A robust anchoring effect in linear ordering

Ulrich von Hecker[1], Karl Christoph Klauer[2],

and André Aßfalg[2]

[1]School of Psychology, Cardiff University, UK.

[2]Albert-Ludwigs-Universität, Freiburg, Germany.

Author Note

Correspondence should be addressed to Ulrich von Hecker, School of Psychology, Cardiff University, Tower Building, Park Place, Cardiff, CF10 3AT, United Kingdom. Electronic mail may be sent to vonheckeru@cardiff.ac.uk.

We would like to thank Dr. Rainer Leonhart for technical help as well as Sascha Topolinski and an anonymous reviewer for helpful comments on an earlier version. We also thank the Editor Dirk Wentura for helpful suggestions. This research was preregistered with Open Science Framework, see https://osf.io/emauf and https://osf.io/6sznj/.

The data on which this research is based are stored alongside the preregistration information under the indicated links.

Abstract

The robustness of effects indicating a spatial component associated with abstract reasoning is tested. Judgments regarding hierarchical orderings tend to be faster and more accurate when the dominant element in any pair from the order (e.g., the older, richer, etc.) is presented on the left of the screen as compared to the right (*left-anchoring* effect). This signature effect is investigated in three conditions (Experiment 1), each implementing a different timing regime for the elements in each pair, during learning. Thereby, the construction of a mental representation of the ordering was exposed to a potentially competing spatial simulation, that is, the well-known "mental timeline" with orientation from left (present) to right (future). First, the left-anchoring effect for order representations remained significant when timeline information was congruent with the presumed left-anchoring process, that is, the dominant element in a pair was always presented first. Second, the same effect remained also significant when the timeline-related information was random, that is, the dominant element being presented either first or second. Third, the same effect was found to be still significant, when the timeline-related information was contrary to the left-anchoring process, that is, the dominant element being presented always second. Experiment 2 replicates the target effect under random timeline information, controlling for color as a stimulus feature. The results are discussed in the context of a theoretical model that integrates basic assumptions about acquired reading/writing habits as a scaffold for spatial simulation, and primacy/dominance representation within such spatial simulations.

(244 words)

Key Words: reasoning, mental models, linear orders, spatial processing

Word Count: 4856 (excluding References, Tables, Figures, and Abstract)

People appear to use spatial representations to learn rank orders, for example, pairs A > B, B > C, etc., with A, B etc. being replaced by first names, and “>” denoting some comparator, replaced, for example, by “older”, or “taller”. Such comparators represent abstract concepts, that is, they do not *a priori* imply any spatial extension, situation, or orientation in space. There exists long-standing experimental evidence suggesting that spatial processes nevertheless might be involved in the learning of such materials. For example, the so-called *symbolic distance effect (SDE)* has often been demonstrated. Queries on pairs of wider distances are usually responded to with greater speed and accuracy than those of narrower distances (e.g., DeSoto, London, & Handel, 1965; Smith & Foos, 1975; for a review Leth-Steensen & Marley, 2000). These classical results have been interpreted as indicative of an analogue representation being formed out of the initial piecemeal information that was learned (e.g., Holyoak & Patterson, 1981). This interpretation has been challenged on the grounds of demonstrations how equivalent experimental effects can be predicted without the assumption of spatial processes taking place (see Leth-Steensen & Marley, 2000). Therefore it is still unclear to what extent a spatial representation of an order is necessary for distance effects to occur, or even, more mildly, to what extent such effects are reliably associated with spatial representations.

The approach taken here rests on the argument that support for the contribution of spatial processes in forming mental representations might consist in the demonstration of lateral asymmetries. There exists evidence for a left-bias, that is, left-to-right oriented representation of numbers (Dehaene et al., 1993; Gevers, Verguts, Reynvoet, Cassens, & Fias, 2006; Ito & Hatta, 2004), and the mental time line in Western participants (Fuhrman & Boroditsky, Gevers, Reynvoet, & Fias, 2003; Gevers, Caessens, & Fias, 2005; Gevers, Reynvoet, & Fias, 2003; Ouellet, Santiago, Lupianez, Perez, & Funes, 2007; Santiago, Funes,

& Lupiáñez, 2010; Tversky, Kugelmass, & Winter, 1991). In the present research, we focus on this *left-anchoring effect*: Participants who are used to read from left to right are predicted to be faster to respond and, perhaps, more accurate, when the dominant element of a linear order (e.g., the oldest, fastest etc.) appears on the left side, compared to the right side, in a pair of elements. We predict a spatial bias to occur for the learning of abstract orderings as mentioned above. Crucially, we investigate two potentially determining factors for such a bias: a) dimensional magnitude ("older", or "taller", etc.) and b) the mental time line, as explained below.

Basic assumptions

In an earlier series of experiments, we (von Hecker, Klauer, Wolf, and Fazilat-Pour, 2016) provided initial evidence for spatial processes associated with the learning of (and subsequent reasoning about) linear orders, such as *A is older than B, B is older than C, C is older than D*, and so forth. As conclusion, spatial processing was seen as a part of constructing a mental representation of such orders. This claim rested on the finding that when later prompted to indicate, for example, the *older* one in any pair such as AB, AC, AD, BC, BD, etc., participants were quicker (and sometimes more accurate) when the *older* person was presented on the *left side* of the computer screen, as compared to the right side (*left-anchoring effect*).

This bias was explained by one general, and two more specific assumptions. As a *general assumption*, the rank order is mentally represented by a horizontal line, that is, a spatial mental model as suggested in classical literature (e.g., DeSoto, London, & Handel, 1965; Smith & Foos, 1975; for a review Leth-Steensen & Marley, 2000). *Specific Assumption 1* explains the directionality of that line: There is a tendency to use learned reading/writing habits as a scaffold for determining the origin, or starting point, of the simulated order (i.e., *left* for Westerners; the effect reversed to a *right side* advantage when tested in Iranian samples with right-to-left reading/writing background, von Hecker et al., 2016). The line is then constructed in the direction away from the origin, that is, from left to right in Westerners. *Specific Assumption 2* explains the semantics of the line: To determine which end of the dimension (e.g., *old* or *young*, *poor* or *rich*, etc.) to place at the origin, one uses *primacy* in a very general sense. Primacy within linear orders is usually derived from dominance in magnitude (e.g., which element in a pair represents the greater magnitude, for example, of *age*, *power*, *wealth*, or even *green-ness,* compared to the other, see von Hecker et al., 2016)[1]. Response times should be faster in trials that show the dominant element on the screen in the same spatial orientation as it has within the mental representation (congruent trials), as compared to trials in which both orientations are opposite (incongruent trials).

