

Detection theory analysis of scaling and discrimination tasks: Responses to noxious thermal stimuli

CHEE-WEE TAN

Queen Margaret University, Edinburgh, Scotland

SHEA T. PALMER

University of the West of England, Bristol, England

DENIS J. MARTIN

University of Teesside, Middlesbrough, England

AND

PATRICIA A. ROCHE

Queen Margaret University, Edinburgh, Scotland

This study's main purpose was to examine the sensitivity estimates obtained from scaling and discrimination approaches for nociception assessment in healthy individuals. This investigation may inform future applications in diagnostic procedures for painful conditions. Models of psychophysical judgment based on those of Durlach and Braida (1969), Laming (1984), and Irwin and Whitehead (1991) were used as the common analytical framework. Noxious thermal contact stimuli were used. The results show that the scaling approach produced lower detection theory sensitivity estimates than did the discrimination approach. The additional judgment variance in scaling tasks could explain this lowered sensitivity. The relative judgmental variance value of 2.18 obtained in this study is lower than variance values found in previous investigations. This discrepancy is probably due to the relatively smaller stimulus range employed in this study. The authors propose that the theoretical framework used in this study may be used in future studies to investigate the different dimensions of pain perception.

Psychophysical measurements are commonly used to quantify the judgments of noxious stimuli. There are two main psychophysical approaches in pain research. The first, the *direct scaling method*, involves having participants estimate their perception of the intensity of the noxious stimulus. An example of such an approach using a bounded scale is the visual analog scale (Price, 1994). The second, the *discrimination method*, involves having participants discriminate between two stimuli of different noxious intensities. An example is the yes-no experiment, in which the participant is required to state whether the stronger (or more noxious) stimulus was presented. Both direct scaling and discrimination methods are used frequently in pain research—for example, in brain imaging studies (Pertovaara et al., 2004), neurophysiological studies (Nahra & Plaghki, 2005), and clinical studies (Kempf et al., 1997). The topic of interest for our research group is the potential use of psychophysical methods as differential diagnostic tools for painful conditions. For example, a clinical scenario may consist of two similar patient groups, within a broader patient grouping, mani-

festing different responses to scaling or discrimination methods. This may provide the basis for further studies investigating whether the differences involve biological, affective, or cognitive dissimilarities (Petersen & Rowbotham, 2006). In order to proceed with this program, the responses of healthy individuals, as measured by scaling and discrimination methods intended for clinical application, need to be compared. However, critics have argued that the discrimination method may not quantify the perceived noxiousness of the stimuli; rather, it may only compare the perceived intensity magnitudes of the stimuli (Craig & Rollman, 1999; Rollman, 1977). Discrimination methods may therefore lack construct validity in measuring the perceived noxiousness of stimuli.

To investigate whether discrimination methods do measure the perceived noxiousness of the stimuli and whether their measurements are comparable with those of the direct scaling methods, Irwin and Whitehead (1991) and Irwin, Hautus, Dawson, Welch, and Bayly (1994) used signal detection theory as a common analytical framework for the data obtained from both methods. The index used

for estimating the sensitivity was d' . The theories, underlying the analytical framework, were originally proposed by Braida and Durlach (1972) and Laming (1984, 1997). This analytical framework provided a means of comparing different psychophysical methods in sensory perception.

Irwin et al. (1994) found that the psychometric function obtained by the discrimination method was similar to that obtained by the direct scaling method. This provided some evidence to support the use of discrimination in the measurement of noxious stimuli perception. However, the sensitivities obtained by the direct scaling method were reduced compared with those obtained by the discrimination method. This may be partially explained by the way judgments are made for both methods. Laming (1984) argued that the magnitude that participants assigned to physical stimuli may be considered to be only nominal. Therefore, information obtainable from these judgments is limited to the response frequencies assigned to different stimuli. Using the information from the response frequencies, the discriminability between the stimuli may be estimated. From this perspective, the sensitivity estimate obtained through sensory judgment is a rough estimate of sensory discrimination. According to Braida and Durlach's (1972) theory, the reduction in d' obtained by the direct scaling method resulted from the way participants judged the stimuli presented during the task. For the direct scaling method, participants compared the sensation of a given stimulus with the range of sensations of all the stimuli presented in the course of the experiment (Braida & Durlach, 1972). In addition to that, the responses made by participants for the direct scaling method may be autocorrelated to the previous response (Laming, 1984). These potential sources of judgment variance may degrade the d' for the direct scaling method. These judgment variances may not be present, or may be minimized, in the discrimination method.

