Assessment and interpretation of bias in 2AFC stimulus comparison through chronometric analysis

Hellström, Åke; Patching, Geoffrey

ASSESSMENT AND INTERPRETATION OF BIAS IN 2AFC STIMULUS COMPARISON THROUGH CHRONOMETRIC ANALYSIS

Åke Hellström

Department of Psychology, Stockholm University, SE-106 91 Stockholm, Sweden
hellst@psychology.su.se
Geoffrey R. Patching
Department of Psychology, Lund University, Box 213, SE-221 00 Lund, Sweden
geoffrey.patching@psychology.lu.se

Abstract

Random-walk and diffusion models for two-choice comparison of paired successive or simultaneous stimuli focus on response time (RT), modeled as the time needed to reach one or the other barrier, and its relation to the response probabilities. Logit $P_1 = ln[P_1/(1-P_1)]$, *where P_1 is the probability of responding "first greater," can be seen as a measure of subjective stimulus difference,* d. Signed response speed (*SRS*), ±1/*RT with the sign of the response, yields another* d *measure. The two measures are highly correlated and, importantly, the intercept in the regression of* logit P_1 *on mean SRS estimates the asymmetry of the starting point relative to the barriers, that is, the bias. New analyses of data from Patching, Englund, and Hellström (2011) show that this bias helps explain the variability of the time-and space order errors. Possible connections of the bias with the parameters in Hellström's (2003) sensation-weighting (SW) model are explored.*

What are the processes that lead to the response in a two-choice situation, such as forced-choice comparison of paired successive or simultaneous stimuli? Several models focus not only on the response probabilities, but also on the response time (RT), seen as comprising the time needed to accumulate sufficient information to reach one or the other response criterion (bound, barrier), where the accumulation is conceived as being stepwise (Link, 1975, 1992; Link & Heath, 1975) or continuous (Ratcliff, 1978, 2002). One purpose of this paper is to introduce a simple analysis method for timed two-alternative forced-choice (2AFC) tasks, developed within the "diffusion-to-bound framework" described by Shadlen et al. (2007; see also Palmer, Huk, & Shadlen, 2005), and demonstrate how it can detect initial bias, the offset of the starting point for the accumulation, and how its results relate to those of an advanced method of fitting the Ratcliff diffusion model (Vandekerckhove et al., 2011). Another purpose is to explore the nature of this bias and how it might contribute to the time- and space-order errors (TOEs, SOEs) in stimulus comparison.

Just as travel between two points is conveniently described, not only by the time required, but also by the speed and direction of travel, a very useful alternative concept and tool for describing a timed response is the *signed response speed (SRS)* = ±1/*RT*, with the sign of the response: + for *1ˢᵗ greater* or *left greater* and - for *2ⁿᵈ greater* or *right greater* (Hellström, 2008). The individual value of *SRS* represents the speed (μ) and direction (+ or -) of the accumulation process (ignoring the nondecision time, τ). This process is not going on forever even when μ = 0. Instead, its randomness makes it likely that a spurious deviation makes it hit one or the other barrier (*A*, -*A*) after an average number of steps that increases with A^2 and decreases with σ^2 as well as with |μ| (Shadlen et al., 2007, Eqs. 10.32-10.35). This makes very long RTs (near-zero values of *SRS*) unlikely and yields a bimodal distribution of *SRS* for a given stimulus pair, which on each side of zero is seen to be symmetric, resembling the normal distribution, and thus more user-friendly than the skewed distribution of RT.

