Atten Percept Psychophys (2012) 74:1668–1680
DOI 10.3758/s13414-012-0334-8

Predictions from a model of global psychophysics about differences between perceptual and physical matches

Ragnar Steingrimsson · R. Duncan Luce

Published online: 19 July 2012

Abstract A well-known phenomenon is that "matched" successive signals do not result in physical identity. This phenomenon has mostly been studied in terms of how much the second of two signals varies from the first, which is called the *time-order error* (TOE). Here, theoretical predictions led us to study the more general question of how much the matching signal differs from the standard signal, independent of the position of the matching signal as the first or second in a presentation. This we call *non-equal matches* (NEM). Using Luce's (*Psychological Review*, *109*, 520–532, 2002, *Psychological Review*, *111*, 446–454, 2004, *Psychological Review*, *115*, 601, 2008, *Psychological Review*, *119*, 373–387, 2012) global psychophysical theory, we predicted NEM when an intensity $\mathbf{z}$ is perceived to be "1 times a standard signal x." The theory predicts two different types of individual behaviors for the NEM, and these predictions were evaluated and confirmed in an experiment. We showed that the traditional definition of TOE precludes the observation, and thus the study, of the NEM phenomenon, and that the NEM effect is substantial enough to alter conclusions based on data that it affects. Furthermore, we demonstrated that the custom of averaging data over individuals clearly leads to quite misleading results. An important parameter in this modeling is a *reference point* that plays a central role in creating variability in the data, so that the key to obtaining regular data from respondents is to stabilize the reference point.

Keywords Audition · Loudness · Weighting function · 2IFC · 2AFC · Matching · Method of adjustment · Non-equal matches · Time-order error · Mathematical modeling

R. Steingrimsson (✉) · R. D. Luce
Institute for Mathematical Behavioral Science,
University of California, Irvine,
Irvine, CA 92697-5100, USA
e-mail: ragnar@uci.edu

The time-order error, or *TOE*, is reported by Stevens (1975) as a "constant error discovered long ago by Fechner . . . [where] on average the second of two equal stimuli tends to be judged greater than the first" (p. 139). For additional remarks on the TOE, see Appendix A.

Hellström (2003) lamented how relatively little work had been carried out on the TOE phenomenon. Indeed, we were quite surprised to realize how dramatic these effects can be, and so we second Hellström's (2003) sentiment, adding that these effects are probably only ignored at the peril of arriving at erroneous conclusions. We came to realize that both the definition of the TOE and the prevailing practice of averaging over individual data has precluded the study of a much more fundamental phenomenon: When a respondent provides a perceptual "match" between two successively presented signals differing in intensity only, in general, the signals are not physically identical. We call this general phenomenon *non-equal matches* (NEM).

Let x denote the signal that is presented as a *standard* that is to be matched, and let $\mathbf{z}$ denote the *matching signal*. Any respondent-selected signal is written in boldface (to emphasize that it is a random variable). With any method of matching, the standard x can be in either the temporally first or second interval. When it is first and $\mathbf{z}$ is presented second, we write $\mathbf{z}_2$, whereas when the standard x is presented second and $\mathbf{z}$ is first, we write $\mathbf{z}_1$. The data actually reported are some measure of the central tendency in the data over trials, which is denoted $\overline{z}_m, m = 1, 2$.

From the first mention of the TOE (Fechner, 1860/1966), it has been defined as the amount that the second stimulus is found to differ from the first stimulus. Let

$$\begin{aligned} \theta_{1,\mathrm{dB}} &:= x_{\mathrm{dB}} - \overline{\mathbf{z}}_{1,\mathrm{dB}} \quad \text{when the first signal, } \mathbf{z}_1, \text{ is adjusted} \\ \theta_{2,\mathrm{dB}} &:= \overline{\mathbf{z}}_{2,\mathrm{dB}} - x_{\mathrm{dB}} \quad \text{when the second signal, } \mathbf{z}_2, \text{ is adjusted,} \end{aligned} \tag{1}$$

which is expressed in dB to allow for subtraction of intensities, in line with convention for this literature. Then, if the number of $\mathbf{z}_1$ trials is n_1 and the number of $\mathbf{z}_2$ trials is n_2, the TOE is defined to be

$$\tau_{\mathrm{dB}} := \frac{1}{n_1 + n_2}\left[n_1\left(x_{\mathrm{dB}} - \theta_{1,\mathrm{dB}}\right) + n_2\left(\theta_{2,\mathrm{dB}} - x_{\mathrm{dB}}\right)\right]. \quad (2)$$

The TOE is said to be negative if $\tau_{\mathrm{dB}} < 0$, and zero or positive otherwise. The TOE is frequently studied with data for which $n_1 = 0$, so by no means can we assume that $n_1 = n_2$. The TOE literature does not ordinarily report $\theta_{1,\mathrm{dB}}$ and $\theta_{2,\mathrm{dB}}$ separately because, by definition, the TOE is concerned only with the deviation in perception of the second stimulus from the first, regardless of the actual position of the standard.

Our theory and data suggest that this definition precludes the observation of a more fundamental process—namely that, regardless of what is present as the standard signal, the respondent's own choice of a reference signal will produce quite different results. This observation led us to define NEM as the two components

$$\begin{aligned} \kappa_{1,\mathrm{dB}} &:= \theta_{1,\mathrm{dB}} = x_{\mathrm{dB}} - \bar{\mathbf{z}}_{1,\mathrm{dB}} \quad \text{when the first signal, } \mathbf{z}_1\text{, is adjusted,} \\ \kappa_{2,\mathrm{dB}} &:= -\theta_{2,\mathrm{dB}} = x_{\mathrm{dB}} - \bar{\mathbf{z}}_{2,\mathrm{dB}} \quad \text{when the second signal, } \mathbf{z}_2\text{, is adjusted,} \end{aligned} \quad (3)$$

which focus on the difference between the presented signal and the adjusted signal. We show that this two-component definition of the NEM is justified by theory and is necessary to capture differential predictions of respondents' behavior not captured by the TOE (Eq. 2).

This article has three main purposes:

1. To present a theory for the NEM phenomenon that is based on the theoretical representation arrived at by Luce (2002, 2004, 2008, 2012).
2. To carry out an experiment to evaluate the fit of data to the theoretical predictions for NEM.
3. To explore some implications of our results for the study of time-order phenomena and research methods.

Theory

The NEM theory makes explicit that the respondent has a choice of which of two sequentially presented stimuli to regard as a reference point, and that this choice leads to two distinctly different response patterns. We will show how these predictions and the associated parameters can be evaluated experimentally.

Background: A psychophysical theory

Luce (2002, 2004, 2008, 2012) developed a theory of global psychophysics, from which the present theory is derived. This theory is not domain specific, but for the sake of concreteness, we outline the necessary background in the context of audition because it is the experimental domain of this article. Suppose that x' denotes the physical intensity of the pure tone presented to the left ear, and that ε_{l} denotes the threshold intensity of the signal to that ear; then, we consider the signal $x = x' - \varepsilon_{\mathrm{l}}$—that is, the signal intensity x' actually presented, minus the left ear's threshold intensity ε_{l}. Similarly, using the same frequency and phase for the right ear, we denote the signal intensity as $u = u' - \varepsilon_{\mathrm{r}}$.[1] Pairs of auditory signal intensities (x, u) are ordered by loudness so that

$$(x, u) \gtrsim (y, v) \text{ iff } (x, u) \text{ is perceived as being at least as loud as } (y, v).$$

In principle, we could study the full binaural case, but in practice that greatly increases the theoretical complexities, and since it matters not for the essential purpose of this article, we decided not to pursue the binaural case further. Therefore, we will focus only on the two monaural cases $(x, 0)$ and $(0, u)$. Recall, in this notation, that 0 denotes the threshold intensity.

