

# Time-order errors and standard-position effects in duration discrimination: An experimental study and an analysis by the sensation-weighting model

Åke Hellström<sup>1</sup> · Thomas H. Rammsayer<sup>2</sup>

Published online: 17 June 2015  
© The Psychonomic Society, Inc. 2015

**Abstract** Studies have shown that the discriminability of successive time intervals depends on the presentation order of the standard (St) and the comparison (Co) stimuli. Also, this order affects the point of subjective equality. The first effect is here called the *standard-position effect* (SPE); the latter is known as the *time-order error*. In the present study, we investigated how these two effects vary across interval types and standard durations, using Hellström’s sensation-weighting model to describe the results and relate them to stimulus comparison mechanisms. In Experiment 1, four modes of interval presentation were used, factorially combining interval type (filled, empty) and sensory modality (auditory, visual). For each mode, two presentation orders (St–Co, Co–St) and two standard durations (100 ms, 1,000 ms) were used; half of the participants received correctness feedback, and half of them did not. The interstimulus interval was 900 ms. The SPEs were negative (i.e., a smaller difference limen for St–Co than for Co–St), except for the filled-auditory and empty-visual 100-ms standards, for which a positive effect was obtained. In Experiment 2, duration discrimination was investigated for filled auditory intervals with four standards between 100 and 1,000 ms, an interstimulus interval of 900 ms, and no feedback. Standard duration interacted with presentation order, here yielding SPEs that were negative for standards of 100 and 1,000 ms, but positive for 215 and 464 ms. Our findings indicate that the SPE can be positive as well as negative, depending on the interval type and standard duration,

reflecting the relative weighting of the stimulus information, as is described by the sensation-weighting model.

**Keywords** Duration discrimination · presentation order · Time-order error · Standard-position effect · Type-B effect

What happens when we compare two successively presented stimuli, X and Y? According to the commonly assumed difference model, the comparison response (e.g., first greater, second greater) is determined by the difference between the momentary subjective magnitudes of the stimuli (Thurstone, 1927a, b). Furthermore, these magnitudes are only dependent on the stimuli, and are independent of such factors as their presentation order (X–Y or Y–X) within the pair. Hence, no systematic under- or overestimation of the first or the second stimulus, relative to the other, should occur. The time-order error (TOE), first noted by Gustav Fechner (1860), violates this prediction but has often been explained, within the framework of the difference model, as being due to some kind of response bias (see, e.g., Alcalá-Quintana & García-Pérez, 2011). However, plenty of experimental evidence, in different sensory continua and using different paradigms, has invalidated the latter explanation of the TOE (e.g., Hellström, 1977, 1978; Jamieson & Petrusic, 1975).

Most often, participants’ abilities to tell the difference between two stimuli are studied by methods in which the stimuli are presented in succession, with one stimulus being held constant (the standard stimulus, St) and the other (the comparison stimulus, Co) varying around it. Discrimination sensitivity is then usually measured by the difference limen (DL), often defined as one-half of the change in either stimulus that is needed to change the response distribution from, for instance, 75% “first greater” to 75% “second greater.” The difference model predicts that discrimination sensitivity, as measured by

✉ Åke Hellström  
hellst@psychology.su.se

<sup>1</sup> Department of Psychology, Stockholm University, SE-106 91 Stockholm, Sweden

<sup>2</sup> University of Bern, Bern, Switzerland

the DL, should be independent of whether the first stimulus is fixed and the second varied (St–Co) or vice versa (Co–St). This prediction is contradicted by what will here be called the *standard-position effect* (SPE): DLs, as well as the percentages of correct responses, typically differ systematically between the orders St–Co and Co–St. This effect was noted for lifted weights by Martin and Müller (1890), who found that more correct answers were given when the order was St–Co than when it was Co–St. The effect was also noted for durations by Rammesayer and Wittkowski (1990), who named it the “constant-position effect.” Grondin and McAuley (2009) and Ulrich and Vorberg (2009) likewise found this effect for duration comparison. Ulrich and Vorberg called it the “Type-B effect” to distinguish it from the TOE, which they called the “Type-A effect.”

The essence of the SPE is that changes in the first- and in the second-presented stimulus do not exert equal impacts on the outcome of the stimulus comparison. Dyjas and Ulrich (2014) defined this effect to be negative in the common case in which the best discrimination is obtained in the order St–Co, so that the first stimulus has less impact on the response than the second. This definition is in analogy with the standard definition of the TOE (Fechner, 1860), according to which the TOE is negative when the first stimulus is perceived as having a lower magnitude than the physically identical second stimulus. The unequal impacts of the compared stimuli can also be noted in experimental designs in which none of the stimuli is held constant: Hellström (1979) found, in comparisons of tone loudness, an effect equivalent to the SPE, which varied in its sign and magnitude across temporal conditions (stimulus duration and interstimulus interval). For stimulus durations and interstimulus intervals of at least 1,000 ms, the impact of the second stimulus on the response (first louder, equal, or second louder) was larger than that of the first, whereas for brief stimulus durations and interstimulus intervals (tone durations of 100–200 ms and interstimulus intervals of 100–700 ms), the first stimulus had a much greater impact than the second. The impact of each stimulus was quantified as a weight coefficient, described below, and (although this was not done in Hellström, 1979) each weight can be translated into a measure of the DL, which is inversely related to the weight.

The existence of the SPE and the TOE leads to the inevitable conclusion that the difference model, which models the subjective difference between two compared stimuli as a simple difference between their subjective stimulus magnitudes—possibly with an added response bias—is unrealistic as a model of stimulus comparison. Still, there is no need to abandon the general notion behind the difference model: that stimulus comparison is based on the perceived difference between monotonic functions of the magnitudes of the two stimuli, and that the buildup of this difference is what the theorist should attempt to model in attempting to understand the comparison process. One useful and straightforward modeling

approach is analogous to a standard model for data analysis in general—namely, linear regression. With this approach, the difference model is replaced by a model analogous to linear regression, with possibly different weight coefficients for the two subjective stimulus magnitudes. One model of this kind, based on Harry Helson’s adaptation-level theory, was introduced by Michels and Helson (1954; see also Helson 1964). A generalized version of this model was successfully fitted to the results from experiments with comparisons of duration (Hellström, 1977) and loudness (Hellström, 1978). In data from loudness comparison in a set of widely varying temporal conditions, as we mentioned above, Hellström (1979) noted a strong relationship, across conditions, between the intercept and the two weight coefficients. The analysis of this relationship led to further development of the stimulus comparison model into the *sensation-weighting model* (Hellström, 1979, 1985, 2000, 2003), which has been successfully used in many studies, and which is able to account for both the TOE and the SPE.

The basic form of the sensation-weighting model is

$$d_{12} = [s_1\psi_1 + (1-s_1)\psi_{r1}] - [s_2\psi_2 + (1-s_2)\psi_{r2}] + b, \quad (1)$$

where  $d_{12}$  is the scaled subjective difference between the first and second stimuli.  $\psi_1$  and  $\psi_2$  are the sensation magnitudes of the stimuli,  $s_1$  and  $s_2$  are weighting coefficients, and  $\psi_{r1}$  and  $\psi_{r2}$  are the subjective magnitudes that correspond to the current reference levels for the respective stimuli. The reference level is formed by the pooling of stimulus information: It is normally near the center of the stimulus range, and may differ somewhat between the two stimuli; see Hellström (1979).  $b$  is a constant term, which captures a possible contribution to  $d_{12}$  that is independent of the weighting mechanism—for instance, a response bias. Notably, the sensation-weighting model thereby includes the difference model with and without bias as special cases.

The sensation-weighting model implies that the SPE and the TOE are two sides of the same coin. With  $s_1 < s_2$ , not only is the SPE negative, but in suitably designed experiments (varying both stimuli or using at least two standards), the familiar inverse relation between the TOE and stimulus magnitude (see, e.g., Needham, 1935) is observed, yielding TOEs that are positive below and negative above the reference levels. With the less common relation  $s_1 > s_2$ , this relation is reversed, and the SPE also reverses and becomes positive. A mean TOE different from zero may arise due to, for instance, extraneous stimulation that enters into the reference levels.

Temporal intervals have often been used in research on stimulus judgment, comparison, and discrimination (see Eisler, Eisler, & Hellström, 2008), because timing and time perception are so important in everyday life, and because their shortcomings, in the forms of failures to keep time and gross misperceptions of the duration of activities and events, are so

obvious. Unlike with other sensory continua, there is no sense organ for time, and the processes underlying time perception remain largely a scientific challenge. The processing of short (in the range of milliseconds) and longer temporal intervals is likely governed at least partly by different mechanisms (see, e.g., Lewis & Miall, 2003a, b; Rammsayer, 1999, 2008; Rammsayer & Troche, 2014): For longer intervals, temporal processing demands cognitive resources, whereas for brief intervals processing is highly sensory in nature and beyond cognitive control. For example, when using a secondary task of a nontemporal and cognitive nature (Rammsayer & Lima, 1991; Rammsayer & Ulrich, 2011), the temporal discrimination of intervals in the 1-s range was markedly impaired, whereas for intervals from 50 to 100 ms, discrimination was not affected (see Mattes & Ulrich, 1998). Also, neuropharmacological studies have suggested temporal processing by a subcortical automatic system controlled by mesostriatal dopaminergic activity for intervals in the range of milliseconds, and by a prefrontal cognitive system for intervals in the 1-s range (for a review, see Rammsayer, 2008).

