AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology: General

ISSN: 0096-3445

2021, Vol. 150, No. 3, 446–465
http://dx.doi.org/10.1037/xge0000907

Spatial Processes in Category Assignment

Ulrich von Hecker
Cardiff University

Karl Christoph Klauer
Albert-Ludwigs-Universität Freiburg

Investigates the hypothesis that spatial processes are involved in judgments on membership in a category. It is argued that membership versus nonmembership of an object or a concept, in a category, is spatially simulated in mental space by a minimal continuum with 2 levels, left for membership and right for nonmembership. In analogy to other embodied dimensions (e.g., time line or number line), the orientation of membership levels on the mental dimension is assumed to follow the acquired reading/writing schema, with procedural primacy implying dominance, hence leftward positioning of dominant elements. This rationale is tested in 7 experiments. A recognition memory paradigm (modified 2AFC paradigm, Experiment 1) revealed that participants were faster indicating the location of an old word on the screen when displayed left within a pair of words, indicating a spatial representation of category membership ("member" = left, "nonmember" = right). For category discrimination (Experiment 2) we found faster and more accurate performance when a target word is presented left as compared with right. This pattern is replicated in Experiments 4a and 4b, with different response alternatives. Discriminating categories in a stimulus-response compatibility paradigm (Experiment 3), participants were faster making correct responses with their left hand than with their right hand in target category trials. In contrast, no differences were found for distractor trials. Experiments 5a and 5b address the spatial bias in spontaneous sorting situations. Overall, this pattern of results across the 7 experiments provides evidence in support of a spatial simulation of category assignment.

Keywords: recognition memory, category discrimination, stimulus-response compatibility, spatial processing, reading direction

In everyday life, our flow of behavior and action frequently demands the making of binary judgments of the form: "Does this stimulus (object, entity, person, etc.) belong to that category (A, B, C, etc.)?" For example, we are asked "Is this bird a puffin?", "Is this one of the cups you bought yesterday?", and so on. In these cases, membership of a salient target in a target category is positive ("yes"), and nonmembership is negative ("no"). The result of such a binary decision is then made available to further reasoning. In this article, we are interested in the question whether decisions of the described kind are spatially supported in cognition, and if so, how.

There are a number of reasons for considering the possibility that spatial processes are involved in binary decisions, such as judgments on category membership. First, from the perspective of mental models (Kaup & Zwaan, 2003; Knauff, 2013; Zwaan & Radvansky, 1998), the case of membership in a category could be simulated in an abstract way by placing something, mentally, onto a location that stands for "member" and placing nothing onto the remaining location that stands for "nonmember." This implies two distinct locations for the two possibilities in mental space (Kosslyn, 1994; Tversky, 1993). Second, the neurophysiology of decision making hints at mechanisms involving one particular structure within the parietal brain, the Inferior Parietal Lobule (IPL). This structure has been identified as important for spatial processing and behavior in space, with recordings of single neurons in this area, as obtained from monkey models, that responded in ways corresponding to spatial stimulation (Andersen, Essick, & Siegel, 1985). At the same time, as more recent research has found, the IPL structure is involved in decision making under uncertainty, as shown in different input modalities, auditory and visual, as well (Vickery & Jiang, 2009). Also, activation changes have been observed in the parietal cortex, as a function of participants' responding in the two-alternative-forced-choice paradigm, as used in recognition memory research (implying "old" vs. "new" decisions). These activation changes were related not only to the formation of a decision, but also to the confidence associated with making it (Kiani & Shadlen, 2009).

Third, in our own research (von Hecker, Klauer, Wolf, & Fazilat-Pour, 2016) we found evidence for spatial support during reasoning about rank orders: When an element A dominated another element B within a previously learned series of transitive ranks, for example, age (i.e., A being older than B), Western participants, using the left and right arrow keys, were quicker to identify the dominating (= older) element when A was presented

This article was published Online First August 13, 2020.

Ulrich von Hecker, School of Psychology, Cardiff University; Karl Christoph Klauer, Department of Psychology, Albert-Ludwigs-Universität Freiburg.

We thank Rainer Leonhart, Insa Borm, Lea Krumpholtz, and Alina Kias for help with data collection, as well as Nikoletta Symeonidou and Zixi Jin for helpful comments on earlier versions. Results from Experiment 1, 2, and 3 were presented at an internal social cognition seminar (weekly lab group) in the School of Psychology, Cardiff, United Kingdom.