Another primitive of the theory is a form of magnitude production in which the respondent is asked to adjust the left ear intensity $\mathbf{z}$ so that it seems to be p times as loud as x. Clearly, $(\mathbf{z}, 0)$ is a function of x and p. For a right ear procedure, the notational changes are trivial.

Luce (2002, 2004, 2008, 2012) provided behavioral invariances (axioms) from which (the axioms) a numerical representation was derived. These assumptions include monotonicity, solvability, decomposition, and linking properties. We do not restate these assumptions in detail here, because they are not material to the present theory; rather, it is sufficient to describe explicitly only the representation. Suffice it to say that the behavioral properties giving rise to this representation have been empirically sustained for loudness (Steingrimsson & Luce, 2005a, b, 2006, 2007), brightness (Steingrimsson, 2009, 2011, 2012c), and perceived contrast (Steingrimsson, 2012a, b).

The first part of the representation is the existence of an order-preserving psychophysical function $\Psi(x, u)$—that is,

$$(x, 0) \gtrsim (y, 0) \Leftrightarrow \Psi(x, 0) \geq \Psi(y, 0), \quad (4)$$

$$\Psi(0, 0) = 0. \quad (5)$$

[1] Auditory signals are usually given in dB, as they are in our experiment, so the signals x and u are presented here as logarithms of $x'/\varepsilon_{\mathrm{l}}$ and $u'/\varepsilon_{\mathrm{r}}$, and therefore in subtractive form.

To simplify the writing, we use the abbreviated notations

$$\psi_1(x) := \Psi(x, 0), \tag{6}$$

$$\psi_r(x) := \Psi(0, x). \tag{7}$$

The second and key result for the present research shows that a numerical distortion function $W(p)$ exists such that the monaural signal presentations satisfies[2]

$$W(p) = \frac{\psi_i(\mathbf{z}) - \psi_i(\rho)}{\psi_i(x) - \psi_i(\rho)}, i = \text{l, r}, x\rho \geq 0, \tag{8}$$

where $\mathbf{z}=\mathbf{z}(x, p, \rho)$ and ρ is a signal parameter that we call a *reference signal*. It is used to establish the "intervals" to be compared. The reference signal ρ may be presented by the experimenter or it may be a "creation" of the respondent. When successfully stabilized, ρ simply becomes a parameter to be estimated.

This article arrives at predictions based on Eq. 8 when $p = 1$ and when the standard is presented in either the first interval, $\mathbf{z}_1$, or the second, $\mathbf{z}_2$.

A theory of NEM

This theory makes use of certain information about the forms of the unknown functions ψ_i and W. This will be discussed before moving on to the predictions for the productions of $\mathbf{z}_1$ and $\mathbf{z}_2$. All proofs are to be found in Appendix B.

Information about ψ_i and W The form of ψ_i: Steingrimsson and Luce (2006) examined one possible mathematical form for ψ_i, i=l, r: namely, power functions

$$\psi_i(x) = \alpha_i x^{\beta_i}, i = \text{l, r}. \tag{9}$$

Such an assumption can be defended in several ways. A qualitative condition, called *multiplicative invariance*, was tested, and the power function form was fully or partially sustained for 19 of 22 respondents (Steingrimsson & Luce, 2006, plus the data presented here). The six respondents who participated in the experiments reported here were screened to have power functions ψ_i.

$W(1) \neq 1$: Steingrimsson and Luce (2007) focused on the form of the weighting function, and one of their important realizations was that it is wrong to assume that $W(1) = 1$, which others, and we, initially, had done. That assumption led to some incorrect inferences (Ellermeier & Faulhammer, 2000; Narens, 1996; Zimmer, 2005), among the most imporant of which was that magnitude production led to data measured only on a subscale of a ratio scale.

Second signal matched to first (z_2) Suppose that for NEM (Eq. 3), but not in dB, we consider using the measure in terms of the psychophysical function—namely,

$$\kappa_{\psi_i}(x) := \psi_i(\mathbf{z}) - \psi_i(x), i = 1, \text{r}. \tag{10}$$

Proposition 1. *Suppose that Eq. 8 holds, and define $\omega := W(1)$. Then*

$$\kappa_{2,\psi_i}(x) = \psi_i(\mathbf{z}_2) - \psi_i(x) = (\omega - 1)[\psi_i(x) - \psi_i(\rho_2)], i = 1, \text{r}, \tag{11}$$

where ρ is subscripted 2 because the second signal is adjusted.

In words, $\kappa_{2,\psi_i}(x)$ is a linear function of $\psi_i(x) - \psi_i(\rho_2)$ with a slope of $\omega - 1$; these facts are of importance in evaluating the theory.[3]

First signal matched to second (z_1) Note that when the second signal is adjusted, $\mathbf{z}_2$ is both the signal that is adjusted and in the second interval. So, there is only one way to fit the model described above. In contrast, matching the first signal to the second is somewhat more complex to analyze, because the representation in Eq. 8 is ambiguous in the following sense: Does the numerator $\psi_i(\mathbf{z}) - \psi_i(\rho)$ of the representation correspond to the second signal presented, or to the signal, in this case $\mathbf{z}_1$, that is adjusted until the respondent perceives it as matching the standard x? As we will see in detail, when the first signal is matched to the second, the model will yield quite different predictions, depending on whether the model numerator corresponds to the second signal or to the adjusted signal. As we will see, they seem to correspond to a difference among respondents. This ambiguity splits respondents into two types.

In the following proposition, the reference signal ρ needs two subscripts. The first identifies that the matching signal is in Interval 1, and the second identifies whether the numerator is the matching signal or the second signal.

[2] For those familiar with utility theory, Eq. 8 has a familiar flavor. Solving for the matching term, $\psi_i(\mathbf{z}) = W(p)\psi_i(x) + [1 - W(p)]\psi_i(\rho)$. With $p \leq 1$, this is the weighted utility representation of a binary gamble in which x occurs with probability p and ρ with probability $1 - p$. Usually in utility theory, $W(1) = 1$, but we definitely do not assume this in the psychophysical context, for unambiguous empirical reasons (see the next section, Information about ψ_i and W).

[3] Hellström (personal communication, 2010) pointed out to us that the form in Eq. 11 is not new. It was suggested as a generalization of his (Hellström, 1979) sensation-weighting function, with ω written as s_1/s_2, which in turn was a generalization of Michels and Helson (1954). Their equations are, from our perspective, ad hoc, whereas our Eq. 11 derives from an axiomatized and successfully tested behavioral theory.

