Human psychophysical functions, an update: methods for identifying their form; estimating their parameters; and evaluating the effects of important predictors.

Diana E Kornbrot[1]

[1]Psychology Department, University of Hertfordshire, UK

Running head: Psychophysical power law

Author Note

All correspondence should be sent to:

Professor Diana E. Kornbrot

Psychology Department

University of Hertfordshire

College Lane

Hatfield, Hertfordshire, AL10 9AB

United Kingdom

Tel: +44 [0] 208 444 2081

Email: d.e.kornbrot@herts.ac.uk

Human psychophysical functions: methods for identifying their form; estimating their parameters; and evaluating the effects of important predictors.

Abstract

Stevens' power law for the judgments of sensation has a long history in psychology and is used in many psychophysical investigations of the effects of predictors such as group or condition. Stevens' formulation $\Psi = aP^n$, where $\Psi =$ is psychological judgment, P is physical intensity, and *n* is the power law exponent, is usually tested by plotting log (Ψ) against log (P). In some, but by no means all, studies, effects on the scale parameter, *a*, are also investigated. This two parameter model is simple but known to be flawed, for at least some modalities. Specifically three parameter functions that include a threshold parameter produce a better fit for many data sets. In addition, direct non-linear computation of power laws often fit better than regressions of log transformed variables. However, such potentially flawed methods continue to be used because of assumptions that the approximations are 'close enough' as to not to make any difference to the conclusions drawn (or possibly through ignorance the errors in these assumptions). We investigate two modalities in detail: duration and roughness. We show that a three-parameter power law is the best fitting of several plausible models. Comparison between this model and the prevalent two parameter version of Stevens' power law show significant differences for the parameter estimates with at least medium effect sizes for duration.

Keywords: magnitude estimation; magnitude production; Stevens' power law; duration; roughness; individual differences.

Human psychophysical functions: methods for identifying their form; estimating their parameters; and evaluating the effects of important predictors.

How loud is that music? How bright is that lamp? How hot is that sauna? These questions have been translated by science into, "How does the subjective experience of intensity or magnitude depend on the objectively measurable physical properties of a stimulus?" One answer lies in the form of the psychophysical function that relates the stated magnitude of *subjective* experience, ψ, to the *objectively* measurable magnitude, P, of an external object. This has been a key project for psychologists for more than 50 years (S. S. Stevens, 1946, 1961; S. S. Stevens & Galanter, 1957) because how the external world affects our internal sensations is a key question for psychology, and indeed philosophy.

The current paper aims firstly to identify the form of the psychophysical function using widely available computational methods for fitting and evaluating non-linear models; and secondly to determine the consequences of using a non-optimal form of the psychophysical function. Many might believe that the first aim was met some time back in 1950s and 60s. Standard texts frequently assert that the psychophysical function is a power law with just two parameters, of the form $\Psi = aP^n$ (where Ψ is the psychological sensation, P is physical magnitude, n is the power law exponent and a is a scaling factor)? Much work on magnitude estimation or production assumes this two parameter power law and uses estimates of n and a from a linear regression of log (Ψ) on log (P). This is in spite of the fact that there is incontrovertible evidence that a better fit to the psychophysical function is obtained with models with more parameters, including Stevens himself (S. S. Stevens, 1975, pp. 289-292); and other workers (G. Borg, Van Den Burg, Hassmen, Kaijser, & Tanaka, 1987; Ekman, 1959; Florentine & Epstein 2006; Marks & Stevens, 1968). A systematic literature search identified 193 items with 'magnitude estimation' in the title published since 2000, and an appeal to the psychophysics community elicited just *two* studies that tested

models with more than two parameters using current computer technology (Allan, 1983; West, Ward, & Khosla, 2000).

So why have deviations from the classic Stevens power function had so little attention? In my view, computing technology is a main reason. Taking logarithms and conducting linear regressions can easily be performed with a spreadsheet, and these procedures are available in every statistical package and many on-line statistical calculators. Meanwhile, deviations from Stevens' law are 'small' and so the assumption is that they 'do not matter', thus discouraging work that looks deeper. This paper explores this optimistic but, as I will argue, fallacious, assumption for two modalities: duration and roughness. There are two different types of consequence of wrong assumptions. Firstly, parameter estimates may be wrong. Because of the nature of power laws, quite small differences in exponents may lead to quite large differences in psychological magnitude for high values of physical intensity. Secondly, the presence and magnitude of differences between groups (e.g. old, young or healthy, diseased) or condition (e.g. high or low cognitive load) may be wrong. Such findings would have major implications for the whole field of psychophysics and the wider discipline of psychology.

This paper uses modern computational methods in the widely available statistical package SPSS to *fully* evaluate two data sets: on duration (Kornbrot, Msetfi, & Grimwood, 2013 in production) and on roughness (Kornbrot, Penn, Petrie, Furner, & Hardwick, 2007). Neither study was designed with the intention of identifying the best psychophysical function; but both provide relevant data at the individual level.

The relation of subjective sensation to objective features of the external world has a long history in modern science, as well as dating back to work we know about in Greece and Rome (and probably work Westerners don't know about in China, India, Mesopotamia, Egypt and S. America). It is an enduring question. In the middle of the 19th century Fechner

postulated that ψ depends on the number of discriminable steps. His hypothesis was supported by a substantial body of work on discriminability, but not on any 'direct' measurement of sensation. Then in the middle of the 20th century S.S. Stevens (S. S. Stevens, 1946, 1961; S. S. Stevens & Galanter, 1957) invented the procedure of magnitude estimation, which requires participants to assign a number to a stimulus that represented its ratio to a standard stimulus; and its inverse, magnitude production, which required a participant to produce [or chose] a stimulus that had the specified ratio of subjective intensity to a standard. Among others, the work of Eisler (H. Eisler, 1976; H. Eisler & Eisler, 1992) is particularly relevant for duration and of Stevens for roughness (J. C. Stevens, 1990).

As an example, in the magnitude estimation of loudness, participants might be first presented with a 1000hz tone of known objective intensity in decibels and told to 'call that standard 100'. Then they are presented with a sequence of 1000Hz tones of different physical intensities and instructed to assign numbers such that, "if it sounds twice as loud as the standard, assign 200, it sounds half as loud assign 50". Magnitude production starts with the same presentation of the standard intensity, but then the participants are required to adjust the volume controls of a generator to match specified intensities, such as 200, 50,30, etc. Using these methods, Stevens showed that the psychophysical function relating, ψ, the numbers representing subjective sensation to P the physical intensity of the sound was a power function $\psi = aP^n$. He did this by showing that a plot of log (ψ) against log (intensity, P) was a straight line. This was revolutionary in two ways. Firstly, it demonstrated that doing 'real science' with messy human experience was possible and productive. Second, it set up the proposition that the power law exponent for a given modality, e.g. sound, was a fundamental property of the human organism just as much as basal internal temperature or blood pressure. Consequently, it was an important project for psychology to determine the power law exponent for all modalities.

So there were many studies in the 60s, 70s, and 80s and beyond aiming at investigating the properties of power law exponents, importantly *assuming* that the power law was indeed the 'correct' psychophysical function, and then exploring other dimensions of the stimuli. How does the exponent depend on content: e.g. the frequency of tones or noise for loudness; whether time interval is filled with visual or auditory material on duration, etc. Another issue was the context of the stimulus, in terms of the range and ensemble of stimuli e.g. (Marks & Stevens, 1966; J. C. Stevens & Marks, 1980; Teghtsoonian, 2012). Context issues also include questions such as whether the exponent is different for ensembles with many loud sounds and ensembles with many quiet sounds. The general result of these studies is that both content and context matter, thus calling into question just how fundamental the power law exponents for each modality really are. There are also workers who challenge the power law approach at a more fundamental level, e.g. (Anderson, 1970).

