity on the left is greater than (>), less than (<), or equal to (=) the reference (standard) image on the right. Raters are not given the actual severities for either image; however, the disease severity of the comparison image is 44% and the reference image is 50%. Both images were generated using the computer program Severity.Pro (Nutter and Litwiller, 1998).

correct, minus the percentage disease severity of the reference stimulus. For example, in the grapevine downy mildew pathosystem, the JND for Rater 10 above the 25% reference stimulus was 32%, minus the 25% reference, for a JND of 7% (Table 3, Figure 4). Weber's fraction was calculated by dividing each JND by the reference stimulus (for this example, 7%/25% = 0.28). For the JND below the 25% reference stimulus, the JND for Rater 10 was 4% (21%−25% = 4%), and the Weber fraction was 0.16 (Table 4, Figure 4).

Crossover points

Crossover points were defined as the mid-point between the last incorrect response and the first response beyond which subsequent comparison

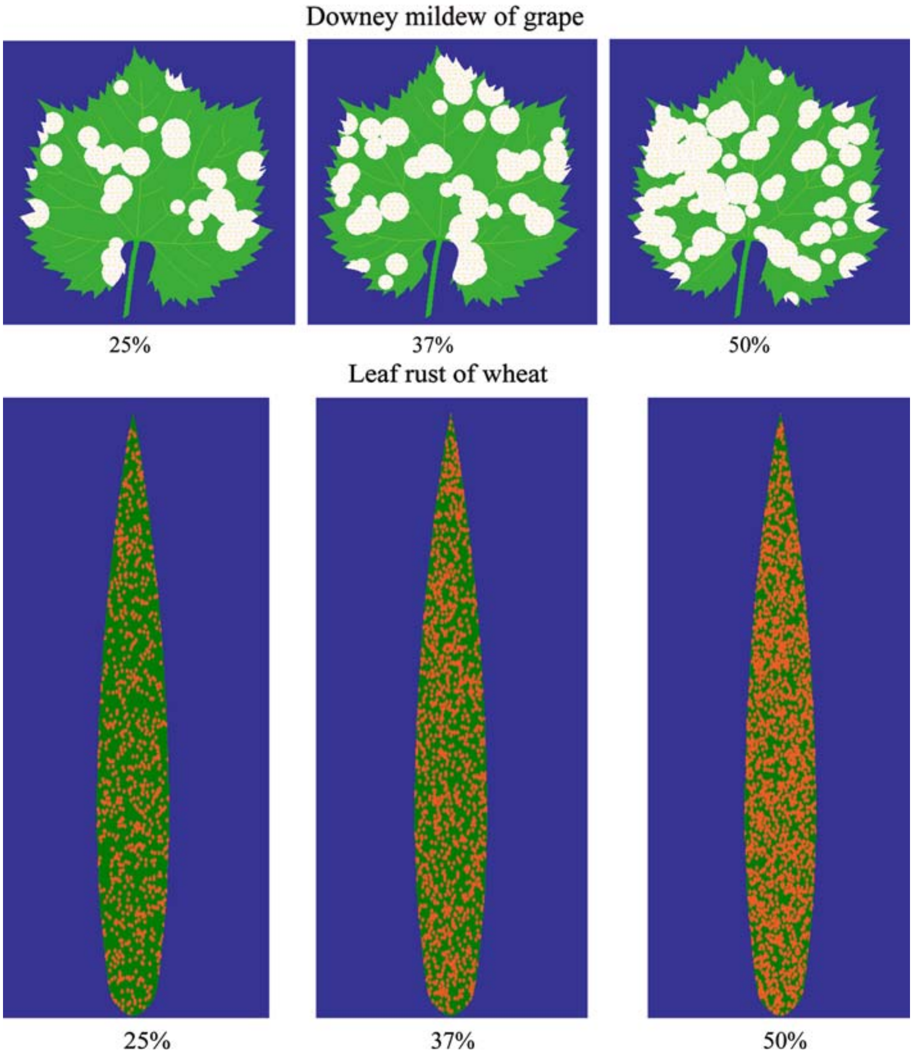

Figure 3. Examples of the three reference stimuli (25%, 37%, and 50%) used for computer-generated images of grapevine downy mildew and wheat leaf rust. Each reference stimulus was randomly paired with one comparison stimulus and comparison stimuli ranged from +/−25% of the reference stimulus in 1% increments (in random order). All images were generated using the computer program Severity.Pro (Nutter and Litwiller, 1998).

responses were correct. For example, in the grapevine downy mildew pathosystem, the crossover point above the 25% reference stimulus for Rater 6 was 30.5% (Figure 4), and the crossover point below the 25% reference stimulus for this rater was 13.5% (Table 4, Figure 4). These data were then used to determine if visual acuity was proportional to the logarithm of the intensity of the stimulus, i.e. to test Horsfall and Barratt's assumption that raters could not accurately discriminate among percentage disease severity levels in the 25–50% range.