Proposition 2. *Assuming the basic representation in Eq. 8 and that the first signal is matched to the second signal,*

1. if the numerator corresponds to the matching signal, then

$$\kappa_{1,\psi_i}(x) = \psi_i(\mathbf{z}_1) - \psi_i(x) = (\omega - 1)[\psi_i(x) - \psi_i(\rho_{1,1})], i = 1, \mathrm{r}. \tag{12}$$

2. if the numerator corresponds to the second presentation, then

$$\kappa_{2,\psi_i}(x) = \psi_i(\mathbf{z}_1) - \psi_i(x) = (1 - \omega//\omega)[\psi_i(x) - \psi_i(\rho_{1,2})], i = 1, \mathrm{r}. \tag{13}$$

In words, $\kappa_{1,\psi_i}(x)$ is a linear function of $\psi_i(x) - \psi_i(\rho_{1,1})$ with the slope $(\omega - 1)$, whereas $\kappa_{2,\psi_i}(x)$ is a linear function of $\psi_i(x) - \psi_i(\rho_{1,2})$ with the slope $(1 - \omega)/\omega$. Note that the slopes and reference points differ for Eqs. 12 and 13, and that Eq. 13 differs from Eq. 11, as well, with respect to both the slope and reference point. These differences are the important facts for developing the testable predictions from the theory.

Testable predictions from Propositions 1 and 2 Proposition 3 is about two concrete predictions for data that derive from Propositions 1 and 2. These predictions concern the slopes and the directions of those slopes for the straight-line equations appearing in those two propositions.

Proposition 3. *Suppose that Propositions 1 and 2 both hold. Then*

1. When the second signal is matched to the first (Eq. 11), and when the first signal is matched to the second signal with the numerator corresponding to the matching signal (Eq. 12), these two straight lines are parallel—that is, have the same slope.

2. When the first signal is matched to the second signal, then the representation in Eq. 8 is either the straight-line Eq. 11 or the straight-line Eq. 13. Both slopes are either 0 or of opposite signs.

Visualizing the NEM By Eq. 9, the form of ψ_i in Eqs. 11, 12, and 13 is a power function with the exponent β. Suppose that data for n different standards x are collected. Then, the NEM measure $\kappa_m^\beta = \mathbf{z}_m{}^\beta - x_n^\beta$ is linearly related to $x_n^\beta - \rho_m^\beta$, $m = 1, 2$ (here, the subscript for ρ is simplified); that is, these data form lines when graphed.

Estimating β, ρ, and ω Despite devising and trying several methods to estimate β from our data for individuals, nothing useful resulted. However, because group averages from a myriad of studies have established $\beta \approx .33$ when measured in intensity units (see Stevens, 1975, p. 15), we assume that value. Presumably the individual fits would be improved a bit by making estimates of individual βs, which would entail an auxiliary experiment. Such issues are discussed in Luce (2012, in Appendix B: Estimation Issues for Binary Representations, subsection on The Production Equation).

The reference signal ρ in each case can easily be estimated by finding the value of x such that

$$\begin{aligned} &\mathbf{z}^\beta - x^\beta = 0 = x^\beta - \rho^\beta \\ &\quad \Leftrightarrow \mathbf{z}^\beta = x^\beta = \rho^\beta. \end{aligned} \tag{14}$$

This usually requires interpolation or extrapolation, which is easy to do when the data plots are approximately linear.

To estimate ω, we need to decide which of the models seems to describe the data best. Denote by $\widehat{S}_1$ the estimated slope of Eq. 12, $\omega - 1$. $\widehat{S}_1$ is positive when $\omega > 1$ but negative when $\omega < 1$. Denote by $\widehat{S}_2$ the estimated slope of Eq. 13, $(1 - \omega)/\omega$, which is positive when $\omega < 1$ and negative when $\omega > 1$. For these two kinds of data—matching the second signal to the first and matching the first signal to the second—we should see for each respondent one of two quite different patterns:

1. Both lines arising in Eqs. 11 and 12 are samples from a single line, both with slope $\omega - 1$, and their estimated slopes should satisfy $\widehat{S}_1 = \widehat{S}_2$.
2. When Eqs. 11 and 13 apply, one slope is positive and the other is negative, and the value of ω is estimated using the calculated product

$$\sigma = \widehat{S}_2\widehat{S}_1 = (\omega - 1)\left(\frac{1 - \omega}{\omega}\right) = -\frac{(\omega - 1)^2}{\omega} \leq 0,$$

which yields

$$\omega = \frac{2 - \sigma \pm \sqrt{\sigma(\sigma - 4)}}{2}. \tag{15}$$

Methods for studying the NEM

To study the NEM, the experimenter must establish the point of subjective equality, *PSE*, of a standard x to another signal $\mathbf{z}_i$, $i = 1, 2$. To study which of the behaviors predicted by the theory the respondent exhibits, the respondent must have access to the information of which signal is the adjusted one; otherwise, the respondent's behavior regarding these predictions is not accessible. Here we use a form of the method of adjustment, but to demonstrate that the choice

of method is not trivial, we note that the oft-used 2IFC/2AFC procedure is not appropriate.

Free adjustment (FA) The FA procedure is a variant on the method of adjustment (which we have used in, e.g., Steingrimsson, 2009, 2011, 2012a, b, c; Steingrimsson & Luce, 2005a, b). To estimate the PSE of standard x to $\mathbf{z}_2$, x is presented followed by $\mathbf{z}_2$; the respondent indicates a change in the intensity of $\mathbf{z}_2$ that brings it closer to a perceptual match to x; the sequence is repeated with the change incorporated. The process is repeated until the respondent is satisfied with the match. The case for $\mathbf{z}_1$ is analogous. In this procedure, the respondent is completely aware when a match is being sought between x and $\mathbf{z}_1$, and when the match is between x and $\mathbf{z}_2$.

Two-interval/alternative forced choice (2IFC/2AFC) For establishing a PSE using a 2IFC procedure, two signals are typically presented, and the respondent is asked to indicate which interval contains the more intense signal, or which signal is the more intense, when viewed as a 2AFC procedure. Several estimates are obtained concurrently by intermingling staircases within a block of trials. Since the respondent makes only the judgment of which of two signals is more intense, he or she does not have access to the information about whether x is in the first or the second interval. Therefore, the 2IFC procedure is not appropriate for the study of the NEM. However, because we were interested in replicating previous TOE studies using our respondents and to explore individual differences, which are far greater than seems to have been appreciated, we collected 2IFC data, which are reported in Appendix D.

Experiment: FA matches

The primary goal of this experiment is to collect NEM data to evaluate the theoretical predictions of the NEM theory using the FA method. Propositions 1 and 2 and Eq. 8 show that the NEM is very dependent on the reference signal ρ. Should this reference signal vary greatly within a session, the NEM will appear quite variable. Conventional wisdom in experimental psychology is that such idiosyncratic errors can be "averaged" out by randomizing both trial types and the signals within experimental blocks of trials. In the context of the NEM theory, such randomization may have more to do with disrupting the stabilization of the reference signal. This fact led us to study two distinct procedures.

> Separated (*S*): A single standard was used in a block of trials. In subsequent blocks, the standards used were in an ascending order of intensity.
>
> Interleaved (*I*): Multiple standards were used and randomly interleaved within a block of trials.

In the separated condition, the changes in standard intensity occurred infrequently as compared to the interleaved condition, in which the respondent experienced frequent changes in the intensity of the standards. We expect this difference to affect the variability of the data, and therefore the measure of the NEM that we calculate from the data.

Method

Respondents This project evolved over several years. A total of 19 students—graduate and undergraduate—from New York University and the University of California, Irvine, as well as the first author (labeled R22),[4] at some point provided pilot data, and six of these respondents provided the data for our final design. All of the respondents reported normal hearing. The data reported here were all collected at UC Irvine. All respondents except the first author received compensation of $10 per session. Each person provided written consent and was treated in accordance with the "Ethical Principles of Psychologists and Code of Conduct" (American Psychological Association, 2002). The consent forms and procedures were approved by New York University's and UC Irvine's Institutional Review Boards.