Correspondence concerning this article should be addressed to Ulrich von Hecker, School of Psychology, Cardiff University, Tower Building, Park Place, Cardiff CF10 3AT, United Kingdom. E-mail: vonheckeru@cardiff.ac.uk

to the left of B on a computer screen, rather than in the opposite spatial positioning. This effect reversed when the experiment was replicated in Iran, with participants who were only exposed to right-to-left scripture, implying that the spatial effect relied on the acquired reading/writing habit. Related to the perspective of the present article, the ordinal information of the individual test item, as used in our series of experiments, may be expressed in a binary way. For example, choosing which test item is the "older" one may be seen as a binary categorization of one item as older and the other one as "not older." In the context of such an individual test item, the ordinal information contained in the overall rank order is condensed into binary information, preserving the order characteristic by seeing older as dominating not older. From this perspective, facilitation occurs if the location of the element identified as older is in the same orientation on the screen as the category older is located in mental space. In contrast to this situation, however, in the present article we are interested in examining similar spatial effects in cases where there is, at the outset, no ordinal information available to frame that binary decision, but just qualitative information, that is, about membership in a category versus nonmembership. In other words, we are investigating whether spatial representations (with order characteristics) are involved in the assignment of a target entity (concept, word, person, object, etc.) to an abstract category.

Spatial Representation of Abstract Concepts: The Dimensional Case

Within a number of dimensional domains, the signature evidence for the involvement of spatial processing in a cognitive task comes from spatial asymmetry effects. For example, Suitner and Maass (2016) defined the so-called "Spatial Agency Bias" as the tendency to imagine the trajectory of movement of an action or event that is perceived as "agentic" (i.e., strong, intense, initiative, vital, etc.) as proceeding rightward from a left origin (see also Chatterjee, 2001). They argue that the dimension of "agency" might be represented in an embodied way, implying a motion schema proceeding from left to right, as adapted from the learned reading/writing schema. The authors also report evidence that this bias reverses when studied among populations that live in a right-left reading/writing environment (e.g., Maass & Russo, 2003; Maass, Suitner, Favaretto, & Cignacchi, 2009; Maass, Suitner, & Nadhmi, 2014; Tversky, Kugelmass, & Winter, 1991). A second domain relevant to left-right asymmetries as a signature of spatial representations is the so-called SNARC effect "spatial numerical association of response codes" (Fias, 1996; Gevers, Caessens, & Fias, 2005). In cultures with left-to-right reading and writing, a mental number line represents numerical magnitude, increasing from left to right, with this direction again reversing in populations with an opposite right-left-system (Dehaene, Bossini, & Giraux, 1993; Zebian, 2005). A third dimensional domain is the mental representation of time as a line with left origin (in left-to-right reading cultures) as the time line begins in the perceived present, and moves to the right as time passes on (Boroditsky & Ramscar, 2002; Fuhrman & Boroditsky, 2010). Distinguishing between words as relating to the past or future is quicker in Western participants if past-related words are shown left and future-related words right on a screen (Ouellet, Santiago, Funes, & Lupiáñez, 2010; Santiago, Lupiáñez, Pérez, & Funes, 2007). Left-hand responses are faster than right-hand responses for months at the beginning of the year as compared with months at the end of the year (Gevers, Reynvoet, & Fias, 2003). Tversky et al. (1991) had children arrange pictures into temporally ordered stories and found a tendency to order these from left to right in English-speaking children, whereas this tendency reversed in Arab and Hebrew children who were used to right-to-left-written languages. However, preliterate kindergarteners did not show any such spatial biases (Dobel, Diesendruck, & Bölte, 2007) which again hints at the tight connection between acquired reading/writing habits and the orientation of the mental time line.

Spatial Representation of Abstract Concepts: The Case of Categories

The concepts of agency, numerical magnitude, and time all are examples for spatial processes supporting abstract reasoning by providing an oriented dimension onto which the abstract concept is projected. Questions now arising are, first, in what way this projection is modeled after the acquired reading/writing habit; and second, what reasons there are to think that similar processes might exist in the case of categorical membership, that is, the assignment of a target to a category.