At the same time, the scope of psychophysical scaling was expanded, first to include other modalities where the physical stimulus could easily be quantified, e.g. the utility of money (W. Edwards, 1954; Galanter, 1962; Kahneman & Tversky, 1979; Kornbrot, Donnelly, & Galanter, 1981; Tversky, 1967); and then to include more abstract stimuli (Galanter, 1990), seriousness of crime (Sellin & Wolfgang, 1964, 1978), pain from disease or injury (Gunnar Borg, 1998; Gunnar Borg, Lindblad, & Holmgren, 1981).

Meanwhile, other work was questioning the *form* of the psychophysical function. Equations 1 – 10 show potential models. Equations 4 and 8 were investigated across modalities (Ekman, 1959) and for thalamic cell responses (Mountcastle, Poggio, & Werner, 1963). Evidence for equation 10 for exertion scales was also found, see (Gunnar Borg, Hassmen, & Lagerstrum, 1987; G. Borg, et al., 1987), which reports earlier work not currently available (G. Borg, 1962). The need to include a *threshold* constant. Has been convincingly demonstrated to be the case for several modalities (Marks & Cain, 1972; Marks

& Stevens, 1968; J. C. Stevens & Marks, 1980). This threshold work fit the non-linear models by 'eye' and trial & error. The superiority of models including a threshold has now been demonstrated convincingly for duration (Allan, 1983) and for brightness and loudness (West, et al., 2000) using easily available modern computational methods. To our knowledge all of the pre-1983 work used specially written programs. Supplementary material provides scripts/dialogues for fitting non-linear models in SPSS.

Lorraine Allan's work was unique (as far as I know) in investigating the difference between a power and a log formulation of the psychophysical law (Allan, 1983). She pointed out that although the power metric form, $\psi = aP^n$ and the logarithmic metric form, $\log(\psi) = n \log(P) + \log(a)$ are *mathematically* equivalent for perfect data, they are not equivalent for noisy real data. Specifically, they are using different loss functions to estimate best fit.. Thus the power metric will fit higher data points better and the log metric will fit low values better. Allan showed that for duration the power metric provides a significantly better fit than the log metric. The current paper extends her important work.

The theoretical underpinnings of the psychophysical power law also attracted attention. There are mathematical arguments for the power law that have empirical support (Steingrimsson, 2011; Steingrimsson & Luce, 2005a, 2005b, 2006, 2007; Steingrimsson & Luce, 2012). Indeed, magnitude estimation is quite a difficult task for many people. Anecdotal reports from colleagues support our own experience that one almost always has to discard the data of a few people where the correlation between psychological and physical magnitude is low (< .75, say).

There are, obviously, other forms of quantitative psychological scale. For many practical reasons, the ubiquitous Likert scales and Likert items are popular (Likert, Roslow, & Murphy, 1993; Norman, 2010). They are prevalent for comparing attitudes and experience of different groups (e.g. age, ethnicity, location) for different products (e.g. soft drinks,

politicians, schools, teachers, medical treatment). Such measures are often useful for these purposes, but are not intended to produce stable parameters for any person modality combination. A multitude of available personality scales also use items and scales do aim to measure stable individual characteristics, but as propensity to act in specified ways rather than as an index of subjective experience or sensation.

Possible Psychophysical Functions

Thus there remains a key role for magnitude estimation and production in quantifying individual psychological experience. In our view this requires the following:

- Identifying the best fitting psychological function for each modality
- Determining the typical range of parameters for each modality

With these aims in mind, following functions will be evaluated:

$\Psi = a_1 P + c_1$	2 parameter linear	Allan 4	1
$\Psi = a_2 P^{n_2}$	2 parameter power	Allan 3	2
$\ln(\Psi) = n_3 \ln(P) + \ln(a_3)$	2 parameter log	Allan 6, West 5	3
$\Psi = a_4 (P - b_4)^{n_4}$	3 parameter power, physical threshold	Allan 7	4
$\ln(\Psi) = n_5 \ln(P - b_5) + \ln(a_5)$	3 parameter log, physical threshold	West 6	5
$\Psi = (P - b_6)^{n_6} + a_6$	3 parameter power, offset		6
$\ln(\Psi) = n_7 \ln((P - b_7)^{n_7} + a_7)$	3 parameter log, offset		7
$\Psi = a_8 P^{n_8} + b_8$	3 parameter, psychological threshold		8
$\ln(\Psi) = \ln(a_9 P^{n_9} + b_9)$	3 parameter log, psychological threshold		9
$\Psi = a_{10} (P - b_{10})^{n_{10}} + c_{10}$	4 parameter, physical & psychological threshold		10

The values for *n* are power law exponent parameters, the values of *a* are here termed scale parameters, and the *b* and *c* values are termed threshold parameters.

Equation 3 is the logarithmic form of Stevens' psychophysical power law, routinely used in most magnitude estimation or production studies to estimate the power law exponent, *n*. Equation 5 was shown, 'by eye', to be superior to equation 3 in several early studies, e.g. (Marks & Stevens, 1968). As noted above, for duration equation 2, the raw form of the two parameter model was superior to equation 3 the log form and equation 4, the three parameter power law, was best of all (Allan, 1983). For loudness, equation 5 was superior to equation 3. Classic studies used log to base 10, here we use log to base e "ln( )". Obviously this makes no difference to the value of ***n*.** However equations 3, 5, 7, 9 will be referred to as log or logarithmic throughout. Most classic studies have been uninterested in the values of *a* or *b*.

This study also investigates the effect of group and condition on the parameters, *a, b, n* for equations 2 to 5. Equation 1 is known not to fit for many modalities, and is in any case a special case of equation 4 with n =1. Preliminary investigations showed that there were no advantages to using any of equations 6 to 10, and they will not be further discussed.

Empirical studies

Two empirical data sets will be reanalyzed, for duration (Kornbrot, et al., 2013 in production); and for virtual roughness (Kornbrot, et al., 2007).

Data Analysis

In the first stage of analysis, individual functions for each model (equation) will be obtained using the SPSS NON-LIN procedure for each participant in each condition. This procedure produces estimates of the parameters and an estimate of adjusted r - squared, R^2_{adj}, as a measure of goodness of fit for each fitted function. Since some participants seem incapable of generating meaningful data in magnitude estimation tasks, minimum R^2_{adj} criteria for the 2 parameter log function were set for each study. For the relatively easy duration task, the criterion was $R^2_{adj} > .90$; for the harder virtual roughness task the criterion was $R^2_{adj} > .75$. These criteria are inevitably arbitrary, but setting some performance criteria

is common in magnitude estimation research, although not always reported. Since R^2_{adj}, is not normally distributed and has a ceiling effect for values near to unity, it is transformed into the Fisher *Z*-score for analysis (where $Z = \text{arctanh}(R_{adj})$. These *Z* scores were normally distributed within group and condition. Thus the output of this first stage on analysis is a set of values of parameters *Z*, *a, n, b* for each participant for each equation (model), categorized as to number of parameters (2, 3) and analysis metric (power, log) for each group and condition.