The NEM theory development assumes a power form, Eq. 9, for the psychophysical function. This assumption was substantiated for each respondent, which involved their satisfying the behavioral invariance property of multiplicative invariance (see Steingrimsson & Luce, 2006, Exp. 1, for details).

Stimulus and apparatus The signal was a 1000-Hz sinusoidal tone presented for 100 ms, which included 10-ms on- and off-ramps. The stimulus consisted of two signals (tones) separated by 375 ms, with a minimum intertrial interval of 1,000 ms; however, there were no other limits on respondents' response times, so the actual intertrial duration could be longer.

Recall that the theory is cast in terms of intensities minus the respondent's threshold intensity. However, because all signals were well above threshold and the respondents were selected for normal hearing, the error from reporting intensities (x', u') in dB SPL (henceforth, abbreviated dB) was negligible.

[4] We judged this to be acceptable because knowledge of the experimental paradigm is not a factor in the perceptual task of judging the relative loudness of two tones. Specifically, there was no clear way to influence the outcome by way of experimental bias toward any desired outcome. R22's data are indeed not exemplary in that respect.

The stimuli were generated digitally using a personal computer and played through a 24-bit digital-to-analog converter (RP2.1 Real-time processor, Tucker-Davis Technology). Presentation levels were controlled by built-in features of the RP2.1, and the stimuli were presented over Sennheiser HD265L headphones to the listener, seated in an individual, single-walled IAC sound booth located in a quiet lab room.

A safety limit of 90 dB was enforced throughout.

Procedure and method The experiments were conducted in sessions that lasted at most 1 h each. The initial session was devoted to explaining the task, running practice trials, and obtaining written consent. All respondents trained for one session on each task. Rest periods were encouraged, with both their frequency and duration under each respondent's control.

When the match is with a standard presented first, the tone sequence is x followed by $\mathbf{z}_2$. The initial intensity for $\mathbf{z}_2$ was selected at random in an interval around the intensity x. The respondent was given the choice of either repeating the tone sequence or increasing or decreasing the intensity of $\mathbf{z}_2$. The respondent pressed a key to select a repetition or to increase or decrease the loudness of the comparison tone in increments of *big*, *medium*, *small*, and *extra-small* steps, which corresponded to changes of 4, 2, 1, and 0.5 dB, respectively. Following the keypress, the signal pair was played with the indicated adjustment incorporated (or no change, for a repetition). The respondent could make as many adjustments in any direction and step size as desired, until he or she was satisfied that the two tones sounded equal in loudness, which was signaled by a keypress. The matching of $\mathbf{z}_1$ to x, was done in an analogous manner.

Three standards were used: $x_1 = 58$, $x_2 = 66$, and $x_3 = 74$ dB.

The two interleaving conditions were run in separate sessions:

- In the *S* condition, 24 matches were collected, in a single block, for the standard x_1, followed by another block of 24 for x_2, and then another 24 for x_3.
- In the *I* condition, the three standards were randomized within the blocks of trials.

The matching conditions, $\mathbf{z}_1$ and $\mathbf{z}_2$, were run in a strictly alternating order within the blocks. The result was a total of four conditions: FA-S_m, $m = 1, 2$, and FA-I_m. A minimum of 30 estimates of each condition were collected, typically requiring one session of practice and three experimental sessions. For each condition, the final estimate was an average of the individual estimates for that condition.

Results and discussion

Data from the six respondents are presented in Fig. 1. In the figure, the data from the FA-*S* and FA-*I* procedures are presented. Plotted is κ_m^β as a function of $x_n^\beta - \rho_m^\beta$—see the Visualizing the NEM section and the predictions from Propositions 1 and 2 above. In the TOE literature, auditory results are almost exclusively reported in dB, so to facilitate comparison with those data, we also present our data in terms of dB in Appendix C.

The reference point ρ and $\omega = W(1)$ are estimated by way of Eqs. 14 and 15. In Table 1, these estimates are given for the FA-*S* data. For R10, R22, and R35, Eq. 15 yielded two estimates for ω, but since for all three $\widehat{S}_{2,2}0$ and $\widehat{S}_{1,2}0$, we concluded that the slope patterns imply that $\omega > 1$.

These data form two distinct patterns, exactly as described in Propositions 1 and 2. The FA-*S* data show that half of the respondents (R10, R22, and R35) seem to accord with the $\mathbf{z}_1$ model described in Eq. 12, whereas the other three (R47, R59, and R60) agree with that described in the subsequent Eq. 13. The FA-*I* data plausibly show the same pattern as the FA-*S*, date, although the pattern is not quite as clear.

Table 1 casts these regularities in the form of estimates of ω and ρ. The slopes are very consistent.

A clear difference between the FA-*S* and the FA-*I* data is that the latter show considerably higher variability than do the former. This difference in variability is exactly in line with the expectation of the reference point ρ having less chance of stabilizing when standards of varying intensities are interleaved within a block of trials. There can be no doubt that we badly need greater understanding—a theory—of how these reference points evolve. Nonetheless, the qualitative analysis gives qualified support for the theory in the case of the FA-*I* data, and strong support in the case of the FA-*S* data.

Discussion and conclusions

A well-established phenomenon is that when two stimuli are matched subjectively, they typically will not agree in intensity, and this has been called the *time-order error* or *TOE*. In this article, we have developed a theory from a well-established psychophysical theory that predicts that when a respondent matches a second stimulus to the first, as compared to matching the first to the second, he or she may in fact exhibit one of two different behaviors that will result in substantially different responses. We call this a *theory of non-equal matches*, or *NEM*. On examination, it is easy to see that the very definition of the TOE makes the recovery of the underlying NEM phenomenon not possible in