Karl von Vierordt (1868) performed basic studies and discovered systematic errors in the reproduction of time intervals. According to Vierordt's law, short intervals are overreproduced, and long intervals are underreproduced. Also, Fechner and his followers noted the idiosyncrasies of temporal discrimination, such as possible violations of Weber's law (see Eisler et al. 2008). In more recent research, it has been found that TOEs occur with time intervals as well as with other kinds of stimuli (e.g., Allan, 1977; Hellström, 1977, 2003). Temporal intervals differ from other stimuli in that their (temporal) magnitude is the same as their duration, so that duration and magnitude, two important factors behind TOEs, are inseparable. Furthermore, two consecutive time intervals can be perceptually concatenated, enabling alternative strategies in the comparison of a pair of successive durations, as is addressed by Eisler (1975, 1981); Eisler et al. (2008) parallel-clock model, according to which what is compared is not the two durations, but the second duration and one-half of the total duration.

A time interval can be represented and demarcated using stimuli of different modalities (e.g., auditory, visual, or tactile) and either can be filled with sensory content or can be empty. One effect of the modality factor is well-known—namely, worse discrimination for filled visual than for filled auditory intervals (Eisler et al. 2008; Grondin & McAuley, 2009; Hellström, 2003; Rammsayer, 2014).

Rammsayer (2014) studied Weber fractions (WFs) as a function of sensory modality (auditory, visual), standard duration, type of interval (filled, empty), and type of task (2AFC, reminder [explained below]). The purpose of the present Experiment 1 was to follow up this study—which was not designed to study the SPE and the TOE—by assessing the SPE—that is, discrimination performance as a function of the presentation order of the standard and comparison

stimuli—as well as the TOEs for four different types of temporal stimuli (intervals), and two interval durations. Another particularly important background study was the one by Hellström and Rammsayer (2004), in which discrimination performance was studied for filled auditory (white-noise) intervals with standard durations of 50 and 1,000 ms. In that study, it was found that with the 1,000-ms standard, discrimination was better (i.e., WFs were smaller) for the presentation order St–Co than for Co–St, whereas the reverse relation was true with the 50-ms standard. This is an effect of the standard's position in the pair—that is, an SPE—and, according to the definition of Dyjas and Ulrich (2014), it was negative with the 1,000-ms standard, but positive with the 50-ms standard.

According to Jamieson and Petrusic (1976, 1978) the practice of providing correctness feedback leads to response bias that may be one factor behind the TOE. Hellström and Rammsayer excluded the feedback in some conditions and found that the main effect of this was to greatly impair discrimination for some participants. It was considered of importance to further investigate the possible impact of feedback by making its presence or absence an additional factor.

Negative SPEs were highlighted by Dyjas and Ulrich (2014), whose internal reference model for stimulus comparison contends that the second stimulus is compared with an internal reference, successively built up as a weighted average (weights  $g$  and  $1 - g$ ,  $0 < g < 1$ ) of the current stimulus and the previous value of the reference. The modeling of Dyjas and Ulrich is based on methods in which one stimulus, the first or the second, is always at the standard level and the other varied around it. For this kind of situation, the internal reference model predicts a negative SPE and would have an explanation neither for a positive SPE nor for any TOE.

One problem seems to be that most researchers concentrate, be it by convenience or by tradition, on a few experimental conditions (in particular by avoiding brief interstimulus intervals) in which the internal reference model and similar models are usually able to describe the results. One exception is Hellström and Rammsayer (2004), who, using a 50-ms standard and interstimulus intervals of 100 through 2,700 ms, obtained consistently positive SPEs. Another exception is the recent study by Bausenhardt, Dyjas, and Ulrich (2015), using an adaptive method, in which negative SPEs were found for filled auditory intervals with 100-ms and 1,000-ms standards and an interstimulus interval of 1,000 ms. With a 100-ms standard, the SPE vanished when the interstimulus interval was shortened to 300 ms, but it did not reverse into positive except for a few participants.

A model related to the internal reference model is that of Raviv, Ahissar, and Loewenstein (2012) who suggested, on the basis of the observer employing Bayesian inference, that a heuristic is used in which “decision in the two tone [frequency] discrimination task is based on a comparison between the

second tone and an exponentially decaying average of the first tone and past tones” (p. 1). With this kind of model, the authors account for the “contraction bias” in their experiment—that is, TOEs that are positive for low, and negative for high, magnitudes. (In their experiment, unlike those of Bausenhardt et al. 2015, both stimuli were varied, which made the TOEs visible.) This explanation may be plausible in most experimental conditions, but not in those of Hellström (1979, 2003) with brief, but by no means extreme, stimulus durations and interstimulus intervals. In such conditions, a straightforward application of Bayesian inference regarding the first-presented stimulus does not seem fruitful. The sensation-weighting model, built on the conception of the comparison process as “a two-way, interactive process that involves the memory representations of *both* stimuli” (Hellström, 1979, p. 469) was instead developed to also account for this kind of results. This development was based on a wide range of conditions, and with a design more suitable for quantitative modeling than are designs in which one of the stimuli is always at the standard level.

SPEs are known from earlier research: The psychometric functions and DLs for the orders St–Co and Co–St often differ, as measured with adaptive as well as with nonadaptive psychophysical methods. Also, as was shown by, for instance, Rammsayer (2014), using a random order of standard and comparison stimuli (“2AFC task”) yields increased DLs as compared to using only the St–Co order (“reminder task”). Ulrich and Vorberg (2009) demonstrated that measures of the DL obtained with a random order become biased in the presence of TOEs and SPEs; they therefore strongly recommended determining DLs separately for the two stimulus orders. That estimates of the DL can differ considerably between the two orders was also demonstrated by Hellström (2000) and by Hellström and Rammsayer (2004) (see Eisler et al. 2008).

Against this background, we considered it important to further explore the comparison and discrimination of time intervals, and in particular how the SPEs might depend on duration range (around 100 ms and around 1,000 ms), sensory modality (auditory, visual), and interval type (filled, empty). Of particular theoretical importance was attempting to confirm the previous (Hellström & Rammsayer, 2004) observations of positive SPEs as well as of TOEs, because these are excluded by both the internal reference model and the difference model.

As in the study by Hellström and Rammsayer (2004), a quantitative modeling approach was used, based on the sensation-weighting model. This approach by no means excluded the possibility of ending up by selecting the difference model (with or without bias), the internal reference model, or some variant of the latter, including the Michels–Helson model. All of these models can be seen as special cases of the more general sensation-weighting model.

## Experiment 1

In this experiment, auditory as well as visual interval durations were compared using an adaptive psychophysical method. Standard durations of 100 and 1,000 ms were used, with an interstimulus interval of 900 ms. The between-groups factors were Interval Type (filled, empty) and Correctness Feedback (with, without). The within-participants (blocked) factors were Interval Modality (auditory, visual), Standard Duration (100 ms, 1,000 ms), and the Magnitude Relation of the standard and comparison stimuli, as is described under Procedure.

## Method

### Participants

The participants were 21 male and 91 female undergraduate psychology students at the University of Bern, ranging in age from 18 to 47 years ( $M = 22.8$ ,  $SD = 4.59$ ). They received course credit for participating in this study. All participants were naïve as to the purpose of this study and reported normal hearing and normal or corrected-to-normal vision. The study was approved by the local ethics committee.

### Apparatus and stimuli

The time intervals were marked by auditory or visual stimuli. The presentation of the intervals and the recording of participants’ responses were computer controlled. The timing accuracy of stimulus presentation was better than  $\pm 1$  ms. The filled-auditory stimuli were white-noise bursts presented binaurally through headphones (Sony CD 450) at an intensity of 66 dBA. The empty-auditory stimuli were bounded by 3-ms white-noise bursts at an intensity of 88 dBA. These different levels of intensity were chosen to achieve equal loudnesses in the two conditions, on the basis of the results of a pilot experiment in which 12 participants were asked to adjust the loudness of a 3-ms white-noise burst until it matched that of a 100-ms burst. Visual stimuli were generated by a red light-emitting diode (LED) (diameter  $0.38^\circ$ , viewing distance 60 cm, luminance  $68 \text{ cd/m}^2$ ) positioned at the eye level of the participant. The intensity of the LED was clearly above threshold, but not dazzling. Empty visual intervals were marked by 3-ms LED flashes.

### Procedure

Using a weighted up–down procedure (Kaernbach, 1991; see Hellström & Rammsayer, 2004), the participants compared the durations of two successive intervals. The between-groups factors were Interval Type (filled, empty) and Feedback (with, without); the within-participants factors were Modality (auditory, visual) and Standard Duration (100 ms, 1,

000 ms). Twenty-eight participants were randomly assigned to each of the four experimental conditions—filled–feedback, filled–no feedback, empty–feedback, and empty–no feedback. Each participant participated in one session, consisting of eight blocks with a 1-min break in between. A block was composed of 64 trials; four consecutive blocks were presented with auditory, and four with visual, stimuli. Half of the participants started with the four auditory blocks, and the other half with the four visual blocks. The auditory and visual parts of the session were preceded by six practice trials, in order to ensure that the participants understood the instructions and to make them familiar with the stimuli.

In two blocks for each sensory modality, the standard was 100 ms, and in two, 1,000 ms. For each combination of modality and standard, in one block the comparison stimulus was longer than the standard (L-block), and in the other it was shorter (S-block). The order of these block types was balanced across participants. Presentation orders of the standard and comparison stimuli were randomized and balanced within each block; each order was used in half of the trials. Using a notation in which the order of standard (St) and comparison (Co) stimuli indicates their presentation order, and “<” and “>” indicate their magnitude relation, an L-block was, equally often,  $St < Co$  or  $Co > St$ , and an S-block was, equally often,  $St > Co$  or  $Co < St$ .

For empty auditory intervals, only the 100-ms standard was used. This was done because in a pilot study, many participants had difficulties with empty auditory intervals in the 1,000-ms range. They reported primarily perceiving a sequence of four auditory clicks rather than two empty intervals separated by an interstimulus interval. This was probably due to confusion of the 1,000-ms stimulus interval with the 900-ms interstimulus interval (cf. Grondin, Meilleur-Wells, Ouellette, & Macar, 1998), transforming the task into one of rhythm perception.