Results

The JND for disease severity assessments of leaf rust of wheat

The JND above the 25%, 37%, and 50% reference stimuli for leaf rust of wheat ranged from 1% to 7%, 3% to 12%, and 4% to 14%, respectively (Table 1). The JND above the 25% reference stimulus was 5% or less for 8 of the 10 raters and the JND was 7% or less above the 37% reference stimulus for 8 of the 10 raters. The JND above the

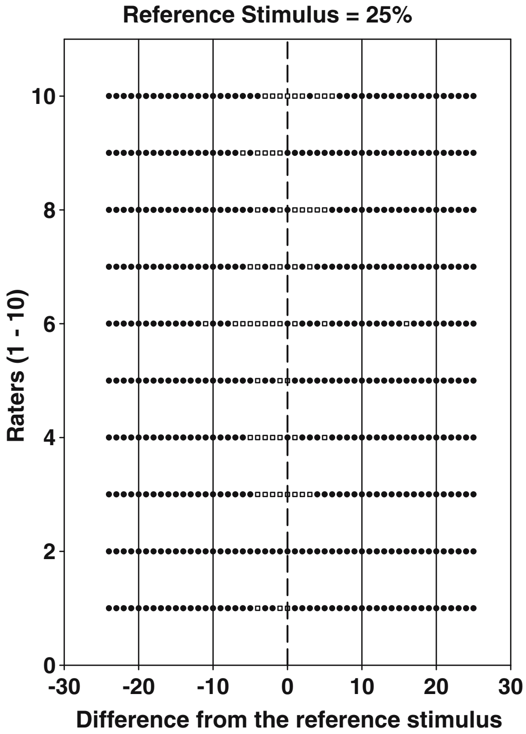

Figure 4. Graphical representation of the results for 10 raters who participated in the method of comparison stimuli study. Each of the 51 comparison stimuli were paired with a reference standard of 25% disease severity. Closed circles indicate a correct decision was made as far as the comparison stimulus being <, >, or = to reference standard stimulus, whereas open circles indicate an incorrect decision. The pathosystem was grapevine downy mildew.

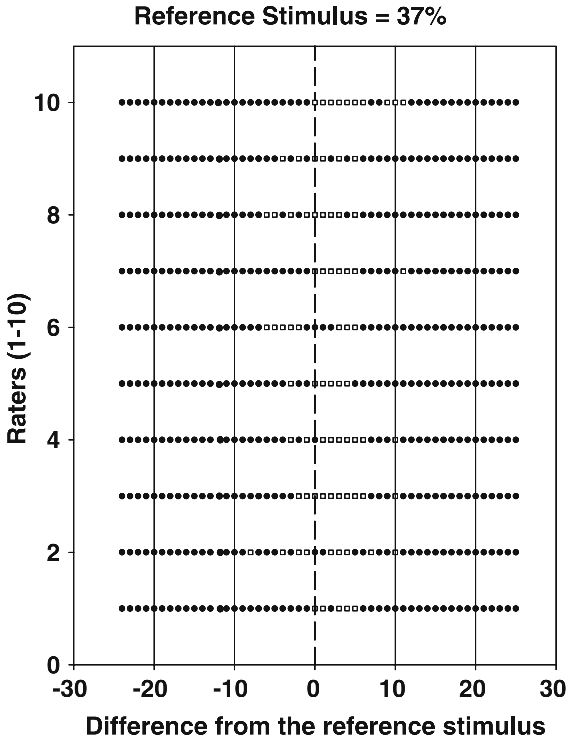

Figure 5. Graphical representation of the results for 10 raters who participated in the method of comparison stimuli study. Each of the 51 comparison stimuli were paired with a reference standard of 37% disease severity. Closed circles indicate a correct decision was made as far as the comparison stimulus being <, >, or = to reference standard stimulus, whereas open circles indicate an incorrect decision, The pathosystem was grapevine downy mildew.

50% reference stimulus was 10% or less for 8 of the 10 raters. The mean JND for the 10 raters was 4.0 ± 2.31% above the 25% reference stimulus, 6.5 ± 2.72% above the 37% reference stimulus, and 8.3 ± 2.95% above the 50% reference stimulus. Thus, the JNDs increased as the intensity of the reference stimuli increased. The JNDs below the 25%, 37% and 50% reference stimuli ranged from 1% to 9%, 1% to 11%, and 3% to 8%, respectively (Table 2). The JND below the 25% reference stimulus was 5% or less for 8 of the 10 raters and the JND below the 37% stimulus was 7% or less for 6 of the 10 raters. All 10 raters had a JND of 8% or less below the 50% reference stimulus. The mean JNDs below the 25, 37, and 50% reference stimuli for leaf rust of wheat were 4.7 ± 2.41%, 6.1 ± 3.45%, and 6.9 ± 1.66%, respectively.

Weber's fraction

Weber's law that the JND divided by the reference stimulus is a constant fraction was found to be true (Table 1). Weber's fraction for JNDs divided by their respective reference stimuli (25%, 37%, and 50%) were 0.160, 0.176, and 0.166, respectively, and were not significantly different ($P \le 0.05$) from one another (Table 1). Weber's fractions below the 25%, 37%, and 50% reference stimuli were 0.208, 0.165, and 0.138 (Table 2). Although somewhat more variable than the Weber's fractions above the thresholds reported in Table 1, Weber's fractions below the 25%, 37%, and 50% were not significantly different from one another, again indicating that Weber's law holds within the disease severity range of 25–50% for leaf rust of wheat.

