# GOODNESS-LEVEL DEPENDENT WORD-ORDER EFFECT IN PREFERENCE COMPARISON: SEMANTICS IS AN IMPORTANT FACTOR

Mats P. Englund, Åke Hellström, and Vladimir Calderon
*Department of Psychology, Stockholm University, SE-106 91 Stockholm, Sweden*
*mats.englund@psychology.su.se*

## Abstract

*Hellström (2003) and Englund (2008) found a goodness-level dependent word-order effect (WOE) for preference judgment–a tendency to prefer the left (first read) of two good alternatives and the right (second) of two bad ones. Stimuli were spaced horizontally, and participants indicated preference by choosing one of several written statements (e.g., "apple I like more than pear"). The results were described as being due to a higher weight for the left/first stimulus than for the right/second. In the present study, Experiment 1 was similar to the previous studies, except that the stimuli were spaced vertically. In Experiment 2, stimuli were spaced horizontally, but preference was indicated by symbols instead of statements. The results of Experiment 1 essentially replicated the earlier findings, with a higher weight for the upper (first read) stimulus, but those of Experiment 2 did not. These results suggest that the semantic structure of the preference statements is an important factor behind the goodness-level dependent WOE.*

Fechner (1860) was the first to note that two physically equivalent stimuli tend to be perceived as different when presented separated by time or by space. He named these two effects the *time-order error* (TOE) and *space-order error* (SOE), which he defined as positive (negative) when the first/left (second/right) stimulus is overestimated relative to the other. The magnitude and sign of the order effect has been found to vary with, among other things, the magnitude level of the specific stimuli (for reviews, see, e.g., Guilford, 1954; Hellström, 1985). Presentation-order effects are not restricted to physically defined stimuli, however, but have also been found in cases where the stimuli were presented verbally, either as list-based descriptions or by single nouns. For example, the presentation order of the choice alternatives has been found repeatedly to affect responses in preference choices of consumer goods (e.g., Houston & Sherman, 1995) and everyday phenomena (Englund, 2008; Hellström, 2003).

Participants in Hellström's (2003) study chose between pairs of stimuli by agreeing to one of six written preference statements (e.g., "apple is much tastier than pear"). Participants also rated their general opinion on the stimuli in each pair, that is, the goodness of the stimuli. The results showed that the sign and magnitude of the order effect depended on the goodness level of the stimuli. Specifically, participants tended to prefer the first alternative out of two good ones and the second out of two bad ones. Hellström explained his results in terms of the sensation weighting (SW) model (Hellström, 1979; 1985). According to this model, the subjective difference between two stimuli, separated by time or space, is described by

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

where $k$ is a scale constant, $s_1$ and $s_2$ are weighting coefficients, $\psi_1$ and $\psi_2$ are subjective stimulus magnitudes, $\psi_{r1}$ and $\psi_{r2}$ are magnitudes corresponding to the current reference levels

(ReLs; potentially different). The WOE equals $d$ when $\psi_1 = \psi_2$; assuming $\psi_{r1} = \psi_{r2} = \psi_r$ Equation 1 reduces to

$$\text{WOE} = d = k(s_1 - s_2)(\psi - \psi_r). \quad (2)$$

In terms of Equation 2, the linear relationship between the WOE and the goodness level in Hellström's (2003) study was caused by sensation weighting with $s_1 > s_2$ (subscripts 1 and 2 denoting left and right stimulus). Hellström's results were later replicated by Englund (2008).

The results of Hellström (2003) and Englund (2008) raise the question whether the WOE should be considered as a SOE or as a TOE. In both these studies, there was a greater weight for the left/first read stimulus than for the right/last read stimulus, and the stimuli were presented in the order left-right, separated spatially by a list of preference statements denoting the preference relations between the stimuli. The stimuli may also be seen as presented in a temporal order, because one stimulus was read before the other. A third option is that the WOE is a different, but analogous, effect where the weight difference is due to the comparison having a direction induced semantically by the preference statements. That is, in every case, the preference statements pointed out one stimulus, the left/first read, to be compared to the right/last read. Under such conditions, the left/first read stimulus is in focus as the *subject* of the comparison and the right/last read stimulus acts as the passive *referent* to which the subject is to be compared (cf. Tversky, 1977; Wänke, 1996). In the present paper we present two experiments that were designed to investigate whether the goodness-level dependent WOE is due to the semantic pointing out of one stimulus as the focus of the comparison, or merely a typical SOE or TOE.