The participant was seated at a table with a keyboard and a computer monitor in a sound-attenuated and dimly lit room. To initiate the first trial, the participant pressed the space bar; the first stimulus was then presented after 900 ms. After the second stimulus, the participant responded by pressing one of two designated keys on a computer keyboard, labeled “first interval longer” and “second interval longer.” Accuracy, not speed, was emphasized in the instructions. In the feedback conditions, “+” (correct) or “-” (false) was displayed on the monitor screen for 1,500 ms after the response. The next trial started 900 ms after presentation of the feedback, and 900 ms after the participant’s response in the no-feedback conditions.

When the standard was 100 (or 1,000)ms, the initial duration of the comparison stimulus was 30 (or 400)ms below or above that of the standard. The comparison stimulus was then changed using an adaptive rule, the weighted up–down method (Kaernbach, 1991), to estimate the upper and lower 75% difference thresholds. When the comparison stimulus

was shorter than the standard, it was increased by 6 (or 25) ms when it was judged to be shorter than the standard, and decreased by 18 (or 75)ms when it was judged to be longer than the standard; its longest possible value was 1 ms below the standard. When the comparison stimulus was longer than the standard, it was decreased by 6 (or 25)ms after it was judged to be longer than the standard, and increased by 18 (or 75)ms after it was judged to be shorter than the standard; its shortest possible value was 1 ms above the standard.

#### Data analysis

**Discrimination performance** With the weighted up–down procedure of Kaernbach (1991), using the ratio 3:1 between the size of an upward and a downward step, performance settles at 75% correct responses. [This happens because, given the current value of the comparison stimulus,  $Co_i$ , its predicted value at the next trial,  $Co_{i+1}$ , is equal to  $p(Co_i - x) + (1 - p)(Co_i + 3x) = Co_i + 3x(.75 - p)$ , where  $x$  and  $3x$  are the step sizes, and  $p$  is the probability of a correct response.] The corresponding duration difference ( $D_{75}$ ) between the comparison stimulus and the standard was computed, for each block and presentation order, as the mean absolute difference in milliseconds (i.e., the mean of the upper and lower 75% thresholds) between the comparison stimulus and the standard across the last 20 trials. To measure discrimination uncontaminated by order effects, each participant’s arithmetic mean of  $D_{75}$  for  $St < Co$  and  $St > Co$ , and for  $Co < St$  and  $Co > St$ , were then computed. These averages indicate the DLs,  $DL_{St-Co}$  and  $DL_{Co-St}$ .

**Modeling approach** To model the participants’ comparison behavior, Hellström’s (1979, 1985, 2000, 2003) sensation-weighting model was applied and adapted to the experimental situation. It was assumed that the psychophysical function was linear,  $\psi = \phi$ , over the range of comparison stimuli for each standard duration (no assumption was made concerning the shape of the psychophysical function across standard durations). It was also assumed that for each block type the two reference levels were equal,  $\phi_{r1} = \phi_{r2} = \phi_r$  (see Hellström, 2000), but that  $\phi_r$  might differ between L- and S-blocks (where in L-blocks, the mean duration of the comparison stimulus was about two DL units longer than in S-blocks, and thereby—since the standard was presented in each trial—the mean stimulus duration was about one DL unit longer). This yielded

$$d_{12} = s_1\phi_1 - s_2\phi_2 + (s_2 - s_1)\phi_{rT} + b, \quad (2)$$

where  $T = L$  or  $S$ , according to block type.

It was assumed that each measured DL corresponds to the same absolute value in the scale of the subjective duration difference  $d_{12}$ , and thereby, according to our assumption  $\psi = \phi$ ,

also in the scale of the equivalent physical duration difference between the comparison stimulus and the standard. This physical difference,  $\Delta\phi$ , then represents the “true” DL, unaffected by order effects. It can be expressed in terms of Eq. 2—expressing each comparison stimulus as a function of the standard and the pertinent DL—as follows:

$$L\text{-block}, St < Co : -\Delta\phi \tag{3}$$

$$= s_1St - s_2(St + DL_{St < Co}) + (s_2 - s_1)\phi_{rL} + b,$$

$$L\text{-block}, Co > St : \Delta\phi \tag{4}$$

$$= s_1(St + DL_{Co > St}) - s_2St + (s_2 - s_1)\phi_{rL} + b,$$

$$S\text{-block}, St > Co : \Delta\phi \tag{5}$$

$$= s_1St - s_2(St - DL_{St > Co}) + (s_2 - s_1)\phi_{rS} + b,$$

$$S\text{-block}, Co < St : -\Delta\phi \tag{6}$$

$$= s_1(St - DL_{Co < St}) - s_2St + (s_2 - s_1)\phi_{rS} + b.$$

Setting  $s_1 = s_2 = 1$  shows that in the absence of weighting and bias, each DL would be equal to  $\Delta\phi$ .

**Difference limens** For each participant, the DLs for the orders standard–comparison (St–Co) and comparison–standard (Co–St) were calculated as follows:

$$DL_{St-Co} = (DL_{St < Co} + DL_{St > Co})/2, \tag{7}$$

$$DL_{Co-St} = (DL_{Co < St} + DL_{Co > St})/2. \tag{8}$$

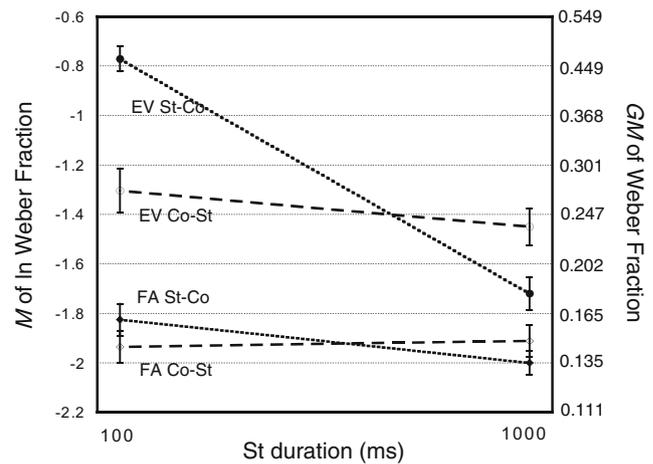
Due to the positive skewness of the DL distributions across participants, and for comparability of the effects between the two standard durations, the Weber fraction (WF) was computed as  $DL/St$ , and  $\ln WF$  was used to assess discriminability. To assess differential discriminability in the two orders, the SPE was quantified as  $100 \cdot \ln(DL_{St-Co}/DL_{Co-St})$ , which can be interpreted as a symmetric percent difference (Graff, 2014). This measure assigns the correct sign to the SPE.

**Results<sup>1</sup>**

*Natural log of the Weber fraction (lnWF)*

The means of  $\ln WF$  are shown in Figs. 1 and 2. Due to the missing values for the 1,000-ms standard in the empty-auditory condition, an overarching analysis of variance

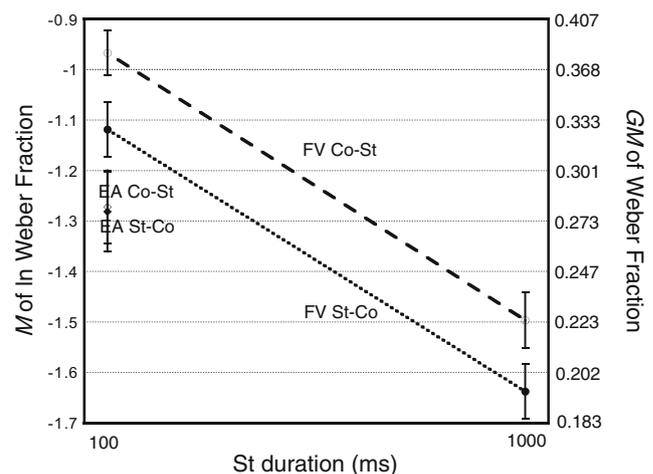
<sup>1</sup> In the Results section, the abbreviations St and Co are used for the standard and comparison stimuli, respectively.



**Fig. 1** Experiment 1. Empty visual and filled auditory intervals: Mean of natural log Weber fraction versus standard duration. Error bars indicate standard errors of the means. The right axis shows the corresponding geometric means of the Weber fractions. The plots illustrate positive standard-position effects (SPEs) for the 100-ms standard duration and negative SPEs for the 1,000-ms standard duration

(ANOVA) for all conditions could not be conducted, nor an ANOVA for all auditory conditions. For the visual conditions, the values of  $\ln WF$  for each type of interval (filled, empty) were submitted to a mixed ANOVA with Standard Duration (100 ms, 1,000 ms) and Order (St–Co, Co–St) as within-participants factors and Feedback (with, without) as a between-participants factor.

For the filled auditory intervals (see Fig. 1), the effect of feedback was significant,  $F(1, 54) = 5.355, p = .025, \eta_p^2 = .090$ . With feedback,  $\ln WF$ s were lower ( $M = -2.022, SE_M = 0.063$ ) than without feedback ( $M = -1.815, SE_M = 0.064$ ), corresponding to a DL reduction by 18.7%. The Duration  $\times$  Order interaction was also



**Fig. 2** Experiment 1. Filled visual and empty auditory intervals: Mean of natural log Weber fraction versus standard duration. Error bars indicate standard errors of the means. The right axis shows the corresponding geometric means of the Weber fractions. The plots illustrate negative standard-position effects (SPEs)

significant,  $F(1, 54) = 7.515, p = .008, \eta^2_p = .122$ , but not the other interactions, nor the main effects of duration and order. Separate analyses with mixed ANOVAs (with the factors Order and Feedback) were conducted for  $St = 100$  ms and  $St = 1,000$  ms. For  $St = 100$  ms, the effect of order was significant,  $F(1, 54) = 5.101, p = .028, \eta^2_p = .086$ : lnWFs were higher for St–Co ( $M = -1.826, SE_M = 0.060$ ) than for Co–St ( $M = -1.935, SE_M = 0.065$ ), yielding a DL ratio of 1.115 and an SPE of + 10.91% ( $SE = 4.83\%$ ). For  $St = 1,000$  ms, the effect of order was not significant,  $p = .167$ .

