TIME-ORDER EFFECTS FOR AESTHETIC PREFERENCE

Åke Hellström
Department of Psychology, Stockholm University, S-106 91 Stockholm, Sweden
Email: hellst@psychology.su.se

Abstract

Participants compared successive color patterns (Exp. 1) or jingles (Exp. 2), selecting the preferred one. Results were well described by Hellström's sensation-weighting model, with a greater weight for the second stimulus than for the first. Mean time-order errors were negative, which can be explained as a consequence of this stimulus weighting and of a reference level for aesthetic attractiveness, lower than that of the average stimulus; this level seems to reflect the low aesthetic value of the visual or auditory stimulus background.

Fechner (1860) was the first to notice the *Zeitfehler*, that is, time-order error (TOE) in comparisons of successive stimuli. Also, he introduced experimental aesthetics (Fechner, 1876). Apparently Fechner did not combine those two subjects. However, as we shall see, the results of this combination should have been of interest to him.

Koh (1967) had different participants compare the pleasantness of paired successive musical excerpts, which had been rated for pleasantness by similar participants. Out of two equally pleasant excerpts participants preferred the second (a positive TOE); out of two equally unpleasant excerpts, the first (a negative TOE). On average, there was a slight negative TOE. These effects resemble findings for, for instance, lifted weights (Hellström, 2000).

Hellström (1979) studied in detail the effects of the stimulus magnitude on the TOE for loudness; this led to the explanation of the TOE as a side effect of *sensation-weighting*: the subjective difference, d, between two compared stimuli is not the simple difference between their magnitudes; instead it is the difference between two weighted compounds, one for each stimulus, where the stimulus and a reference level (ReL) enter with weights s and (1-s):

$$d = k\,\{[s_1\,\psi_1 + (1 - s_1)\,\psi_{r1}] - [s_2\,\psi_2 + (1 - s_2)\,\psi_{r2}]\}, \tag{1}$$

where d is the scaled subjective difference, k a scale constant, ψ_1 and ψ_2 the sensation magnitudes of the stimuli, s_1 and s_2 weighting coefficients, and ψ_{r1} and ψ_{r2} the subjective magnitudes of the ReLs (possibly different for the two stimuli). The weighting-in of the ReLs substitutes averaged magnitudes for stimulus magnitudes that are missing or noisy due to, for instance, memory loss; this improves stimulus discriminability (Hellström, 1985, 1989).

The TOE can be defined in subjective units as the value of d in a pair of equal stimuli (Hellström, 1985). Setting $\psi_1 = \psi_2 = \psi$ and simplifying by assuming $\psi_{r1} = \psi_{r2} = \psi_r$ yields

$$(\text{TOE} =)\ d = k\,(s_1 - s_2)\,(\psi - \psi_r). \qquad (2)$$

In earlier research (with ISIs of several seconds) the TOE was generally negative, more so the higher the stimulus magnitude level in the series, and positive only for stimuli of low magnitude. Eq. 2 explains this as a consequence of the weight relation $s_1 < s_2$ and the stimulus magnitudes in the pair being below the ReL $(\psi - \psi_r < 0)$ (cf. Hellström, 2000). For brief stimuli and ISIs Hellström (1979, 1992) found the opposite effect of stimulus magnitude, and interpreted this as being due to the weight relation $s_1 > s_2$.

In Koh's (1967) experiment the stimuli were of rather long duration, and the ISI was kept constant at 6 s. No account was taken of individual differences. The present study aimed at studying the phenomenon by using the SW model and individual scaling of the stimulus magnitudes, and investigating the possible effect of the ISI for relatively brief stimuli.

Experiment 1: Color patterns

Method

Participants. Undergraduate psychology students, 9 men and 23 women, age 19-50 years (mean: 27.4) participated to fulfil a course requirement.

Apparatus and procedure. The participant sat in a quiet and softly lighted room, in front of a Commodore Amiga 1000 computer with a Commodore 1081 color display screen equipped with a reflex-damping mesh filter. (Each participant took part in one session with 1-5 experiments with different kinds of stimuli; see Hellström, 1992.) The preferred stimulus was indicated by pressing a keyboard key – "1" for "first," "2" for "second," "0" for "cannot decide" – and then "Enter." The response could be corrected before entering it.