Link (1992) and Shadlen et al. (2007) discuss in terms of the average number of steps to reach either border. Instead, we here partition the *SRS* values into positive and negative values, SRS^+ and SRS^-, corresponding to responses of *1st greater* and *2nd greater*. We also denote the corresponding mean accumulation rates by μ^{*+} and μ^{*-}. This yields $E(SRS^+) = \mu^{*+}/A$ and $E(SRS^-) = \mu^{*-}/A$. Thus, $\mu = P\,\mu^{*+} + (1-P)\,\mu^{*-}$, and $E(SRS) = P\,\mu^{*+}/A + (1-P)\,\mu^{*-}/A$. The individual *SRS* values estimate either μ^{*+}/A or μ^{*-}/A, but, as $\mu^{*+} = -\mu^{*-}$ (Shadlen et al., Eq. 10.38), $E(SRS) = \mu/A$, so that M_{SRS} for a given stimulus pair, over replications, yields an estimate of μ/A. Further, Shadlen et al.'s Eq. 10.38 yields logit $P = 2\,\mu A/\sigma^2$. Thus, from the slope *b* of the plot of logit *P* against M_{SRS} one can estimate logit $P\,/\,M_{SRS} = (2\,\mu A/\sigma^2)\,/\,(\mu/A) = 2A^2\,/\,\sigma^2$. This plot is usually linear near the origin with a slight inverted-*S* shaped deviation towards the ends; linearity can be improved by estimating τ and computing $SRS_{adj} = 1/(RT-\tau)$.

Effects of initial bias. In the model above, with no bias, and thus with bounds *A* and -*A*, it is clear that with $P = 1/2$, yielding logit $P = 0$, $E(SRS)$ also becomes 0. However, with an initially biased starting point *c*, the bounds, relative to this point, are *A-c* and -(*A+c*). So,

$$E(SRS) = P\,E(SRS^+) + (1-P)\,E(SRS^-) = P\,\mu^{*+}\,/\,(A-c) + (1-P)\,\mu^{*-}\,/\,(A+c). \qquad (1)$$

When bounds are symmetrical, $\mu^{*+} = -\mu^{*-}$ should hold. With bounds *A-c* and -(*A+c*), $E(SRS^+) = \mu^{*+}\,/\,(A-c)$ and $E(SRS^-) = \mu^{*-}\,/\,(A+c)$. Assuming that, still, $\mu^{*+} = -\mu^{*-}$, $E(SRS) = 0$ yields $P\,\mu^{*+}\,/\,(A-c) + (1-P)\,\mu^{*-}\,/\,(A+c) = 0$, so that $P/(1-P) = (A-c)/(A+c)$ and thus

$$\text{logit } P_{E(SRS)\,=\,0} = \ln[(1-q)/(1+q)], \qquad (2)$$

where $q = c/A$. Thus, with logit $P_{E(SRS)\,=\,0} = a$, $q = (1-e^a)/(1+e^a)$. For *a* small, *q* is close to -*a*/2, and for $a = 0$, $q = 0$. So, the intercept in the plot of logit *P* against M_{SRS} estimates the bias *c* relative to *A*. As the influence of τ vanishes when $SRS \rightarrow 0$, this estimate should be independent of τ. It can also be shown that under our assumptions, logit $P_{\mu\,=\,0} = -\text{logit } P_{E(SRS)\,=\,0}$.

Diffusion model analysis. Validation of the chronometric results is available by comparison with results from the fitting of Ratcliff's (2002) diffusion model to the data from our experiments. Computational estimation of the parameters of this model requires a large number of alternate responses (i.e., > 50) to each stimulus pair. Recently, however, Vandekerckhove et al. (2011) detail a hierarchical Bayesian fitting method, which analyzes the data from all participants while allowing for differences between participants. Thus, in addition to the analyses based on logit *P* and M_{SRS}, the Ratcliff Diffusion Model was fitted to the timed response data using the program WinBUGS (a Microsoft[TM] **Win**dows program for **B**ayesian inference **U**sing **G**ibbs **S**ampling). For each participant the fitting yielded one common value of the response criterion *A* and one value of the relative initial bias *q* for each ISI or spatial separation.