Braida and Durlach's (1972) theory states that it is possible to estimate the relative extent of the judgment variance inherent in the direct scaling method compared with the variance associated with the discrimination method. According to the standard model of signal detection theory, the sensitivity is $d'_D = (\mu_2 - \mu_1)/\sigma_D$, where d'_D is the discriminability between the two adjacent classes of stimuli, μ_1 and μ_2 are the means of the normal probability densities, and σ_D is their common standard deviation. When the standard model is extended to encompass the additional variance inherent in the direct scaling method, then $d'_S = (\mu_2 - \mu_1)/(\sigma_D^2 + \sigma_S^2)^{1/2}$, where d'_S is the discriminability between the two adjacent classes of stimuli in the direct scaling method, σ_D^2 is the stimulus variance associated with the discrimination method, and σ_S^2 is the judgmental variance associated with the direct scaling method (Macmillan & Creelman, 2005, p. 134). An estimate of the additional judgmental variance relative to the stimulus variance can be obtained by

$$\sigma_S^2 / \sigma_D^2 = (d'_D / d'_S)^2 - 1 \quad (1)$$

(Durlach & Braida, 1969; Macmillan & Creelman, 2005).

One of the aims of the present investigation was to extend the analytical framework proposed by Irwin et al.

(1994) to psychophysical measurements of responses to noxious thermal stimuli. The relevance of noxious thermal stimuli in pain research has been established at several levels. Such stimuli are among the most commonly used physical stimuli for evoking experimental pain (Gracely, 2005). Neurobiologically, a thermal stimulus activates a known narrow range of primary afferent fiber nociceptors—namely C-fiber and Type I and Type II A-fiber nociceptors (Meyer & Campbell, 1981; Treede, Meyer, Raja, & Campbell, 1995). At the molecular level, a noxious thermal stimulus activates a nonselective cation channel, the transient receptor potential vanilloid-1 receptor, which is a potential therapeutic target for pharmacological management of pain (Caterina et al., 1997).

Another aim was to verify previous findings by Braida and Durlach (1972), Irwin and Whitehead (1991), and Irwin et al. (1994) that the direct scaling method produced decreased d' compared with the discrimination method. The present study also estimated the relative amounts of judgment variance in both methods.

METHOD

Participants

The participants were recruited from among the students and staff of Queen Margaret University, Edinburgh, using convenience sampling. Six healthy volunteers (4 women and 2 men) took part in the experiment. The participants' median age was 28 years (range: 21–35 years).

Ethical approval. This study was approved by Queen Margaret University's research ethics committee. All of the participants provided written, informed consent for participation in this experiment.

Inclusion and exclusion criteria. The inclusion criteria were (1) age of 18 years or more and (2) ability to provide consent for participation in the study. The exclusion criteria were (1) the presence of medical conditions that caused anesthesia to the tested limb or the consumption or application of medication that caused analgesia or anesthesia to the tested limb, and (2) any wounds or injury to the tested limb.

Apparatus and Stimuli

The thermal stimuli were applied on the ventral surface of both forearms. A Thermotest (Somedic AB, Sweden) was used to administer heat stimuli via a contact thermode (with surface measuring 25 mm × 50 mm). Heating and cooling of the contact surface were achieved through a Peltier element housed within the thermode. The stimulus sets (45°C, 46°C, 47°C, and 48°C) were preprogrammed using EXPOSURE software (Somedic AB, Sweden).

Procedure

There were two tasks: a magnitude description task (MDT), representing the direct scaling method, and an intensity resolution task (IRT), representing the discrimination method. Each task was performed on different forearms for each participant, chosen at random without replacement. All randomizations within this experiment were performed using an online randomization plan generator (www.randomization.com). Each participant completed both the MDT and the IRT within the same day. Twenty practice trials, similar to the actual trials, were presented at the beginning of every task for familiarization.

Magnitude description and intensity resolution tasks. The one-interval rating task was used for both the IRT and the MDT. Each trial began with the experimenter instructing the participant to place his or her forearm on the thermode (preset at the relevant testing temperature). Each trial's observation period lasted 3 sec. An

automated auditory signal indicated to the participant to remove the forearm from the thermode after the 3 sec had elapsed. If participants were not able to tolerate the full length of stimulus application, they were allowed to lift their forearms away from the thermode, although no participants did so during the study. There was an interstimulus interval (ISI) of 10 sec before the next trial started.

The stimulus set for both tasks consisted of four temperatures: 45°C, 46°C, 47°C and 48°C. For the IRT trials, each trial presented one of two temperatures. There was equal probability of presentation for either of the two temperatures. There were three stimulus pairs in total: 45°C and 46°C, 46°C and 47°C, and 47°C and 48°C. The stimulus pair presentation was randomized. The three stimulus pairs of the IRT clocked a total of 240 trials per participant (80 trials for each pair). For the MDT trials, each trial presented one of four temperatures. Again, there was equal probability of presentation for any one of the four temperatures. There were a total of 160 trials per participant clocked for the MDT (40 trials for each of the four temperatures). The order of trial presentation for both tasks was randomized.

The participants verbally indicated their judgments to the experimenter, and these were recorded. For both tasks, responses were made according to response sets with six categories (see Figures 1A and 1B). The MDT required the participants to estimate the perceived magnitude of the stimulus presented according to six descriptions of sensory quality: warm, hot, faint pain, painful, very painful, and severe pain (Figure 1A). For the IRT, the participants rated their degree of confidence about whether the stimulus presented was the higher or lower intensity of a pair of stimulus intensities (Figure 1B).