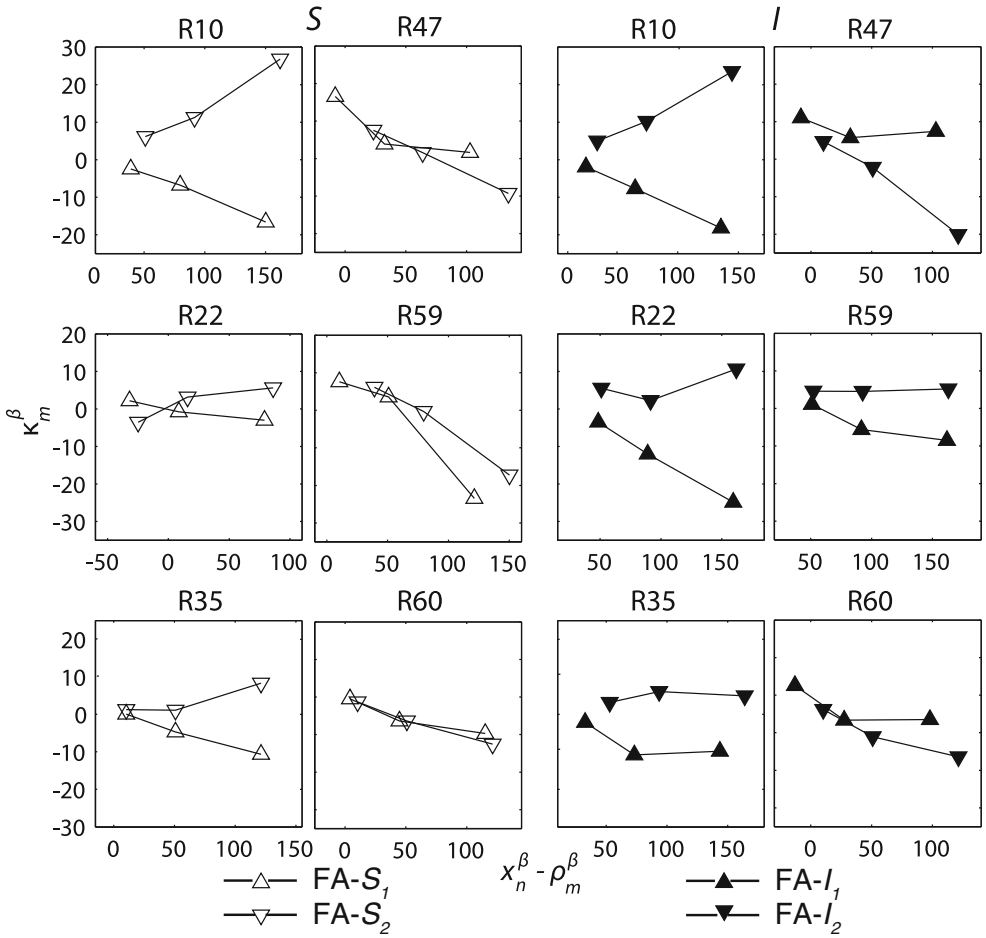

Fig. 1 Results for six respondents in the FA-*S* and FA-*I* conditions. Graphed is the NEM (in intensity) as a function of the intensity range minus the reference point for each condition. The data are graphed in intensity units, where the ranges of the abscissa and ordinate are the same for all of the individual graphs. The Open symbols indicate the separated condition and the filled symbols indicate the interleaved condition. Upward-pointing triangles indicate the results for $\bar{\mathbf{z}}_1$, and the downward-pointing triangles indicate the results for $\bar{\mathbf{z}}_2$

general, which is presumably why it has escaped observation until now. A derived prediction from the NEM theory is that respondents will exhibit one of two response patterns. This prediction was confirmed using auditory stimuli.

Stevens (1957) included the existence of TOE as one of the criteria for labeling an attribute as *prothetic*. There is, therefore, a certain circularity in saying that the TOE has been found for all prothetic scales studied, but it is certainly true that this effect has been obtained for all of those scales that have been thought to be prothetic for other reasons (Hellström, 2003). Here, we studied the NEM only for audition, but because of the relation that the TOE has to NEM, we suspect that similar results will be found with other prothetic attributes.

Table 1 Estimated slope ω and reference signal ρ (in dB) from the FA-*S* data for each respondent

Respondent	ω	ρ in dB
10	1.17	30.0
22	1.06	64.0
35	1.08	55.0
47	0.86	53.4
59	0.75	47.5
60	0.91	56.5

The estimates were obtained using Eqs. 15 and 14, respectively.

Our data demonstrate clearly that the NEM is an important factor in experiments, and one potentially large enough to alter conclusions radically, especially because there are important individual differences in the behavior of the slope of the matching functions. It is an effect that can be neither averaged nor randomized out of consideration. An interesting question for future work will be whether or not it is possible to devise a procedure to shift a person from their "natural" behavior when the first signal is adjusted to the other behavior—that is, to exchange Statements 1 and 2 of Proposition 2. From these verified predictions, it is clear that

the TOE in effect averages over two distinct response patterns, which is really inappropriate for studying order effects in perceptual matching.

Of the two procedures that we have run, it is very clear that FA-*S* yields data that are well described by our global theory, for which there is considerable independent supporting evidence. We believe that the FA-*S* procedure induces a rather more stable reference point ρ than does the FA-*I* procedure. We are in need of a far better theoretical understanding of the nature of reference point selection and how best to stabilize it in experimental practice. These are major open problems.

Appendix D reports data for the same respondents using the 2IFC procedure. However, this procedure has the limitation that it cannot be used to uncover the full NEM. However, those data appear to show that when stimuli are even more mixed within a block, as is the case for the FA-*I* over the FA-*S* data, the reference point is so unstable as to result in very irregular results for most respondents. Of course, the FA-*S* method runs counter to what is widely believed to be an important practice of randomizing conditions. In this method, the respondent finds him- or herself in a homogeneous intensity environment for about 10 min at a time, and the subsequent change in environment is always ascending with relatively small intensity changes. In this sense, each point is collected in a homogeneous intensity environment, which is not true using the 2IFC procedure.

Early work clearly assumed that the TOE was symmetric, in the sense that by randomly mixing the placements of the standard in the first or the second presentation, the TOE would be eliminated (Needham, 1934b). Because our data reject this assumed symmetry, this is a further reason to study the full NEM rather than just the TOE.

In our empirical work (e.g., Steingrimsson & Luce, 2005a, b, 2006, 2007), we built averaging over experimental conditions into our procedures, in part because of the assumption that doing so would wash out possible TOE effects. In what we expected to be a simple routine check of this assumption, we came to realize that the issue is far more complex than that. Steingrimsson and Luce (2007) verified that our results to that date were unaffected by this realization, but the effort did lead us to develop a systematic exploration of the TOE, which in turn led us to the NEM.

We have demonstrated that the NEM has numerous consequences for the evaluation of data potentially affected by it. For instance, we showed that individual responses fall into two very distinct groups of results in which linear fits to the data have slopes of opposite signs, a difference that obviously vanishes when averaged over. For the traditional TOE, we showed that it too can have opposite signs from one individual to another, which obviously is obscured by averaging, and that there is considerable variability in individual responses, which as we demonstrated is also obscured when results are averaged; see the appendices for additional detail. From these results alone, the assumption that time-order effects are "averaged out" is more than a bit optimistic. Because most of the published work on the TOE has been based on averaged data, the caveats mentioned easily apply to this entire body of work. It would be practically impossible to compile an exhaustive list of results in the literature that might be reinterpreted in the light of our results, but we conclude that the NEM appears to be a novel observation with underexplored experimental effects.

Author note This research was initially supported by National Science Foundation Grant BCS-0720288, and concluded under Air Force Office of Scientific Research Grant FA9550-08-1-0468 to the University of California, Irvine. Any opinions, findings, and conclusions or recommendations expressed in this article are those of the authors and do not necessarily reflect the views of either the National Science Foundation or the Air Force Office of Scientific Research. Additional financial support was provided by the School of Social Sciences and the Department of Cognitive Sciences at UC Irvine. Discussions with several people have contributed to our efforts, and for this reason we thank Joetta Gobell, A. A. J. Marley, Louis Narens, and Hal Stern. The comments of three referees on a previous submission, and especially those of Åke Hellström, have greatly influenced this version. The suggestions of the most recent three referees led to a greatly improved presentation.

Appendix A: Background on the TOE

In his comprehensive review of the TOE, Hellström (1985) summarized:

> Among factors that have been shown to be of importance for the direction and magnitude of the TOE are level of stimulus magnitude (e.g., Bartlett, 1939; Needham, 1935; Woodrow, 1933), length of ISI (e.g., Needham, 1935), and intensity of stimulation interpolated into the ISI (e.g., Ellis, 1973b; Lauenstein, 1933). These factors interact in a complex way with amount of training (Köhler, 1923; Needham, 1934a; Woodrow, 1933), and stimulus duration (Inomata, 1959), as well as with the particular set of ISIs used in the experiment (Wada, 1937). . . . The picture is thus very complicated, and the outcome of a TOE experiment can indeed be hard to predict. Besides, the TOE effects are often (but not always) rather small, and they vary considerably from subject to subject. (p. 36)

Moreover, Hellström (1985) noted the following refinement of Stevens's characterization: In general, for two consecutively presented stimuli, the second tends to be judged as

greater than the first at low intensities, with the effect decreasing as intensity grows until, eventually, at high intensity the second stimulus is judged to be less intense than the first. This clearly overturns the claim that the TOE is a constant error.