Starting from these assumptions, new questions arise because primacy can clearly be derived from other indicators, too. For example, sequential steps in counting may influence the orientation of the mental number line from small numbers (left) to larger numbers (right) in Western cultures (see Dehaene, Bossini, & Giraux, 1993; Maass & Russo, 2003; Tversky, Kugelmass, & Winter, 1991; Zebian, 2005). An influence from the sequence of action elements in time (e.g., counting steps) would point to a potential role of the time line in determining the mental representation of order.

Experiment 1

For the left-anchoring effect, it is unclear to what extent the assumed spatial construction is genuinely related to the concept of dimensional magnitude. Alternatively, the constructive process could be guided or at least be influenced by other mental representations which are salient at the time of learning, causing the left-anchoring effect. Time is a candidate concept here. It is well established that time is itself represented as a dimension stretching out from left to right (from early to late) in Western participants (see above). It is therefore possible that participants would take their pre-existing, well-overlearned time line as a model or scaffold for the actual order dimension they are in the process of constructing. Hence, the experiment to be reported addresses the possible influence of the mental time line on the construction of an abstract linear order.

Three conditions were run in a between-groups design. In Group 1, the primacy of elements in each presented pair during learning was consistently identical for time and magnitude, i.e., the element presented first was always the dominant element as in von Hecker et al. (2016). In Group 2, primacy of elements in a pair was random with regard to time, that is, the first or the second element in a pair was randomly chosen to be the dominant element. In Group 3, the second element in a pair was always presented as dominant element, thus pitting time and dominance against each other, as competing for primacy. In a subsequent test phase, two elements were presented horizontally on the screen, with the dominant element being on the left or on the right. Participants had to quickly press a button on the side where they thought the dominant element appeared (see below). Responses were expected to be faster and more accurate when the side of the presented dominant element in a pair corresponded to the side where the maximum of the dimension was anchored within the hypothetical mental model. Responses were expected to be slower when there was no such correspondence.

We predicted that if spatial processing in the service of constructing the actual linear order is genuinely tied to the left-anchoring effect, and primacy derived from dimensional dominance alone, then such an effect should be robust across all three group conditions. If however spatial construction is influenced or even entirely determined, at stimulus acquisition, by the pre-existing orientation of the time line, and primacy derived from it, then there should be left-anchoring in Group 1, no anchoring effect in Group 2 (dominant element in random position), and a right-anchoring effect in Group 3. Whether or not the group factor moderates the left-anchoring effect and, in particular, the question whether any anchoring effect occurs in Group 2 (when temporal priority is controlled for) is of central theoretical interest for interpreting previously obtained anchoring effects (e.g., von Hecker et al., 2016) as being independent from other salient dimensions with ties to the horizontal line. Therefore, the present intention was to test for main effects associated with the side of dominant element, as well as for an interaction between that factor and the group factor.

With the current assumptions, we predicted that participants usually construct an analogue spatial mental model as suggested in classical literature (e.g., DeSoto, London, & Handel, 1965; Smith & Foos, 1975; for a review Leth-Steensen & Marley, 2000). For any pair of two elements from such a spatial mental model, a display exhibiting a spatial orientation which is incongruent with the spatial orientation in the mental model (e.g., the dominant element appearing on the right side) should cause interference, and slow down the response. In order to test the construction of the model, and to replicate the SDE, we will introduce pair distance as a factor in all analyses. According to the SDE, responses are expected to be more accurate, and also quicker, for pairs at a wider as compared to a narrower distance within the spatial model.

Method

This research was pre-registered with the Open Science Framework (OSF, von Hecker, Klauer, & Aßfalg, 2017).

Participants

In the series of experiments reported in von Hecker et al. (2016) we obtained a medium-sized effect of Cohen's $d_z = .48$ on average for the anchoring effect (Borenstein, Hedges, Higgins, & Rothstein, 2009) with meta-analytic procedures[2]. Since we planned to be able to analyse groups separately, we conducted a power analysis based on the above effect size, a t-test between two dependent measures (dominant element left or right), one-tailed testing, stipulating a power of 1 - β = .80 and an alpha-level of .05 [deleted: "…. a correlation between repeated measures of r = .8"] (GPower 3.1.3., Faul, Erdfelder, Buchner, & Lang, 2009), yielding a minimal required sample size of n = 29 per group. Taking a conservative approach on the basis of this analysis, we planned for a minimum of 40 participants per group, and 120 participants in total[3].

All participants were sampled from the subject pool of the second and third authors' department, mostly comprising undergraduate students with German-spoken backgrounds. Participants received course credit or €5.00 for their participation. They were randomly assigned to one of the three groups. Inclusion criteria were: Age between 17 and 45 years, mother tongue German, normal or corrected-to-normal visual acuity. These criteria were monitored by the experimenters but sometimes, for logistical reasons, could be checked only after the experiment had been completed. In such cases we excluded participants violating the inclusion criteria after participation. Furthermore, participants with incomplete data due to computer problems were also excluded. In total, data from 132 participants were collected (Group 1: 33 female, 10 male; Group 2: 30 female, 15 male; Group 3: 35 female, 9 male; total mean age = 24.4 years). See Results section below for the impact of exclusion procedures on that number.

Materials

Six German adjectives were used to denote six semantically different order relations (German translations of *older, richer, taller, smarter, stronger,* and *faster*), to be used in the six experimental blocks (one adjective per block, and "*more sportive*" for a practice block). In each block, a separate set of five German first names was randomly assigned to that block (here denoted as A, B, C, D, and E), out of a large pool of names, matched for frequency of recent use. The set of names used matched the participant's gender, in order to avoid differences between in- versus out-group perceptions with respect to some of the semantic comparators we used. All possible pairwise combinations ($n = 10$) of the five names constituted the stimuli in each block. In the learning phase, these pairs were presented centrally, one name after the other, whereas in the test phase, pairs were presented horizontally with a gap between them. Presentation was in blue, green, or black letters on a white background.