It is important to note that R^2_{adj} has been adjusted for the number of parameters, so that larger values for the 3 parameter models already take into account the loss of 1 df. Indeed for some individual participants R^2_{adj} is higher for the 2 parameter than the equivalent 3 parameter model [see supplementary information].

In the next stage, the best fitting model is identified using a mixed ANOVA analysis of *Z*, with group as a between subjects factor and condition, metric and number of parameters as repeated measures factors. Because we are interested in results at the individual level, descriptive statistics for each group/condition/model comprise: mean, SD, minimum and maximum; as well as mixed ANOVA model derived mean and 95% confidence levels. The frequency of participants' pattern of functions is also investigated, with fitted functions classified according to whether R^2_{adj} is statistically significant and whether the estimated parameters are different from those predicted for complete accuracy.

Once the best model has been identified, planned comparisons between the best fitting model and the prevalent two-parameter log model are also conducted. Separate analyses are performed for *a* and *n* as response variables. Mixed ANOVAs were conducted with model and judgment as repeated measures predictors and mood as a between group predictor. Investigations focused on whether the effects of predictors depended on (interacted with) model and whether *a* and *n* differed significantly from unity, according to model. The log and

power versions of the three *parameter* models were compared in a similar way for *a, b*, and *n*. All parameter estimates give means with associated 95% confidence limits in brackets. Inferential statistics use 95% confidence levels and report effect sizes.

Methods

The description of the method in the empirical studies is limited to a brief overview as full details are available in the relevant papers.

The Duration Study

Duration Method

Student participants were classified as to their current *mood* state using the Beck Depression Inventory (BDI: Beck, Ward , Mendelson, Mock, & Erbaugh, 1961) as: mildly depressed, high BDI, with BDI =>,7, (22 participants); or non-depressed, low BDI, BDI < 7, (24 participants). All participants made 10 subjective time *judgments* of durations ranging from 2 to 65 seconds. In the *estimation* condition they made a single magnitude estimate of 5 different experimenter presented unfilled durations; whereas in the *production* condition, they produced 5 unfilled durations specified by the experimenter, by pressing the space bar at the start and end of the specified interval. The judgment conditions were blocked, so that each participant performed in an estimation block and a production block, with presentation times randomized within block, and the order of blocks counterbalanced. Times in the estimation and production conditions were similar, but multiples of 5 were not used in estimation [see supporting material]

Duration Results

Goodness of fit for duration

There were 4 high BDI and 3 low BDI participants with $R^2_{adj,} < .90$ in the 2 parameter log analysis, and these were excluded from all further analyses, leaving 18 high and 21 low BDI participants.

A mixed ANOVA was conducted with the goodness of fit measure, Fisher's Z, as the response variable, number of parameters, K, (2,3), and metric (raw, log) as repeated measures predictors. The mixed procedure in SPSS takes into account the correlation between estimation and production judgments at the individual participant level.

Table 1 shows summary statistics for the goodness of fit measure Z with 95% confidence levels and adjusted R^2 values equivalent to mean Z. Higher Z values indicate superior fit. The columns in Table 1 for p, and η^2 come from the mixed ANOVA comparison of the values between the horizontal lines. The 3 parameter models that includes a threshold parameter, *b*, fit better than the 2 parameter models with a substantial effect size, partial eta squared, $\eta^2 = .57$ ($\eta^2 = .14$ is a 'large' effect size by convention). The power metric models fit better than the log models overall, However, post hoc analyses, following up the interaction, show that this superiority of power metric models is only present for the two parameter models, $F(1,38) = 27.8$, $p < .0005$, $\eta^2 = .42$; and not significant for the three parameter models $F(1,38) = 1.0$, $p = .333$. In summary, the best fitting model is the three-parameter power model, although it is not reliably superior to the three-parameter log model.

Figure 1 shows examples of fitted functions for the four models together with the observed data.

Parameters for the duration model

Parameter estimates and the effects of mood and judgment for the best model (3 parameter power) and most prevalent model (2 parameter log) were evaluated. Separate mixed model ANOVAs were conducted with *a* and *n,* as response variables, model (2 parameter log, 3 parameter power) and judgment (estimation, production) as repeated factor predictors and mood (low BDI, high BDI) as a between factor predictor.

For the power law exponent, *n*, the mixed ANOVA gave no significant main effects or interactions, all p –values $> .12$. However, it is noteworthy that $n = .97$ (.94, .99) for the

two parameter log model, i.e. significantly less than 1; while n = .99 (.93, 1.04) for the three parameter power model with confidence limits spanning 1. At the individual level, the two parameter log model had 28/39 estimation functions with n < 1 (2 significantly so), exact p = .002, but 19/39 production functions with n < 1 (5 significantly so). Conversely, the three parameter power law had 19/39 estimation functions with n < 1 (4 significantly so), but 29/39 production functions with n < 1 (4 significantly so), p = .001. Thus at the individual level the significant results are different for the two models. Furthermore, there is some evidence for a preponderance of people with n < 1, even with the three-parameter power model. It should be further noted that if n is reliably <1, then the predicted number of functions with n < 1 is 1. The probability of obtaining 4/39 with n < 1 is then .003.

Table 2 summarizes the effect of model, mood and judgment on the parameter a. The mixed ANOVA on a, gives a main effect of model, F(1,37) = 5.39, p =.026, η^2 = .13, a mood by judgment interaction, F(1,37) = 6.99, p = .012, η^2 =.16, and a model by mood by judgment interaction, F(1,37) = 7.30, p = .01, η^2 =.16. The main effect of model shows that a = 3.86 (3.62, 3.53) is higher for the three parameter power model than the two parameter log model, a = 3.32 (3.13, 3.53).

Separate two-way ANOVAs carried out for the different models show that the overall three-way interaction is due to there being no significant effects with the two parameter log model (maximum p = .20); but a significant mood by judgment interaction for the three parameter power model, F(1,37) 7.98, p = .008, η^2 =.18. Thus an effect that is 'large' using the better fitting three parameter power model is not significant at all, F(1,37) =.02, p = .894, when using the prevalent two parameter log model.

There is no significant effect of judgment on threshold parameter, b, F(1,38) = .23, p = .635 and the mean b = -.65 (-1.58, .29) is not significantly different from zero. There is a wide range of b values, varying from -27.1 to 5.0 and a high SD = 4.5.

Additionally, separate mixed model ANOVAs were conducted with *a* and *n,* as response variables, model (3 parameter log, 3 parameter power) and judgment (estimation, production) as repeated factor predictors and mood (low BDI, high BDI) as a between factor predictor. There was no model (equation) main effect, or any interaction that included equation for either *n* or *a*. Consequently, for duration, the same results would have been found for three parameter models, whether one used the power or the log formulation.

Correlations between parameters

For both estimation and production separately, all models show a strong negative correlation, r, between *n* and *a*, |*r*| at least = .64. Similarly, three parameter models show a strong negative correlation between *n*, and b, , |*r*| at least = -.81; while correlations between *a* and *b*, are positive, , |*r*| at least = .55. With 39 participants analyzed separately for estimation and production, all p(null) < .0005.