Method

In each of four microcomputer-controlled experiments, forty different participants compared successive or simultaneous paired circular spots of light. In Exp. 1, the spots had nine luminance levels from 3.5 to 5.9 cd/m^2 in steps of 0.3 cd/m^2. Their diameter was 5 mm. In Exp. 2, the luminance was 4.7 cd/m^2 and the diameter varied from 5.1 to 6.7 mm in steps of 0.2 mm. The temporal separation (ISI) between the light spots was 400, 800, 1600, or 3200 ms. Each spot was presented for 200 ms. Response time was measured from the onset of the 2nd stimulus. The intertrial interval was 3000 ms. In Exps. 3 and 4, the two light spots were presented simultaneously for 200 ms with a spatial separation of 10, 20, 40, or 80 mm. In Exp. 3, the

luminance ranged from 1.5 to 7.9 cd/m^2 in steps of 0.8 cd/m^2, and the diameter was 5 mm. In Exp. 4, the diameter ranged from 4.3 to 7.5 mm in steps of 0.4 mm, and the luminance was 4.7 cd/m^2. After Hellström (1978) the nine levels of luminance or diameter were combined factorially in their mean and difference, to create 25 different stimulus pairs. In Exps. 1 and 2, half of the participants were to press the left key if the first spot was the larger and the right key if the second spot was the larger (response assignment 1, RA1), and the other half had the reverse assignment (RA2). In Experiment 3, all participants were to press the left key if the left spot was the brighter and the right key if the right spot was the brighter (RA1). In Exp. 4, half of the participants were to press the left key if the left spot was the larger and the right key if the right spot was the larger (RA1), and the other half were to press the left key if the left spot was the smaller and the right key if the right spot was the smaller (RA2).

Results and Discussion

E*stimates of initial bias.* The q values from the chronometric (CM) and WinBUGs (WB) analyses were found to be linearly related, with the following rs, in order from the lowest to the highest ISI or separation: Exp. 1: .865, .858, .812, .813; Exp. 2: .755, .803, .782, .869; Exp. 3: .531, .707, .625, .609; Exp. 4: .822, .609, .541, .724. All rs were significantly > 0 (ps $< .001$). In linear regressions with the WB values as DV, the intercepts were *ns* (ps $> .05$), and the slopes were on average 2.67 (Exp. 1), 2.50 (Exp. 2), 1.43 (Exp. 3), and 1.70 (Exp. 4). Apparently, the CM and WB estimated q values measure the same thing, although on different scales, the reason for this discrepancy being unclear. We here favor the WB estimates as they are built on more data and seem more reliable.

    WB-estimated q values. (Means are shown as "intercept" in Figs. 2a-d.) For Exp. 2, means were significantly < 0 for each ISI (ps .025, .022, .007, .001). For Exps. 1, 3, and 4, means were not significantly $\neq 0$ for any ISI or separation.

    Mean TOE or SOE. For Exps. 1, 2, and 4, $M_{TOE/SOE}$ (estimated using M_{SRS}) was not significantly $\neq 0$ for any ISI. For Exp. 3, M_{SOE} was > 0 for 40 mm separation ($p = .018$).

Modeling Stimulus Comparison and the TOE/SOE

Sensation weighting. Hellström's (e.g., 2003) sensation weighting (SW) model accounts for effects of time- or space-order in stimulus comparison. In the simplified version of the model, the subjective difference d is described by this equation:

$$d = W_1 \psi_1 - W_2 \psi_2 + U, \tag{3}$$

where k is a scale constant, $W_1 = k s_1$, and $W_2 = k s_2$ (cf. Eq. 4). When ψ_1 and ψ_2 are at their mean values, U becomes a measure of the TOE or SOE.

    For Exp. 1, W_1-W_2 was < 0 for all ISIs, significantly so for 800 and 3200 ms (ps $= .001$), and with $p = .052$ for 1600 ms. For Exp. 2, 400 ms, W_1-W_2 was > 0 ($p = .038$). For Exp. 3, W_1-W_2 was not significantly $\neq 0$ for any separation. For Exp. 4, 80 mm, W_1-W_2 was > 0 ($p = .022$). Thus, differential sensation weighting is in operation.

Initial Bias in Stimulus Comparison

    Empirical relation of initial bias to TOE/SOE. Correlations were computed (for each ISI, over participants) between TOE/SOE (estimated using M_{SRS}) and q (estimated by WB).