The participants were told the temperature of the administered stimulus at the end of each trial—that is, trial-by-trial feedback was provided for both tasks. Participants' judgments may be biased by the comparison of observations with a weighted average of stimulus effects. This is also known as the *adaptation level effect* (Helson, 1964). Feedback was introduced to minimize this bias. An unpublished pilot study conducted by our group determined that participants were apprehensive about making judgments when no feedback was given. Feedback thus reassured the participants and encouraged the use of the entire response set.

For the MDT, the participants received the following instructions:

In this experiment you will be asked to judge the intensities of heat stimuli presented to you. The judgment method involves assigning categories with descriptions to match the intensities

of the heat sensations you will experience [Figure 1A shown to the participant]. There are six categories of intensities. Verbally indicate to the experimenter the category number with a description that matches most closely to the sensation you experienced. After you have done this, you will be told the temperature of the heat stimulus just presented to you.

For the IRT, the participants received the following instructions: In this experiment you will be asked to determine which one of two heat stimuli was presented to you. One stimulus is hotter than the other. Your task is to indicate whether the presented stimulus was the higher or the lower intensity and how confident you are in making that decision. There are six categories to describe your decision [Figure 1B shown to the participant]. Verbally indicate to the experimenter the category number with a description that matches most closely to your decision. After you have done this, you will be told the temperature of the heat stimulus just presented to you.

Prevention of hyperalgesia, heat injury, and windup. Two specific procedures were implemented to prevent hyperalgesia onset and heat injury of the test sites. The first procedure involved instructing the participant to shift the position of the thermode to an adjacent forearm skin area at the beginning of a new trial. The second procedure involved the enforcement of an ISI of 10 sec. The latter procedure also minimized the effect of the perceived noxiousness in latter trials increasing as a result of temporal summation. This phenomenon of noxious temporal summation is termed *windup* (Price, Hu, Dubner, & Gracely, 1977; Staud, Price, Robinson, Mauderli, & Vierck, 2004). Windup is usually maintained when the ISI is less than 3 sec. The participant's forearm was checked by the experimenter for signs of heat injury after every 20 trials or if there was a concern that heat injury might have occurred. Signs of heat injury or hyperalgesia, shown by profound erythema with pain or hypersensitivity of the skin, were identified as criteria for withdrawal from the study. No participants suffered any form of heat injury during this study.

Analysis

The receiver operating characteristic (ROC) curves of each stimuli pair and task were also plotted for every participant. The Gaussian unequal variance model was fitted to the data using the RScorePlus software written by Lewis Harvey. RScorePlus is derived from Dorf-

1	2	3	4	5	6
Warm	Hot	Faint Pain	Painful	Very Painful	Severe Pain

1	2	3	4	5	6
I am absolutely certain the weak stimulus was presented	I am fairly certain the weak stimulus was presented	I am somewhat certain that the weak stimulus was presented	I am somewhat certain that the strong stimulus was presented	I am fairly certain the strong stimulus was presented	I am absolutely certain the strong stimulus was presented

Figure 1. (A) Magnitude description scale representing the scaling method. This scale was presented to the participants during the magnitude description task (MDT) for judgment. The participants verbally provided the number that matched the description of the magnitude of sensation felt. (B) Intensity resolution scale representing the discrimination method. This scale was presented to the participants during the intensity resolution task (IRT) for judgment. The participants verbally provided the number that matched the description of their degree of confidence about which of the two stimuli (stronger or weaker) was presented.

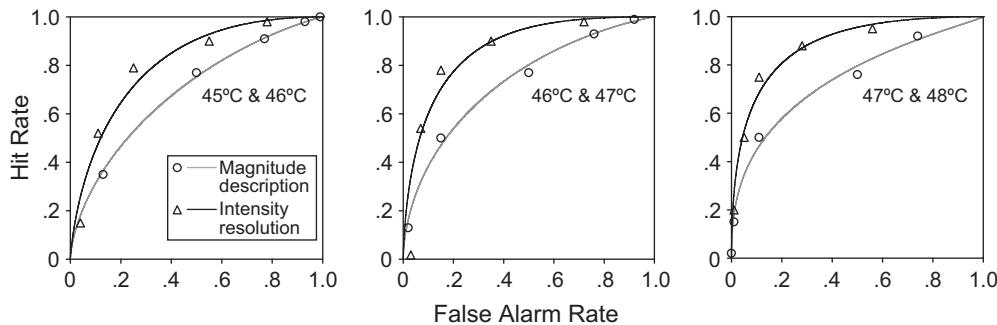


Figure 2. Receiver operating characteristic (ROC) curves fitted using a jackknifed procedure utilizing the pooled ratings of all 6 participants. Each panel shows the ROC curves of the MDT and the IRT for each stimulus pair.