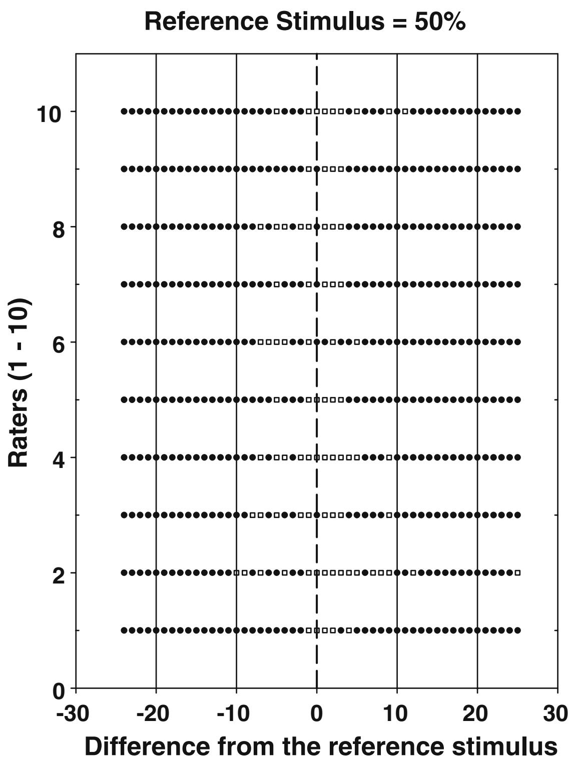

Figure 6. Graphical representation of the results for 10 raters who participated in the method of comparison stimuli study. Each of the 51 comparison stimuli were paired with a reference standard of 50% disease severity. Closed circles indicate a correct decision was made as far as the comparison stimulus being <, >, or = to reference standard stimulus, whereas open circles indicate an incorrect decision. The pathosystem was grapevine downy mildew.

Crossover points for leaf rust of wheat

Raters could detect that percentage severity levels were 'just noticeably different' below a reference stimulus of 25% with a mean crossover point of 20.3 ± 2.30% below the 25% reference stimulus (Table 2), and 28.6 ± 2.47% above the 25% reference stimulus (Table 1). Thus, severity levels below 18% and severity levels above 31% were recognized as being significantly different from the 25% reference stimulus. The mean crossover point below the standard reference stimulus of 37% was 31.3 ± 3.46% (Table 2), and the mean crossover point above the 37% reference stimulus was 43.1 ± 2.63%; thus, raters could recognize that disease severity levels less than 27.8% and above 45.7% were significantly different from the 37% reference stimulus. The mean crossover point below the 50% reference stimulus was 43.3 ± 1.93%, while above the 50% reference stimulus, it was 57.8 ± 2.95%. Therefore, disease severity levels below 41.4% and above 60.8% were recognized as being significantly different from a reference stimulus of 50%.

Most critical to this study was the question of whether or not raters could accurately discriminate between 25% and 50% disease severity. Quite clearly, the answer is yes, thereby refuting Horsfall and Barratt's contention that raters cannot discriminate among disease severity levels between 25% and 50% (Horsfall and Barratt, 1945; Horsfall and Cowling, 1978). The mean crossover point above the 25% disease severity reference stimulus was 28.6 ± 2.47% and the mean crossover point below the reference stimulus of 50% disease severity was 43.6 ± 1.93% (Figure 7). Furthermore, this strongly suggests that another 'level' could be added to the Horsfall–Barratt scale and that this level lies between 31% and 41.4%. If a grade at approximately 36% was added to the Horsfall–Barratt scale, this would make the scale linear between 25% and 50% disease severity. This indicates that the human eye does not perceive the increase in disease severity in logarithms, because much less than a doubling in percentage disease severity was required to perceive a difference in the level of disease severity. So, for small lesion (pustule) sizes that are orange–brown in colour against a green (wheat leaf) background, Weber's law appears to be valid, but Fechner's law does not, and raters could clearly discriminate among disease severity levels between 25% and 50% for the leaf rust of wheat pathosystem.