Stevens (1957) required four criteria to be satisfied for a perceptual dimension to be called *prothetic*. One of these was that there should be a TOE. Indeed, the TOE has been demonstrated in multiple domains, all of which are considered prothetic—including time (Eisler, Eisler, & Hellström, 2008), brightness (Ono, 1950; Woodworth & Schlosberg, 1954, p. 228, citing three studies), heaviness (Fechner, 1860/1966; Hellström, 2000), loudness (Stevens, 1957, 1975), and many more—leading to the probable conclusion that the TOE of sequential signals indeed obtains for all prothetic continua.

Hellström (1985) noted that "the TOE effects . . . vary considerably from subject to subject" (p. 36). While these individual differences have been known for a long time (e.g., Needham, 1934a, 1935) and have been further noted more recently (e.g., Hellström, 1979, 2003), both empirical studies and modeling efforts have centered on population data exclusively (e.g., Hellström, 2003; Michels & Helson, 1954; Stevens, 1975). Thus, we were surprised to realize that the individual variations, as seen in our experiment, are well beyond simple variability imposed on a single representation; rather, two quite distinct patterns appear in our data, which preclude sensible interpretations of population averages. It is well known that when individual functions are not linear, the evaluation of models cannot be based on averaged data, but only on individual data (e.g., Luce, 1995, p. 20). Our analysis of individual data clearly demonstrates that the preconditions for working with population-averaged data simply are not fulfilled. Thus, we have exclusively reported individual behavior.

Appendix B: Proofs

Proposition 1

Proof Set $p = 1$ in Eq. 8. Then, simple algebra yields Eq. 11. QED.

Proposition 2

Proof

1. Assume that the matching signal, $\mathbf{z}_1$, corresponds to the numerator. Then, the calculation is identical to Eq. 11, but with $\mathbf{z}_2$ replaced by $\mathbf{z}_1$, and the reference signal is written $\rho_{1,1}$.
2. Because the adjusted signal $\mathbf{z}_1$ is presented prior to presenting the standard x, we must invert Eq. 8, to yield

$$\frac{1}{\omega} = \frac{\psi_i(\mathbf{z}_1) - \psi_i(\rho_{1,2})}{\psi_i(x) - \psi_i(\rho_{1,2})}.$$

A calculation similar to that of Eq. 11 yields the linear relation in Eq. 13. QED.

Proposition 3

Proof

1. Both slopes are $\omega - 1$.
2. Multiply together the slopes of Eqs. 11 and 13. Taking into account the fact that $\omega > 0$, we see that

$$(\omega - 1)\left(\frac{1 - \omega}{\omega}\right) = -\frac{(\omega - 1)^2}{\omega} \leq 0. \qquad \text{(B1)}$$

Thus, either $\omega = 1$, which means that both slopes are 0, or the two slopes must be of opposite signs—one positive and one negative. QED.

Appendix C: FA data, graphed in dB

Most of the auditory literature on the TOE has reported results in dB. To facilitate comparing our results with those in the literature, we present the FA data in terms of dB in Fig. C1.

Note that Fig. C1 thoroughly masks the marked individual differences that are so apparent in Fig. 1.

Appendix D: 2IFC data

In an auditory realization of the 2IFC procedure, the TOE (Eq. 2) is studied by determining the point of subjective equality between two sequentially presented tones. Upon hearing the sequence, the respondent indicates which of the two tones was experienced as the louder one; this is the forced choice. One tone is the standard x, and the other is a tone varied by the software in response to the respondent's previous choice. In one condition, x is presented first, followed by $\mathbf{z}_2$, where the subscript indicates the position in the sequence of the varied tones; in the reverse order, the varied tone is $\mathbf{z}_1$. Each pair x and $\mathbf{z}_m$, $m = 1, 2$, is a condition. A predetermined number of trials, n, of each condition are presented within a session in a staircase fashion in which the intensity of $\mathbf{z}_m$ is varied according to a simple up–down method (Levitt, 1971). Typically, two or more staircases are intermingled within a session, so respondents generally are not aware of the condition to which any given trial belongs. The latter observation is

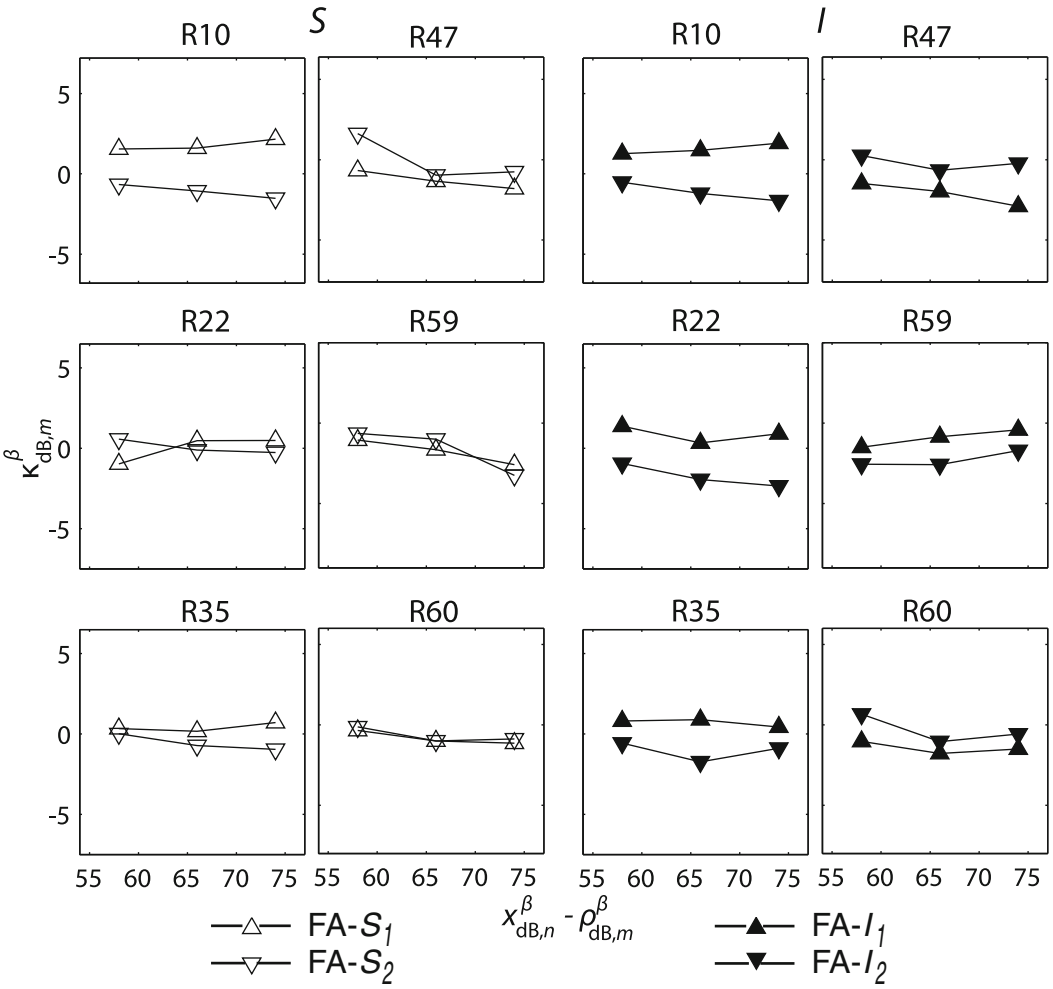

Fig. C1 Results for six respondents in the FA-*S* and FA-*I* conditions. Graphed is the NEM (in dB) as a function of the intensity range (also in dB). The data are graphed in intensity units, where the ranges of the abscissa and ordinate are the same for all of the individual graphs. The open symbols indicate the separated condition, and the filled symbols indicate the interleaved condition. Upward-pointing triangles indicate the results for $\bar{\mathbf{z}}_1$, and downward-pointing triangles indicate the results for $\bar{\mathbf{z}}_2$