Summary of Duration Findings

1. Three parameter models that include a threshold parameter, *b (equations 4 and 5)*, fit substantially better than two parameter models (equations 2 and 3).
2. The three-parameter power model, equation 4, gives the best fit, but is not reliably superior to the three-parameter log model, equation 5. Which three parameter formulation is chosen makes no significant difference to the parameter estimates or the effects of predictors
3. The best fitting three-parameter power model, equation 4, shows no effect of mood or judgment on the power law exponent *n*, or the threshold parameter *b***.**
4. The best fitting three parameter power model shows a mood by judgment effect for the scale parameter *a*, such that non-depressed participants have higher mean *a* for estimation but depressed participants have a higher mean *a* for production. Using a two parameter log model instead of a three parameter power model would

significantly underestimate the scale parameter *a*, but would not significantly alter the estimate of *n*.

5. Using the two-parameter log model instead of the three-parameter power model would miss a mood by judgment interaction that is statistically 'large' ($\eta^2 = .18$).

In summary using the extremely prevalent two parameter log model leads to an underestimate of an important parameter, *a*, and missing a psychologically important interaction.

The Virtual Roughness Study

Roughness Method

The purpose of this study was to evaluate various haptic probes in virtual reality for purposes of producing haptic virtual interfaces for visually impaired users. There were 2 groups of participants labeled according to their visual status: 10 registered blind and 13 with normal or corrected to normal vision (sighted). All made magnitude estimates of the roughness of 11 virtual surfaces with two different probes, a stylus and a thimble.

Roughness Results

Goodness of fit

Equation 2, the 2 parameter log function was fit for all participants with both probes. There were 2 blind and 3 sighted participants with $R^2_{adj} < .75$ for at least 1 probe. These participants were omitted from further analyses, leaving 8 blind and 10 sighted participants.

Then, a mixed model ANOVA was conducted with Fisher's *Z* as the response variable, and the number of model parameters, K, (2, 3), and metric (power, log) as repeated measures predictors

The statistically significant results of the mixed ANOVA on *Z* are summarized in Table 3. There were main effects for the number of parameters and metric, as shown in Table3. The three-parameter models fit better than the two-parameter models; and the power

metric fit better than the log. Post hoc analyses show that the three-parameter power model is statistically significantly better than the three-parameter log model.

Figure 2 shows examples of fitted functions for the four models together with the observed data. The differences are small, but as with the duration data, the three parameter models fit the deceleration with increasing roughness, as a lower exponent is compensated by a higher threshold parameter.

Parameters for roughness models

Separate mixed ANOVAs were conducted for the three parameter power model for response variables *a, n,* and *b*, with visual status as a between factor predictor and model and judgment as repeated predictor factors. There were model effects for all three response variables. Consequently, post hoc mixed ANOVAs were conducted for *a, n,* and *b* comparing the three parameter log and the three parameter power model; and for *a, n,* comparing the prevalent two parameter log with the optimal three parameter power model. Table 4 gives descriptive statistics for *a, n,* and *b*, as a function of model.

For the power law exponent, *n*, the mixed ANOVA comparing the three parameter power with the two parameter log model had a large and significant model effect, $F(1,16) = 8.54$, $p = .010$, $\eta^2 = .35$, with no other main effects or interactions. Similarly, comparing the three parameter log with the three parameter power model gave a main effect of model, $F(1,16) = 18.39$, $p = .001$, $\eta^2 = .55$. Thus the value of $n = .63$ (.48, .78) for the optimal model is substantially less than the value of *n* for the other two models (see Table 4). In keeping with this value of *n* for the best model, well below 1, all 18 participants in the stylus condition had $n < 1$ (10 significantly so); and 15/18 participants had $n < 1$ in the probe condition (10 significantly so).

For the scale parameter, *a*, the mixed ANOVA comparing the three parameter power with the two parameter log model had no significant effects at all. Thus the value $a = 1.39$

(1.09, 1.70) for the three parameter power law is larger than for the two parameter log (see Table 4), but not significantly so, $F(1,16) = 3.19$, $p = .093$. However, comparing the three parameter log with the three parameter power model gave a strong main effect of model, $F(1,16) = 13.03$, $p = .002$, $\eta^2 = .55$.

The threshold parameter, $b = .09$ (-.05, .24) for the three parameter power model is significantly larger than that for the three parameter log model (see Table 4), $F(1,16) = 6.04$, $p = .026$, $\eta^2 = .55$. As with duration, the threshold parameter is not significantly different from zero, even though the three parameter power model fits substantially better than other models. However, at the individual level, for the stylus condition, 13/18 individual functions have $b > 0$ (8 significantly so), exact $p = .045$; while for the thimble condition, 15/18 individual functions have $b > 0$ (7 significantly so), exact $p = .004$.

Correlations between parameters

The three parameter power law and the two parameter log model show no significant correlation between *a* and *n* for either probe; while the three parameter log shows a negative correlation, $r = -.53$, $p = .017$, for the stylus only. However both three parameter models show significant negative correlations of at least -.55, $p < .02$ between *n* and *b* for both probes. There are more modest positive correlations between *a* and *b*: for the three parameter power model stylus $r = .46$, $p = .056$, for the thimble, $r = .56$, $p = .015$; while for the three parameter log the correlations are numerically larger, at least $r = .63$, $p < .005$.

Summary Roughness

1. Including a threshold parameter improves the model fit.
2. The three-parameter power model is substantially better than either the two parameter log or three parameter log model.
3. The pattern of behavior of predictors is same for all models with no effect of visual status or probe.

4. The value of parameters depends strongly on the model. The exponent, *n*, is lowest for the best fitting three parameter power; while *a* and *b* are both highest for the three parameter power model.

In summary, for this experiment on roughness, the model affected the value of parameters, but not the pattern of results.

General Discussion

This paper exhaustively investigates model fitting and parameters from two studies that used two very different modalities. It is not the first time that the importance of thresholds has been raised (Marks & Stevens, 1968; J. C. Stevens, 1974; J. C. Stevens & Marks, 1999). Nor is it the first time that it has been suggested that fitting the power law version of the psychophysical function might be superior to fitting a log version (A. L. Edwards, 1983; West, et al., 2000), although Allan is the only scientist, I know of, who has actually provided data (though just for duration) to support this suggestion. However, this systematic study has provided such data and has produced results that seriously challenge the whole domain of magnitude scaling.

Comparing the best fitting model three parameter power model with the most prevalent two parameter log model and with the three parameter log model shows that:

- The most prevalent model of magnitude estimation and production is seriously flawed for the modalities of roughness and time, as it is missing a key ingredient, a third parameter. This is the case whether the model used the log or the power metric.
- The neglect of the threshold can lead to incorrect estimates of fundamental psychophysical parameters. For duration, there was no effect on the power law exponent, but a quite large effect on the scale parameter. For roughness, the prevalent model seriously overestimates *n* and underestimates *a* and *b*. The effects are mostly statistically “very large” or “large” and always bigger than “medium”.

- The neglect of the threshold can lead to missing psychologically important patterns of effects of predictor variables. For duration, an interaction that was quite large with the best model, $\eta^2 = .18$ was completely missed by the most prevalent model.
- For roughness, the formulation of the three parameter model matters. The power and log formulations give substantially different parameter estimates. For duration, there was no significant difference between the two formulations.

These results do not imply that a third parameter is needed for every modality. That is an empirical question that needs to be answered for each modality separately. Rather they strongly suggest that whenever a plot of log (psychological magnitude) regressed on log(physical magnitude) shows systematic deviations from linearity at the individual level, models with a third parameter should be considered. This would involve comparing goodness of fit of two and three parameter models at the individual level using the methods described here. Some workers may be satisfied that the exponents derived from the two parameter models are 'near enough' for practical and even theoretical purposes. That is obviously an individual scientific judgment, but it is one that should be informed by the work reported here.