One way how the left-right-oriented mental dimension can be modeled after the culturally learned reading/writing schema is that the reading/writing schema is abstracted to, or yields the model for, a general schema of sequential action (Chatterjee, Southwood, & Basilico, 1999; Patro, Nuerk, & Cress, 2015; Suitner & Maass, 2016). Such a schema, as a dimension, will serve two purposes. First, it allows one to represent order of magnitude, because different locations on the dimension will reflect different levels of magnitude in terms of an embodied simulation (Hegarty & Just, 1993; Niedenthal, Barsalou, Winkielman, Krauth-Gruber, & Ric, 2005). Second, the notion of sequential action, abstracted from the reading/writing process, itself implies higher *primacy* of elements the closer to the origin of the dimension they are situated. We use the term "primacy" here in the sense of a generalized notion of "what comes first/has most immediacy (therefore: importance)," referring to a sequence in processing. Thus, if one assumes that the origin of the dimension has highest primacy, the maximum of the dimension is likely to be located at the origin, as there is reason to assume that magnitude/dominance is easily conflated and metaphorically blended with the notion of primacy (Casasanto, 2009): "The first in a series (or a pair) is the leftmost (. . .). Linguistic expressions like 'the prime example' conflate primacy with goodness (i.e., this phrase can mean the first example, the best example, or both). Speakers of languages like English may be predisposed to consider the leftmost item to be the first and therefore the best" (Casasanto, 2009, p. 362).[1] In this example, it is assumed that goodness is a magnitude or an asset, and that more goodness dominates less goodness. From this perspective, the answer to our first question is that in general, the projection of an abstract concept onto a dimension in mental space can be modeled after the acquired reading/writing habit because the latter provides an overlearned, familiar schema of primacy that allows for an ordered

[1] "Conceptual Blending" is defined here, following the linguistic literature, as the integration of information from disparate domains. When two concepts are blended, overlapping aspects of their individual meanings form the core of a new, enhanced, or pragmatically more useful, integrated meaning (see Casasanto, 2009; Coulson & Oakley, 2005; Fauconnier & Turner, 1998).

differentiation along the dimension, of degrees in magnitude (Lourenco, Ayzenberg, & Lyu, 2016; Lourenco & Longo, 2010; Walsh, 2003). However, this projection also allows for the *anchoring* of the dimensional maximum at the left side; that is, on the basis of a "metaphorical blend" between primacy and dimensional dominance (Casasanto, 2009; see also Pecher, Van Dantzig, Boot, Zanolie, & Huber, 2010). In support of this, in one of our earlier experiments (von Hecker et al., 2016, Experiment 4b) we found that participants, after having learnt a series of stimuli of different heights, were faster to indicate which was the "taller" one within a horizontally presented pair, via left and right arrow keys, in trials that showed the taller one on the left, as compared with the taller one on the right. This bias was not observed, however, when learning and testing took place the same way, but using "shorter" to describe differences in height. In describing levels of height, "taller" as the unmarked expression (Hamilton & Deese, 1971) represented a *positive magnitude* and was more likely to be blended with primacy; thus, located at the left side where the dimension was anchored, as compared with "shorter" which, as the marked expression, represented a *lack of magnitude*.[2]

Our second question relates to the assignment of a target to a category. On what grounds can we generalize the above reasoning to this case? First, we argue that in the context of category assignment, membership can be described as a (minimal) order dimension with two levels, member and nonmember, whereby member is mentally represented by simulating a positive entity/magnitude, and nonmember by a lack of magnitude. This brings an ordinal characteristic to the dimension, as member dominates nonmember. Therefore, second, we argue that metaphorical blending will produce a tendency to mentally place the category member to the left, for greater primacy, and the category of nonmember to the right. As a general hypothesis pertaining to the experiments reported below, we expect an interaction between the assumed mental representation of category membership and the experimental display of stimuli, to gain evidence for the spatial support of category assignment. In other words, if a participant makes a judgment about a target object/word/entity as being a member of a particular category, this is tantamount to the participant placing the target onto that particular location in mental space where "positive membership" is simulated. Therefore, if the display orientation of the target (as left vs. right on a screen) is congruent with the orientation of the hypothesized categories of member versus nonmember, then a response should be made more quickly than when the display orientation is incongruent with that mental representation of category membership.

Experiment 1: Recognition Memory

Recognition memory is chosen as a first example to demonstrate the above ideas, mainly for methodological reason as will become clear immediately. Still, it is necessary to shortly recapitulate some of the theoretical discussion within this literature, to clarify the relevant methodological point. Generally, the pertaining models (for a review, see Malmberg, 2008) either assume that a latent familiarity variable determines a recognition-memory judgment, as signal detection theory (SDT) does (see Swets, Tanner, & Birdsall, 1961). Alternatively, some models assume discrete processes, such as the two-high-threshold model (e.g., Bröder & Schütz, 2009), or dual-process SDT models (e.g., Yonelinas & Parks, 2007). We do not enter this discussion here (but see Kellen & Klauer, 2014, for an approach aimed at comparing the two principal alternatives), as the present article deals with the representational outcomes of such processes, and not the processes themselves. Instead we will focus on a methodological issue within one particular paradigm that is often used to inform these debates, which is the two-alternatives-forced-choice paradigm (2AFC).