of particular importance, because it means that no separate predictions can be made for $\mathbf{z}_1$ and $\mathbf{z}_2$, as is the case for the FA method (see Propositions 1 and 2). The consequence is that the 2IFC procedure is not suited to studying the NEM phenomenon. Initially, we too did not realize the existence of the NEM, and therefore also did not understand this limitation of the 2IFC. Hence, we did collect a substantial amount of data on the TOE using this method. These data are reported here.

Method

The stimulus and procedural aspects of this experiment are described in the text of the main experiment.

1. A staircase consisted of 65 trials for a given pair x and $\mathbf{z}_m$. The intensity of $\mathbf{z}_m$ was varied in 1-dB steps, according to the rules of the up–down method (Levitt, 1971). For instance, for a $\mathbf{z}_1$ trial in which the respondent indicated the first tone, $\mathbf{z}_1$, to be louder than the second tone, x, the intensity of $\mathbf{z}_1$ was decreased by 1 dB. This change would occur the next time that the software picked a trial from this staircase to be presented.
2. Two ranges for the standard were used.
   - a broad, "b," range: $x_1 = 58$, $x_2 = 64$, $x_3 = 70$, $x_4 = 76$, $x_5 = 82$ dB.
   - a narrow, "n," range: $x_1 = 70$, $x_2 = 73$, $x_3 = 76$, $x_4 = 79$, $x_5 = 82$ dB, which is the broad range clipped below 70 dB.

These range conditions were run in separate sessions.

3. Two interleaving conditions (see the Experiment) were used.
   - In the *S* condition, trials from the two staircases with standards of the same intensity were interleaved

within a block. These blocks were ordered within a session from the lowest intensity of the standard to the highest.

- In the *I* condition, all ten staircases (two stimulus orders, five standards) were run interleaved within a session.

4. A PSE estimate, $\bar{z}_m$, for each staircase was taken as the average of the $n = 65$ trials in a staircase, excluding the first 15 trials. The TOE was calculated as in Eq. 2.

This means that there were four session types: two for each range condition by two for each interleaving condition. These four session types are labeled 2IFC-I_b, 2IFC-I_n, 2IFC-S_b, and 2IFC-S_n, where the "b" and "n" indicate the range conditions.

Results

Data from the six respondents are reported in Fig. D1.

R10 continued to exhibit a fairly consistent pattern, independent of the procedure: The TOE is large as compared to the other five respondents, and R10 is the only respondent with a clearly smaller TOE for the narrow than for the broad condition. Three respondents (R10, R22, and R35) show a positive ($\tau_{dB} > 0$) TOE, whereas two (R47 and R60) show a negative TOE; the remaining respondent, R59, exhibits a very irregular pattern of responses.

In contrast to the individual data, the averaged data exhibit a quite regular positive TOE, which is consistent with data in the literature.

Lu, Williamson, and Kaufman (1992) hypothesized that respondents' reference gravitated to the central tendency of the intensity environment of the experiment. They collected data using a 2IFC procedure much like ours. A direct prediction from this hypothesis is that the maximal magnitude of the TOE should diminish as the stimulus range becomes smaller. The 2IFC-*S* condition places a respondent in a nearly constant stimulus range environment for an extended period of time. Therefore, the TOE should be substantially diminished in this condition. Again, for R10, this prediction holds and, although the slopes for 2IFC-*S* are smaller than those for 2IFC-*I*, the effect is neither consistent nor large. Another prediction is that the narrower range should simultaneously show a smaller TOE as well as a

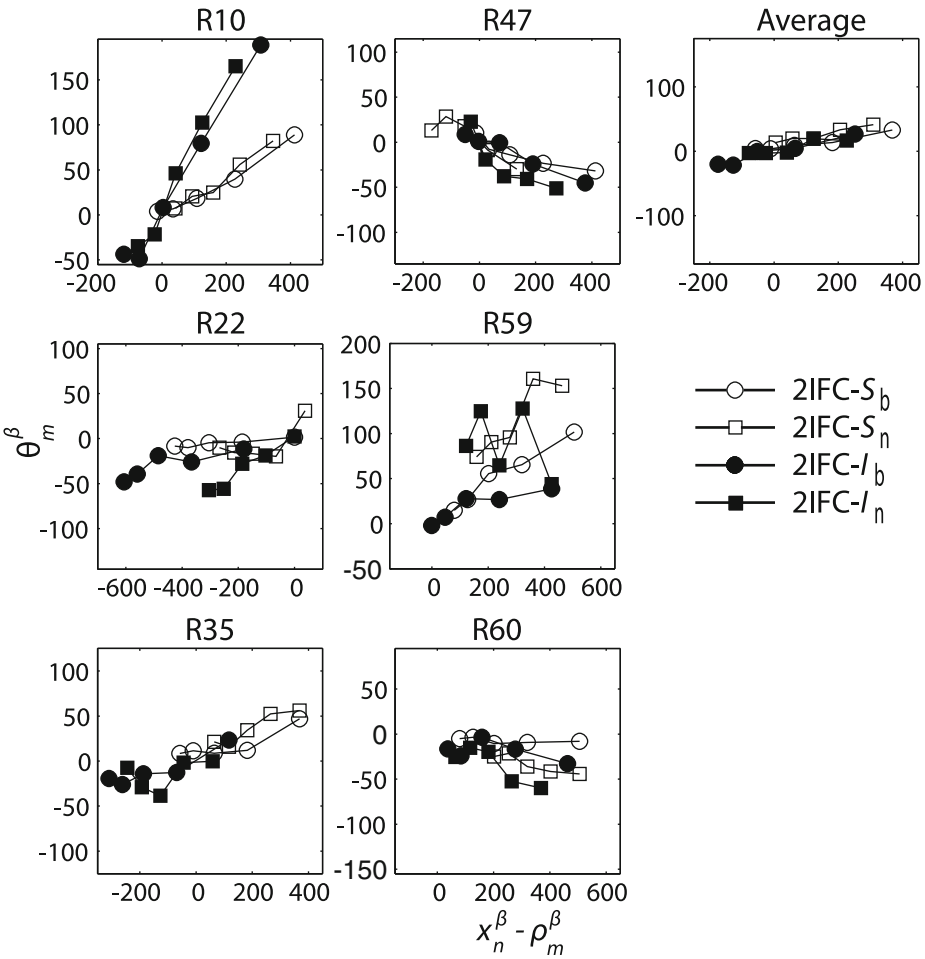

Fig. D1 Results for the six respondents and their averaged data in the 2IFC procedure. Graphed is the TOE (in intensity) as a function of the intensity range minus the reference point for each condition. The data are graphed in intensity units, where the ranges the abscissa and ordinate are the same for all of the individual graphs. The open symbols indicate the separated condition, and the filled symbols indicate the interleaved condition. Circles show the data for the broad range condition, and squares the data for the narrow range condition

shift to a higher reference point. Again, the data are not broadly consistent with this prediction.