The JND, Weber's law, and crossover points in the downy mildew of grape pathosystem

The assessment of white, random-sized lesions (grapevine downy mildew) against a green (grape leaf) background supported the findings for leaf rust of wheat. Although JND values in Tables 3 and 4 for grapevine downy mildew are slightly more variable than for leaf rust of wheat, the Weber fractions above (0.164, 0.219, and 0.144) and below (0.264, 0.115, and 0.144) the 25%, 37%, and 50% reference stimuli (respectively) are not significantly different from one another ($P \leq 0.05$)

Table 1. The just noticeable difference (JND[a]) above the reference stimulus, Weber's fraction[b], and crossover points[c] as determined using the method of comparison stimuli for 10 raters assessing computer-generated images of leaf rust of wheat

Rater	JND (%) Reference stimulus			Weber's fraction Reference stimulus			Crossover points Reference stimulus		
	25	37	50	25	37	50	25	37	50
1	1	5	8	0.04	0.14	0.16	25.5	41.5	57.5
2	4	5	8	0.16	0.14	0.16	28.5	41.5	57.5
3	1	7	8	0.04	0.19	0.16	25.5	43.5	57.5
4	7	7	11	0.28	0.19	0.22	31.5	43.5	60.5
5	4	12	14	0.16	0.32	0.28	28.5	48.5	63.5
6	5	6	9	0.20	0.16	0.18	29.5	42.5	58.5
7	5	3	10	0.20	0.08	0.20	29.5	39.5	59.5
8	5	10	4	0.20	0.27	0.08	29.5	46.5	53.5
9	7	4	6	0.28	0.11	0.12	32.5	41.5	55.5
10	1	6	5	0.04	0.16	0.10	25.5	42.5	54.5
Mean	4.0	6.5	8.3	0.160	0.176	0.166	28.60	43.10	57.80
Standard deviation	2.31	2.72	2.95	0.092	0.072	0.059	2.47	2.63	2.95

[a]The JND is the number of 1% increments above a reference (standard) stimulus (25%, 37%, or 50% for this study) needed for a rater to correctly detect a difference in percent disease severity between a standard stimulus and a comparison stimulus.
[b]Weber's fraction is Weber's law JND/stimulus = constant fraction.
[c]Crossover points are defined as the mid-point between the last incorrect response and the first response beyond which all subsequent comparison responses between a comparison stimulus and a reference stimulus were correct.

(Tables 3 and 4). Again, raters could accurately discriminate between 25% and 50% disease severity, with a mean crossover point above the 25% reference stimulus of 30.0 ± 3.66% and a mean crossover point below the 50% reference stimulus of 44.2 ± 2.75% (Figure 8). As with leaf rust of wheat, another disease level that lies between 33.6% and 41.5% (~37%) could be added to the Horsfall–Barratt scale, thus making this scale linear between 25% and 50% disease severity for grapevine downy mildew. Thus, for grapevine downy mildew, Weber's law also holds for refer-

Table 2. The just noticeable difference (JND[a]) below the reference stimulus, Weber's fraction[b], and crossover points[c] as determined using the method of comparison stimuli for 10 raters assessing computer-generated images of leaf rust of wheat

Rater	JND (%) Reference stimulus			Weber's fraction Reference stimulus			Crossover points Reference stimulus		
	25	37	50	25	37	50	25	37	50
1	5	6	8	0.20	0.16	0.16	20.5	30.5	40.5
2	5	4	8	0.20	0.11	0.16	20.5	33.5	42.5
3	2	5	7	0.28	0.14	0.14	18.5	32.5	42.5
4	5	6	7	0.20	0.16	0.14	20.5	31.5	43.5
5	9	9	7	0.36	0.24	0.14	16.5	28.5	43.5
6	4	9	5	0.16	0.24	0.10	21.5	28.5	45.5
7	8	11	8	0.32	0.30	0.16	17.5	26.5	42.5
8	4	9	8	0.16	0.24	0.16	21.5	28.5	42.5
9	4	1	8	0.16	0.03	0.16	21.5	36.5	42.5
10	1	1	3	0.04	0.03	0.06	24.5	36.5	47.5
Mean	4.7	6.1	6.9	0.208	0.165	0.138	20.3	31.3	43.3
Standard deviation	2.41	3.45	1.66	0.012	0.091	0.033	2.30	3.46	1.93

[a]The JND is the number of 1% increments above a reference (standard) stimulus (25%, 37%, or 50% for this study) needed for a rater to correctly detect a difference in percent disease severity between a standard stimulus and a comparison stimulus.
[b]Weber's fraction is Weber's law JND/stimulus = constant fraction.
[c]Crossover points are defined as the mid-point between the last incorrect response and the first response beyond which all subsequent comparison responses between a comparison stimulus and a reference stimulus were correct.

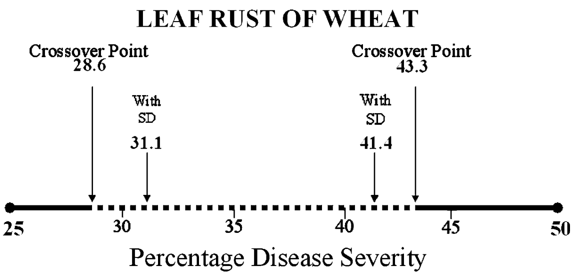

Figure 7. Crossover points for disease severity of leaf rust of wheat above 25% and below 50%, plus standard deviations, indicating that raters could accurately discriminate among disease severity levels between 25% and 50% and that another disease level at 37% could be added, thus making the Horsfall–Barratt scale linear for this range.

ence stimuli of 25%, 37%, and 50%, but Fechner's law does not hold because a doubling of the intensity of the stimulus (i.e. a log increase in disease severity) was not needed for raters to perceive a difference in the estimated response between 25% and 50% disease severity.