The differences due to metric are equivocal and these results do not suggest that nay major issue hangs on whether one uses a power or a log metric. Furthermore, fitting the low end of a power function may entail mathematical problems that have nothing to do with sensory systems.

The number of parameters is another matter. From the present findings, one can only conclude that the findings of *any* studies, of the effects of group differences, or of experimental manipulations on magnitude estimation in any modality, might be flawed and misleading if the prevalent two parameter log form of the psychophysical function is used. Goodness of fit is, of course, not the only criterion for choosing a model, (S. S. Stevens,

1975). In particular, coherence of parameters, particularly the power law exponent, across modalities as measured by cross modality matching is an important criterion. So is the relation of power law exponents to other modality parameters, including range of sensitivity (dynamic range [DR]) and resolving power (the capacity to resolve small changes in stimulus intensity), as argued by Teghtsoonian (2012). Indeed power law exponents derived from the three parameter model may provide superior coherence for cross modality matching and for Teghtsoonian's models.

The interpretation of any third parameter also demands theoretical attention. The term 'threshold' has been used conforming to earlier work e.g. (Marks & Stevens, 1968). However, the common occurrence of negative values for *b* calls this interpretation into question. An approach that starts with equation 10, following Borg (1987), and then tests whether b_{10} and c_{10} are zero may be a fruitful topic for further research. However, fitting a four parameter model requires a considerable range and number of physical values (11 was insufficient to get a sensible fit for roughness, in the study reported here).

Do these results imply then that much of psychophysics needs rewriting or re-analyzing? Recall that there were nearly 200 articles published since 2000 with "magnitude estimation" in the title. The roughness study, already published, and reanalyzed here is a case in point. Due consideration suggests that we should not be *too* worried about the pattern of those particular results, as the best model would still have no effect of probe or visual status. However, comparisons with exponents reported in other studies in the literature e.g. using real not virtual surfaces, or sandpaper as opposed engineered grooves, become hard to interpret. Moreover, there are no gold standard studies of roughness, which include optimal analysis, out there to our knowledge. The data for duration do not suggest much or any change in power law exponent, but the effects on *a* for those who are interested in *a* are very substantial. Thus literature summaries and parameter estimations across many modalities

need to be revisited. One cannot be confident either about the parameter estimates, *or* about the effects of predictors declared significant, or about the effects of predictors declared non-significant.

In addition, individual differences are important. In the analyses reported here, summaries included not only means and confidence levels but also the *range* of values obtained and the proportion of people who conform to the mean results. Here this was reported in terms of whether individual exponents and offsets are lower than 1. It is also possible to analyze what proportion of people show a significant effect (e.g. for duration is there a judgment effect, for roughness is there a probe effect). These are important questions, but require larger samples for a meaningful level of power.

It is also the case that many studies only report power law exponents and ignore the scale parameter. This is regrettable as *a* and *n* index different aspects of psychological experience. For example, our analyses show a dissociation between the effect of predictors on *a* and *n* that merits further investigation. Speculatively, one might identify *n*, with sensory effects see (Teghtsoonian, 2012) and *a* with bias,. Possible interpretations for *a* are discussed by G. Borg and Marks (1983). Clearly, single parameter scales, such as Likert scales are inherently unable to separate sensation and bias.

In terms of the future of psychophysical methodology and analysis, currently, there is no single accepted source that summarizes psychophysical functions across modalities,. Stevens and Galanter (1957) is often cited e.g. (Lindsay & Norman, 1977; "Stevens' Power Law,"), and there is a later, but far from recent, summary (Teghtsoonian, 1971). However, there is no summary that includes all quality studies for each modality and whether the exponents are averages of individual exponents or group functions from average judgment over participants for each physical value. This is in spite of the extensive valuable work in many separate modalities cited in the introduction. Consequently, a handbook giving up to

date values for power law exponents and other parameters is sorely needed. Following this study, parameters should be based on the best model for each modality, easily found with modern analytical methods and tools. We certainly have the technology. We surely have the data (scattered across the planet). So, in my view, performing the necessary time consuming work to collect all the currently available data would enormously enrich the evidence base for sophisticated mathematical models e.g. (Steingrimsson & Luce, 2012), which also summarizes current theoretical issues. A thorough analysis of the empirical evidence for physical modalities should also provide a base for more abstract modalities. These might include as utility of money and feelings, e.g. (McGraw, Larsen, Kahneman, & Schkade, 2010) for recent work and review; seriousness of crime (Sellin & Wolfgang, 1964) where the potential has never been exploited; annoyance e.g. (Fucci, Petrosino, Hallowell, Andra, & Wilcox, 1997),all of which have important social consequences. The psychophysical project of relating psychological sensation to physical magnitude thus remains a key goal for the psychology of perception. The analyses reported here are a clarion call to reinvigorate the project and generate results that will truly stand the test of time.

References

Allan, L. G. (1983). Magnitude estimation of temporal intervals. *Perception, and Psychophysics, 33*(1), 29-42. http://dx.doi.org/10.3758/BF03205863

Anderson, N. H. (1970). Functional measurement and psychophysical judgment. *Psychological Review, 77*(3), 153.

Beck, A., Ward , C., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry, 4*(6), 561-571. http://dx.doi.org/10.1016/S0010-440X(61)80020-5

Borg, G. (1962). *Physical performance and perceived exertion.* (Vol. XI). Lund, Sweden: Gleerup.

Borg, G. (1998). *Borg's perceived exertion and pain scales*: Human Kinetics Publishers.

Borg, G., Hassmen, P., & Lagerstrum, M. (1987). Perceived exertion related to heart rate and blood lactate during arm and leg exercise. *European journal of applied physiology and occupational physiology, 56*(6), 679-685. http://link.springer.com/article/10.1007%2FBF00424810

Borg, G., Lindblad, I., & Holmgren, A. (1981). Quantitative evaluation of chest pain. *Acta Medica Scandinavica, 209*(S644), 43-45. http://dx.doi.org/10.1111/j.0954-6820.1981.tb03117.x

Borg, G., Van Den Burg, M., Hassmen, P., Kaijser, L., & Tanaka, S. (1987). Relationships between perceived exertion hr and hla in cycling running and walking. [Article]. *Scandinavian Journal of Sports Sciences, 9*(3), 69-78. <Go to ISI>://BCI:BCI198886009184

Borg, G. V., & Marks, L. (1983). Twelve meanings of the measure constant in psychophysical power functions. *Bulletin of the Psychonomic Society, 21*(1), 73-75. http://dx.doi.org/10.3758/BF03329958

Edwards, A. L. (1983). *Techniques of attitude scale construction*. New York: Irvington.

Edwards, W. (1954). The theory of decision making. *Psychological Bulletin, 51*(4), 380-417.

Eisler, H. (1976). Experiments on subjective duration 1868-1975: A collection of power function exponents. *Psychological Bulletin, 83*(6), 1154-1171. http://dx.doi.org/10.1037/0033-2909.83.6.1154

Eisler, H., & Eisler, A. D. (1992). Time perception: effects of sex and sound intensity on scales of subjective duration. *Scandinavian Journal of Psychology, 33*(4), 339-358. http://dx.doi.org/10.1111/j.1467-9450.1992.tb00923.x

Ekman, G. ñ. (1959). Weber's law and related functions. *The Journal of Psychology, 47*(2), 343-352.