## Experiment 1

Hellström (2003) and Englund (2008) used a left-right presentation order. In Experiment 1, the presentation order of the alternatives was, instead, up-down. If the goodness-level dependent WOE were to still appear, this would suggest that the WOE is not just a standard SOE and that the semantics of the preference statements is an important factor behind the goodness-level dependent WOE.

### *Method*

There were 121 participants (79 women, 42 men, $M_{age}$ = 28.6, $SD_{age}$ = 10.6); 66 were volunteers, and the rest were students at the Department of Psychology, Stockholm University, taking part to partially fulfill a course requirement. Participants were randomly assigned to one of the two groups AB (59 participants) and BA (62 participants). Participants received a three-section booklet consisting of: (a) a preference judgment task, (b) a personality test (to conceal the purpose of the study), and (c) a stimulus goodness-rating task. In the preference section, there were 20 stimulus pairs to be compared, two on each page. The two stimuli in each pair were printed horizontally centered above one another, with six response alternatives on single lines in between the two stimuli A and B: "[A/B] I generally like [much better/better/somewhat better/somewhat worse/worse/much worse] than [B/A]". The two groups (AB and BA) received opposite within-pair word orders. The order of the pairs was randomized individually for each participant. The second section of the booklet consisted of a personality test with 20 multiple-choice questions. In the third section of the booklet, the participants made goodness ratings of the same items as in the first section, but now one at a time on a seven-step scale. This resulted in 40 items of the form "[A/B] I generally [*like greatly*, *like*, *like somewhat*, *neither like nor dislike*, *dislike somewhat*, *dislike*,

*dislike greatly*]". The stimulus words were centered horizontally above the response alternatives, which were placed in the same way as in the preference section. The order of the stimuli was randomized individually for each participant with the restriction that there had to be at least five stimuli from other stimulus pairs between two stimuli from the same pair.

*Results and Discussion*

Data from nine participants were excluded due to incomplete responses. Preference ratings were scaled as 2.5 (strongest preference for the upper stimulus) to -2.5 in steps of 1. The goodness ratings were scaled from 3 (highest positive goodness) to -3 in steps of 1.

Using Equation 1, the predicted subjective difference between the two stimuli, with the preference ratings for Group AB and Group BA ($d_{AB}$ and $d_{BA}$), can be expressed as (Hellström, 2003)

$$d_{AB} = k\{[s_U \cdot \psi_A + (1 - s_U)\,\psi_{r,U}] - [s_L \cdot \psi_B + (1 - s_L)\,\psi_{r,L}]\} \quad (3a)$$

$$d_{BA} = k\{[s_U \cdot \psi_B + (1 - s_U)\,\psi_{r,U}] - [s_L \cdot \psi_A + (1 - s_L)\,\psi_{r,L}]\}, \quad (3b)$$

where subscripts *U* and *L* denotes upper and lower stimulus, respectively. Hence, WOE can be written as

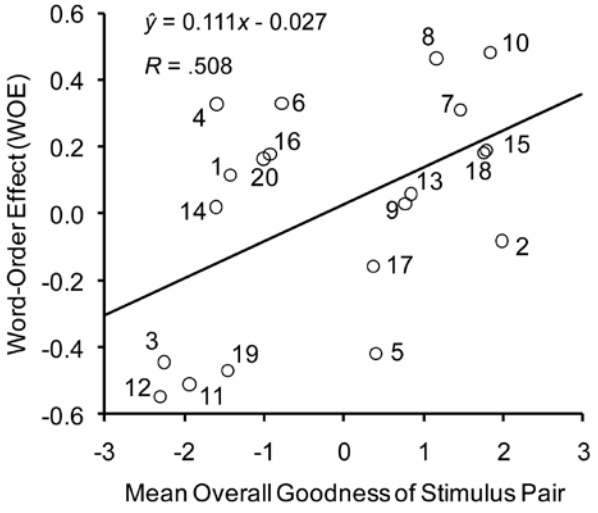
$$\mathrm{WOE} = (d_{AB} + d_{BA})/2. \quad (4)$$