Discussion

For 60 years, plant pathologists have been told that the laws of psychophysics dictate that rating scales used to visually assess disease severity should be logarithmic (Horsfall and Barrett, 1945). This is because the 'accidental rediscovery' of the Weber–Fechner law by Horsfall and Barratt (1945) led to their declaration that 'visual acuity is proportional to the logarithm of the stimulus' when visually estimating disease severity (Horsfall and Cowling, 1978). No supporting data from the field of psychophysics was presented by Horsfall and Barratt (or has yet been presented) to support their conclusion that raters perceive increasing disease severity in logarithmic steps (according to the Weber–Fechner law). There is no Weber–Fechner law *per se*, but two separate laws, Weber's and Fechner's. Horsfall and Barratt did not test Fechner's two assumptions needed to derive Fechner's law (Fechner, 1860). Using the classical psychophysics method of comparison stimuli, we have shown here that Weber's law holds true for reference stimuli of 25%, 37%, and 50% disease severity, but Fechner's law does not. Our study is also the first to demonstrate that raters can discriminate another level of disease severity between 25% and 50%, thus, calling into question the validly of employing logarithmic scales while assessing disease severity. Furthermore, we have refuted the presumption that another disease severity level (class) between 25% and 50%

Table 3. The just noticeable difference (JND[a]) above the reference stimulus, Weber's fraction[b], and crossover points[c] as determined using the method of comparison stimuli for 10 raters assessing computer-generated images of downy mildew of grape

Rater	JND (%) Reference stimulus			Weber's fraction Reference stimulus			Crossover points Reference stimulus		
	25	37	50	25	37	50	25	37	50
1	1	6	5	0.04	0.16	0.10	31.5	42.5	54.5
2	8	11	13	0.20	0.30	0.26	25.5	47.5	62.5
3	4	11	10	0.16	0.30	0.20	30.5	47.5	59.5
4	6	11	10	0.24	0.30	0.20	28.5	47.5	59.5
5	1	5	4	0.04	0.14	0.08	30.5	41.5	53.5
6	6	6	6	0.24	0.16	0.12	30.5	42.5	55.5
7	4	6	4	0.16	0.16	0.08	30.5	48.5	53.5
8	6	6	4	0.24	0.16	0.08	28.5	42.5	53.5
9	1	6	4	0.04	0.16	0.08	38.5	42.5	53.5
10	7	13	12	0.28	0.35	0.24	25.5	49.5	61.5
Mean	4.4	8.1	7.2	0.164	0.219	0.144	30.0	45.2	56.7
Standard deviation	2.63	3.00	3.65	0.093	0.082	0.073	3.66	3.13	3.65

[a]The JND is the number of 1% increments above a reference (standard) stimulus (25%, 37%, or 50% for this study) needed for a rater to correctly detect a difference in percent disease severity between a standard stimulus and a comparison stimulus.
[b]Weber's fraction is Weber's law JND/stimulus = constant fraction.
[c]Crossover points are defined as the mid-point between the last incorrect response and the first response beyond which all subsequent comparison responses between a comparison stimulus and a reference stimulus were correct.

Table 4. The just noticeable difference (JND[a]) below the reference stimulus, Weber's fraction[b], and crossover points[c] as determined using the method of comparison stimuli for 10 raters assessing computer-generated images of downy mildew of grape

Rater	JND (%) Reference stimulus			Weber's fraction Reference stimulus			Crossover points Reference stimulus		
	25	37	50	25	37	50	25	37	50
1	5	1	2	0.20	0.03	0.04	20.5	36.5	48.5
2	11	9	11	0.44	0.24	0.22	14.5	28.5	39.5
3	5	3	9	0.20	0.08	0.18	20.5	34.5	44.5
4	6	4	8	0.24	0.11	0.16	19.5	33.5	42.5
5	5	4	6	0.20	0.11	0.12	20.5	33.5	44.5
6	12	7	8	0.48	0.19	0.16	13.5	30.5	42.5
7	6	1	6	0.24	0.03	0.12	19.5	36.5	44.5
8	5	7	8	0.20	0.19	0.16	20.5	30.5	42.5
9	7	5	2	0.28	0.14	0.04	18.5	32.5	48.5
10	4	1	6	0.16	0.03	0.24	21.5	36.5	44.5
Mean	6.6	4.2	6.6	0.264	0.115	0.144	18.9	33.3	44.20
Standard deviation	2.72	2.82	2.86	0.109	0.075	0.067	2.72	2.82	2.75

[a]The JND is the number of 1% increments above a reference (standard) stimulus (25%, 37%, or 50% for this study) needed for a rater to correctly detect a difference in percent disease severity between a standard stimulus and a comparison stimulus.
[b]Weber's fraction is Weber's law JND/stimulus = constant fraction.
[c]Crossover points are defined as the mid-point between the last incorrect response and the first response beyond which all subsequent comparison responses between a comparison stimulus and a reference stimulus were correct.

should not be added to the Horsfall–Barratt scale (Horsfall and Cowling, 1978).