Procedure

A PsychoPy (Peirce, 2007) script used for running the experiment is available in the OSF repository for download (https://osf.io/cn89q/). After reading instructions, participants were seated approximately 60 cm in front of a computer screen. First, they underwent a practice block consisting of four pairs to be learned, each after a 1 s fixation cross. A pair was presented for practice at the centre of the screen, one name after the other, each name appearing for 1.5 s and separated by a blank screen for 800 ms. Instructions asked participants in all three groups to consider the name written in *blue* letters as the dominant one (e.g., the older person). Within each pair trial, either the first or the second name was shown in blue colour (and the other name in green colour). Immediately after presenting a learning pair, a test item was prompted for practice. In it, the two names were presented horizontally, with the dominant name randomly assigned to the left or the right. Participants were instructed to press one of two horizontally arranged response buttons according to the side of the dominant name. There was a 2 s blank screen separation between the practice trials. When a participant committed an error during practice, they were presented with the practice instructions and the four practice trials again, until they achieved an error-free practice run.

After practice, participants were then presented with the learning phase of the first experimental block. In the learning phase, participants viewed all ten possible pairs twice, in two cycles such that all possible pairs occurred once before all of them were presented again. The sequence of pairs was determined randomly for each cycle, excluding a repetition of pairs at the connection between the two cycles. The timing of stimuli was identical to the way described above, and 2 s blank screen intervals occurred between individual learning pairs. In all groups, the dominant and non-dominant names appeared in blue and green letters, respectively. In Group 1, the dominant name appeared always first. In Group 2, either the first or the second name was presented in blue letters as dominant, determined randomly for each pair, but with the order of the two names within a pair held constant between the two cycles of pairs during learning (as in the two other groups as well). In Group 3, the dominant name appeared always second. As a result of this part of the procedures, participants had been given all necessary information in order to construct a mental model of the hierarchy, incorporating all elements (i.e., names A to E) in an ordered sequence from maximum (i.e., the oldest) to minimum (i.e., the least old). Whether or not such an order had been indeed constructed was ascertained by testing for replication of the SDE (see below), expecting more accurate and faster responses to pairs from wider than narrower distances in rank position.

After each learning phase, a test phase followed immediately. Each test pair trial began with a 1 s fixation stimulus (“X”) at the centre of the screen. After that interval, the pair was presented in black letters as described above, with an open response interval. There was a 2000 ms blank screen interval between any two consecutive test trials. A test phase consisted of 40 items, that is, all ten possible combinations were presented four times. Test cycles, invisible to the participant, were programmed such that all ten pairs were presented before the next cycle began. Left- vs. right-orientation of the dominant person was determined randomly for each pair across the four cycles such that each pair appeared twice with the dominant person left and twice with the dominant person right. Two specially prepared mouse devices were used for responses. These devices had marked buttons to indicate the use of the forefinger of each hand, that is, on the right mouse button of the left mouse, and on the left mouse button of the right mouse. This enabled participants to hold their hands in a comfortable position using both mouse devices simultaneously, while being also able to quickly respond. Before each block started, an instruction screen told participants to keep the two index fingers on the marked mouse buttons during the experiment. The participants' task in each test phase was to indicate as quickly and accurately as possible the side where the dominant person appeared (i.e., the older, taller, etc.).

Learning and test phases as described occurred six times, corresponding to the six order dimensions *older, richer, taller, smarter, stronger,* and *faster,* assigned to blocks in random order. In between blocks, a series of four easy arithmetic problems was solved as interpolated task, to clear participants' short-term memory from the previous set of names. One session lasted between 30 and 40 minutes, including debriefing.

Results

Data preparation

Amongst the 132 initial participants (43 in Group 1, 45 in Group 2, and 44 in Group 3), three had eyesight problems (two in Group 2 and one in Group 1) and were excluded from the analysis. Two participants were further excluded because German was not their native language (one each from Groups 1 and 2). Average accuracy and latency was continuously monitored via Boxplot tests while targeting the pre-registered sample sizes for the three groups. In this process, that is, prior to any analyses, participants were excluded when they were extreme outliers according to Tukey's criterion (i.e., three times the interquartile range above or below the upper or lower quartile, respectively, in the current participant sample's distribution of average accuracy rates or average correct response latencies). Using this method, three participants were excluded on the grounds of low accuracy (two from Group 2 and one from Group 3), as well as two participants on the grounds of slow responding (two from Group 2), such that the remaining sample sizes were 42 for Group 1, 38 for Group 2, and 42 for Group 3. Note that the exclusion criteria had been preregistered (von Hecker et al., 2017).

With respect to the hypotheses, we analysed latencies, as well as accuracy data for correct responses. In previous studies (von Hecker et al., 2016), the target anchoring effects, although mostly occurring in latencies, sometimes also occurred in accuracies, or in both. Latencies were trimmed within each participant according to the Tukey criterion based on excluding outliers with values larger (smaller) than the upper (lower) quartile plus (minus) 1.5 times the interquartile range in the individual's distribution of latencies (see Clark-Carter, 2004, Chapter 9). All data were analysed using t-tests and linear mixed models (Jaeger, 2008; Judd, Westfall, & Kenny, 2012), using the package afex (Singmann & Bolker, 2014) within the programming language R (R Core Team, 2013). Cohen's d_z will be reported for those effects that are relevant to the hypotheses. We also report Bayes factor values for critical comparisons related to the hypotheses, as calculated using the package Bayes Factor in R (Morey, Rouder, & Jamil, 2015)[4].

Latency and accuracy data were each analysed in two steps. Linear mixed models were estimated (for the accuracy data: generalized linear mixed models with logistic link function) with *participants* as random factor, and it was first determined which random structure would best fit the data. Subsequently, a final model with appropriate random effects was used to evaluate fixed effects (see Jaeger, 2008; Judd, Westfall, & Kenny, 2012). Such models provide more test power as compared to the conventional ANOVA approach by including, if statistically appropriate, random slopes for within-Ss factors, as a function of participants, into the predictive part of the model. The strategy for selecting a model with appropriate random-effects structure is described in the Appendix, along with information about the particular random-effects structure adopted for each model in each experiment. We report effect sizes (Cohen's d_z) for those effects that are interpreted as relevant to the main hypothesis. Here, Cohen's d_z refers to the standardized mean difference between the conditions with dominant elements on the left vs. right, across all other conditions, while accounting for correlations of dependent measures (Morris & DeShon, 2002).