For group data, the WOE for a particular stimulus pair thus becomes half the sum of the mean preference ratings; two stimuli equally rated regardless of their presentation order yields a WOE of zero. Figure 1 displays the WOE values plotted against mean overall goodness and the best fitted regression; the slope was significant, $t(18) = 2.50$, $p = .022$, but the intercept was not, $t(18) = 0.404$, $p = .69$. These results essentially replicate those of Hellström (2003) and Englund (2008), and suggest that a horizontal spatial separation of the stimuli is not a necessary condition for the goodness-level dependent WOE.

Equation 1 can be rewritten as

$$WOE = d = W_U \cdot \psi_U - W_L \cdot \psi_L + A, \quad (5)$$

where $W_U = k \cdot s_U$, $W_L = k \cdot s_L$, and $A = k(\psi_{rU} - s_U \cdot \psi_{rU} - \psi_{rL} + s_L \cdot \psi_{rL})$. According to Equation 5, estimates of the *W* values can be achieved by regressing the preference ratings on the individual goodness ratings for the upper and the lower stimulus, respectively. The mean *R* across participants was 0.648 ($SD = 0.164$). The *W*-value for the upper stimulus ($W_U = 0.654$) was of significantly greater magnitude than that for the lower stimulus ($W_L = 0.591$), $t(110) = 3.68$, $p < .001$, indicating that $s_U > s_L$. This is in accordance with the hypothesis that the weight difference found by Hellström (2003) and Englund (2008), $s_{left} > s_{right}$, was due to the preference statements semantically dictating the comparison direction, and the left stimulus (here, upper) being the subject in the comparison (cf. Tversky, 1977).




| No. | Stimulus pair |
|---|---|
| 1. | Freezing-Sweating |
| 2. | Birthday presents-Christmas presents |
| 3. | Having fever-Vomiting |
| 4. | Feeling angry-Feeling sad |
| 5. | Bus-Subway |
| 6. | Banana peel-Apple core |
| 7. | Flower scent-Perfume scent |
| 8. | Board games-Card games |
| 9. | Circles-Triangles |
| 10. | Sea blue-Forest green |
| 11. | Headache-Stomachache |
| 12. | Cough-Runny nose |
| 13. | Fall-Winter |
| 14. | Cash queue-Telephone queue |
| 15. | Reading-Writing |
| 16. | Snakes-Spiders |
| 17. | Pasta-Rice |
| 18. | Paper bags-Plastic bags |
| 19. | Radio noise-Traffic noise |
| 20. | Rain-Snowfall |

*Figure 1*. Experiment 1, word order effect plotted against mean goodness ratings.

## Experiment 2

The results of Experiment 1 essentially replicated those of Hellström (2003) and Englund (2008) in that there was a goodness-level dependent WOE, and a higher weight for the first read stimulus than for the last read one. This suggests that the goodness-level dependent WOE does not require that the stimuli are presented in the order left-right. In Experiment 2, the aim was to test further this hypothesis by an attempt to remove the semantically dictated comparison direction, but still using a left-right stimulus presentation. Consequently, if the goodness-level dependent WOE was not found in Experiment 2, this would indicate that the WOE is not a SOE, but rather an analogous effect due to the semantically induced directed comparison.