All *r*s were positive [Exp. 1: ISI = 400 ms: .920; 800 ms: .875; 1600 ms: .753; 3200 ms: .624 (*p*s < .001); Exp. 2: .578; .569; .651; .639 (*p*s < .001); Exp. 3: separation 10 mm: .500, *p* = .001; 20 mm: .524, *p* =.001; 40 mm: .348, *p* = .028; 80 mm: .418, *p* = .007. Exp. 4: .675; .650; .691; .556 (*p*s < .001)]. Similar results, but with somewhat lower *r*s, were obtained for the *U* values estimated from logit *P*. Thus, it seems clear that the measured initial bias contributes to the TOE/SOE. It remains to find out how.

The full equation specified by Hellström's (e.g., 2003) SW model may be written,

$$d = k [s_1 (\psi_1 - \psi_{r1}) - s_2 (\psi_2 - \psi_{r2}) + (\psi_{r1} - \psi_{r2})] + b, \qquad (4)$$

where *d* is the scaled subjective difference between the compared stimuli, and *k* is a scale constant. ψ_1 and ψ_2 are the sensation magnitudes of the stimuli, s_1 and s_2 weighting coefficients, and ψ_{r1} and ψ_{r2} the subjective magnitudes that correspond to the current *reference levels* (ReLs). *b* is a constant term, which captures a possible contribution to *d* independent of the weighting mechanism; here, any kind of bias that is independent of accumulated evidence.

Possible nature of initial bias (q). (1) *Response preference*: One possibility is that *q* reflects the SW model's bias term, *b*, which can be seen as describing a general preference for one response over the other. If so, *q* should be constant across ISIs or stimulus distances, and thus highly intercorrelated across conditions within participants. (2) *ReL-primed evidence accumulation*: Another conjecture is that evidence accumulation starts from the reference levels with the null assumptions $\psi_1 = \psi_{r1}$ and $\psi_2 = \psi_{r2}$, so that with $\mu = 0$, ψ_1 and ψ_2 are identified with their respective ReLs, ψ_{r1} and ψ_{r2}. This would save response time if the ReLs are near the current values of ψ_1 and ψ_2. $k (\psi_{r1} - \psi_{r2})$ then becomes the initial bias, *q*, in the CM or WB analysis, and μ will only reflect the remainder of the right member of Eq. 4. According to the CM analysis, for $\mu = 0$, logit *P* = -ln [(1+*q*)/(1-*q*)], which should thus estimate $k (\psi_{r1} - \psi_{r2})$. When $\mu \neq 0$, this term is added to the measure of μ, and [as shown by numerical exploration of Link's (1992) Eq. 11.2] 'amplified' to a degree that increases with μ.

According to both suggested accounts of *q*, the *q* value for each ISI or spatial separation should reflect the same underlying bias. To check this, for each experiment the four WB-estimated *q* values for each participant were submitted to repeated measures ANOVAs (using SPSS 19 GLM with polynomial contrasts and multivariate tests) to check for a possible effect of ISI or spatial separation. For Exp. 1, the effect of ISI was significant, *p* = .016 (*p* = .002 for the quadratic effect). For Exp. 2, the effect of ISI was *ns*. For Exp. 3, the effect of spatial separation had *p* = .051. For Exp. 4, this effect was *ns*.

Thus, the evidence on the constancy of mean *q* across conditions is mixed and does not favor the notion that *q* reflects a general response preference. Also, in order to identify *q* with $k (\psi_{r1} - \psi_{r2})$, we would have to accept the idea that the inter-ReL distance, $\psi_{r1} - \psi_{r2}$, can vary greatly with the stimulus separation. A more likely possibility seems to be that *q* reflects a response readiness that is dependent on the temporal or spatial stimulus separation.