Response latencies

Table 1 lists the response latencies. The final model had fixed effects for group (1 – 3), block (1 – 6), side of dominant element (left vs. right) and pair distance (1 step, …, 4 steps). A main effect of block showed that average responses got quicker across blocks 1 to 6 (M_1 = 1033 ms; M_2 = 995 ms; M_3 = 982 ms; M_4 = 966 ms, M_5 = 941 ms; M_6 = 948 ms) reflecting increased practice with the task, $F(5, 131.35) = 2.33$; $p = .05$. Relevant to the main hypothesis, side of dominant element had a significant effect, $F(1, 24269.64) = 28.83$; $p < .001$, $d_z = 0.34$. Participants were faster responding to the dominant stimulus in a pair if that stimulus was presented on the left side (M_{left} = 963 ms) than the right side (M_{right} = 992 ms). There was a main effect of pair distance, $F(3, 213.41) = 198.53$; $p < .001$, showing again the symbolic distance effect, ($M_{1\ step}$ = 1109 ms; $M_{2\ steps}$ = 1045 ms; $M_{3\ steps}$ = 938 ms; $M_{4\ steps}$ = 815 ms), that is, participants responded faster when the distance in steps between the elements of a pair was large compared to small. In terms of interactions, group significantly interacted with block, $F(10, 131.35) = 2.18$; $p = .02$, suggesting that the practice effect across blocks was most pronounced in Group 3 (dominant element always second). Also, there was a significant interaction between block and pair distance, such that the symbolic distance effect, measured as the difference in latency (ms) between the narrowest (1 step) and the widest pair distance (4 steps), increased from Block 1 to a plateau in Blocks 4 - 6 (Block 1: 189 ms; Block 2: 295 ms; Block 3: 308 ms; Block 4: 341 ms; Block 5: 320 ms; Block 6: 311 ms). No further significant effects emerged. Notably, there was no significant interaction between group and side of dominant element, $F(2, 24270.73) = 1.39$; $p = .25$, which suggests that the left-anchoring effect was not moderated by the group factor. The Bayes factor[4] in favour of a model identical to the one specified above, but without the group x side of dominant element interaction, against the full model as initially specified, was 381.91, indicating "extreme evidence" (Jeffreys, 1961) in favour of the superior, more parsimonious model, which constitutes support against incorporating this particular interaction as a part of the model.

To test robustness of the left-anchoring effect in each group, we further examined the side of dominant element effect, running models of similar structure as above for each group separately. In Group 1, the crucial main effect of dominant side was significant, $F(1, 8523.63) = 11.30$; $p = .0008$, $d_z = 0.40$ ($M_{left} = 963$ ms; $M_{right} = 990$ ms) which was also the case in Group 2, $F(1, 7556.76) = 17.44$; $p < .0001$, $d_z = 0.48$ ($M_{left} = 933$ ms; $M_{right} = 974$ ms), as it was in Group 3, $F(1, 8229.65) = 3.31$; $p = .07$, $d_z = 0.21$ ($M_{left} = 991$ ms; $M_{right} = 1009$ ms) which corresponds, in line with our power-planning (see above) to a one-tailed t-test result at $p = .035$[5].

Accuracy

Table 1 lists the mean accuracies. The overall error level was 11%. The final model had the same fixed effect structure as the one for latencies. Block yielded a significant effect, $\chi^2(5) = 91.62$; $p < .001$, showing an increase in accuracy over the first five blocks of the experiment ($M_1 = .851$; $M_2 = .914$; $M_3 = .926$; $M_4 = .933$, $M_5 = .946$; $M_6 = .930$), probably reflecting a practice effect. Pair distance had a significant effect, $\chi^2(3) = 226.85$; $p < .001$, replicating the classical symbolic distance effect (see above), that is, participants were more accurate when the distance in steps between the elements of a pair was large compared to small ($M_{1\ step} = .841$; $M_{2\ steps} = .916$; $M_{3\ steps} = .947$; $M_{4\ steps} = .962$). The interaction between group and block was significant as well, $\chi^2(10) = 20.29$; $p < .03$, indicating that the presumed practice effect was most pronounced in Group 2, which is plausible because the task was comparatively most complex in this group (dominant element in random position). Lastly, the interaction between block and pair distance yielded a significant effect, $\chi^2(15) = 48.68$; $p < .001$. Inspecting the differences in percentage correct between distances reflecting the narrowest (1 step) and the widest (4 steps) pair distance, it appears that the symbolic distance effect is more pronounced in the first half of the experiment than in the second half (Block 1: 15%; Block 2: 13%; Block 3: 10%; Block 4: 9%; Block 5: 15%; Block 6: 11%). In terms of the main hypothesis, there are no significant effects in the accuracy data, neither in terms of a side of dominant element main effect, $\chi^2(1) = 0.0$; $p > .99$; Bayes factor: 6.48 (moderate evidence in favour of the full model without side of dominant element as predictor and its interactions, compared to the full model itself), nor in terms of the particular interaction between group and side of dominant element, $\chi^2(2) = 1.45$; $p = .48$; Bayes factor: 169.89 (extreme evidence in favour of the full model without that particular interaction versus the full model itself). This outcome is quite common. In previous research, the left-anchoring effect has been visible mainly in response latencies, and only sporadically in accuracies (see von Hecker et al., 2016).

Experiment 2

Experiment 2 was conducted for two reasons. First, we aimed at a replication of the left-anchoring effect especially under the conditions of Group 2, where the mental time line could not serve as diagnostic scaffold for model construction, since first and second elements within the presented series of pairs were randomly chosen to be the dominant one. In this situation, with no other cue than dimensional dominance, a spontaneous left-anchoring is predicted and should be replicable. Second, as prompted by a reviewer's query, we wanted to investigate whether the color of the dominant element could play any role. For this purpose, and different from the method used for Group 2 in Experiment 1, half of the participants in the present experiment received displays in which dominance was indicated by names written in blue, whereas dominance was indicated by names written in green for the remainder of participants.