Florentine, M., & Epstein , M. (2006, 25-28 July 2006). *To honor stevens and repeal his law (for the auditory system.* Paper presented at the Fechner Day 2006. Proceedings of the 22nd Annual Meeting of the International Society for Psychophysics, St. Albans, UK. http://www.ispsychophysics.org/fd/index.php/proceedings/article/view/332/324

Fucci, D., Petrosino, L., Hallowell, B., Andra, L., & Wilcox, C. (1997). Magnitude estimation scaling of annoyance in response to rock music: Effects of sex and listeners' preference. *Perceptual and Motor Skills, 84*(2), 663-670. <Go to ISI>://WOS:A1997WR76600045

Galanter, E. (1962). The Direct Measurement of Utility and Subjective Probability. *The American Journal of Psychology, 75*(2), 208-220. http://links.jstor.org/sici?sici=0002-9556%28196206%2975%3A2%3C208%3ATDMOUA%3E2.0.CO%3B2-M

Galanter, E. (1990). Utility Functions for Nonmonetary Events. *The American Journal of Psychology, 103*(4), 449-470. http://www.jstor.org/stable/1423318

Kahneman, D., & Tversky, A. (1979). Prospect theory. *Econometrica, 47*, 263. http://dx.doi.org/10.2307/1914185

Kornbrot, D. E., Donnelly, M., & Galanter, E. (1981). Estimates of Utility Function Parameters from Signal-Detection Experiments. [Article]. *Journal of Experimental Psychology-Human Perception and Performance, 7*(2), 441-458. http://dx.doi.org/10.1037/0096-1523.7.2.441

Kornbrot, D. E., Msetfi, R. M., & Grimwood, M. (2013 in production). Time perception and depressive realism: judgment type, psychophysical functions and bias. *PLOS one, 2013 in press*. http://dx.doi.org/10.1371/journal.pone.0071585

Kornbrot, D. E., Penn, P., Petrie, H., Furner, S., & Hardwick, A. (2007). Roughness perception in haptic virtual reality for sighted and blind people. [Article]. *Perception & Psychophysics, 69*, 502-512. http://dx.doi.org/10.3758/BF03193907 | http://hdl.handle.net/2299/2959

Likert, R., Roslow, S., & Murphy, G. (1993). A simple and reliable method of scoring the Thurstone attitude scales. *Personnel Psychology, 46*(3), 689-690. http://dx.doi.org/10.1111/j.1744-6570.1993.tb00893.x

Lindsay, D. R. J., & Norman, D. A. (1977). *Human information processing* (3 ed.). London: Academic.

Marks, L. E., & Cain, W. S. (1972). Perception of intervals and magnitudes for three prothetic continua. *Journal of Experimental Psychology, 94*(1), 6-17. http://dx.doi.org/10.1037/h0032746

Marks, L. E., & Stevens, J. C. (1966). Individual brightness functions. *Perception & Psychophysics, 1*(1), 17-24. http://dx.doi.org/10.3758/BF03207815

Marks, L. E., & Stevens, J. C. (1968). The form of the psychophysical function near threshold. *Perception, & Psychophysics, 4*(5), 315-318. http://dx.doi.org/10.3758/BF03210523

McGraw, A. P., Larsen, J. T., Kahneman, D., & Schkade, D. (2010). Comparing Gains and Losses. *Psychological Science, 21*(10), 1438-1445. http://dx.doi.org/10.1177/0956797610381504

Mountcastle, V. B., Poggio, G. F., & Werner, G. (1963). The relation of thalamic cell response to peripheral stimuli varied over an intensive continuum. *Journal of Neurophysiology, 26*, 807-834. <Go to ISI>://MEDLINE:14065329

Norman, G. (2010). Likert scales, levels of measurement and the "laws" of statistics. *Advances in health sciences education, 15*(5), 625-632. http://dx.doi.org/10.1007/s10459-010-9222-y

Sellin, T., & Wolfgang, M. E. (1964). *The measurement of delinquency*.

Sellin, T., & Wolfgang, M. E. (1978). *The measurement of delinquency* (2 ed.). New York: John Wiley.

Steingrimsson, R. (2011). Evaluating a model of global psychophysical judgments for brightness: II. Behavioral properties linking summations and productions. *Attention Perception & Psychophysics, 73*(3), 872-885. http://dx.doi.org/10.3758/s13414-010-0067-5

Steingrimsson, R., & Luce, R. D. (2005a). Evaluating a model of global psychophysical judgments. I: Behavioral properties of summations and productions. *Journal of Mathematical Psychology, 49*(4), 290-307. http://dx.doi.org/10.3758/APP.71.8.1916

Steingrimsson, R., & Luce, R. D. (2005b). Evaluating a model of global psychophysical judgments. II: Behavioral properties linking summations and productions. *Journal of Mathematical Psychology, 49*(4), 308-319. http://dx.doi.org/10.1016/j.jmp.2005.03.001

Steingrimsson, R., & Luce, R. D. (2006). Empirical evaluation of a model of global psychophysical judgments: III. A form for the psychophysical function and intensity filtering. *Journal of Mathematical Psychology, 50*(1), 15-29. http://dx.doi,org/10.1016/j.jmp.2005.11.005

Steingrimsson, R., & Luce, R. D. (2007). Empirical evaluation of a model of global psychophysical judgments: IV. Forms for the weighting function. *Journal of Mathematical Psychology, 51*(1), 29-44. http://dx.doi.org/10.1016/j.jmp.2006.08.001

Steingrimsson, R., & Luce, R. D. (2012). Predictions from a model of global psychophysics about differences between perceptual and physical matches. *Attention Perception & Psychophysics, 74*(8), 1668-1680. http://dx.doi.org/10.3758/s13414-012-0334-8

Stevens, J. C. (1974). Families of converging power functions in psychophysics. In H. R. Moskowitz, B. Scharf & J. C. Stevens (Eds.), *Sensation and measurement: Papers in honor of S. S. Stevens.* Oxford, England: D. Reidel.

Stevens, J. C. (1990). Perceived roughness as a function of body locus. *Attention, Perception, & Psychophysics, 47*(3), 298-304. http://dx.doi.org/10.3758/BF03205004

Stevens, J. C., & Marks, L. E. (1980). Cross-modality matching functions generated by magnitude estimation. *Perception & Psychophysics, 27*(5), 379-389. http://dx.doi.org/10.3758/BF03204456

Stevens, J. C., & Marks, L. E. (1999). Stevens power law in vision: exponents, intercepts, and thresholds. *Fechner Day, 99*, 82-87.

Stevens, S. S. (1946). On the theory of scales of measurement. *Science, 103*, 677-680. http://www.jstor.org/stable/1671815

Stevens, S. S. (1961). To honour Fechner and repeal his law. *Science, 133*, 80-86. http://dx.doi.org/10.1126/science.133.3446.80

Stevens, S. S. (1975). *Psychophysics*. New York: Wiley.

Stevens, S. S., & Galanter, E. H. (1957). Ratio scales and category scales for a dozen perceptual continua. *Journal of Experimental Psychology, 54*, 377-411. http://dx.doi.org/10.1037/h0043680

. Stevens' Power Law. Retrieved 18/8/2013, from http://en.wikipedia.org/wiki/Stevens%27_power_law

Teghtsoonian, R. (1971). On the exponents in Stevens' law and the constant in Ekman's law.