Component structure of initial bias. The latter conclusion makes it of interest to determine if the variability of *q* across separations might have a component structure. Therefore, principal component analyses were conducted, using all eight *q* values, from WB as well as CM analyses. For each of Exps. 1, 2, 3, and 4, two components, with eigenvalues 5.34 and 1.42; 6.06 and 0.66; 3.81 and 1.84; 4.21 and 1.31, were extracted. In Fig. 1, the unrotated component loadings are plotted against the temporal or spatial separation. *q* can be seen as a linear combination of two components. The plots all show one component, *C1*, that loads

similarly on q for all separations, and another, *C2*, whose loading is negative for small, and positive for large separations. Next, the contribution of *C1* and *C2* to the TOE and SOE was investigated by linear regression of U (estimated by M_{SRS}) on the unrotated component scores. The results are shown in Fig. 2. About half of the variance of U can be ascribed to the influence of *C1* and *C2*. When contributing to the TOE/SOE (by the regression coefficients in Fig. 2), *C1* and *C2* are weighted-in differently than when determining q (by the loadings in Fig. 1).

One might expect that the *response assignment* (RA1 vs. RA2) could influence the response criterion *A*. For Exps. 1 and 2, WB-estimated *A* was higher for RA2, $p = .027$ (Exp. 1) and $p = .001$ (Exp. 2). (For Exp. 4, *A* was likewise higher for RA2, but with $p = .165$, *ns*). Thus, it takes a little extra accumulated evidence to respond using the less straightforward response assignment, RA2.
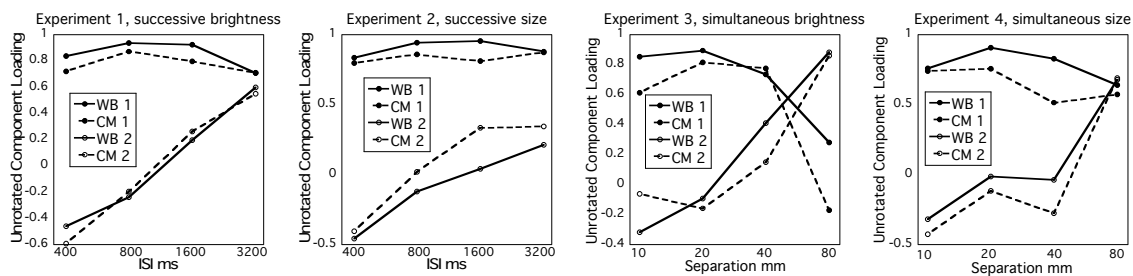

Fig. 1. Unrotated loadings from principal component analyses of q estimates from WinBUGs and chronometric analyses, plotted against temporal or spatial stimulus separation.
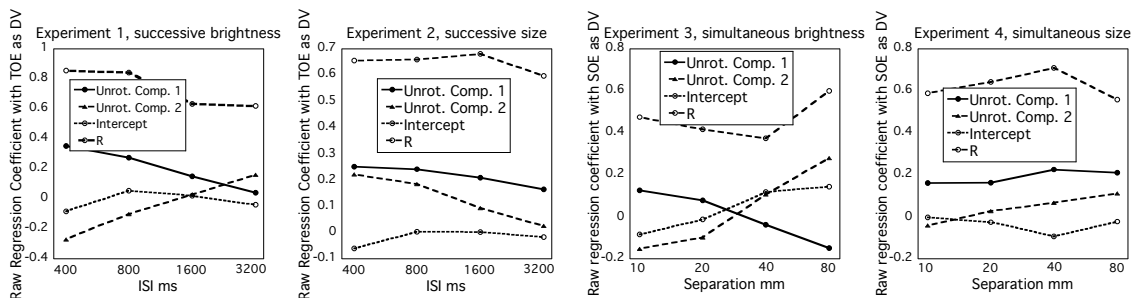

Fig. 2. Regression coefficients, with TOE or SOE as dependent variable, for unrotated scores from principal component analyses of q estimates.