Method

This research was also pre-registered with the Open Science Framework (OSF, von Hecker, Klauer, & Aßfalg, 2018).

Participants

All participants were sampled from the subject pool of the second and third authors' department, mostly comprising undergraduate students with German-spoken backgrounds. Forty-nine participants were recruited who received course credit or €5.00 for their participation (27 female, 22 male; total mean age = 24.7 years). The inclusion criteria were the same as in Experiment 1.

Materials and Procedure

All materials and experimental procedures were the same as in Experiment 1, except the assignment of color to stimulus dominance, which was *blue* for one half of the sample, and *green* for the other half. Instructions during the practice trials were modified accordingly.

Results and Discussion

Data preparation

Amongst the 49 initial participants three were excluded from the analysis, as outliers after a boxplot analysis for overall accuracy, so the remaining sample size was $N = 46$. All data preparation in terms of applying the Tukey criterion to latencies were identical to Experiment 1.

Response latencies

Table 2 lists the response latencies. Color was included as an additional fixed factor in the modelling. The final model had fixed effects for color (dominant: blue vs. green), block (1 – 6), side of dominant element (left vs. right) and pair distance (1 step, …, 4 steps). A main effect of block showed that average responses got quicker across blocks 1 to 6 (M_1 = 1128 ms; M_2 = 1097 ms; M_3 = 1047 ms; M_4 = 1089 ms, M_5 = 1058 ms; M_6 = 996 ms) again reflecting increased practice with the task, $F(5, 50.60) = 3.30$; $p = .01$. Relevant to the main hypothesis, side of dominant element had a significant effect, $F(1, 71.72) = 6.72$; $p = .01$, d_z = 0.25. Participants were faster responding to the dominant stimulus in a pair if that stimulus was presented on the left side (M_{left} = 1053 ms) than the right side (M_{right} = 1085 ms). There was also a main effect of pair distance, $F(3, 78.56) = 100.32$; $p < .001$, showing again the symbolic distance effect, ($M_{1\ step}$ = 1222 ms; $M_{2\ steps}$ = 1129 ms; $M_{3\ steps}$ = 1026 ms; $M_{4\ steps}$ = 895 ms), that is, participants responded faster when the distance in steps between the elements of a pair was large compared to small. No further effects were significant, in particular, the interaction between color and side of dominant element was insignificant, $F(1, 71.72) = .76$; $p = .39$.

The Bayes factor in favor of the above linear model minus the interaction between color and side of dominant element against the full model was 18.27, indicating “strong evidence” against this interaction (Jeffreys, 1961).

Accuracy

Table 2 lists the mean accuracies. The overall error level was 9%. The final model had the same fixed effect structure as above for latencies. Block yielded a significant effect, $\chi^2(5) = 56.98$; $p < .001$, showing a tendency of increasing accuracy over the first five blocks of the experiment (M_1 = .865; M_2 = .940; M_3 = .937; M_4 = .941; M_5 = .937; M_6 = .951), reflecting a practice effect. Pair distance had a significant effect, $\chi^2(3) = 85.74$; $p < .001$, replicating the classical symbolic distance effect (see above), that is, participants were more accurate when the distance in steps between the elements of a pair was large compared to small ($M_{1\ step}$ = .861; $M_{2\ steps}$ = .928; $M_{3\ steps}$ = .960; $M_{4\ steps}$ = .964). The interaction between block and pair distance was also significant, $\chi^2(15) = 26.33$; $p = .03$, showing that with increasing practice, the SDE tended to be smaller. This same effect of diminishing SDE tended to be more pronounced when the color assigned to dominant names was *green*, compared to *blue*, as a significant triple-interaction between color, block and distance indicated, $\chi^2(15) = 29.44$; $p < .01$. No further effect was significant, in particular, the interaction between color and side of dominant element was insignificant, $\chi^2(1) = 0.0$; $p > .99$.

The Bayes factor in favour of the above linear model minus the interaction between color and side of dominant element, against the full model, was 19.37, reflecting “strong evidence” against this interaction (Jeffreys, 1961). Overall, Experiment 2 provided a replication of the left-anchoring effect under the condition of no valid sequential cue, that is, the mental time line could not serve as diagnostic scaffold for model construction. Color was shown to have no influence on the target effect as it did not interact with the side of the dominant element.

General Discussion

Participants who are used to read from left to right are faster to respond when, at a later test trial, the dominant element within a pair from a linear order (e.g., the older, faster etc.) appears on the left side, compared to the right side. This *left-anchoring* effect, as reported earlier (von Hecker et al., 2016) proved replicable and robust under three conditions designed to address alternative influences during the mental construction of the rank order. We asked the question whether a well-researched, strong, spatial simulation, that is, the mental time line

(Fuhrman & Boroditsky, Gevers, Reynvoet, & Fias, 2003; Gevers, Caessens, & Fias, 2005; Ouellet, Santiago, Funes, & Lupianez, 2010; Santiago, Lupianez, Perez, & Funes, 2007; Tversky, Kugelmass, & Winter, 1991) would interfere with the construction of a linear order by degrees of dominance. Assuming a salient mental time line, the prediction in favour of such an influence (*time line influence*) is that any order element that is acquired *first* in a pair has primacy, thus is the one to be assigned, in terms of a spatial representation, to the left side of the second element that is acquired *later* in time. The extreme formulation of this hypothesis is that the mental time line, or indeed any other spatial simulation that is salient during construction, is the main or even the exclusive determinant of the left-anchoring effect. The competing prediction (*primacy influence*) is that even in the presence of an available alternative spatial simulation during the construction phase, reasoning about comparative levels of magnitude is largely independent of such alternative influences. According to this argument, construction follows the overlearned reading/writing habits and therefore unfolds, for Westerners, from left to right. Additionally it is assumed that dimensional magnitude within the order translates into dominance, therefore primacy, in terms of spatially represented order position, such that the dimensional maximum is placed on the left.