Even before Stevens published his Power law in 1957, a number of researchers in the field of psychophysics had openly questioned the validity of Fechner's law (Baird and Noma, 1978; Goldstein, 1989). Stevens (1936, 1957) specifically questioned Fechner's second assumption of accepting the JND as the basic unit of measurement as early as 1936 (Stevens, 1936; Goldstein, 1989). Opponents' arguments centred upon the fact that many human senses simply did not obey Fechner's law and this was particularly exemplified in the case of line

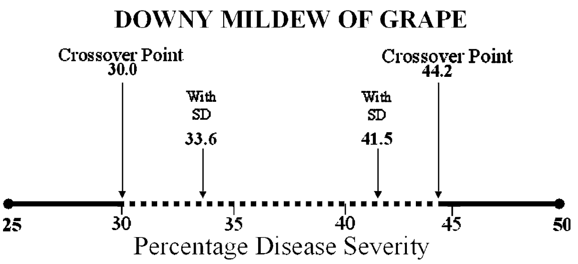

Figure 8. Crossover points for disease severity of grapevine downy mildew above 25% and below 50%, plus standard deviations, indicating that raters could accurately discriminate among disease severity levels between 25% and 50% and that another disease level at 37% could be added, thus making the Horsfall–Barratt scale linear for this range.

length estimation (Goldstein, 1989). Since area measurements are derived from line lengths (e.g. length × width, radius of a circle) one may infer that increasing the intensity of one area measurement (lesion or disease area) by another area measurement (total leaf area or the area of a sampling unit) × 100 would also result in a perceived linear relationship between actual disease severity (X) and estimated disease severity (Y).

Drawing from the practical experience of the two current authors in teaching and training in plant disease epidemiology and assessment, a large volume of student/participant data has been obtained over the past 20 years. In more than 90% of the pre-training test results obtained from hundreds of students, the relationship between the actual disease severity (X) and the estimated severity (Y) was strongly linear, with R^2 values ranging from 80% to 98% (Nutter and Parker, 1997; Nutter et al., in press). Compared to the log X–linear Y (Fechner) or log X–log Y (Stevens), models, the linear (X)–linear (Y) model provided the best fit in more than 90% of cases, based upon the F-statistics for the overall models, the coefficients of determination (R^2), and the standard errors of the estimate for Y (Nutter and Schultz, 1995). Fechner's law (Horsfall and Barrett, 1945), therefore, does not apply.

A number of other research studies have also reported that raters perceive increases in disease severity in a linear fashion. A linear relationship between actual and estimated severity has been shown in the case of dollarspot of bentgrass (Nutter et al., 1993), the Stagonospora leaf spot of orchard grass (Sherwood et al., 1983), peanut late leaf spot (Nutter and Worawitlikit, 1990), and Cercospora leaf spot of lettuce pathosystems (Gomes et al., 2004). Moreover, O'Brien and van Bruggen (1992) found that the use of the Horsfall–Barratt scale to assess the severity of Corky root of lettuce did not improve the precision of assessment in the 20–80% severity class range, as would be expected if the Weber–Fechner law was applicable to this pathosystem. Finally, in a more recent paper by Nita et al. (2003), it was reported that the Horsfall–Barratt scale was not more reliable (precise) and accurate than the direct estimation of disease severity and that the relationship between estimated and actual severity was linear for the Phomopsis leaf blight of strawberry pathosystem.

It has been previously shown for a number of pathosystems that there is a strong linear relationship between percentage disease severity (X) and percentage reflectance at 610 nm (visible green wavelength band). Thus, as a crop becomes more diseased the canopy absorbs less radiation at 610 nm wavelength, and consequently, more radiation at this wavelength is reflected upwards. In the infrared region, the opposite relationship exists in that the healthier the crop canopy (lower disease severity), the higher the percentage reflectance (Nutter and Littrell, 1996; Guan and Nutter, 2003). This reflected radiation can then be measured and recorded using a remote sensing device (such as a multispectral radiometer), as well as visually by the human eye (but for a narrower wavelength range) (Nutter et al., 1993; Nutter and Schultz, 1995). Since estimated disease severity and percentage reflectance data were both found to increase linearly as the actual severity increased, this suggests that actual severity and estimated severity also have a linear relationship, which was found to be true for a number of pathosystems (Nutter et al., 1993; Nutter and Esker, 2001; Guan and Nutter, 2003, 2004).