Stimuli. Rectangles 70 (horizontal) by 100 (vertical) pixels were divided into four rectangles with two colors, A and B, in the pattern $\begin{smallmatrix} A & B \\ B & A \end{smallmatrix}$. The following five patterns (P1-P5) were used (defining A and B by the Amiga's 16 levels, 0-15, of red, green, and blue): P1: (6 12 2) (8 14 10); P2: (13 15 9) (4 4 5); P3: (14 3 0) (1 14 1); P4: (5 0 15) (5 15 15); P5: (12 4 14) (7 14 6). With four sets, one for each ISI, of the 20 pairs of different patterns, in both orders, 80 pairs were presented. The pattern duration was 100 ms, and the ISIs were 100, 300, 900, and 2700 ms. The pairs were presented in random order (different for each participant) with ISIs intermixed. The session, except instructions, lasted on average 9.13 min (SD = 0.76).

Scaling. For each pair the subjective difference d was scaled by d^*: +100 for "1," –100 for "2," and 0 for "0." The preference value, p^*, for each pattern was obtained by scoring +100 for each choice of the pattern and -100 for each choice of the other pattern in a pair, and averaging over the 32 occurrences. For each participant and set, a linear regression was computed with d^* as the dependent variable and p^* for the compared stimuli as independent variables.

Data treatment. Eq. 1 simplifies to

$$d^* = B_1\,\psi_1 - B_2\,\psi_2 + A, \qquad (3)$$

where $B_1 = k\,s_1$, $B_2 = k\,s_2$, and

$$A = \psi_{r1} - \psi_{r2} + s_2\,\psi_{r2} - s_1\,\psi_{r1}. \qquad (4)$$

The subjective TOE was computed as the d^* value predicted from Eq. 3 for a pair of stimuli, both equal to the mean value of p^*, that is, zero. The measure is equal to A, and equivalent to $D\%$, the difference between the percentages of responses "1" and "2." It is termed *TOE%*.

Results and Discussion

The mean p^* values for the patterns (SDs in parentheses) were: P1: -1.61 (8.36); P2: -4.64 (10.79); P3: -0.42 (10.43); P4: 5.59 (8.15); P5: 1.09 (9.20). The mean p^* value across patterns is, by definition, zero. p^* had a mean intraindividual range of 27.03. The mean value, across participants and ISIs, of the multiple R of d^* with the p^* values was .81. The mean B values are shown in Fig. 1. As is seen, for each ISI $B_1 < B_2$ and $TOE\% < 0$.

Statistical analysis. The B values were analyzed by an ANOVA (multivariate for repeated measures) with order (1st, 2nd) and ISI (100, 300, 900, 2700 ms) as within-subject factors. The effect of order was significant, $F(1,31) = 5.93$, $p = .021$) but not those of ISI, $F(3,29) = 1.33$, $p = .285$, and Order x ISI, $F(3,29) = 1.45$, $p = .250$. Separate ANOVAs with ISI as within-subject factor yielded nonsignificant effects of ISI on B_1, $F(3,29) = 1.48$, $p = .241$, and on B_2, $F(3,29) = 1.00$, $p = .406$. For *TOE%* the mean difference from zero was nonsignificant, $F(1,30) = 3.17$, $p = .085$, as well as the effect of ISI, $F(3,29) = 0.26$, $p = .851$.

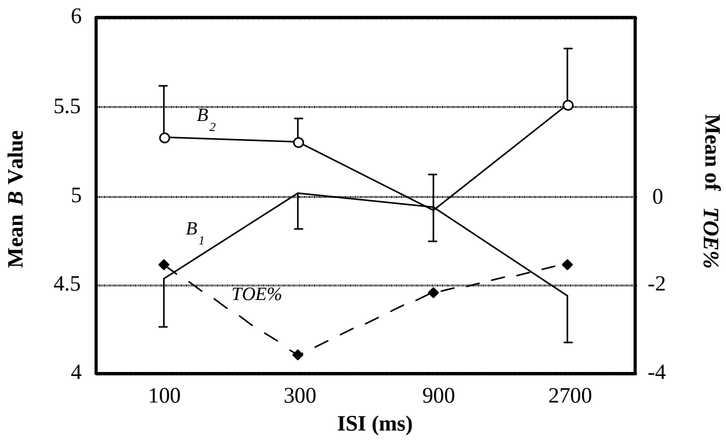

Figure 1. Color patterns (Exp. 1). Mean values of B_1, B_2, and *TOE%* as a function of ISI. Bars indicate standard error of mean.