For the *empty* auditory intervals (see Fig. 2), only a 100-ms standard was used. The lnWFs were submitted to a mixed ANOVA, with Order (St–Co, Co–St) as a within-participants factor and Feedback (with, without) as a between-participants factor. No significant effects were found (all  $ps > .05$ ).

For the filled visual intervals (see Fig. 2), the effect of feedback was significant,  $F(1, 54) = 9.443, p = .003, \eta^2_p = .149$ . With feedback, lnWFs were lower ( $M = -1.416, SE_M = 0.051$ ) than without feedback ( $M = -1.195, SE_M = 0.051$ ), corresponding to a DL reduction of 19.8%. We observed a significant effect of duration,  $F(1, 54) = 147.929, p < .001, \eta^2_p = .733$ . For  $St = 100$  ms, lnWFs were higher ( $M = -1.043, SE_M = 0.041$ ) than for  $St = 1,000$  ms ( $M = -1.567, SE_M = 0.043$ ), corresponding to a WF ratio of 1.69. We also found a significant effect of order,  $F(1, 54) = 10.651, p = .002, \eta^2_p = .165$ . For St–Co, lnWFs were lower ( $M = -1.378, SE_M = 0.043$ ) than for Co–St ( $M = -1.232, SE_M = 0.042$ ). None of the interactions were significant.

For the empty visual intervals (see Fig. 1), the effect of feedback was nonsignificant. A significant effect of duration emerged,  $F(1, 54) = 93.375, p < .001, \eta^2_p = .634$ : For  $St = 100$  ms, lnWFs were higher ( $M = -1.038, SE_M = 0.057$ ) than for  $St = 1,000$  ms ( $M = -1.585, SE_M = 0.064$ ). There was also a significant effect of order,  $F(1, 54) = 6.163, p = .016, \eta^2_p = .102$ , but the interactions were nonsignificant, except for Duration  $\times$  Order, which was significant,  $F(1, 54) = 54.306, p < .001, \eta^2_p = .501$ . Separate analyses with mixed ANOVAs (with the factors Order and Feedback) were conducted for  $St = 100$  ms and  $St = 1,000$  ms. For  $St = 100$  ms, the effect of order was significant,  $F(1, 54) = 34.818, p < .001, \eta^2_p = .392$ . The lnWFs were higher for St–Co ( $M = -0.772, SE_M = 0.050$ ) than for Co–St ( $M = -1.304, SE_M = 0.089$ ), yielding a DL ratio of 1.703 and an SPE of + 53.24% ( $SE_M = 9.02\%$ ). For  $St = 1,000$  ms, the effect of order was likewise significant,  $F(1, 54) = 21.653, p < .001, \eta^2_p = .286$ . The lnWFs were lower for St–Co ( $M = -1.720, SE_M = 0.067$ ) than for Co–St ( $M = -1.450, SE_M = 0.074$ ), yielding a DL ratio of 0.763 and an SPE of – 27.01% ( $SE = 5.81\%$ ).

For all of the *auditory* conditions with standard duration 100 ms, the lnWF values were submitted to a mixed

ANOVA with Order (St–Co, Co–St) as a within-participants factor and Interval Type (filled, empty) and Feedback (with, without) as between-participants factors. The significant effects were those of order,  $F(1, 220) = 11.691, p < .001, \eta^2_p = .050$ ; interval type,  $F(1, 220) = 18.684, p < .001, \eta^2_p = .078$ ; feedback,  $F(1, 220) = 6.375, p = .012, \eta^2_p = .028$ ; and the Order  $\times$  Interval Type interaction,  $F(1, 220) = 16.209, p < .001, \eta^2_p = .069$ .

For all of the *visual* conditions, the lnWF values were submitted to a mixed ANOVA with Standard Duration (100 ms, 1,000 ms) and Order (St–Co, Co–St) as within-participants factors and Interval Type (filled, empty) and Feedback (with, without) as between-participants factors. The significant effects were those of feedback,  $F(1, 108) = 8.966, p < .003, \eta^2_p = .077$ ; and standard duration,  $F(1, 108) = 226.722, p < .001, \eta^2_p = .677$ ; the significant interactions were Order  $\times$  Interval Type,  $F(1, 108) = 16.037, p < .001, \eta^2_p = .129$ ; Standard Duration  $\times$  Order,  $F(1, 108) = 33.564, p < .001, \eta^2_p = .237$ ; and Standard Duration  $\times$  Order  $\times$  Interval Type,  $F(1, 108) = 35.408, p < .001, \eta^2_p = .247$ . The latter interaction effect confirmed the difference between filled and empty visual intervals regarding the effect of standard duration on the SPE.

For all of the *filled* conditions, the lnWF values were submitted to a mixed ANOVA with Modality (auditory, visual), Standard Duration (100 ms, 1,000 ms), and Order (St–Co, Co–St) as within-participants factors and Feedback (with, without) as a between-participants factor. The significant effects were those of feedback,  $F(1, 54) = 8.321, p = .006, \eta^2_p = .135$ ; modality,  $F(1, 54) = 348.352, p < .001, \eta^2_p = .865$ ; duration,  $F(1, 54) = 56.630, p < .001, \eta^2_p = .512$ ; and order,  $F(1, 54) = 4.279, p = .0043, \eta^2_p = .073$ ; along with the Modality  $\times$  Duration,  $F(1, 54) = 64.467, p < .001, \eta^2_p = .544$ ; and Modality  $\times$  Order,  $F(1, 54) = 7.197, p = .0097, \eta^2_p = .118$ , interactions. The Modality  $\times$  Duration  $\times$  Order interaction did not reach significance ( $p = .088$ ).

For all of the conditions with  $St = 100$  ms, the lnWF values were submitted to a mixed ANOVA with Modality (auditory, visual) and Order (St–Co, Co–St) as within-participants factors and Interval Type (filled, empty) and Feedback (with, without) as between-participants factors. The significant effects were those of modality,  $F(1, 108) = 185.039, p < .001, \eta^2_p = .631$ ; order,  $F(1, 108) = 16.123, p < .001, \eta^2_p = .130$ ; interval type,  $F(1, 108) = 18.542, p < .001, \eta^2_p = .147$ ; and feedback,  $F(1, 108) = 6.326, p = .013, \eta^2_p = .055$ ; and the significant interactions Modality  $\times$  Interval Type,  $F(1, 108) = 57.231, p < .001, \eta^2_p = .346$ ; Order  $\times$  Interval Type,  $F(1, 108) = 22.353, p < .001, \eta^2_p = .171$ ; Modality  $\times$  Order,  $F(1, 108) = 4.157, p = .044, \eta^2_p = .037$ ; and Modality  $\times$  Order  $\times$  Interval Type,  $F(1, 108) = 33.940, p < .001, \eta^2_p = .239$ .

Time-order error

The TOE was defined as the difference in physical stimulus magnitudes ( $\phi_2 - \phi_1$ ) that yielded  $d = 0$ . It was estimated as follows for the two presentation orders:

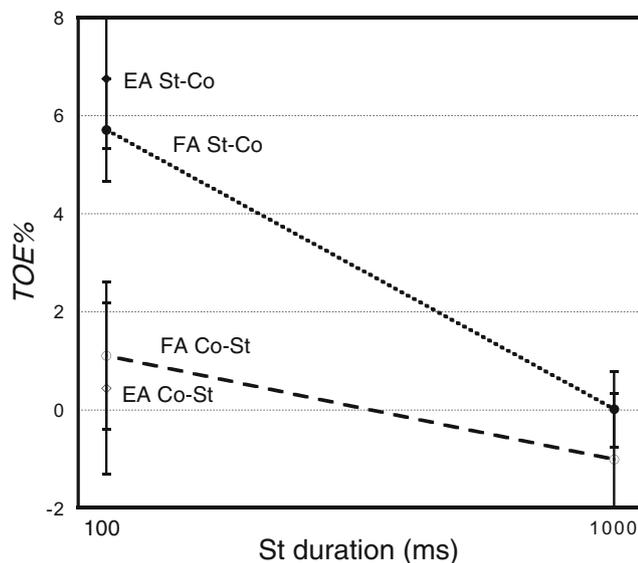
$$TOE_{St-Co} = (DL_{St < Co} - DL_{St > Co})/2, \tag{9}$$

$$TOE_{Co-St} = (DL_{Co < St} - DL_{Co > St})/2. \tag{10}$$

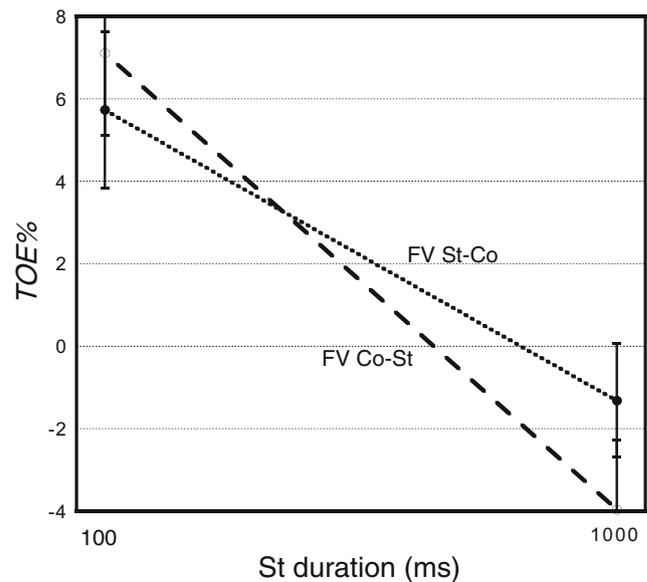
The mean of  $TOE_{St-Co}$  and  $TOE_{Co-St}$ , expressed as a percentage of the standard, was named  $TOE\%$ . The means of  $TOE\%$  are plotted in Figs. 3 and 4. For all interval types, the mean  $TOE\%$  was negative for  $St = 1,000$  ms, and positive for  $St = 100$  ms. The values of  $TOE\%$  were analyzed by mixed ANOVAs with Duration and Order as within-participants factors and Feedback as a between-participants factor.