### *Method*

A total of 194 undergraduate psychology students at Stockholm University participated to fulfill a partial course requirement. There were 34 men (one with unstated age) and 161 women ($M_{age}$ = 25.6, $SD_{age}$ = 6.8). The method was similar to that of Experiment 1, with two important differences. First, the stimuli in each pair were placed in the left and right margin, respectively, of the booklet pages. Second, the preference sentences were replaced by preference relations consisting of arrows coupled with short preference statements, representing different degrees of preference without dictating a comparison direction semantically. Specifically, there were six preference relations printed on separate lines, each with an arrow pointing to the left or to the right stimulus, respectively, and to the side of the arrow there was the partial preference statements *prefer strongly*, *prefer*, and *prefer somewhat*. For half of the participants, the topmost arrow pointed to the left stimulus, and for half it pointed to the right stimulus.

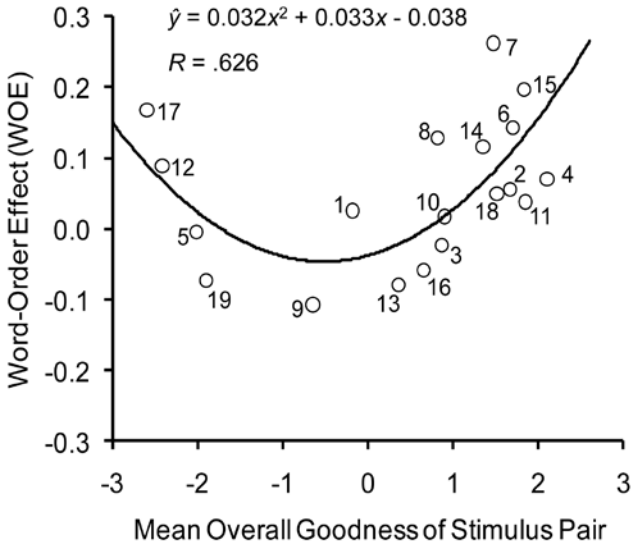
*Results and Discussion*

Participants' responses were scaled in the same way as in Experiment 1, and WOE values for the respective stimulus pair was calculated in the same way as in Experiment 1 (using Equation 5). Figure 2 displays the WOE values plotted against mean overall goodness for the respective pair. As can be seen in Figure 2, there was not the linear relationship between the WOE and the general goodness level that was found in Experiment 1, by Hellström (2003), and by Englund (2008). Instead, as shown in Figure 2, there was a curvilinear relationship, well described by a second-order polynomial ($p = .017$).

A reasonable explanation for the relationship displayed in Figure 2 was found by performing analyses on the individual data. For each participant, the relative preference was regressed on the goodness ratings of the respective stimuli in each stimulus pair in order to find the $W$ values that are proportional to the weights ($W = k \cdot s$). The $W$ values were entered as within-subjects variable (stimulus position: left, right) into a repeated-measures ANOVA with word order (AB, BA) and scale order (uppermost arrow pointing left, uppermost arrow pointing right) as between-subjects factors. Not unexpectedly, there were no significant effects of scale order on the weighting of the stimuli (both $F$s < 1). The main effect of stimulus position was not significant, $F(1, 188) = 0.29$, $p = .59$, suggesting that there was no differential weighting of the stimuli. However, the Stimulus position x Word order interaction was significant, $F(1, 188) = 9.78$, $p = .002$.

Simple analyses showed that the weight for the left stimulus ($W_L = 0.757$) was of greater magnitude than the right stimulus ($W_R = 0.729$) for Group AB, $t(97) = 2.06$, $p = .042$, and that the opposite was the case for group BA, $W_L = 0.707$, $W_R = 0.746$, $t(95) = -2.39$, $p = .019$. That is, no matter the within-pair presentation order, participants tended to weight the A-stimuli higher than the B-stimuli. Furthermore, there were no significant differences between the two groups AB and BA regarding the weights for stimulus set A, $t(193) = 0.29$, $p = .77$, nor regarding the weights for stimulus set B, $t(193) = 0.64$, $p = .53$.