The implications of refuting 60 years of Horsfall and Barratt's claim that the human eye can only perceive a change in actual disease severity in log steps (3–6%, 6–12%, and 25–50%) has far reaching implications in disease assessment theory and practice. The 1945 abstract by Horsfall and Barratt is one of the most widely cited publications in the phytopathology literature, and the Horsfall–Barratt scale has been used in countless studies to estimate disease severity and as the basis to create standard area diagrams in logarithmic, as opposed to linear gradations. Are the studies that utilized the Horsfall–Barratt scale invalid? Of course not. Would estimates of disease severity between 25% and 50% have been more accurate and more precise using a linear scale? Possibly, since conversion tables are needed to obtain percentage severity estimates when using the Horsfall–Barratt graded intervals. One may find it cumbersome to examine a sampling unit (leaf), mentally estimate the percentage severity, convert this to one of the Horsfall–Barratt grades, and then use conversion tables to get back to a mean value of percentage disease severity that represents an entire grade. Kranz (1988) stated that the number of classes can influence the accuracy of disease severity estimates by raters and proposed a maximum of seven severity classes (grades), compared to the 12 classes comprising the Horsfall–Barratt scale. Furthermore, Herbert (1982) questioned whether the logarithmic function that underlies the Horsfall–Barratt scale was really valid and stated that linearity of response should be tested before designing scales and standard area diagrams.

Amanat (1977) has raised another important issue with regards to the ability (or inability) of raters to accurately assign disease severities to the twelve Horsfall–Barratt classes (grades). Amanat (1977) found that out of 203 disease assessments performed by raters, only 15.8% were assigned to their appropriate classes and that 23.2% were misclassified to a lower class and 61.1% were mistakenly assigned to a higher class (which would result in an overestimation of the actual level of disease severity). Horsfall and Cowling (1978) acknowledged that 'a stranger to the Weber–Fechner law has difficulty at first in believing that a rater reads 3–6% with the same accuracy as 25–50%' and 'when the rater discovers this, the rater is apprehensive that they cannot do better between 25% and 50%'. They further stated that 'by examining a large number of plants (sampling units), the mean amount of disease turns out to be accurate' (although accuracy was never operationally defined by the authors). In practical terms, however, increasing the number of sampling units increases both time (labour) and money (Nutter and Gaunt, 1996). For example, if estimated severity was 44%, i.e. close to the end of Grade 5 in the Horsfall–Barratt scale (25–50%), the conversion back to percentage disease severity would be approximately 38%. If estimated severity of a second assessment was 52%, i.e. at the low end of Horsfall–Barratt Grade 6 (50–75%), the conversion back to percentage severity would be 63%. Assuming the rater can correctly classify the estimated severities into their respective grades, the mean of the two assessments converted back to percentage disease severity would be approximately 50% (38 + 63 divided by 2). Since the estimated severities were 44% and 52%, respectively, the mean of these two assessments would be close to the mean that was calculated using the Horsfall–Barratt scale (48% vs. 50%), but the precision (standard deviation) of the two assessment methods would be quite different, which could affect the sample size needed to achieve a predetermined level of precision (Campbell and Madden, 1990; Nutter, 2001). It would be interesting to see if more sampling units are needed using the Horsfall–Barratt scale to obtain a predetermined level of precision (e.g. a coefficient of variation less than 10%), compared to just estimating percentage disease severity using a 0–100 scale in 1% increments, or a 0–100 scale in 5% or 10% increments. With computer-generated images that accurately depict true percentage disease severity levels, such questions can now be addressed.

Acknowledgements

We thank Chad Conover and Jie Guan for help in developing the computerized disease assessment training tools used in these studies. We also thank all participants for their willingness to participate in these studies and Dr. Rosalee Coelho Netto, National Institute for Research in the Amazon (INPA), Manaus, Brazil for help in preparing tables and figures.

References

Amanat P (1977) Modellversuche zur Ermittlung individueller und objektabhängiger Schätzfehler bei Planzenkrankheiten. Dissertation, University of Giessen, Germany.

Baird JC and Noma E (1978) Fundamentals of Scaling and Psychophysics, John Wiley and Sons, New York.

Bordens KS and Abbott RB (1996) Research Design and Methods – A Process Approach, Mayfield Publishing Company, Mountain View, CA.

Campbell CL and Madden LV (1990) Introduction to Plant Disease Epidemiology, John Wiley, New York.

DeVellis RF (1991) Scale development. Applied Social Research Methods Series, Sage Publications, Newbury Park, CA.

Fechner GT (1860) Elemente der Psychophysik (Vol. 1). Leipzig: Breitkopf and Harterl. HE Adler, DH Howes, and EG Boring (Trans.). Holt, Rinehart, and Winston.

Forbes GA and Jeger MJ (1987) Factors affecting the estimation of disease intensity in simulated plant structures. Zeitschrift fur Pflanzen Krankheiten und Pflanzenschutz 94: 113–120.

Gaunt RE (1995) New philosophies and technologies in disease measurement and appraisal. Canadian Journal of Plant Pathology 17: 185–189.

Goldstein EB (1989) Sensation and perception, 3rd edn. Wadsworth Publishing Co, Belmont, CA.

Gomes AMA, Michereff SJ and Mariano RLR (2004) Development and validation of a diagrammatic key for Cercospora leaf spot of lettuce. Summa Phytopathologica 30: 38–42.

Guan J, Nutter FW and Jr. (2003) Quantifying the intra-rater repeatability and inter-rater reliability of visual disease and remote sensing assessment methods in the alfalfa foliar disease pathosystem. Canadian Journal of Plant Pathology 25: 143–149.