man and Alf's (1969) RScore program, and it provides a maximum-likelihood fit of the signal detection model to the rating data. A total of 36 ROCs (6 participants \times 3 stimuli pairs \times 2 tasks) were generated for analysis. For the MDT, the adjacent temperatures were paired for analysis. This yielded the same number of stimulus pairs as did the IRT. Data from both tasks were analyzed in a similar manner. The detection theory index of discriminability, d_a (Macmillan & Creelman, 2005; Simpson & Fitter, 1973), and the slopes of the ROCs based on three stimulus pairs, s , were computed. The index d_a assumes an unequal variance model and is numerically equal to d' in the equal variance case. The Gaussian equal variance index, d' , was to be adopted if s for the discrimination and scaling data did not systematically depart from unity. When extreme response frequencies were present (i.e., categories containing proportions of zero), the categories were collapsed for analysis.

Cumulative sensitivity function. The d' values of adjacent stimuli for both tasks were cumulated so that the total sensitivity across the temperature range could be visualized. Durlach and Braida (1969) named the resultant plots *cumulative sensitivity functions* (CSF). The lines of best fit through the origin were plotted using the least-squares method for the CSF of both tasks. The Weber fraction was calculated for each using the CSF. The Weber fraction, in this context, may be defined as the stimulus difference that is needed to produce a performance of $d' = 1$ as the just noticeable difference.

Relative judgmental variance. Equation 1 was used to estimate the relative variance, which is reproduced here as $\sigma_{\text{MDT}}^2 / \sigma_{\text{IRT}}^2 = (d'_{\text{IRT}} / d'_{\text{MDT}})^2 - 1$, where d'_{MDT} is the sensitivity between the two adjacent temperatures in the MDT, d'_{IRT} is the sensitivity between the two temperatures in the IRT, σ_{MDT}^2 is the judgmental variance associated with the MDT, and σ_{IRT}^2 is the stimulus variance associated with the IRT.

RESULTS

Receiver Operating Characteristics

Of the 36 total ROC curves obtained, two differed significantly from the unequal variance model at the .05 significance level according to the chi-square goodness-of-fit statistic. The individual data from all participants were jackknifed, following the approach by Dorfman and Berbaum (1986), to generate six additional ROC curves to summarize the results of all stimulus pairs in both tasks. These ROC curves are shown in Figure 2. The jackknife procedure aims to avoid the common drawbacks of conventional averaging of sensitivity estimates (Macmillan & Kaplan, 1985); one of these drawbacks is obtaining a lower estimate of sensitivity compared with the sensitivity estimates that would be obtained from the original data if no averaging was used.

The ROC slopes for the discrimination and scaling methods based on the three stimulus pairs were 1.01 ($SE = .09$) and 1.05 ($SE = .13$), respectively. The slopes for both tasks did not depart systematically from unity; therefore, the Gaussian equal variance d' was used instead of d_a .

Discriminability Results

Figure 3 summarizes the discriminability of the stimulus pairs within each task. Although the data were jackknifed to generate the ROC curves shown in Figure 2, the conventional averaging of the sensitivity means was retained in Figure 3 to show the actual data for the 6 participants. Figure 3 shows that the average discriminability of the IRT was always higher than that of the MDT. This observation is as predicted by the analytical framework. It also agrees with results from previous studies using a noxious electrocutaneous stimulus (Irwin et al., 1994; Irwin & Whitehead, 1991; Rollman, 1983). Also, the discriminability of both tasks increased with an elevation of the temperatures of the stimulus pair. A repeated measures ANOVA (2 tasks \times 3 stimulus pairs) performed on the discrimination ability data showed a significant main effect of task [$F(1,5) = 24.98, p = .004$]. There was also a

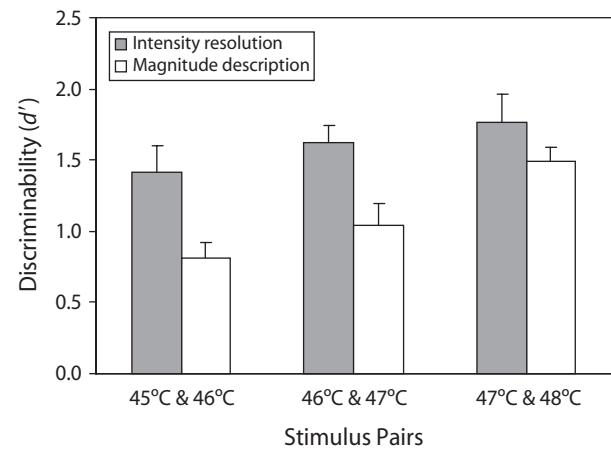


Figure 3. Mean discriminability, obtained through conventional averaging, of the MDT and IRT methods for all stimulus pairs. The error bars depict standard errors of the means.

significant main effect of stimulus pairs [$F(2,10) = 5.37, p = .026$]. Contrasts showed that sensitivity estimates for the 46°C–47°C stimulus pairs were not significantly higher than were those for the 45°C–46°C stimulus pairs with a large effect size [$F(1,5) = 2.63, p = .166, r = .59$]. The contrast also showed that sensitivity estimates for the 47°C–48°C stimulus pairs were significantly higher than those for the 45°C–46°C stimulus pairs with a large effect size [$F(1,5) = 7.529, p = .041, r = .60$]. However, the interaction effect between task and stimulus pair was not significant ($p = .152$).