Teghtsoonian, R. (2012). The Standard Model for Perceived Magnitude: A Framework for (Almost) Everything Known About It. *American Journal of Psychology, 125*(2), 165-174. http://dx.doi.org/10.5406/amerjpsyc.125.2.0165

Tversky, A. (1967). Additivity, utility and subjective probability. *Journal of Mathematical Psychology, 4*, 175-202.

West, R., Ward, L., & Khosla, R. (2000). Constrained scaling: The effect of learned psychophysical scales on idiosyncratic response bias. *Attention, Perception, & Psychophysics, 62*(1), 137-151. http://dx.doi.org/10.3758/BF03212067

Acknowledgements

I would like to thank Joseph Glicksohn, Robert Teghtsoonian, and a third anonymous reviewer together with Editor Alberto Maydeu-Olivare for extremely perceptive and constructive comments on earlier versions of this Ms. I would also like to thank members of the International Society for Psychophysics for helpful discussions and email comments.

Table 1

Duration goodness of Fit Measure Z for Equations 2-5

Equation	K	Metric	Mean Z	LCL Z	UCL Z	SD Z	R^2	p	η^2
2 & 3	2		3.46	3.26	3.65	.77	.9960	.00000	.58
4&5	3		3.84	3.62	4.06	.90	.9981		
2&4		Power	3.72	3.50	3.94	.89	.9977	.00389	.20
3&5		Log	3.57	3.38	3.77	.83	.9969		
2	2	Power	3.58	3.38	3.79	.81	.9969	.00002	.39
3		Log	3.33	3.13	3.53	.72	.9949		
4	3	Power	3.85	3.62	4.09	.94	.9982		
5		Log	3.82	3.60	4.04	.87	.9981		

Note. K is the number of model parameters. SD is the raw standard deviation for the relevant group. LCL is lower 95% confidence level, UCL is upper 95% confidence level from model fitted in MIXED. R^2 is adjusted value equivalent to mean Z. p is the probability of the null hypothesis, and h2 is the effect size, from MIXED analysis F with df (1,34) for the comparison within horizontal lines

Table 2

Duration summary statistics for scaling parameter a as a function of model, mood and judgment for prevalent 2 parameter log and best fitting 3 parameter power models

Equation	Model	Mood	Judgment	Mean	LCL	UCL	SD
3	2 Parameter log	Normal	Estimation	1.11	.93	1.30	.43
			Production	1.00	.87	1.13	.30
		Depressed	Estimation	1.18	.99	1.38	.39
			Production	1.04	.90	1.18	.30
	3 Parameter power	Normal	Estimation	2.07	1.36	2.78	2.09
			Production	1.12	.72	1.52	.59
		Depressed	Estimation	.96	.20	1.73	.72
			Production	1.65	1.22	2.08	1.17

Note. SD is the raw standard deviation for the relevant group. LCL is lower 95% confidence level, UCL is upper 95% confidence level from model fitted in MIXED.

Table 3

Goodness of fit measure, Fisher's Z, for roughness

Equation	K	Metric	Mean Z	LCL Z	UCL Z	SD Z	R^2	p	η^2
2 & 3	2		1.71	1.52	1.90	.42	.8780	.00050	.54
4&5	3		1.82	1.62	2.02	.46	.9000		
2&4		Power	1.81	1.59	2.03	.47	.8980	.01870	.30
3&5		Log	1.72	1.55	1.90	.41	.8800		

Note. K is the number of model parameters. SD is the raw standard deviation for the relevant group. LCL is lower 95% confidence level, UCL is upper 95% confidence level from model fitted in MIXED. p is the probability of the null hypothesis, and η^2 is the effect size, from MIXED analysis F with df (1,34).

Table 4

Summary of Roughness Parameters as a Function of Model

Equation	Model	Mean	LCL	UCL	SD
		Power law exponent, *n*			
4	Power3	.63	.48	.78	.37
5	Log3	.92	.72	1.12	.59
3	Log2	.84	.69	1.00	.41
		Scale parameter, *a*			
4	Power3	1.39	1.09	1.70	.58
5	Log3	1.18	.86	1.49	.75
3	Log2	1.28	1.07	1.49	.41
		Threshold parameter, *b*			
4	Power3	.09	-.05	.24	.37
5	Log3	-.27	-.63	.09	1.32

Note. LCL is lower 95% confidence limit, UCL is upper 95% confidence level, SD is standard.

Figure Captions

Fig 1 Example duration fits. Top left low BDI estimation: Equation 2 blue solid, r= .967 n= .92 a=1.30; Equation 3 green long dashes, r= .955 n= .99 a= 1.02; Equation 4 red dotted, r=. 969 n= .60 a= 4.40 b= 3.33; Equation 5 purple short dashes, r= .986 n= .45 a= 7.40 b= 4.58. Top right production: Equation 2 blue solid r= .997 n= .93 a= 1.30; Equation 3 green long dashes, r= .997 n= .94 a= 1.25; Equation 4 red dotted r= .999 n= 1.54 a= .09 b= -10.41; Equation 5power r= .998 n= 1.16 a= .51 b= -2.34. Bottom left high BDI estimation: Equation 2 blue solid, r= .996 n= .89 a= 1.98; Equation 3 green long dashes, r= .995 n= .98 a= 1.45; Equation 4 red dotted, r= .998 n= 1.62 a= .08 b= -9.70; Equation 5 purple short dashes, r= .996 n= 1.19 a= .55 b= -1.56. Bottom right high BDI production: Equation 2 blue solid, r= .988 n= 1.01 a= 1.30; Equation 3 green long dashes, r= .987 n= 1.04 a= 1.17; Equation 4 red dotted, r= .995 n= .63 a= 4.57 b= 4.12; Equation 5 purple short dashes, r= .998 n= .71 a= 3.43 b= 3.39

Fig 2 Example roughness fits. Top left blind stylus: Equation 2 blue solid, r= .952 n= .63 a= 1.13; equation 3 green long dashes, r= .921 n= .74 a= 1.15; equation 4 red dotted, r= .977 n= .30 a= 1.31 b= .34; equation 5 purple short dashes, r= .951 n= .34 a= 1.33 b= .33. Top right blind thimble: equation 2 blue solid, r=. 971 n= .66 a= 1.13; equation 3 green long dashes, r= .966 n= .77 a=1 .16; equation 4 red dotted, r= .991 n= .34 a= 1.34 b= .33; equation 5 purple short dashes, r= .985 n= .38 a= 1.35 b= .31. Bottom left sighted stylus: equation 2 blue solid, r= .905 n= .73 a= 1.49; equation 3 green long dashes, r= .922 n= .97 a= 1.57; equation 4 red dotted, r= .946 n= .36 a= 1.83; equation 5 purple short dashes, r= .944 n= .52 a= 1.91. Bottom right sighted thimble: equation 2 blue solid, r= .943 n= .89 a= 1.27; equation 3 green long dashes, r= .923 n= 1.18 a= 1.34; equation 4 red dotted, r= .973 n= .41 a= 1.61; equation 5 purple short dashes, r= .973 n= .42 a= 1.60