Likewise, the RA might influence the initial bias q, or its components. For Exp. 1, mean WB-estimated q values for each ISI were negative for RA1 but positive for RA2. A repeated-measures ANOVA with multivariate tests showed a significant effect of RA, $F(1,38) = 6.20$, $p = .017$, but *ns* effects of ISI and of RA x ISI. Likewise, *C1* was negative for RA1 but positive for RA2, $p = .012$ for the difference. These results suggest that there was a general preference for using the right key, regardless of whether this meant judging the first or the second stimulus as the brighter. (All participants in Exp. 1 were right-handed.) However, the effect is not present for Exps. 2 and 4 (in Exp. 3, only RA1 was used). For Exp. 2, *C2* was higher for RA1, $p = .044$.

Effects of gender and age were also explored. *Gender*: For Exp. 1, *C1* was higher for females than for males, $t(38) = 2.025$, $p = .0499$. For Exp. 4, *C2* was higher for males than for females, $t(38) = 2.21$, $p = .033$. For Exp. 2, *A* (WB estimated) was higher for males than for females, $t(38) = 2.76$, $p = .009$. *Age*: For Exp. 1 *A* increased with age ($r = .35$, $p = .027$). For Exp. 4, *A* likewise tended to increase with age, $r = .31$, *ns* ($p = .052$), and *C2* decreased with

age, $r = -.415$, $p = .008$. Although these effects of gender and age are unpredicted and might be spurious, they suggest that individual response tendencies, depending on personal characteristics, exist. This comes out more clearly for *C1* and *C2* than for *q*, for which there were no significant effects. Whereas *C1* is largely independent of the ISI, C2 indexes an initial bias tendency used in opposite ways with a small versus a large separation. One might like to think that this tendency is somehow rational, using stimulus separation as a cue.

Conclusion

Although the initial bias, *q*, is not the main factor behind the TOE/SOE, it contributes to it. The nature of *q* could not be fully determined from the present data. In some types of experiments it may be partly due to a hand preference and thus sensitive to the response assignment. Yet, *q* also has a component that depends, in a participant-specific manner, on the temporal or spatial stimulus separation.

Acknowledgement

References

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

Hellström, Å (2008). Time- and space-order effects in stimulus comparison in the light of response-time data. In B. A. Schneider, B. M. Ben-David, S. Parker, & W. Wong (Eds.). *Fechner Day 2008. Proceedings of the 24th Annual Meeting of the International Society for Psychophysics* (pp. 139-144). Toronto: International Society for Psychophysics.

Link, S. W. (1975). The relative judgment theory of two choice response time. *Journal of Mathematical Psychology, 12*, 114-115.

Link, S. W., & Heath, R. A. (1975). A sequential theory of psychological discrimination. *Psychometrika, 40*, 77-105.

Link, S. W. (1992). *The wave theory of difference and similarity*. Hillsdale, NJ, USA: Lawrence Erlbaum Associates.

Palmer, J., Huk, A. C., & Shadlen, M. N. (2005). The effect of stimulus strength on the speed and accuracy of a perceptual decision. *Journal of Vision, 5*, 376-404.

Patching, G. R., Englund, M. P., & Hellström, Å. (2011). *Time- and space-order effects in timed brightness and size discrimination of paired visual stimuli.* Manuscript in revision.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review, 85*, 59-108.

Ratcliff, R. (2002). A diffusion model account of response time and accuracy in a brightness discrimination task: Fitting real data and failing to fit fake but plausible data. *Psychonomic Bulletin & Review, 9*, 278-291.

Shadlen, M. N., Hanks, T. D., Churchland, A. K., Kiani, R., & Yang, T. (2007). The speed and accuracy of a simple perceptual decision: A mathematical primer. In K. Doya, S. Ishii, A. Pouget, Alexandre, & R. P. N. Rao (Eds.), *Bayesian brain: Probabilistic approaches to neural coding* (pp. 209-237). Cambridge, MA, USA: MIT Press.

Vandekerckhove, J., Tuerlinckx, F., & Lee, M. D. (2011). Hierarchical diffusion models for two-choice response times. *Psychological Methods, 16,* 44-62.