Cumulative sensitivity functions. The CSFs were obtained using the jackknifed sensitivity estimates. The d' values of adjacent stimuli were cumulated. The successive cumulative sensitivities provided coordinates on the y -axis for plotting the CSF. Figure 4 shows the CSFs for this study. The linear functions were fitted to the data using the least-squares method, with the functions passing through the origin. There is a difference between the slopes of the two CSFs. The slope for the IRT is steeper than that of the MDT, indicating that the overall discriminability of the IRT was better than that of the MDT. Since the linear fit of these functions was adequate, it may be said that the averaged discrimination performances of the participants were in accordance with Weber's law. The Weber fractions were found to be 0.026 for the MDT and 0.015 for the IRT.

Relative judgmental variance. Using Equation 1, the additional variance in the MDT was calculated to be 2.18 times greater than the variance in the IRT. This number was calculated using the cumulated sensitivity values obtained from the CSFs for the MDT and IRT.

DISCUSSION

This study found that the MDT yielded decreased sensitivities compared with the IRT for noxious thermal stim-

uli. The amount of additional judgmental variance in the MDT was 2.18 times greater than that in the IRT. These results are consistent with Durlach and Braida's predictions (1969) and results from previous studies (Irwin et al., 1994; Irwin & Whitehead, 1991; Rollman, 1983).

The Contribution of Judgmental Variance to a Poorer Sensitivity in MDT

This study was conducted to integrate the direct scaling and discrimination methods (MDT and IRT, respectively) under a common framework. The present finding of lower sensitivity estimates yielded by the MDT as compared with those yielded by the IRT supports the prediction that an additional component of variance may be attributed for direct scaling methods. Our results suggest that outcomes yielded by discrimination methods and direct scaling methods may be related. Therefore, this finding adds evidence to the assertion that discrimination methods are suitable for measuring responses from noxious stimuli. However, our results would have to be interpreted under the framework and assumptions of Durlach and Braida's (1969) theory. Irwin et al. (1994) stated that if this same analytical framework were extended to the method of magnitude estimation, similar results could be expected. It could be argued that the MDT used in this study is an example of scales that involve judgments of sensation magnitude (Braida & Durlach, 1972).

Lower Relative Variance in This Study Compared With Relative Variance in Previous Studies

It is interesting to note that the relative variance between the two methods found in the present study was 2.18, lower than that found by Irwin and Whitehead (1991). The relative variances in their description task (similar to the MDT) and identification task were 5.4 and 2.22 times more, respectively, than the variance in the discrimination task. Our result is, perhaps, not unexpected, and has two possible explanations. The first explanation may be associated with the use of relatively lower numbers of trials in this study, and the second may be connected with the stimulus range used for the MDT.

Influence of Lower Trial Numbers on Judgment Variance

The use of lower numbers of trials would inevitably increase both the variability of the responses and the likelihood of extreme proportions. This response variability may contribute considerable statistical bias to the sensitivity estimates (Hautus, 1997). An unpublished study by our research group found that when the number of trials per intensity in a one-interval confidence-rating task was decreased from 40 to 17, the amount of variance for the sensitivity estimates of the 17-trial task increased 1.74 times.

One might argue that higher numbers of trials could be used to suppress the amount of variance in the sensitivity estimates, and we acknowledge that using more trials should be done as much as is practically possible. There are, however, other factors to consider when large numbers of trials are used, such as the onset of heat injury

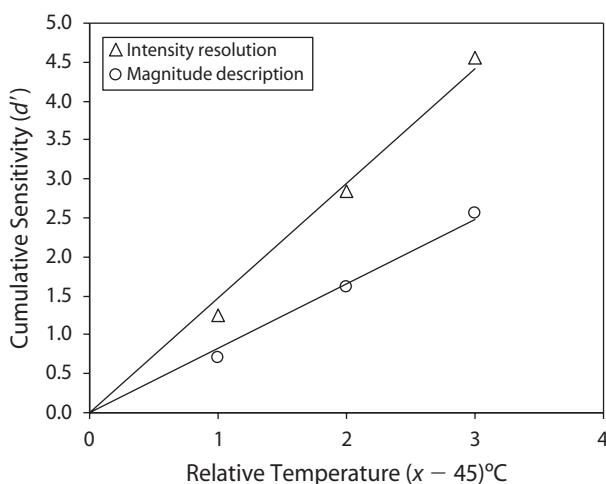


Figure 4. Cumulative sensitivity functions for the MDT and IRT as a function of relative temperature. The jackknifed d' values were used to obtain the cumulative sensitivity functions. The relative temperatures were obtained by subtracting 45°C from the higher temperature of each stimulus pair.

and hyperalgesia (Pedersen & Kehlet, 1998), the ethical acceptability of prolonged noxious stimulation (Charlton, 1995), and, ultimately, the transferability of the laboratory protocol to clinical studies. All of these factors should be carefully considered when deciding on numbers of trial presentations.