For the *filled auditory intervals* (see Fig. 3), the effect of duration was significant,  $F(1, 54) = 23.781, p < .001, \eta^2_p = .306$ . The effect of order was also significant,  $F(1, 54) = 4.137, p = .047, \eta^2_p = .071$ . None of the interactions was significant except for Duration  $\times$  Order,  $F(1, 54) = 4.160, p = .046, \eta^2_p = .072$ . Separate analyses with mixed ANOVAs (with the factors Order and Feedback) were conducted for  $St = 100$  ms and  $St = 1,000$  ms. For  $St = 100$  ms, only the effect of order was significant,  $F(1, 54) = 5.126, p = .028, \eta^2_p = .087$  ( $TOE\%_{St-Co} > TOE\%_{Co-St}$ ). For  $St = 1,000$  ms, we observed no significant effects.

For the empty auditory intervals (see Fig. 3) ( $St = 100$  ms only), only the effect of order was significant,  $F(1, 54) = 6.765, p = .012, \eta^2_p = .111$ . For the *filled visual intervals* (see Fig. 4), only the effect of duration was significant,  $F(1, 54) =$



**Fig. 3** Experiment 1. Filled and empty auditory intervals: Mean of  $TOE\%$  versus standard duration. Error bars indicate standard errors of the means



**Fig. 4** Experiment 1. Filled visual intervals: Mean of  $TOE\%$  versus standard duration. Error bars indicate standard errors of the means

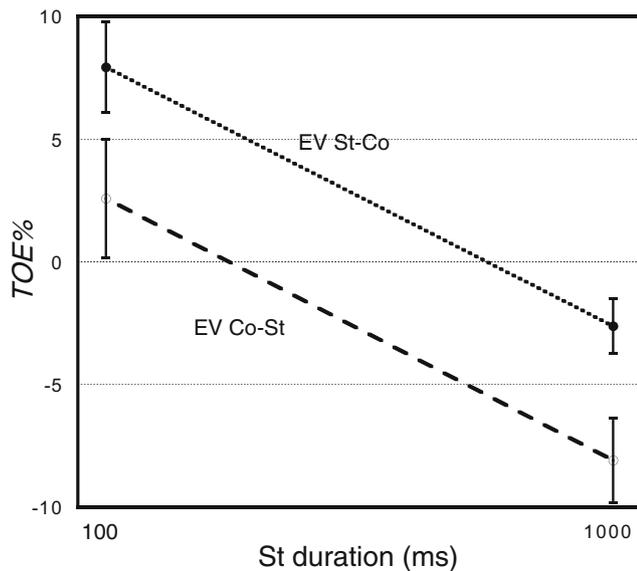
$22.827, p < .001, \eta^2_p = .297$ . For the *empty visual intervals* (see Fig. 5), the effect of duration was significant,  $F(1, 54) = 28.724, p < .001, \eta^2_p = .347$ , as well as the effect of order,  $F(1, 54) = 7.943, p = .007, \eta^2_p = .128$ , but none of the interactions.

For all of the auditory conditions with standard duration 100 ms, the  $TOE\%$  values were submitted to a mixed ANOVA with Order ( $St-Co, Co-St$ ) as a within-participants factor and Interval Type (filled, empty) and Feedback (with, without) as between-participants factors. The only significant effect was that of order,  $F(1, 220) = 10.308, p = .002, \eta^2_p = .045$ .

For all of the visual conditions, the values of  $TOE\%$  were submitted to a mixed ANOVA with Standard Duration (100 ms, 1,000 ms) and Order ( $St-Co, Co-St$ ) as within-participants factors and Interval Type (filled, empty) and Feedback (with, without) as between-participants factors. The significant effects were those of standard duration,  $F(1, 108) = 51.486, p < .001, \eta^2_p = .323$ ; and order,  $F(1, 108) = 7.209, p = .008, \eta^2_p = .063$ ; and the Order  $\times$  Interval Type interaction,  $F(1, 108) = 4.495, p = .036, \eta^2_p = .040$ .

For all of the filled conditions, the  $TOE\%$  values were submitted to a mixed ANOVA with Modality (auditory, visual), Standard Duration (100 ms, 1,000 ms), and Order ( $St-Co, Co-St$ ) as within-participants factors, and Feedback (with, without) as a between-participants factor. The duration effect was significant,  $F(1, 54) = 33.659, p < .001, \eta^2_p = .384$ , as were the Modality  $\times$  Duration,  $F(1, 54) = 7.631, p = .008, \eta^2_p = .124$ ; and Modality  $\times$  Duration  $\times$  Order,  $F(1, 54) = 5.947, p = .018, \eta^2_p = .099$ , interactions.

For all of the conditions with  $St = 100$  ms, the  $TOE\%$  values were submitted to a mixed ANOVA with Modality



**Fig. 5** Experiment 1. Empty visual intervals: Mean of TOE% versus standard duration. Error bars indicate standard errors of the means

(auditory, visual) and Order (St–Co, Co–St) as within-participants factors, and Interval Type (filled, empty) and Feedback (with, without) as between-participants factors. The significant effects were those of modality,  $F(1, 108) = 3.968, p = .049, \eta^2_p = .035$ ; and order,  $F(1, 108) = 8.243, p = .005, \eta^2_p = .071$ .

*Estimation of parameters in the sensation-weighting model*

**Estimation of  $s_1/s_2$  from TOEs** From Eqs. 3 through 6, it follows that

$$TOE_{St-Co} = (1/s_2) \{ (s_1 - s_2) [St - (\phi_{rL} + \phi_{rS})/2] + b \} \tag{11}$$

and

$$TOE_{Co-St} = (1/s_1) \{ (s_1 - s_2) [St - (\phi_{rL} + \phi_{rS})/2] + b \}, \tag{12}$$

so that

$$TOE_{St-Co} / TOE_{Co-St} = s_1/s_2, \tag{13}$$

which in principle yields an unbiased estimate of  $s_1/s_2$ , regardless of the relation between  $\phi_{rL}$  and  $\phi_{rS}$ , and of a possible judgment bias ( $b$ ). This estimate is, however, not very useful, especially for individual participants, because the two TOE values are very variable, are often close to zero, and may even have opposite signs. The ratio of the group means of  $TOE_{St-Co}$  and  $TOE_{Co-St}$  may still be used to roughly estimate  $s_1/s_2$  in conditions in which these means are sufficiently reliable and at a reasonably safe distance from zero. This is illustrated below.

**Estimation of  $s_1/s_2$  from DLs** It follows from the sensation-weighting model that unless  $s_1 = s_2$ , the DLs differ between the presentation orders St–Co and Co–St, which means that a SPE occurs. Combining Eqs. 3 through 6 yields

$$DL_{St-Co} = (s_1/s_2)DL_{Co-St} + (1 - s_1/s_2)(\phi_{rL} - \phi_{rS}), \tag{14}$$

where  $DL_{St-Co} = (DL_{St > Co} + DL_{St < Co})/2$  and  $DL_{Co-St} = (DL_{Co > St} + DL_{Co < St})/2$ .

A blocked design was used, so that virtually only the stimuli in the particular block should determine the  $\phi_r$ . As was said before, with the present design, the mean stimulus duration in L-blocks is about one DL unit longer than in S-blocks, which gives reason to hypothesize that  $\phi_{rL}$  is longer than  $\phi_{rS}$  in some proportion,  $u$ , to this amount.

We now express  $\phi_{rL} - \phi_{rS}$  as  $u(DL_{St-Co} + DL_{Co-St})/2$ . This yields  $DL_{St-Co} = DL_{Co-St} [2k + u(1 - k)] / [(2 - u(1 - k))]$ , where  $k = s_1/s_2$ . For  $u = 1$ , this yields  $DL_{St-Co} / DL_{Co-St} = 1$ , and for  $u = 0$ —that is, for  $\phi_{rL} = \phi_{rS}$ —

$$DL_{St-Co} / DL_{Co-St} = s_1/s_2. \tag{15}$$

If  $0 < u < 1$ , this DL ratio will be between  $s_1/s_2$  and 1.

In conditions that display sizable and reliable effects of stimulus order, Eqs. 11–14 are very useful to help interpret the WF and TOE data, in order to estimate the relations (i) of  $s_1$  to  $s_2$ , (ii) of  $\phi_{rL}$  to  $\phi_{rS}$ , and (iii) of St to  $\phi_r$ . This estimation will here be illustrated only for the empty-visual condition, in which  $s_1/s_2$  was estimated, estimating the TOE ratio (Eq. 13), as 3.074 and 0.323 for St = 100 ms and St = 1,000 ms, respectively. Using Eq. 14, together with the DL estimates for the orders St–Co and Co–St,  $\phi_{rL} - \phi_{rS}$  was estimated as 17.9 and 152.5 for St = 100 ms and St = 1,000 ms, respectively. These values can be compared with the corresponding geometric means of the DLs—35.4 and 205.0 ms—and indicate  $u$  values of .51 and .74. Finally, Eq. 11, together with the estimate of  $TOE_{St-Co}$ , was used to estimate  $(\phi_{rL} + \phi_{rS})/2 - b/(s_1 - s_2)$  as 96.2 and 961.3. Assuming  $b = 0$ , we then arrive at these estimates: for St = 100 ms,  $\phi_{rL} = 105.1, \phi_{rS} = 87.2$ ; for St = 1,000 ms,  $\phi_{rL} = 1,037.6, \phi_{rS} = 885.0$ . The resulting estimates of  $(\phi_{rL} + \phi_{rS})/2$ , 96.2 and 961.3 ms, are about 4% below the corresponding standard duration. Interestingly, the TOEs in this case resemble the usual pattern in TOE studies—positive for St = 100 ms, negative for St = 1,000 ms—but this has a different cause according to the sensation-weighting model: namely, a positive, nearly constant relative difference between the standard duration and reference level, together with differing weight ratios ( $s_1/s_2$ ).