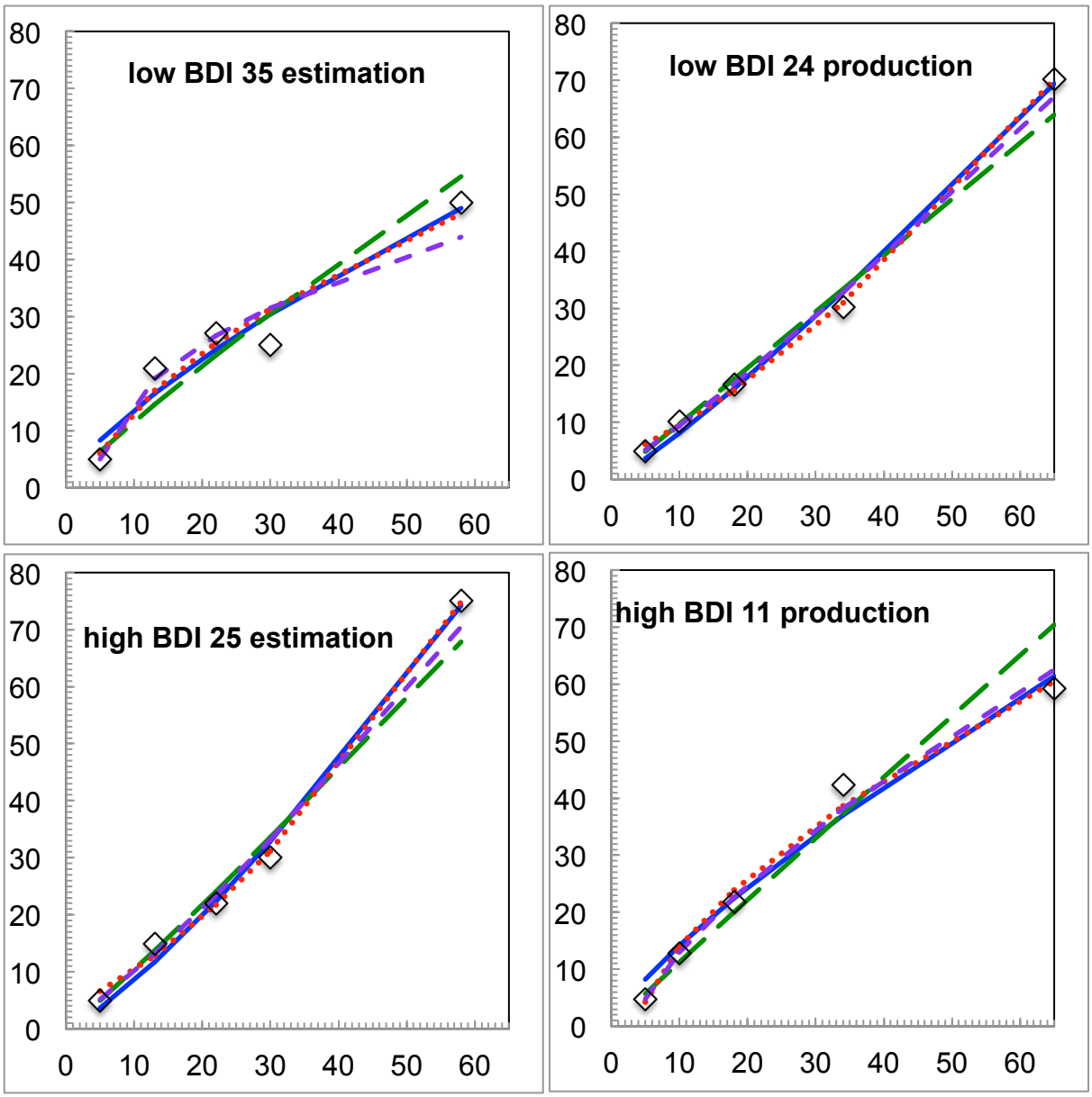

Fig 1 Example duration fits.

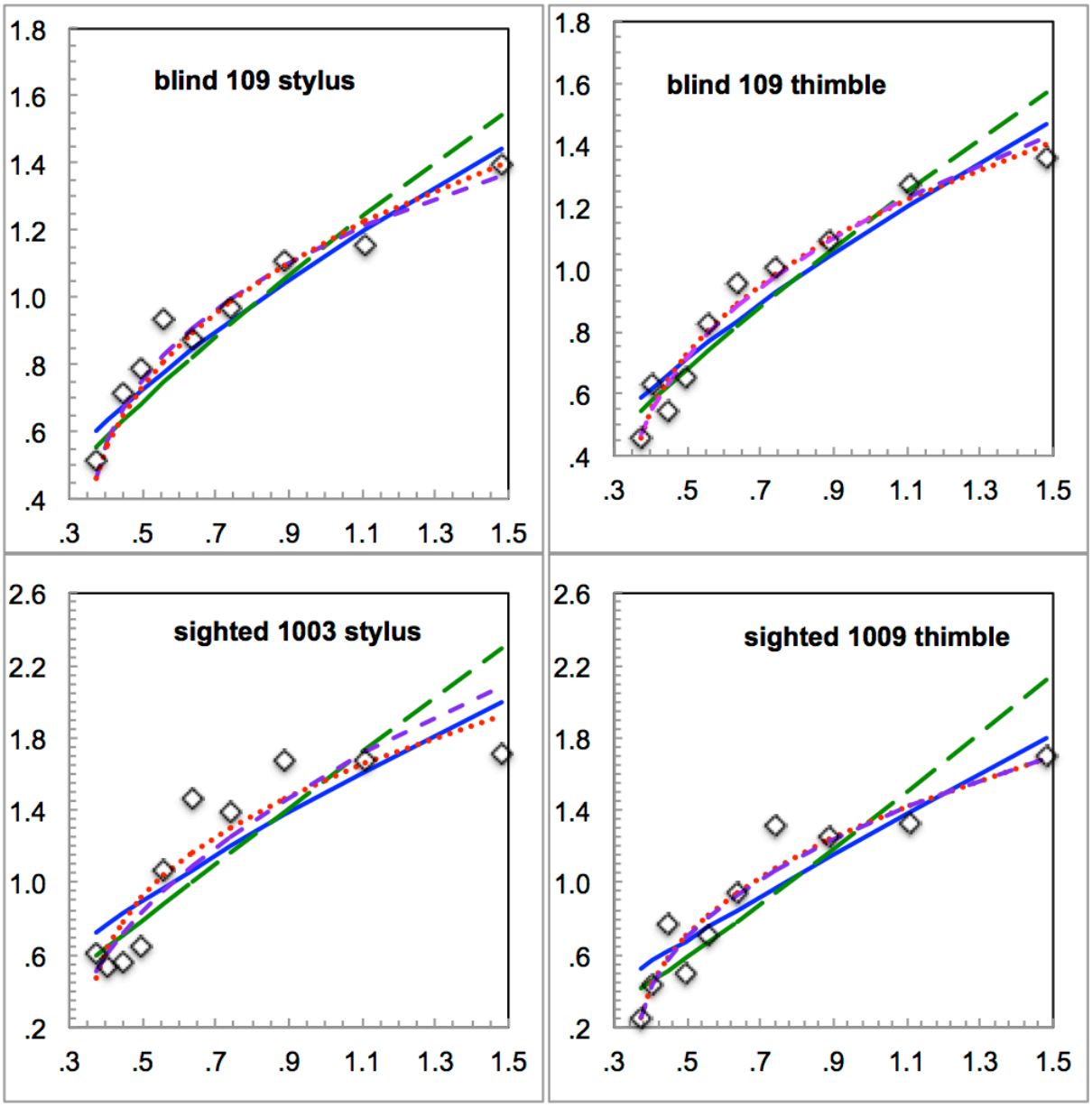

Fig 2 Example roughness fits.