Influence of Stimulus Range on Judgment Variance

Another factor that may have influenced the amount of the judgment variance for the MDT was the range of stimuli judged. For the IRT, the participant was required to concentrate only on the difference between the two stimuli presented in a pair. This is in contrast to the MDT, which may have allowed participants to also compare the sensation magnitude of the presented stimulus to the context of the stimulus range, in spite of the trial-by-trial feedback provided to participants. A similar explanation was also offered by Rollman (1979), based on adaptation level theory (Helson, 1964). Durlach and Braida (1969) theorized that if the stimulus range for the scaling task were large, the task would become more difficult for the participants, leading to lowered sensitivity estimates. Since according to Durlach and Braida's theory, discrimination tasks are easier because the judgmental component is absent, performance on such tasks will always be better than performance on direct scaling tasks. However, Durlach and Braida predicted that for a small stimulus range, the contribution of the judgment variance in direct scaling would become almost negligible, and performance on the scaling task would be similar to that on the discrimination task. This prediction was generally supported by Pynn, Braida, and Durlach's (1972) study on auditory intensity discrimination. This raises another possible reason for the lower additional variance observed for the MDT in our investigation compared with other studies: the possibility that the relatively narrow temperature range for this study caused smaller values of the judgment variance to be found, as predicted by Durlach and Braida's theory. Nevertheless, further studies need to be conducted to confirm this conjecture.

CSF As a Potential Tool for Investigating Suprathreshold Sensitivities

The perception of noxious experimental stimuli has also been studied with methods that obtain point estimates of the transition from innocuousness to painfulness (Graven-Nielsen, Sergerdahl, Svensson, & Arendt-Nielsen, 2001). An example of the use of point estimates in pain research is the determination of the pain threshold using the method of limits. The effectiveness of pain relief treatments has been evaluated largely on the basis of the lowering of this threshold. This method does not illuminate the effects of pain relief treatments on the suprathreshold sensitivities in which pain, the construct of interest, resides. This is especially important for suprathreshold sensitivities in studies examining nociception. The same criticism could be leveled at the sensitivity estimates obtained for individual stimulus pairs in the present study. The sensitivity esti-

mates provided information confined to only one specific stimulus pair, which reveals little about the sensitivities contained within the sensory range of interest. This problem was solved, for the purposes of this study, through the use of CSFs. CSFs may provide additional information on the suprathreshold range of sensitivities and the effects of intervention on them (Gracely, 2005), and they may be a valuable tool for future studies investigating the description and influence of interventions on suprathreshold sensitivities.

The Relevance of Our Findings for Future Clinical Studies

Our findings show that when humans judged the intensities of thermal stimuli, decisions were made within the context of the type of task and the stimulus range. Since most experimental measures of pain in clinical studies use variants of the direct scaling method, our findings suggest that the responses of clinical participants may contain a component of judgment variance. In order to prevent diminishing the participants' discrimination ability by the effect of stimuli range comparisons (Poulton, 1989), the responses of clinical participants may be examined using the discrimination method, within the framework proposed by Irwin and Whitehead (1991).

Nevertheless, clinical pain is a multidimensional experience involving affective, cognitive, and sensory components (Melzack, 1999). Chronic pain experienced by patients may be associated with changes in their empirical pain thresholds or self-reported pain intensity (Kosek, Ekholm, & Hansson, 1996). Affective, cognitive, and sensory responses to pain may interact to alter the amount of variance within the psychophysical responses. Studies have provided some evidence that signal detection theory measures of pain may be influenced by affective disorders (Dworkin, Clark, & Lipsitz, 1995; Kemperman et al., 1997). These studies have used the direct scaling method within the framework of signal detection theory. Since the direct scaling method yields an additional variance on participant responses, it would be interesting to establish, in future studies, the interaction between affect and pain response in the absence of additional variance (i.e., using the discrimination method). A clinical study examining this question in chronic pain sufferers is currently being conducted by our research group.