**“True” DL obscured by weighting** By adding the expressions for  $\Delta\phi$ , the “true” DL, in Eqs. 3–6, we obtain  $4\Delta\phi = s_1(DL_{Co < St} + D_{75Co > St}) + s_2(DL_{St < Co} + DL_{St > Co})$ . This yields

$$\Delta\phi = (s_1 DL_{Co-St} + s_2 DL_{St-Co})/2, \quad (16)$$

which means that each obtained DL depends not only on its underlying, “true” value, but also on the comparison procedure. For instance, in stimulus comparison with a constant interstimulus interval, as in the present experiments, the time interval between the onsets of the first and second stimuli (i.e., the stimulus onset interval) shortens when the stimuli become briefer. As was suggested by earlier research (Hellström, 1979), this may lead to lower weights ( $s_1$  and  $s_2$ ) for the briefer stimuli. This, in turn, leads to increased DLs, so that if Weber’s law holds for the “true” DLs, the obtained WFs no longer would be constant, but would be larger for the briefer stimuli (see Hellström & Rammsayer, 2004).

## Discussion

For TOEs, the results affirm the commonly found effect of stimulus level—negative TOEs for long, but positive TOEs for short, durations. However, this effect occurred despite the fact that the 100-ms and 1,000-ms standards were presented in separate blocks. For the empty visual condition, the detailed analysis above indicates that the effect was due to different weighting patterns for the 100-ms and 1,000-ms durations—also visible by causing SPEs with different signs—rather than to the position of the standard duration in relation to a common reference level. Whatever the detailed reason for the TOE in each condition, it clearly excludes an explanation by simple judgment bias (see Hellström, 1978).

For the filled and empty visual intervals, but not for the filled auditory intervals,  $\ln WF$  was markedly higher for the 100-ms than for the 1,000-ms standard. This suggests that although the mechanisms behind the temporal perception of auditory and of visual intervals may be the same (Grondin, 1993), those behind their comparison and discrimination may differ.

For filled auditory intervals, the *TOE%* and  $\ln WF$  data confirmed earlier results (Hellström & Rammsayer, 2004). Feedback lowered WFs by 15% – 20% but did not change the pattern of results, and all interactions involving feedback were nonsignificant. WFs were affected by whether the first or the second stimulus was varied: For filled auditory and empty visual intervals, with the 100-ms standard, varying the first stimulus (i.e., Co–St) had a higher impact on the discriminative response, as was shown by smaller WFs, than varying the second stimulus (i.e., St–Co). This means that a positive SPE was found. Negative SPEs were obtained, as well: For filled

auditory and, in particular, for empty visual intervals, WFs were higher with the presentation order Co–St than with St–Co for the 1,000-ms standard. Correspondingly, the Order  $\times$  Duration interactions were significant ( $p = .008$  and  $p < .001$ , respectively). For empty auditory intervals (in which only the 100-ms standard was used), the WFs did not differ between orders.

The results of Experiment 1 are incompatible with the classic difference model, which is built on simple subtraction, thereby assuming that variation of each of the stimuli exerts an equal impact on the response, and therefore allows no SPEs. The results are also incompatible with the internal reference model, which (like the quite similar model of Michels & Helson, 1954) specifies a lower impact for the first stimulus, and therefore predicts only negative SPEs. Instead, the results suggest a more flexible ratio of impacts (i.e., weights) for the first and second stimuli, as is described by the sensation-weighting model (Hellström, 1979, 2003). The latter model is able to account for the positive as well as the negative SPEs. In addition, it accounts for the variation of TOEs across conditions, whereas the difference model and the internal reference model would have to invoke some kind of response bias as the basis for the TOE, and seek an explanation for the variation of this bias.

## Experiment 2

In Experiment 1, for filled auditory and empty visual intervals we found a marked difference in results for the 100-ms and 1,000-ms standards, in that for  $\ln WF$  an interaction emerged between standard duration and stimulus order. These results confirmed those from Hellström and Rammsayer (2004), suggesting a transition between different weighting patterns somewhere between the duration levels of 100 and 1,000 ms. Therefore, Experiment 2 was conceived to investigate the weighting patterns for four, logarithmically equally spaced, filled auditory intervals between 100 and 1,000 ms, using no feedback in order to exclude its possibly biasing influence (Jamieson & Petrusic, 1975, 1976). The number of participants included was expected to give a level of measurement precision similar to that in Experiment 1.

## Method

### Participants

The participants were 14 male and 42 female undergraduate psychology students ranging in age from 19 to 40 years ( $M = 22.25$ ,  $SD = 3.44$ ). They received course credit for participating in this study. All participants were naïve about the purpose of this study and reported normal hearing and normal or

corrected-to-normal vision. The study was approved by the local ethics committee.

*Apparatus and stimuli*

The same apparatus was used as in Experiment 1. Filled auditory intervals (white-noise bursts at 66 dBA) were presented with standard durations of 100, 215, 464, and 1,000 ms. No correctness feedback was given. The durations were logarithmically equally spaced:  $\log_{10}(\text{duration in ms}) = 2.00, 2.33, 2.67, \text{ and } 3.00$ .

*Procedure*

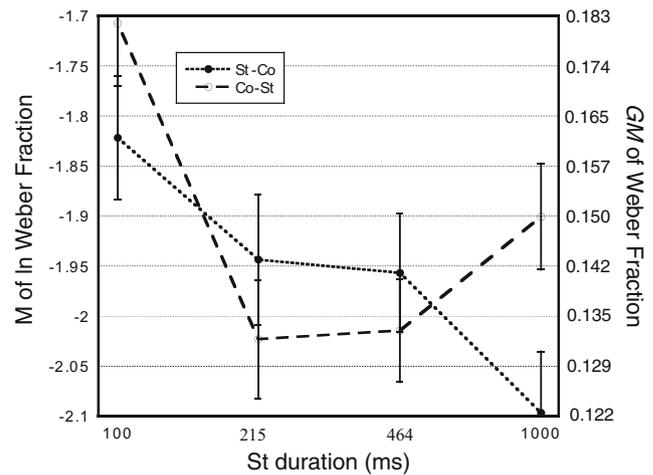
The basic blocked procedure for stimulus presentation and threshold measurement was the same as in Experiment 1. However, here there were four block pairs, each with one of the four durations, with the duration order balanced across participants. In each block pair, one was an L-block and one an S-block, as we defined for Experiment 1. For half of the participants, each block pair always started with an L-block, and for the other half, always with an S-block. The session was initiated by six practice trials.

**Results**

The basic calculations were the same as for Experiment 1. Two participants were excluded: one due to failure to follow the instructions, the other due to extremely poor discrimination for the 100-ms standard duration (upper threshold > 200 ms).

*Natural log Weber fractions*

The mean lnWFs are displayed in Fig. 6. The lnWFs were submitted to a repeated measures ANOVA, using the multivariate approach with Pillai’s tests, with Duration (100, 215, 464, 1,000 ms) and Order (St–Co, Co–St) as within-participants factors. The main effect of duration was significant,  $F(3, 51) = 11.386, p < .001, \eta^2_p = .401$ , as were the linear trend,  $F(1, 53) = 22.363, p < .001, \eta^2_p = .297$ , and the quadratic trend,  $F(1, 53) = 6.811, p = .012, \eta^2_p = .114$ . The main effect of order was not significant. The Duration  $\times$  Order interaction was significant,  $F(3, 51) = 5.852, p = .002, \eta^2_p = .256$ ; notably, the Quadratic  $\times$  Linear trend, mirrored by the double intersection of the curves, was significant,  $F(1, 53) = 17.471, p < .001, \eta^2_p = .248$ . Separate ANOVAs were therefore conducted for the two presentation orders. For the order St–Co, the effect of duration was significant,  $F(3, 51) = 6.915, p < .001, \eta^2_p = .289$ ; only the linear trend was significant,  $F(1, 53) = 18.286, p < .001, \eta^2_p = .257$ . For the order Co–St, the effect of duration was likewise significant,  $F(3, 51) = 10.136, p < .001, \eta^2_p = .374$ ; the linear trend was significant,  $F(1, 53) =$



**Fig. 6** Experiment 2. Filled auditory intervals: Mean of natural log Weber fraction versus standard duration (durations are logarithmically equally spaced). Error bars indicate standard errors of the means. The right axis shows the corresponding geometric means of the Weber fractions

$8.541, p = .005, \eta^2_p = .139$ , as was the quadratic trend,  $F(1, 53) = 17.952, p < .001, \eta^2_p = .253$ .

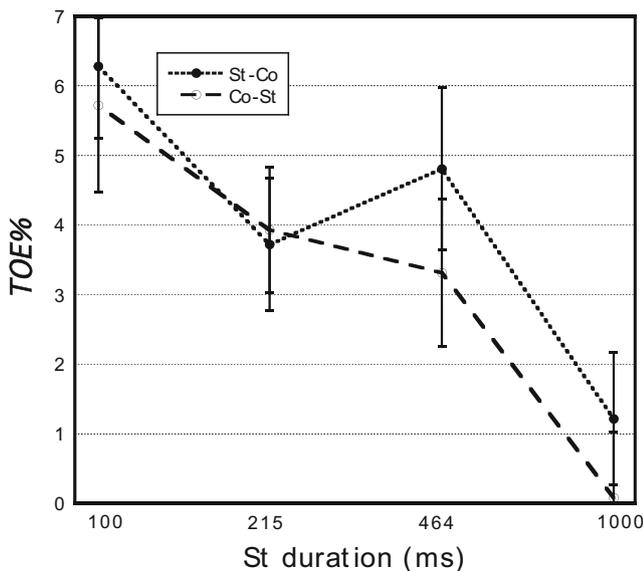
With pairwise *t*tests, the effect of order reached significance only for St = 1,000 ms,  $t(53) = -3.246, p = .002$ , but just approached significance for St = 100 ms,  $t(53) = 1.936, p = .058$  (in both cases,  $\ln WF_{St-Co} < \ln WF_{Co-St}$ , indicating negative SPEs).