The obtained results from both studies presented here speak in favour of the latter set of assumptions, that is, primacy influence. As manipulation check, the symbolic distance effect (e.g., Smith & Foos, 1975) was replicated, suggesting that participants formed analogue representations of rank orders out of the initial piecemeal information that was learned (Holyoak & Patterson, 1981; Leth-Steensen & Marley, 2000). Importantly, in Experiment 1 we found strong evidence for left-anchoring of the orders in Groups 1 (dominant-first) and 2 (dominant-random). Whilst in Group 1 processes 1 (time line influence) and 2 (primacy influence) were congruently applicable, in Group 2 time line influence was not applicable because the information pertaining to the mental time line, and therefore to the positioning of elements on that line, was randomised. Experiment 2 replicated left-anchoring under conditions of Group 2 of Experiment 1, now also controlling for color which was shown to have no influence on this effect (see above for Bayes factors from both accuracy and latency sections). Therefore, in order to explain the strong left-anchoring effect observed in Group 2 of Experiment 1 and in Experiment 2, we need to refer to primacy influence and the above assumptions regarding reading/writing, magnitude and primacy, as plausible mechanisms. The result in Group 3 (Experiment 1) again supports the robustness of primacy influence. This is a condition in which the mental time line is salient and provides an alternative scaffold for construction during the learning phase. But this scaffold is incongruent with the one implied by primacy influence. The significant left-anchoring effect even in this condition means that participants tended to follow the implications of reading/writing, magnitude and primacy, despite the presence of an alternative scaffold.

It is important to note similarities and differences between this and existing work on mental model construction that has already shown left-to-right bias. For example, Jahn, Knauff, and Johnson-Laird (2007) demonstrated that their participants constructed models of input information such as “TV – table – chair” working from left to right, linking this tendency to a cultural bias to scan in the trained direction of reading and writing (Chan & Bergen, 2005; Spalek & Hammad, 2005). In a similar way, work by Román, El Fathi, and Santiago (2013) shows that auditorily presented materials such as “the table is between the lamp and the TV” were graphically represented by their participants in a way consistent with their learned reading/writing direction. However, the input information used in the above examples is genuinely spatial in the first place, whereas this is not the case in our materials. Rather, the present approach suggests two things. First, dimensions that a priori do *not* imply any spatial extension, situation, or orientation (e.g., *older, richer, smarter* etc.) are nevertheless adapted to a (linear) spatial representation, as a dimension. Second, the abstract relation of dominance, that is, the meaning of this dimension, is spatially represented by left-anchoring the maximum magnitude at the left side of the mental model, and constructing the dimension rightwards with the constraint that any element situated to the left of another will be the dominant one when comparing the two. Therefore, whereas the literature cited above shows reading / writing habits to have an effect on the construction of mental models of spatial layouts, our approach suggests that reading /writing habits have an effect on non-spatial, abstract dimensions of reasoning as well.

To the extent that, overall, substantial evidence in support of left-anchoring is accepted from the present studies, our general interpretation is that there is a spontaneous, largely time-independent tendency to construct rank-orders from left to right, and that temporal influences incongruent to primacy influence, even if made salient and presented consistently, as in Group 3 (Experiment 1), are not strong enough to abolish or reverse the effect. Consequently, we posit that primacy influence, that is, defining the origin for model construction according to reading/writing habits, and placing the element highest in dominance at that origin, is sufficiently robust, or fundamental, to be impervious to such presentational influences. We are not adopting a strictly encapsulated view of primacy influence, as being a part of some “central reasoning system” (Fodor, 1983; Sloman, 1996; Pylyshyn, 1999; Anderson, 2007). Instead, our view is closer to what Boroditsky and Ramscar (2002) have called a “milder view” of embodiment: the roots of the process lie in physical experiences of some action dynamic (reading/writing), but its application later on, after consolidation, proceeds largely independent of physical parameters, including presentational conditions. The present research thereby not only attests to the robustness of the left-anchoring effect, but also contributes to the fascinating question of how presentational factors and more centrally driven, basic mechanisms of model construction interact.

References

Anderson, J. R. (2007). *How can the human mind occur in the physical universe?* New York, NY: Oxford University Press. doi: 10.1093/acprof:oso/9780195324259.001.0001

Bates, D., Maechler, M., Bolker, B. M., & Walker, S. (2015). "Fitting Linear Mixed-Effects Models using lme4." ArXiv e-print; Journal of Statistical Software. Retrieved from http://arxiv.org/abs/1406.5823

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. (2009). *Introduction to meta-analysis*. Chichester: John Wiley & Sons.

Boroditsky, L., & Ramscar, M. (2002). The roles of body and mind in abstract thought. *Psychological Science*, *13*(2), 185-189.

Casasanto, D. (2009). Embodiment of abstract concepts: Good and bad in right- and left-handers. *Journal of Experimental Psychology: General, 138,* 351-367.

Chan, T. T., & Bergen, B. (2005). Writing direction influences spatial cognition. In B. Bara, L. W. Barsalou, & M. Bucciarelli (Eds.), *Proceedings of the 27th Annual Conference of the Cognitive Science Society* (pp. 412-417). Mahwah, NJ: Erlbaum.

Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of parity and number magnitude. *Journal of Experimental Psychology: General, 122,* 371–396.

De Soto, C. B., London, M., & Handel, S. (1965). Social reasoning and spatial paralogic. *Journal of Personality and Social Psychology, 2*, 513-521.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149-1160.

Fodor, J. A. (1983). *Modularity of mind*. Cambridge, MA: MIT Press.

Gevers, W., Reynvoet, B., & Fias, W. (2003). The mental representation of ordinal sequences is spatially organized. *Cognition, 87,* B87-B95.

Gevers, W., Caessens, B., & Fias, W. (2005). Towards a common processing architecture underlying Simon and SNARC effects. *European Journal of Cognitive Psychology, 17*, 659–673.

Gevers, W., Verguts, T., Reynvoet, B., Caessens, B., & Fias,W. (2006). Numbers and space: A computational model of the SNARC effect. *Journal of Experimental Psychology: Human Perception and Performance, 32*, 32–44.

Holyoak, K. J., & Patterson, K. K. (1981). A positional discriminability model of linear-order judgments. *Journal of Experimental Psychology: Human Perception and Performance, 7,* 1283-1302.

Ito, Y., & Hatta, T. (2004). Spatial structure of quantitative representation of numbers: Evidence from the SNARC effect. *Memory & Cognition*, *32*, 662-673.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language, 59,* 434-446.