An Alternative Interpretation: The Dimensional Hypothesis

Although we have interpreted our findings on the basis of a theory of judgment, it is possible that the results could be due to dimensional overlap between the responses of both tasks. This would mean that the judgment theory might have to be revised for nociception, since it assumes perceptual one-dimensionality (Durlach & Braida, 1969; Macmillan & Creelman, 2005, pp. 113–115). This alternative interpretation of dimensional overlap could be tested in several ways. The first method would be to utilize a multidimensional analytical approach. For example, Clark, Yang, Carroll, and Janal (1986) and Clark, Ferrer-

Brechner, Janal, Carroll, and Yang (1989) analyzed the dimensions of both experimental and clinical pain using individual differences scaling procedures. Another method would be to observe the directional shifts of sensitivity from both discrimination and direct scaling methods when an analgesic or anesthetic procedure has been performed (Rollman, 1983). If the anesthetic procedure led to similar directional shifts in sensitivity for both tasks, this would provide some evidence that responses from both tasks existed in similar dimensions. A disconfirmation test for the perceptual dimension similarity hypothesis may also be investigated in painful clinical conditions. That is, some characteristics of the painful condition may interact with the experimental stimulus, which would then yield opposite shifts in sensitivity between discrimination and direct scaling methods. Even so, disconfirmation does not negate the potential usefulness of both tasks for diagnostic purposes. In fact, the underlying basis for the opposite shifts in sensitivity, be it biological or cognitive in nature, could be elucidated and applied as a powerful clinical diagnostic tool for future treatment of painful conditions.

CONCLUSION

This study demonstrated that the discrimination approach is comparable to the direct scaling approach. Bridging the two approaches was made possible by analyzing the data under the theoretical framework of Durlach and Braida (1969), on the basis of the assumption of perceptual one-dimensionality. Our results are consistent with Durlach and Braida's prediction that an additional component of judgment variance contributes to the decreased sensitivity in the direct scaling approach. This finding is useful for clinical pain studies that employ psychophysical methods of testing, and it may also inform diagnostic procedures for painful conditions. Regardless of the type of psychophysical method used in clinical studies, it is possible to relate and compare findings. This would also suggest that discrimination methods are admissible as psychophysical procedures for pain studies. Therefore, this framework may serve as a potentially useful tool for evaluating the often complex processes of pain perception.

AUTHOR NOTE

We thank Michael Hautus, Gary Rollman, and all the reviewers for their invaluable comments on the manuscript. Correspondence concerning this article should be addressed to C.-W. Tan, Queen Margaret University, Duke Street, Edinburgh EH6 8HF, Scotland (e-mail: ctan@qmu.ac.uk).