Considering only the durations 100 and 1,000 ms, as in Experiment 1, the effect of duration was significant,  $F(1, 53) = 26.502, p < .001, \eta^2_p = .333$  ( $\ln WF_{1,000 \text{ ms}} < \ln WF_{100 \text{ ms}}$ ). The effect of order was also significant,  $F(1, 53) = 13.247, p < .001, \eta^2_p = .200$  ( $\ln WF_{St-Co} < \ln WF_{Co-St}$ ), but not the Duration  $\times$  Order interaction,  $p = .334$ . Considering only the durations 215 and 464 ms, only the effect of order was significant,  $F(1, 53) = 4.193, p = .046, \eta^2_p = .073$  ( $\ln WF_{St-Co} > \ln WF_{Co-St}$ , indicating positive SPEs).

*Time-order error*

The means of TOE%*s* are given in Fig. 7. All obtained mean values were positive. The TOE% values were ANOVA-analyzed similarly to the lnWFs. Only the effect of duration was significant,  $F(3, 51) = 11.030, p < .001, \eta^2_p = .394$ ; likewise, the linear trend was significant,  $F(1, 53) = 22.442, p < .001, \eta^2_p = .297$ , and the cubic trend, mirrored in the S-shape of the plots, also reached significance,  $F(1, 53) = 5.840, p = .019, \eta^2_p = .099$ . Considering only the durations of 100 and 1,000 ms—for comparison with Experiment 1—again, only the effect of duration on TOE% was significant,  $F(1, 53) = 27.693, p < .001, \eta^2_p = .343$ .

For a further comparison between Experiments 1 and 2, an ANOVA was conducted on the lnWF values (only for the no-



**Fig. 7** Experiment 2. Filled auditory intervals: Mean of TOE% versus standard duration (durations are logarithmically equally spaced). Error bars indicate standard errors of the means

feedback condition) with Standard Duration (100 ms, 1,000 ms) and Order (St–Co, Co–St) as within-participants factors and Experiment (1, 2) as a between-participants factor. The effect of duration was significant,  $F(1, 80) = 11.472, p < .001, \eta^2_p = .125$ . The effect of experiment was not significant,  $F(1, 80) = 0.635, p = .428, \eta^2_p = .008$ . The Order  $\times$  Experiment interaction was significant,  $F(1, 80) = 6.006, p = .016, \eta^2_p = .050$ , as was the Duration  $\times$  Experiment interaction,  $F(1, 80) = 4.204, p = .044, \eta^2_p = .070$ , but the triple, Duration  $\times$  Order  $\times$  Experiment interaction was not,  $F(1, 80) = 0.198, p = .658, \eta^2_p = .002$ . The effect of order, like the Duration  $\times$  Order interaction, did not reach the 5% level of statistical significance.

For TOE%, the results agree quite well between the experiments, in that the values were less positive for the order Co–St than for the order St–Co, and more so for St = 1,000 ms than for St = 100 ms. Unlike in Experiment 1, TOE% stayed on the positive side.

An ANOVA similar to that on the data from both experiments for the lnWF values was conducted on the TOE% values. Only the effect of order was significant,  $F(1, 80) = 4.993, p = .028, \eta^2_p = .059$ , whereas the Duration  $\times$  Order interaction failed to reach statistical significance.

## Discussion

For WFs, the results of Experiment 2 differed from those of Experiment 1, in that the effect of the order of the standard and comparison stimuli on lnWF was the same for standard durations 100 and 1,000 ms ( $\ln WF_{St-Co} < \ln WF_{Co-St}$ ), whereas for the intermediate durations 215 and 464 ms, the relation was the opposite ( $\ln WF_{St-Co} > \ln WF_{Co-St}$ ). In terms of SPEs, we

observed a transition from a negative effect at 1,000 ms to a positive effect at 464 and 215 ms, which unexpectedly reversed at 100 ms. We cannot yet offer an explanation for the difference in the effects of standard duration between Experiments 1 and 2. It may have been due to an unknown effect of the differences in the sets of interval types (Exp. 1, auditory and visual; Exp. 2, only auditory) and standard durations encountered by the participants. Still, the body of results confirms the occurrence of both types of SPE—positive and negative—and that this effect does interact with stimulus duration: a pattern of results that is compatible with the sensation-weighting model but not with the difference model, nor with the internal reference model.

## General discussion

The aim of the present experiments was to investigate the comparison and discrimination of intervals with different durations and in different representation forms, as an extension of the studies by Hellström and Rammsayer (2004) and Rammsayer (2014). For this purpose, we employed two different types of intervals (filled and empty ones) presented in the auditory and visual sensory modalities. The sensation-weighting model was used as an analytical and descriptive tool. It should be noted, however, that this by no means excluded the possibility of ending up by selecting the difference or the internal reference model. Duration discrimination sensitivity was shown to depend on the order of the standard and comparison stimuli (the SPE), in a fashion that depended on the type of interval as well as on the standard duration. Although they are incompatible with the difference model and the internal reference model, these results are compatible with the sensation-weighting model, which suggests a flexible strategic weighting of the two stimuli within the comparison process. With the employed interstimulus interval of 900 ms, one common strategy was to give a lower weight to the first-presented stimulus, reflecting the partial retention loss of its duration information, and causing a negative SPE. For some types of brief intervals, however, the opposite strategy was used. This might reflect the fact that with a constant interstimulus interval, the interval between the onsets of the two stimuli decreases with the stimulus duration, and may become brief enough that proactive stimulus interference sets in (see Hellström, 1979). Such flexibility of stimulus weighting is in line with the recent finding by Dyjas, Bausenhardt, and Ulrich (2014) that indicating to the participant which stimulus was the comparison eliminated the SPE, thereby showing that the weighting of stimulus information can be put under cognitive control.

It seems pertinent, at this point, to quote Helson (1964):

In classical psychophysics and for many modern workers effects of order of stimulation are merely “errors” of judgment which should be eradicated; but for us TOE

and SOE [space-order effect] have deeper significance. So-called time-order and space-order errors arise from the decentered position of adaptation level and are therefore manifestations of lawful underlying mechanisms. ... The systematic nature of such effects rules out the “error” view of their origin. (p. 159)

The sensation-weighting model as well as the internal reference model (Dyjas et al. 2014) are both in agreement with Helson (1964) in rejecting the difference model. What distinguishes their theoretical positions is that the internal reference model assumes that only the first stimulus is replaced by an internal reference (as was also assumed by Michels & Helson, 1954), whereas the sensation-weighting model is more general, since it also assumes an analogous process for the second stimulus. This generalization (Hellström, 1979) is built on the notion that, at the moment of comparison, both stimuli are in memory, not just the first one. Many of the results of, for example, Hellström (1979) could never be explained without the notion that weighting also affects the second stimulus. Furthermore, within the conceptual framework of the sensation-weighting model, stimulus weights  $> 1$  (indicating contrastive rather than assimilative integration of stimulus and reference level) are allowed, and indeed seem to occur in some stimulus conditions (see Fig. 3 of Hellström, 1979). The model is thereby able to account for a wide range of phenomena in stimulus comparison, notably the TOE, for which the sensation-weighting model has an explanation similar to Helson’s (1964), whereas the internal reference model has none. As is suggested by the results of the present experiments, factors such as sensory modality (auditory, visual), type of interval (filled, unfilled), and standard duration modulate the parameters of the sensation-weighting model.

TOEs need not always appear in an experiment with only one standard per block (see Hellström, 2000). According to the sensation-weighting model, this is because, apart from bias effects, the TOE is proportional to the distance from the reference level. It is important to note that the terminologies in the models differ in a crucial way: In the internal reference model, the *internal reference* represents the current stimulus weighted together with all the previous experimental stimulation (with a constant weight across trials). Extra-experimental stimulation is ignored in the internal reference model, which is a troublesome limitation, in that it excludes any TOEs. Except for this, the internal reference model is quite similar to Michels and Helson (1954); see also Helson 1964, pp.209–215) TOE model, in which “ $s$  and  $t$  are the relative weights of  $S$  [the first-presented stimulus] and  $SAL$  [the series adaptation level] in determining  $A_I$  [the comparative adaptation level; i.e., the internal reference] and are subject to the condition:  $t = 1 - s$ ” (p. 329). (Please note that Michels & Helson, 1954, applied the  $s$

and  $t$  weights to the logarithms of the stimulus magnitudes.) In the internal reference model, this kind of weighting is applied at every trial, specifying how the experimental stimuli build up the internal reference in the course of the experiment. In the sensation-weighting model, which arose as a generalization of the Michels–Helson model (see Hellström, 1979), the reference level is similar to what Michels and Helson called the “series adaptation level,” in that it represents all of the influential stimulation, which may be seen as a weighted compound of experimental, background, and residual stimulation (see Helson, 1964).

The sensation-weighting model makes no detailed assumptions as to the buildup processes. Rather, the subjective magnitudes of the stimulus and the pertinent reference level are combined by weights to form the subjective magnitudes, the difference of which produces the subjective difference underlying comparison and discrimination. As was discussed in detail in Patching, Englund, and Hellström (2012), this weighting mechanism seems to be optimized to serve as an adaptive strategy aimed at maximizing the discriminability of small relative changes in the compared stimuli. TOEs arise as by-products. According to this view, SPEs and TOEs originate from the same process, although this becomes clearer when using designs in which both stimuli are varied (e.g., Hellström, 1979; Patching et al. 2012).