Jahn, G., Knauff, M., & Johnson-Laird, P. N. (2007). Preferred mental models in reasoning about spatial relations. *Memory & Cognition, 35*, 2075-2087.

Jeffreys, H. (1961). *Theory of probability* (3rd Ed.). Oxford, UK: Oxford University Press.

Judd, C. M., Westfall, J., & Kenny D. A. (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problema. *Journal of Personality and Social Psychology, 103,* 54-69.

Leth-Steensen, C., & Marley, A. A. J. (2000). A model of response time effect in symbolic comparison. *Psychological Review, 107,* 62–100.

Lumley, T. (2015). Package "rmeta". Version 2.16. Retrieved from *cran.r-project.org/web/packages/rmeta/rmeta.pdf*

Makel, M. C., Plucker, J. A., & Hegarty, B. (2012). Replications in psychology research: How often do they really occur? *Psychological Science, 7,* 537-542.

Maass, A., & Russo, A. (2003). Directional bias in the mental representation of spatial events: Nature or culture? *Psychological Science, 14,* 296–301.

Makel, M. C., Plucker, J. A., & Hegarty, B. (2012). Replications in psychology research: How often do they really occur? *Psychological Science, 7,* 537-542.

Morey, R. D., Rouder, J. N., & Jamil, T. (2015). Computation of Bayes factors for common designs. Version 0.9.2+, http://bayesfactorpcl.r-forge.r-project.org/.

Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-group designs. *Psychological Methods, 7,* 105-125.

Ouellet, M., Santiago, J., Funes, M. J., & Lupianez, J. (2010). Thinking About the Future Moves Attention to the Right. *Journal of Experimental Psychology: Human Perception and Performance, 36*(1), 17-24. doi: 10.1037/a0017176

Peirce, J. W. (2007). PsychoPy - Psychophysics software in Python. *Journal of Neuroscience Methods, 162,* 8-13.

Pylyshyn, Z. (1999). Is vision continuous with cognition? The case for cognitive impenetrability of visual perception. *Behavioral and Brain Sciences*, 22, 341–423. doi: 10.1017/S0140525X99002022

R Core Team. (2013). R: *A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria*. Retrieved from http://www.R-project.org/

Román, A., El Fathi, A., & Santiago, J. (2013). Spatial biases in understanding descriptions of static scenes: the role of reading and writing direction. *Memory & Cognition, 41*, 588-599.

Santiago, J., Lupianez, J., Perez, E., & Funes, M. J. (2007). Time (also) flies from left to right. *Psychonomic Bulletin & Review,* 14, 512–516.

Singmann, H., & Bolker, M. (2014). afex: Analysis of Factorial Experiments. *R package version 0.12–135*.

Singmann, H., Bolker, B., Westfall, J., & Aust, F. (2018). afex: Analysis of Factorial Experiments. R package version 0.20-2. *https://CRAN.R-project.org/package=afex*

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin,* 119, 3–22. doi: 10.1037/0033-2909.119.1.3

Smith, K. H., & Foos, P. W. (1975). Effect of presentation order on the construction of linear orders. *Memory and Cognition, 3,* 614–618.

Spalek, T. M., & Hammad, S. (2005). The left-to-right bias in inhibition of return is due to the direction of reading. *Psychological Science, 16*, 15-18.

Tversky, B., Kugelmass, S., & Winter, A. (1991). Cross-cultural and developmental trends in graphic productions. *Cognitive Psychology, 23,* 515–557.

von Hecker, U., Klauer, K. C., & Aßfalg, A. (2017, December 17). Spatial construction of orders 2 (mental time line). Retrieved from https://osf.io/emauf

von Hecker, U., Klauer, K. C., & Aßfalg, A. (2018, December 18). Spatial construction of orders 3 (replication). Retrieved from https://osf.io/6sznj/

von Hecker, U., Klauer, K. C., & Sankaran, S. (2013). Embodiment of social status: Verticality effects in multi-level rank-orders. *Social Cognition, 31*, 387-402.

von Hecker U., & Klauer, K. C, Wolf, L., & Fazilat-Pour, M. (2016). Spatial processes in linear ordering. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42,* 1003-1033.

Zebian, S. (2005). Linkages between number concepts, spatial thinking and directionality of writing: the SNARC effect and the REVERSE SNARC effect in English and in Arabic monoliterates, biliterates and illiterate Arabic speakers. *Journal of Cognition and Culture. Special Issue: on Psychological and Cognitive Foundations of Religiosity, 5,* 165-190.

Footnotes

1

Deriving primacy as an abstract and generalised dimension from more proximal, empirically more accessible magnitude is germane to the process of *metaphoric blending* as discussed in Casasanto (2009).

2

The meta-analysis across the 8 experiments in von Hecker et al. (2016) was conducted with the R package "rmeta" (Lumley, 2015).

3

We ran an alternative power analysis based on d_z= 0.48, focusing on the alternative hypothesis that the left-anchoring effect is caused by temporal order that leads one to expect an interaction between group and the factor "side of the dominant element". Conceptualizing the interaction as a between-participants comparison between the three groups with respect to the left-anchoring effect, assuming *alpha* = .05, to achieve 1-*beta* = .80, we used the menu "ANOVA: Fixed effects, omnibus, one-way" (GPower 3.1.3., Faul, Erdfelder, Buchner, & Lang, 2009), stipulating effect sizes of .48, .0, and -.48, for Groups 1, 2, and 3, respectively. This yielded a total *N* of 66. Given the a priori assumed effect size, the planned total *N* of 120 ensures sufficient power to detect the interaction effect, as conceptualized in the above way.

4

Bayes factors were computed with function lmBF of the R-package BayesFactor, version 0.9.2+ (Morey, Rouder, & Jamil, 2015), with parameters rscaleRandom and rscaleFixed set to the default values of “nuisance” and “medium”, respectively.

5

Completing the intended methods of analysis as preregistered (von Hecker et al., 2017), we also used a composite measure, as aggregated across accuracy and inverse latency (response speed) data, after z-standardisation of both types of variables. These results closely mirror the previously reported results.

Appendix: Modelling of effects

Experiment I

In order to determine which random effect structure to assume, we used generalized linear mixed models with random effects for *participants* for accuracy data, and linear mixed models with random effects for *participants* for the latency data.