One possible explanation to these results is that one stimulus set, A or B, is slightly more prototypical than the other, and that the stimuli of this or the other set thus spontaneously received the status of being the subject in the comparison. According to Tversky (1977), the more prototypical stimulus usually receives the status of referent in similarity comparisons. Given the semantic dictation of comparison direction in Hellström's




| No. | Stimulus pair |
|---|---|
| 1. | Aftonbladet-Expressen |
| 2. | Apple-Pear |
| 3. | Bus-Subway |
| 4. | Cinema-Consert |
| 5. | Cough-Runny nose |
| 6. | Deciduoud forest-Coniferous forest |
| 7. | Dogs-Cats |
| 8. | Fall-Winter |
| 9. | Filth brown-Dust gray |
| 10. | Finland-Norway |
| 11. | Forest green-Sea blue |
| 12. | Headache-Stomachache |
| 13. | High jump-Long jump |
| 14. | Pasta-Rice |
| 15. | Piano-Guitar |
| 16. | Plastic bag-Paper carrier |
| 17. | Toothache-Nausea |
| 18. | TV news-Paper news |
| 19. | Wasps-Mosquitoes |

*Figure 2*. Experiment 2, word order effect plotted against mean goodness ratings.

(2003) and Englund's (2008) studies, their results suggest that the subject is the stimulus with the highest weight. Combining the results of Hellström and of Englund and the suggestion by Tversky suggests that the stimuli denoted A were slightly less prototypical than those denoted B, due to the larger weight for A. However, Tversky discussed similarity comparisons, and those principles may not apply to preference comparisons.

## General Discussion

The results of the two experiments presented here show that the goodness-level dependent WOE is not a regular SOE. Instead, it seems that the differential sensation weighting behind this effect results from the directedness of the comparison, where one stimulus (the subject) is being compared to the other (the referent). By inducing a direction of the stimulus comparison, judges place a greater weight on the subject in the comparison than on the referent. However, the experimentally induced comparison direction is not imperative. One important facet of the present results is that despite the effort to induce a supposedly nondirected comparison in an attempt to annihilate the systematic differential weighting of the stimuli, participants may still weight the stimuli differently. Without the experimentally induced comparison direction, subtle effects of the stimulus set may appear and cause strong goodness-level dependent WOEs for stimuli in the extreme ends of the goodness continuum.

The results presented here demonstrate that researchers seeking the least biased preference measure for verbal stimuli need to use both presentation orders and calculate the arithmetic mean. This is because of the differential weighting of the two stimuli (cf. Equations 3a-b), which yields $d_{AB} = [(s_L + s_R)/2](\psi_A - \psi_B)$. This is also relevant in contexts where the researcher has not dictated, by experimental manipulation, a comparison direction, because a direction may be induced by, for example, the stimuli being differently prototypical.

## References


Englund, M. P. (2008). Sensation weighting in preference judgments creates a goodness-level dependent word-order effect. In Schneider, B. A., Ben-David, B. M., Parker, S., & Wong, W. (Eds.), *Fechner Day 2008. Proceedings of the 24th Annual Meeting of the International Society for Psychophysics,* Toronto, Canada: The International Society for Psychophysics (pp. 133-138).

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

Guilford, J. P. (1954). *Psychometric methods* (2nd ed.). New York: McGraw-Hill.

Hellström, Å. (2003). Word-order effects in preference comparison. In B. Berglund & E. Borg (Eds.), *Fechner Day 2003. Proceedings of the 19th Annual Meeting of The International Society for Psychophysics, Larnaca Bay, Cyprus, October 19-23, 2003.* Stockholm: International Society for Psychophysics (pp. 121-126).

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 errors and its relatives: Mirrors of cognitive processes in comparing. *Psychological Bulletin, 97*, 35-61.

Houston, D. A., & Sherman, S. J. (1995). Cancellation and focus: The role of shared and unique features in the choice process. *Journal of Experimental Social Psychology, 31*, 357-378.

Tversky, A. (1977). Features of similarity. *Psychological Review, 84*, 327-352.

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