ReL, Weighting, and TOE. Simplifying by assuming $\psi_{r1} = \psi_{r2} = \psi_r$, Eq. 3 becomes

$$A = k\,(s_2 - s_1)\,\psi_{r\cdot} = (B_1 - B_2)\,\psi_{r\cdot} \quad (5)$$

From Eq. 5 ψ_r was roughly estimated, on the same scale as the p^* values, as $A/(B_1 - B_2)$ using the mean values of A, B_1, and B_2 across ISIs and participants. The estimated ψ_r is –3.93.

The negative ψ_r reflects an aesthetic value lower than for the average pattern; most likely this low value reflects the grey background, The effect of the weighting can be described as an assimilation of the first pattern to this background (Koh, 1967), which yields a negative TOE.

Experiment 2: Jingles

Method

Participants. 17 men and 13 women, normal-hearing, mostly psychology students–undergraduate (fulfilling a course requirement) and graduate (volunteering)–age 21-64 (mean: 32.8).

Apparatus, stimuli, and procedure. The stimuli were five different jingles, sequences of seven sine-tone notes, played through the built-in loudspeaker of a Commodore Amiga 2000 at a comfortable level. The tempered scale with A_4 = 440 Hz was used (subscript indicates octave). The jingles were: (J1) D_5-$C^\#_5$-H_4-A_4-G_4-$F^\#_4$-E_4; (J2) E_5-F_4-$F^\#_4$-G_4-E_4-C_5-G_4; (J3) D_5-C_5-D_5-E_5-F_4-E_4-F_4; (J4): C_6-H_5-G_5-C_6-E_5-G_5-D_5; (J5) $F^\#_6$-E_6-D_6-$C^\#_6$-H_5-A_5-H_5.

The notes lasted 200 ms each, and succeeded each other immediately. Thus the duration of each jingle was 1400 ms. The ISIs were 500, 1000, 2000, and 4000 ms. Eighty pairs (one set for each ISI of the 20 pairs of different jingles, using both time orders) were presented in a random order (different for each participant) with ISIs intermixed. The laboratory environment was similar, and the response mode and the preference scaling the same, as in Exp. 1. The session, except instructions, lasted 12.37 min on average (SD = 0.68).

Results and Discussion

The mean p^* values for the jingles (SDs in parentheses) were: J1: -3.73 (6.75); J2: -1.58 (8.51); J3: 3.50 (7.43); J4: 3.47 (6.40); J5: -1.65 (9.00). The mean intraindividual range of p^* was 21.05. The mean R (across participants and ISIs) of d^* with the p^* values was .72.

Statistical analysis. The B and *TOE%* values, which are shown in Fig. 2, were analyzed by ANOVAs like for Exp. 1. The effect of order (1st, 2nd) on the B values was significant, F (1,29) = 7.59, p = .0101) but not the effects of ISI, F (3,27) = 2.20, p = .111, and Order x ISI, F (3,27) = 2.67, p = .067. However, the interaction of order with the linear effect of ISI was significant, t (27) = 2.41, p = .023. The effect of ISI on B_2 was significant, F(3,27) = 3.55, p = .028 (linear effect: t (27) = 3.37, p = .002) but not the effects on B_1, F(3,27) = 1.95, p = .145, and on *TOE%*, F(3,26) = 1.34, p = .283. For *TOE%*, the mean difference from zero was non-significant, F(1,29) = 0.91, p = .349. ψ_r was estimated, in the same way as in Exp. 1, as –2.12.

ReL, Weighting, and TOE. Plausibly, the background (mainly a slight rumble from the Amiga's fan) had a lower aesthetic value than the average jingle and therefore lowered the ReL to a negative value. The weight relation $s_1 < s_2$ therefore yielded a negative mean *TOE*.

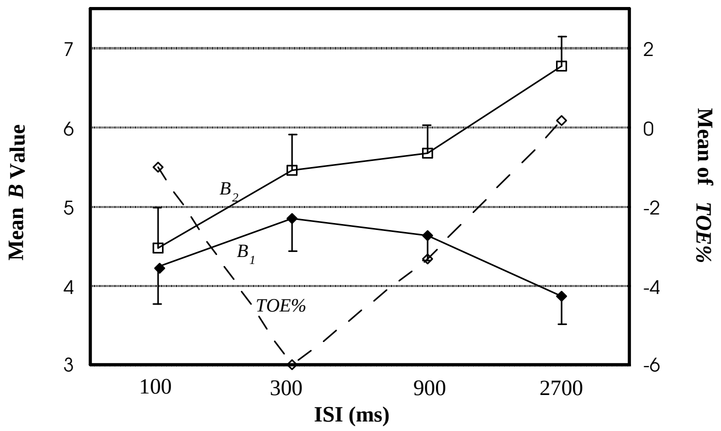

Figure 2. Jingles (Exp. 2). Mean values of B_1, B_2, and *TOE%* as a function of ISI. Bars indicate standard error of mean.