Model comparisons were performed in a two-steps procedure: In the first step, we fitted four models for each data type (a1, a2, a3, a4 for accuracy data, and tm1, tm2, tm3, and tm4 for latency data). All of these models had the same fixed effect structure, that is, presentation side of the dominant element, pair distance and block as well as their interactions. All models had a random intercept for participants. Models a4 and tm4 had only this intercept, so these models are minimal. Models a1 / tm1 also had a random slope for block as function of participant, whereas a2 / tm2 had a random slope for dominant side instead, and a3 / tm3 had a random slope for pair distance instead. These models were then compared using the Chi square difference statistic $\Delta\chi^2$. Models of a given type 1, 2, or 3 were compared with the corresponding model of type 4, the minimal model. If there was a significant difference in fit, the particular type of random slope as specified in the non-minimal model under comparison was then retained for the final model, afinal, resp., tfinal. In a second step, these final models were assembled and run in order to evaluate the respective fixed effect structure from those models (see Jaeger, 2008). This strategy thus considers random intercepts and random slopes for the main effects of the experimental design. Models with more complex random effects structures (e.g., random slopes for interactions) could not be estimated in reasonable amounts of time. The analyses employed the statistical programming language R (R Core Team, 2016), using the package lme4 (Bates, Maechler, Bolker, & Walker, 2015) and afex (Singmann, Bolker, Westfall, & Aust, 2018).

Latencies

Model	*df*	*AIC*	*BIC*	*loglik*	*deviance*	$\Delta\chi^2$	Δdf	*p*
t4	146	26857	28044	-13282	26565			
t1	166	25583	26934	-12626	25251	1313.4	20	< 2.2e-16
t2	148	26859	28063	-13282	26563	1.4934	2	0.4739
t3	155	26763	28024	-13226	26453	111.61	9	< 2.2e-16

tfinal: Random slopes for block and pair distance, as a function of participants, are kept.

Accuracies

Model	*df*	*AIC*	*BIC*	*loglik*	*deviance*	$\Delta\chi^2$	Δdf	*p*
a4	145	16526	17727	-8117.8	16236			
a1	165	16004	17371	-7836.9	15674	561.89	20	< 2.2e-16
a2	147	16527	17745	-8116.6	16233	2.4172	2	0.2986
a3	154	16430	17706	-8061.0	16122	113.59	9	< 2.2e-16 ***

afinal: Random slopes for block and pair distance, as a function of participants, are kept.

AIC = Akaike's Information Criterion; BIC = Bayesian Information Criterion.

Experiment 2

The model comparisons performed for this experiment had the same structure as in Experiment 1.

Latencies

Model	*df*	*AIC*	*BIC*	*loglik*	*deviance*	$\Delta\chi^2$	Δdf	*p*
t4	98	13071	13774	-6437.3	12875			
t1	118	12781	13627	-6272.6	12545	329.49	20	< 2.2e-16 ***
t2	100	13066	13783	-6433.1	12866	8.4733	2	0.01446 *
t3	107	13069	13837	-6427.7	12855	19.315	9	0.02265 *

tfinal: Random slopes for block, dominant side and pair distance, as a function of participants, are kept.

Accuracies

Model	*df*	*AIC*	*BIC*	*loglik*	*deviance*	$\Delta\chi^2$	Δdf	*p*
a4	97	5790.9	6499.9	-2798.4	5596.9			
a1	117	5518.1	6373.3	-2642.1	5284.1	312.78	20	< 2.2e-16 ***
a2	99	5794.0	6517.6	-2798.0	5596.0	0.8663	2	0.6485
a3	106	5774.3	6549.1	-2781.1	5562.3	34.592	9	7.03e-05 ***

afinal: Random slopes for block and pair distance, as a function of participants, are kept.

AIC = Akaike's Information Criterion; BIC = Bayesian Information Criterion.

Table 1 Experiment 1: Mean (SD) accuracies and response latencies by side of dominant element for the three groups, and four levels of pair distance.

Group 1	Dominant	Pair distance									
	person	1 step		2 steps		3 steps		4 steps		Total	
Latency	left	1086	(268)	1022	(.111)	926	(302)	840	(323)	963	(316)
	right	1125	(281)	1048	(285)	946	(296)	912	(368)	990	(327)
Accuracy	left	.878	(.142)	.951	(304)	.968	(.102)	.968	(.137)	.942	(.129)
	right	.836	(.167)	.931	(.127)	.969	(.096)	.962	(.159)	.924	(.150)

Group 2	Dominant person	Pair distance									
		1 step		2 steps		3 steps		4 steps		Total	
Latency	left	1053	(334)	1001	(326)	888	(335)	786	(320)	933	(345)
	right	1102	(346)	1059	(343)	921	(290)	810	(316)	974	(344)
Accuracy	left	.850	(.175)	.920	(.147)	.947	(.134)	.962	(.161)	.920	(.161)
	right	.817	(.178)	.901	(.168)	.930	(.151)	.964	(.151)	.903	(.171)

Group 3	Dominant person	Pair distance									
		1 step		2 steps		3 steps		4 steps		Total	

Accuracy	left	1126	(333)	1076	(380)	951	(345)	806	(326)	991	(367)
	right	1158	(390)	1059	(339)	989	(360)	827	(332)	1009	(375)
Latency	left	.834	(.187)	.904	(.170)	.923	(.180)	.958	(.176)	.905	(.184)
	right	.831	(.187)	.890	(.185)	.940	(.135)	.954	(.157)	.904	(.174)

Table 2 Experiment 2: Mean (SD) accuracies and response latencies by side of dominant element and four levels of pair distance.

	Dominant person	Pair distance: 1 step		2 steps		3 steps		4 steps		Total	
Latency	left	1209	(337)	1109	(362)	1023	(368)	869	(373)	1054	(380)
	right	1236	(375)	1151	(390)	1029	(352)	922	(457)	1085	(412)
Accuracy	left	87.3	(15.6)	94.2	(12.1)	96.4	(11.1)	95.8	(16.8)	93.4	(14.5)
	right	85.0	(16.7)	91.5	(16.6)	95.7	(12.3)	97.1	(13.8)	92.3	(15.7)