REFERENCES

- BRAIDA, L. D., & DURLACH, N. I. (1972). Intensity perception: II. Resolution in one-interval paradigms. *Journal of the Acoustical Society of America*, **51**, 483-502.
- CATERINA, M. J., SCHUMACHER, M. A., TOMINAGA, M., ROSEN, T. A., LEVINE, J. D., & JULIUS, D. (1997). The capsaicin receptor: A heat-activated ion channel in the pain pathway. *Nature*, **389**, 816-824.
- CHARLTON, E. (1995). Ethical guidelines for pain research in humans. *Pain*, **63**, 277-278.
- CLARK, W. C., FERRER-BRECHNER, T., JANAL, M. N., CARROLL, J. D., & YANG, J. C. (1989). The dimensions of pain: A multidimensional scaling comparison of cancer patients and healthy volunteers. *Pain*, **37**, 23-32.
- CLARK, W. C., YANG, J. C., CARROLL, J. D., & JANAL, M. N. (1986). Multidimensional scaling reveals two dimensions of thermal pain. *Journal of Experimental Psychology*, **12**, 103-107.
- CRAIG, J. C., & ROLLMAN, G. B. (1999). Somesthesia. *Annual Review of Psychology*, **50**, 305-331.
- DORFMAN, D. D., & ALF, E., JR. (1969). Maximum-likelihood estimation of parameters of signal-detection theory and determination of confidence intervals: Rating-method data. *Journal of Mathematical Psychology*, **6**, 487-496.
- DORFMAN, D. D., & BERBAUM, K. (1986). RSORE-J: Pooled rating-method data: A computer program for analyzing pooled ROC curves. *Behavior Research Methods, Instruments, & Computers*, **18**, 452-462.
- DURLACH, N. I., & BRAIDA, L. D. (1969). Intensity perception: I. Preliminary theory of intensity resolution. *Journal of the Acoustical Society of America*, **42**, 372-383.
- DWORKIN, R. H., CLARK, W. C., & LIPSITZ, J. D. (1995). Pain responsibility in major depression and bipolar disorder. *Psychiatric Research*, **56**, 181.
- GRACELY, R. H. (2005). Studies of pain in human subjects. In S. B. McMahon & M. Koltzenburg (Eds.), *Wall and Melzack's textbook of pain* (5th ed., pp. 267-289). Philadelphia: Churchill Livingstone.
- GRAVEN-NIELSEN, T., SERGERDAHL, M., SVENSSON, P., & ARENDT-NIELSEN, L. (2001). Methods for induction and assessment of pain in humans with clinical and pharmacological examples. In L. Kruger (Ed.), *Methods in pain research* (pp. 263-290). Boca Raton, FL: CRC Press.
- HAUTUS, M. J. (1997). Calculating estimates of sensitivity from group data: Pooled versus averaged estimators. *Behavior Research Methods, Instruments, & Computers*, **29**, 556-562.
- HELSON, H. (1964). *Adaptation-level theory*. New York: Harper & Row.
- IRWIN, R. J., HAUTUS, M. J., DAWSON, N. J., WELCH, D., & BAYLY, M. F. (1994). Discriminability of electrotactile stimuli after topical anesthesia: Detection-theory measurement of sensitivity to painful stimuli. *Perception & Psychophysics*, **55**, 125-132.
- IRWIN, R. J., & WHITEHEAD, P. R. (1991). Towards an objective psychophysics of pain. *Psychological Science*, **2**, 230-235.
- KEMPERMAN, I., RUSS, M. J., CLARK, W. C., KAKUMA, T., ZANINE, E., & HARRISON, K. (1997). Pain assessment in self-injurious patients with borderline personality disorder using signal detection theory. *Psychiatric Research*, **70**, 175-183.
- KOSEK, E., EKHOLM, J., & HANSSON, P. (1996). Sensory dysfunction in fibromyalgia patients with implications for pathogenic mechanisms. *Pain*, **68**, 375-383.
- LAMING, D. (1984). The relativity of "absolute" judgements. *British Journal of Mathematical & Statistical Psychology*, **37**, 152-183.
- LAMING, D. (1997). The psychophysical primitive. In *The measurement of sensation* (pp. 143-167). Oxford: Oxford University Press.
- MACMILLAN, N. A., & CREELMAN, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). Mahwah, NJ: Erlbaum.
- MACMILLAN, N. A., & KAPLAN, H. L. (1985). Detection theory analysis of group data: Estimating sensitivity from average hit and false-alarm rates. *Psychological Bulletin*, **98**, 185-199.
- MELZACK, R. (1999). From the gate to the neuromatrix. *Pain*, **82** (Suppl. 6), S121-S147.
- MEYER, R. A., & CAMPBELL, J. N. (1981). Evidence for two distinct classes of unmyelinated nociceptive afferents in monkey. *Brain Research*, **224**, 149-152.
- NAHRA, H., & PLAGHKI, L. (2005). Innocuous skin cooling modulates perception and neurophysiological correlates of brief CO₂ laser stimuli in humans. *European Journal of Pain*, **9**, 521-530.
- PEDERSEN, J. L., & KEHLET, H. (1998). Secondary hyperalgesia to heat stimuli after burn injury in man. *Pain*, **76**, 377-384.
- PERTOVAARA, A., MARTIKAINEN, I. K., HAGELBERG, N., MANSIKKA, H., NAGREN, K., HIETALA, J., & SHEINEN, H. (2004). Striatal dopamine D2/D3 receptor availability with individual response characteristics to pain. *European Journal of Neuroscience*, **20**, 1587-1592.
- PETERSEN, K. L., & ROWBOTHAM, M. C. (2006). Quantitative testing scaled up for multicenter clinical research networks: A promising start. *Pain*, **123**, 219-220.
- POULTON, E. C. (1989). *Bias in quantifying judgments*. Hove, U.K.: Erlbaum.
- PRICE, D. D. (1994). Psychophysical measurement of normal and abnormal pain processing. In J. Boivie, P. Hansson, & U. Lindblom (Eds.), *Touch, temperature, and pain in health and disease: Mechanisms and assessments* (pp. 3-25). Seattle: IASP Press.

- PRICE, D. D., HU, J. W., DUBNER, R., & GRACELY, R. H. (1977). Peripheral suppression of first pain and central summation of second pain evoked by noxious heat pulses. *Pain*, **3**, 57-68.
- PYNN, C. T., BRAIDA, L. D., & DURLACH, N. I. (1972). Intensity perception: III. Resolution in small-range identification. *Journal of the Acoustical Society of America*, **51**, 559-566.
- ROLLMAN, G. B. (1977). Signal detection theory methods in pain research. *Pain*, **3**, 187-211.
- ROLLMAN, G. B. (1979). Signal detection theory pain measures: Empirical validation studies and adaptation-level effects. *Pain*, **6**, 9-21.
- ROLLMAN, G. B. (1983). Multiple representations of experimental pain. In J. J. Bonica, U. Lindblom, & A. Iggo (Eds.), *Advances in pain research and therapy* (pp. 865-869). New York: Raven.
- SIMPSON, A. J., & FITTER, M. J. (1973). What is the best index of detectability? *Psychological Bulletin*, **80**, 481-488.
- STAUD, R., PRICE, D. D., ROBINSON, M. E., MAUDERLI, A. P., & VIERCK, C. J. (2004). Maintenance of windup of second pain requires less frequent stimulation in fibromyalgia patients compared to normal controls. *Pain*, **110**, 689-696.
- TREEDE, R. D., MEYER, R. A., RAJA, S. N., & CAMPBELL, J. N. (1995). Evidence for two different heat transduction mechanisms in nociceptive primary afferents innervating monkey skin. *Journal of Physiology*, **483**, 747-758.

(Manuscript received September 20, 2006;
revision accepted for publication February 5, 2007.)