Raviv, Lieder, Loewenstein, and Ahissar (2014) suggested that “sensory processes cannot be studied in isolation, or ‘out of context’: even in the simplest discriminations they involve complex statistical learning that affects participants’ performance” (p.2, Author Summary). This statistical learning is incorporated, albeit differently, in contemporary models such as the sensation-weighting model and the internal reference model, as well as in the older Michels–Helson model. The insight that sensory processes are dependent on their context may help reduce the frustration due to diverging results in seemingly similar experiments, and may provide clues to a deeper understanding of these results.

For temporal stimuli, once they have generated temporal sensations, these seem to be handled by essentially the same comparison mechanisms as do sensations in other modalities—resulting in TOEs and SPEs being obtained in temporal comparison studies such as the present one. Still, the detailed functioning of these mechanisms is dependent on the stimulus conditions, as the present experiments illustrate, but as is also known from studies in other modalities (see Hellström, 1979, 2003). In this respect, time may not be so special after all.

**Author note** Preliminary results from Experiment 1 were presented at Fechner Day 2013, the 29th Annual Meeting of the International Society for Psychophysics, in Freiburg, Germany, October 21–25, 2013.

## References

- Alcalá-Quintana, R., & García-Pérez, M. A. (2011). A model for the time-order error in contrast discrimination. *Quarterly Journal of Experimental Psychology*, *64*, 1221–1248. doi:10.1080/17470218.2010.540018
- Allan, L. G. (1977). The time-order error in judgments of duration. *Canadian Journal of Psychology*, *31*, 24–31.
- Bausenhart, K. M., Dyjas, O., & Ulrich, R. (2015). Effects of stimulus order on discrimination sensitivity for short and long durations. *Attention, Perception, & Psychophysics*, *77*, 1033–1043. doi:10.3758/s13414-015-0875-8
- Dyjas, O., Bausenhart, K., & Ulrich, R. (2014). Effects of stimulus order on duration discrimination sensitivity are under attentional control. *Journal of Experimental Psychology: Human Perception and Performance*, *40*, 292–307.
- Dyjas, O., & Ulrich, R. (2014). Effects of stimulus order on discrimination processes in comparative and equality judgments: Data and models. *Quarterly Journal of Experimental Psychology*, *67*, 1121–1150. doi:10.1080/17470218.2013.847968
- Eisler, H. (1975). Subjective duration and psychophysics. *Psychological Review*, *82*, 429–450. doi:10.1037/0033-295X.82.6.429
- Eisler, H. (1981). Applicability of the parallel-clock model to duration discrimination. *Perception & Psychophysics*, *29*, 225–233.
- Eisler, H., Eisler, A. D., & Hellström, Å. (2008). Psychophysical issues in the study of time perception. In S. Grondin (Ed.), *Psychology of time* (pp. 75–109). Bingley, UK: Emerald.
- Fechner, G. T. (1860). *Elemente der Psychophysik [Elements of psychophysics]*. Leipzig, Germany: Breitkopf & Härtel.
- Graff, C. (2014). Expressing relative differences (in percent) by the difference of natural logarithms. *Journal of Mathematical Psychology*, *60*, 82–85.
- Grondin, S. (1993). Duration discrimination of empty and filled intervals marked by auditory and visual signals. *Perception & Psychophysics*, *54*, 383–394.
- Grondin, S., & McAuley, J. D. (2009). Duration discrimination in crossmodal sequences. *Perception*, *38*, 1542–1559.
- Grondin, S., Meilleur-Wells, G., Ouellette, C., & Macar, F. (1998). Sensory effects on judgments of short time-intervals. *Psychological Research*, *61*, 261–268.
- Hellström, Å. (1977). Time-errors are perceptual: An experimental investigation of duration and a quantitative successive-comparison model. *Psychological Research*, *39*, 345–388.
- Hellström, Å. (1978). Factors producing and factors not producing time errors: An experiment with loudness comparisons. *Perception & Psychophysics*, *23*, 433–444.
- Hellström, Å. (1979). Time errors and differential sensation weighting. *Journal of Experimental Psychology: Human Perception and Performance*, *5*, 460–477. doi:10.1037/0096-1523.5.3.460
- 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. doi:10.1037/0096-1523.26.1.6
- 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
- Hellström, Å., & Rammsayer, T. H. (2004). Effects of time-order, interstimulus interval, and feedback in duration discrimination of noise bursts in the 50- and 1000-ms ranges. *Acta Psychologica*, *116*, 1–20.
- Helson, H. (1964). *Adaptation-level theory*. New York, NY: Harper & Row.
- Jamieson, D. G., & Petrusic, W. M. (1975). Presentation order effects in duration discrimination. *Perception & Psychophysics*, *17*, 197–202. doi:10.3758/BF03203886
- Jamieson, D. G., & Petrusic, W. M. (1976). On a bias induced by the provision of feedback in psychophysical experiments. *Acta Psychologica*, *40*, 199–206.
- Jamieson, D. G., & Petrusic, W. M. (1978). Feedback versus an illusion in time. *Perception*, *7*, 91–96.
- Kaernbach, C. (1991). Simple adaptive testing with the weighted up-down method. *Perception & Psychophysics*, *49*, 227–229. doi:10.3758/BF03214307
- Lewis, P. A., & Miall, R. C. (2003a). Brain activation patterns during measurement of sub- and supra-second intervals. *Neuropsychologia*, *41*, 1583–1592.
- Lewis, P. A., & Miall, R. C. (2003b). Distinct systems for automatic and cognitively controlled time measurement: Evidence from neuroimaging. *Current Opinion in Neurobiology*, *13*, 1–6.
- Martin, L. J., & Müller, G. E. (1890). *Zur Analyse der Unterscheidungsempfindlichkeit [On the analysis of discrimination sensitivity]*. Tübingen, Germany: Laupp.
- Mattes, S., & Ulrich, R. (1998). Directed attention prolongs the perceived duration of a brief stimulus. *Perception & Psychophysics*, *60*, 1305–1317. doi:10.3758/BF03207993
- Michels, W. C., & Helson, H. (1954). A quantitative theory of time-order effects. *American Journal of Psychology*, *67*, 327–334.
- Needham, J. G. (1935). The effect of the time interval upon the time error at different intensive levels. *Journal of Experimental Psychology*, *18*, 539–543.
- Patching, G. R., Englund, M. P., & Hellström, Å. (2012). Time- and space-order effects in timed discrimination of brightness and size of paired visual stimuli. *Journal of Experimental Psychology: Human Perception and Performance*, *38*, 915–940.
- Rammsayer, T. (1999). Neuropharmacological evidence for different timing mechanisms in humans. *Quarterly Journal of Experimental Psychology*, *52B*, 273–286.
- Rammsayer, T. (2008). Neuropharmacological approaches to human timing. In S. Grondin (Ed.), *Psychology of time* (pp. 295–320). Bingley, UK: Emerald.
- Rammsayer, T. H. (2014). The effects of type of interval, sensory modality, base duration, and psychophysical task on the discrimination of brief time intervals. *Attention, Perception, & Psychophysics*, *76*, 1185–1196. doi:10.3758/s13414-014-0655-x
- Rammsayer, T. H., & Lima, S. D. (1991). Duration discrimination of filled and empty auditory intervals: Cognitive and perceptual factors. *Perception & Psychophysics*, *50*, 565–574.
- Rammsayer, T. H., & Troche, S. J. (2014). In search of the internal structure of the processes underlying interval timing in the sub-second and the second range: A confirmatory factor analysis approach. *Acta Psychologica*, *147*, 68–74. doi:10.1016/j.actpsy.2013.05.004
- Rammsayer, T., & Ulrich, R. (2011). Elaborative rehearsal of nontemporal information interferes with temporal processing of durations in the range of seconds but not milliseconds. *Acta Psychologica*, *137*, 127–133. doi:10.1016/j.actpsy.2011.03.010
- Rammsayer, T. H., & Wittkowski, K. M. (1990). Zeitfehler und Positionseffekt des Standardreizes bei der Diskrimination kurzer Zeitdauern [Time-order error and position effect of the standard stimulus in the discrimination of short durations]. *Archiv für Psychologie*, *142*, 81–89.
- Raviv, O., Ahissar, M., & Loewenstein, Y. (2012). How recent history affects perception: The normative approach and its heuristic

- approximation. *PLoS Computational Biology*, 8(e1002731), 1–10. doi:[10.1371/journal.pcbi.1002731](https://doi.org/10.1371/journal.pcbi.1002731)
- Raviv, O., Lieder, I., Loewenstein, Y., & Ahissar, M. (2014). Contradictory behavioral biases result from the influence of past stimuli on perception. *PLoS Computational Biology*, 10(e1003948), 1–10. doi:[10.1371/journal.pcbi.1003948](https://doi.org/10.1371/journal.pcbi.1003948)
- Thurstone, L. L. (1927a). A law of comparative judgment. *Psychological Review*, 34, 273–286. doi:[10.1037/h0070288](https://doi.org/10.1037/h0070288)
- Thurstone, L. L. (1927b). Psychophysical analysis. *American Journal of Psychology*, 38, 368–389.
- Ulrich, R., & Vorberg, D. (2009). Estimating the difference limen in 2AFC tasks: Pitfalls and improved estimators. *Attention, Perception, & Psychophysics*, 71, 1219–1227. doi:[10.3758/APP.71.6.1219](https://doi.org/10.3758/APP.71.6.1219)
- von Vierordt, K. (1868). *Der Zeitsinn nach Versuchen [The time-sense according to experiments]*. Tübingen, Germany: Laupp.