Discussion of the 2IFC data

Several things are notable. Studying the TOE using the 2IFC method means that the more general NEM phenomenon is not reported, and therefore the fact that the respondents fell into two different types under NEM—see the main Experiment—also escapes notice.

Yet several notable observations can be made. From the averaged data, it is clear that our TOE data are very much in line with those reported in the literature (e.g., Hellström, 1985), so there is nothing clearly unusual about our data. However, the data for individual respondents demonstrate clearly that studying only their averaged data misses several notable facts. Among those are that the TOE is neither negative nor positive for individuals, that effects of the range of stimuli (context) can vary among individuals, and that the magnitude of the TOE is highly variable as well. The clear conclusion is that averaging the data from individual respondents is quite misleading, in that important information is lost.

Motivated by Propositions 1 and 2, we suspected that the reference point ρ is highly impacted by context and is more variable as the loudness context is made more variable. This context is more stable in the 2IFC-*S* condition than in the 2IFC-*I* condition, and also with the narrower than with the broader range condition. This expectation was borne out by the data, thus confirming this aspect of the data from the FA method.

The facts that the magnitude of the TOE, expressed in dB, reached ~6 dB and that the TOE behavior for 2IFC-*S* was radically different from the 2IFC-*I* behavior shows how dependent the value of TOE is on the experimental context. Randomization of conditions (along with other experiment-specific considerations) is commonly thought to be a way to avoid context effects in psychological research; however, our results show unequivocally that randomization of conditions is simply not a "cure" for the TOE.

A further worry is that because the TOE can be so large and so varied, depending upon the context, an experimental outcome can potentially be radically altered by the simple act of adding or removing conditions or by altering the ordering of conditions.

While we only studied the NEM for the FA task, it is easy to see that were the TOE calculated for those data, it would differ in both magnitude and nature from the TOE obtained using the 2IFC procedure. This means that the magnitude and nature of the TOE is not independent of the data collection method. This is yet another reason that experimental results need to be evaluated in light of the effect that the TOE (and the NEM) may be having on them.

References

American Psychological Association. (2002). Ethical principles of psychologists and code of conduct. *American Psychologist*, *57*, 1060–1073. doi:10.1037/0003-066X.57.12.1060

Eisler, H., Eisler, A. D., & Hellström, Å. (2008). Psychophysical issues in the study of time perception. In S. Grondin (Ed.), *Psychology of time* (pp. 75–110). Bingley, U.K.: Emerald.

Ellermeier, W., & Faulhammer, G. (2000). Empirical evaluation of axioms fundamental to Stevens's ratio-scaling approach: I. Loudness production. *Perception & Psychophysics*, *62*, 1505–1511. doi:10.3758/BF03212151

Fechner, G. T. (1966). *Elements of psychophysics* (H. E. Adler, Trans.). New York, NY: Holt, Rinehart & Winston. (Original work published 1860)

Hellström, Å. (1979). Time errors and differential sensation weighting. *Journal of Experimental Psychology: Human Perception and Performance*, *5*, 460–477.

Hellström, Å. (1985). The time-order error and its relatives: Mirrors of cognitive processes in comparing. *Psychological Bulletin*, *97*, 35–61. doi:10.1037/0033-2909.97.1.35

Hellström, Å. (2000). Sensation weighting in comparison and discrimination of heaviness. *Journal of Experimental Psychology: Human Perception and Performance*, *26*, 6–17.

Hellström, Å. (2003). Comparison is not just subtraction: Effects of time- and space-order on subjective stimulus difference. *Perception & Psychophysics*, *65*, 1161–1177. doi:10.3758/BF03194842

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America*, *49*, 467–477.

Lu, Z.-L.,Williamson, S. J., & Kaufman, L. (1992). Behavioral lifetime of human auditory sensory memory predicted by physiological measures. *Science*, *258*, 1668–1670.

Luce, R. D. (1995). Four tensions concerning mathematical modeling in psychology. *Annual Review of Psychology*, *46*, 1–26.

Luce, R. D. (2002). A psychophysical theory of intensity proportions, joint presentations, and matches. *Psychological Review*, *109*, 520–532. doi:10.1037/0033-295X.109.3.520

Luce, R. D. (2004). Symmetric and asymmetric matching of joint presentations. *Psychological Review*, *111*, 446–454. doi:10.1037/0033-295X.111.2.446

Luce, R. D. (2008). "Symmetric and asymmetric matching of joint presentations": Correction to Luce (2004). *Psychological Review*, *115*, 601. doi:10.1037/0033-295X.115.3.601

Luce, R. D. (2012). Predictions about bisymmetry and cross-modal matches from global theories of subjective intensities. *Psychological Review*, *119*, 373–387. doi:10.1037/a0027122

Michels, W. C., & Helson, H. (1954). A quantitative theory of time-order effects. *American Journal of Psychology*, *67*, 327–334.

Narens, L. (1996). A theory of ratio magnitude estimation. *Journal of Mathematical Psychology*, *40*, 109–129.

Needham, J. G. (1934a). The time error as a function of continued experimentation. *American Journal of Psychology*, *46*, 558–567.

Needham, J. G. (1934b). The time error in comparison judgments. *Psychological Bulletin*, *31*, 229–243.

Needham, J. G. (1935). The effect of the time interval upon the time-error at different intensive levels. *Journal of Experimental Psychology*, *18*, 530–543.

Ono, S. (1950). On the positive time-error in the successive comparison of brightness. *Japanese Journal of Psychology*, *20*, 6–15.

Steingrimsson, R. (2009). Evaluating a model of global psychophysical judgments for brightness: I. Behavioral properties of summations and productions. *Attention, Perception, & Psychophysics*, *71*, 1916–1930. doi:10.3758/APP.71.8.1916

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

Steingrimsson, R. (2012a). *Evaluating a model of global psychophysical judgments for perceived contrast: I. Behavioral properties of summations and productions*. Manuscript in preparation.

Steingrimsson, R. (2012b). *Evaluating a model of global psychophysical judgments for perceived contrast: II. Behavioral properties linking summations and productions*. Manuscript in preparation.

Steingrimsson, R. (2012c). *Evaluating a model of global psychophysical judgments for brightness: III. Forms for the psychophysical and the weighting function*. Manuscript in preparation.

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*, 290–307.

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*, 308–319.

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*, 15–29.

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*, 29–44.

Stevens, S. S. (1957). On the psychophysical law. *Psychological Review*, *64*, 153–181. doi:10.1037/h0046162

Stevens, S. S. (1975). *Psychophysics: Introduction to its perceptual, neural, and social prospects*. New York, NY: Wiley.

Woodworth, R. S., & Schlosberg, H. (1954). *Experimental psychology* (Revised ed.). New York: Holt, Rinehart & Winston.

Zimmer, K. (2005). Examining the validity of numerical ratios in loudness fractionation. *Perception & Psychophysics*, *67*, 569–579. doi:10.3758/BF03193515