Guan J, Nutter FW and Jr. (2004) Comparison of single-point alfalfa yield models based on visual disease intensity and remote sensing assessments. Canadian Journal of Plant Pathology 26: 314–324.

Herbert TT (1982) The rationale for the Horsfall–Barratt plan disease assessment scale. Phytopathology 72: 1269.

Horsfall JG and Barratt RW (1945) An improved grading system for measuring plant disease. Phytopathology 35: 655 abstr.

Horsfall JG and Cowling EB (1978) Pathometry: The measurement of plant disease. In: Horsfall JG and Cowling EB (eds) Plant Disease: An Advanced Treatise, Vol. 2 (pp 120–136) Academic Press, New York.

Kranz J (1988) Measuring plant disease. In: Kranz J and Rotem J (eds) Experimental Techniques in Plant Disease Epidemiology (pp 35–50) Springer-Verlag, New York.

Nita M, Ellis MA and Madden LV (2003) Reliability and accuracy of visual estimation of Phomopsis leaf blight of strawberry. Plant Disease 93: 995–1005.

Nutter FW and Jr. (1997) Disease severity assessment training. In: Francl L and Neher D (eds) Exercises in Plant Disease Epidemiology (pp 1–7) APS Press, St. Paul, MN.

Nutter FW and Jr. (2001) Disease assessment terms and concepts. In: Maloy OC and Murray TD (eds) Encyclopedia of Plant Pathology (pp 312–323) John Wiley and Sons, Inc, NY.

Nutter FW, Jr. and Esker PD (2001) Disease assessment keys. In: Maloy OC and Murray TD (eds) Encyclopedia of Plant Pathology (pp 323–326) John Wiley and Sons, Inc, NY.

Nutter FW, Jr., Esker PD, Guan J and Conover C (2001) A computer program to quantify the "just noticeable difference" when assessing disease severity. Phytopathology 91: S179.

Nutter Jr. FW, Esker PD and Coelho Netto RA Disease assessment concepts and the advancements made in improving the accuracy and precision of plant disease data. European Journal of Plant Pathology (in press).

Nutter FW, Jr. and Gaunt RE (1996) Recent developments in methods for assessing disease losses in forage/pasture crop. In: Chakraborty S, Leath DT, Skipp RA, Pederson GA, Bray RA, Latch GCM, Nutter FW and Jr. (eds) Pasture and Forage Crop Pathology. American Society of Agronomy (pp 93–118) Crop Science Society of America, and Soil Science Society of America, Madison, WI.

Nutter FW, Jr, Gleason ML, Jenco JH and Christians NC (1993) Assessing the accuracy, intra-rater repeatability and inter-rater reliability of disease assessment systems. Phytopathology 83: 806–812.

Nutter FW, Jr. and Guan J (2001) Disease losses. In: Maloy OC and Murray TD (eds) Encyclopedia of Plant Pathology (pp 340–351) John Wiley and Sons, Inc, NY.

Nutter FW, Jr. and Littrell RH (1996) Relationships among percent defoliation, canopy reflectance, and pod yield in the peanut-late leafspot system. Crop Protection 15: 135–142.

Nutter FW, Jr. and Litwiller D (1998) A computer program to generate standard area diagrams to aid raters in assessing disease severity. Phytopathology 88: S117.

Nutter FW, Jr. and Parker SK (1997) Fitting disease progress curves using EPIMODEL. In: Francl L and Neher D (eds) Exercises in Plant Disease Epidemiology (pp 24–28) APS Press, St. Paul, MN.

Nutter FW, Jr. and Schultz PM (1995) Improving the accuracy and precision of disease assessment: Selection of methods and use of computer-aided training programs. Canadian Journal of Plant Pathology 17: 174–184.

Nutter FW, Jr. and Worawitlikit O (1990) Disease Pro: A computerized disease assessment training program. Proceedings of the American Peanut Research and Ed. Society 22: 38.

O'Brien RD and van Bruggen AHC (1992) Accuracy, precision, and correlation to yield loss of disease severity scales for corky root of lettuce. Phytopathology 82: 91–96.

Redman CE, King EP and Brown Jr. IF (1969) Tables for Conversion of Barratt–Horsfall Rating Scores to Estimated Mean Percentages. Elanco Products Co., Indianapolis, Indiana. Mimeo, unnumbered.

Sherwood RT, Berg CC, Hoover MR and Zeiders KE (1983) Illusions in visual assessment of Stagonospora leaf spot of orchardgrass. Phytopathology 73: 173–177.

Stevens SS (1936) A scale for measurement of a psychological magnitude loudness. Psychological Review 43: 405–416.

Stevens SS (1957) On the psychophysical law. Psychological Review 64: 153–181.

Stevens SS (1961) To honor Fechner and repeal his law. Science 133: 80–88.

Subrahmanyam P, McDonald D, Gibbons RW, Nigam SN and Nevill DJ (1982) Resistance to rust and late leafspot diseases in some genotypes of *Arachis hypogaea*. Peanut Science 9: 6–10.

Weber EH (1834) De pulsu, resorptione, auditu et tactu: Annotationes, anatomical et physiological. Koehler, Leipzig.