General Discussion

The present study demonstrates the continuity of aesthetic judgments with judgments of physical stimuli. In addition to the effect of aesthetic level, there was a general effect that favored the second stimulus (a negative mean TOE). Earlier studies (see Koh, 1967) likewise showed mainly negative TOEs for affective and aesthetic judgments. This seems to be due to the stimuli being “better” than their background. Thus the task is analogous to heaviness comparison (Hellström, 2000) with the same type of weighting, stimuli heavier than their background and, consequently, a negative mean TOE.

This common result pattern partly reflects psychophysicists’ use of stimuli above the background in the judged attribute, presented in a slow tempo. However, the temporal weighting pattern in Fig. 2 (jingles, Exp. 2) resembles those for loudness (Hellström, 1979, 1992) and line length (Hellström, 1992) in that the *B* curves converge for shorter ISIs. Fig. 1 (color patterns, Exp. 1) shows no such tendency. Thus, even for aesthetic preference, time-order effects differ between stimulus types, probably reflecting different modes of processing.

Perspective: Effects of Word Order on Preference Choices

With the present and earlier results in mind, one may suspect that the stimulus-weighting effect is so general that it occurs whenever we face a difficult choice between successive alternatives. Then, for instance, responses to market research questions such as “Would you prefer to get seven free e-mail addresses or 30% increased connection speed?” could depend on the order of the alternatives. In particular, the choice might be more influenced by the second alternative than of the first, and therefore biased by TOEs.

Wänke, Schwarz, & Noelle-Neumann (1995) had university students compare tennis and soccer (A and B, or B and A); the alternatives were “A is much more exciting than B,” “A is more exciting than B,” A is somewhat more exciting than B,” “A and B are equally exciting,” “A is somewhat more exciting than B,” “A is less exciting than B,” A is less exciting than B,” “A is much less exciting than B.” The word order had a striking effect: with the orders tennis-soccer and soccer-tennis, 35% vs. 77% found tennis more exciting than soccer. The students also compared (with a different question wording) their male and female high-school teachers. With the orders female-male and male-female, 12% vs. 55% found male teachers more empa-thic. Wänke (1996) offered the explanation, based on Tversky’s (1977) theory of similarity judgment, that “respondents who are asked to compare X to Y focus on the features of X and ask to see if these features are also present in Y. . . . As a result, comparisons of X to Y are based on a different selection of features than comparisons of Y to X, resulting in different evaluations” (p. 365).

Wänke (1996) gave no explanations for the directions of the order effects for sports and teachers. Assuming that general popularity of sports was above, but that of teachers below "average," the results are consistent with the alternative interpretation that in the comparison the first alternative tends to get a lower weight than the second. This is equivalent to a regression towards the mean for the first alternative, and makes the judge prefer the second out of two attractive alternatives, but the first out of two less attractive ones. If this is correct, then word order may have the same effect as the time order of other kinds of stimuli.

References

Fechner, G. T. (1860). *Elemente der Psychophysik* [Elements of psychophysics]. Leipzig: Breitkopf & Härtel.

Fechner, G. T. (1876). *Vorschule der Ästhetik* [Introduction to aesthetics]. Leipzig: Breitkopf & Härtel.

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.

Hellström, Å. (1989). What happens when we compare two stimuli? In Ljunggren, G., & Dornic, S. (Eds.), *Psychophysics in action.* Berlin: Springer.

Hellström, Å. (1992). Comparative judgments of successive and simultaneous stimuli. In Borg, G., & Neely, G. (Eds.), *Fechner Day 92* (pp. 99-104). Stockholm: International Society for Psychophysics.

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

Koh, S. D. (1967). Time-error in comparisons of preferences for musical excerpts. *American Journal of Psychology, 80*, 171-185.

Tversky, A. (1977): Features of similarity. *Psychological Review, 84*, 327-353.

Wänke, M. (1996). Comparative judgments as a function of the direction of comparison versus word order. *Public Opinion Quarterly, 60*, 400-409.

Wänke, M., Schwarz, N., & Noelle-Neumann, E. (1995). Asking comparative questions: The impact of the direction of comparison. *Public Opinion Quarterly, 59*, 347-372.