In this paradigm, a participant learns a list of items and is later asked to select one of two items in a presented pair as old, that is, as part of the previously learned set. Thus, as two words are presented side by side horizontally on a screen, this paradigm suits the present enquiry well because it implements a possibility of creating displays that are congruent, or incongruent, with the mental representation of category membership. Therefore, if a participant is asked to quickly indicate the side on which the old item is located, and if generating a response is associated with mentally placing the old word to the left side, then that mental representation will be either aligned with, or stand in conflict, with the perceived display. For the latter case, longer response latencies are expected. In terms of accuracies, expectations are less determinate. Empirically, left-anchoring effects, that is, lower accuracy levels in the conflict case, have been observed alongside latency effects in some studies, whereas in others, left-anchoring effects remained limited to latencies (see Suitner & Maass, 2016; von Hecker et al., 2016).

One constraint of the 2AFC paradigm is that one has to declare one and only one of the items old, regardless of whether any or both of them might meet an internal criterion for recognition as old (Egan, 1975; Hicks & Marsh, 1998; Macmillan & Creelman, 1991; Smith & Duncan, 2004). This means that the emphasis is placed on making a *relative* judgment; that is, to compare both items against each other in the sense of reliance on signal-strength difference, rather than *absolute* signal strength level (Glanzer & Adams, 1990; Glanzer & Bowles, 1976; Green & Swets, 1966; Smith & Duncan, 2004). The paradigm has, therefore, been argued to be criterion-free (Egan, 1975; Hicks & Marsh, 1998; Macmillan & Creelman, 1991). However, recent findings provide evidence to the contrary. It appears that participants sometimes generate their judgments on the basis of individual-item recognition; therefore, relying on an absolute criterion of recognition. First, Jou, Flores, Cortes, and Leka (2016) found left-biased choices in horizontally presented pairs, consisting of either one new and one old item (normal), or both-new, or both-old items ("null"). Such left-bias was associated with higher hit and false alarm rates, as well as shorter left- than right-choice response times, for normal pairs. This pattern is consistent with assuming that participants made their decision on the basis of a left-individual-item recognition (presumably according to their reading/writing habit), without taking both items of the pair into account. For null pairs, latencies were longer in case of both-new than both-old pairs, which again supports the idea of

[2] An adjective can be defined as unmarked when it neutralizes to describe the dimension, and as marked when it does not (Proctor & Cho, 2006). In the cited experiment, older, richer, taller, smarter, stronger, and faster were used, each of which would, in its noncomparative form, describe the dimension in a neutral way, as opposed to their marked counterparts younger, poorer, shorter, dumber, weaker, and slower, which do not. Hamilton and Deese (1971) demonstrated that people are reliably able to distinguish these two types of adjectives.

judgments based on absolute, rather than relative, levels of familiarity. Second, in a study using eye-tracking procedures, Starns, Chen, and Staub (2017) provided evidence for absolute judgments, using pairs of old items and lures for testing in a horizontal presentation. They found that participants, especially in correct trials selecting the old item, looked at the old one first and exclusively, which happened most often when the old item was positioned on the left side. In trials when they looked at both items in the pair, participants were more correct when they fixated the old one first than when they fixated the lure first, implying that there was some probability to make false absolute decisions on the basis of the lure.

Thus, the left-bias as found in the 2AFC paradigm is informative in relation to the theoretical assumptions about the process of generating a judgment as being either relative or absolute. That is, the left bias arises as a consequence of decision strategies that process only the left item displayed in an absolute judgment. However, exactly the susceptibility to such biases makes the paradigm problematic, if unmodified, as a tool for investigating the spatial representation of the judgment. The left-bias originating from absolute versus relative response generation would be superimposed on, or conflated with, any potential left-bias stemming from processing a spatial representation of category membership. In this experiment, we implement a modified methodology that eliminates spatial bias that would be unrelated to a possible effect genuinely arising from the display orientation being spatially congruent or incongruent with the mental representation of category membership (see above).

Method

General. Data in this and all following main experiments (except Experiment 5) were collected in Germany. The present research used only procedures that are exempt from formal ethical approval under the ethical guidelines of the Deutsche Gesellschaft für Psychologie (German Psychological Society). All participants gave written consent and were formally debriefed.

Participants. Forty undergraduate students from the University of Freiburg (27 female, 13 male, mean age = 26.9 years) took part in the experiment, all with German-spoken backgrounds. They received course credit or €5.00 for their participation.

Materials. The word pool used for the experiment contained 644 neutral German nouns taken from Lahl, Göritz, Pietrowsky, and Rosenberg (2009) ranging from four to eight letters in length. According to the ratings obtained by Lahl et al. (2009), the words were all of medium valence (ranging from 3.5 to 6.5 on an 11-point scale) and low in arousal (ranging from 0.5 to 4.5 on an 11-point scale). Furthermore, all words were of approximately equal word frequency in common German, as indicated by log frequency ratings obtained for each word via WordGen (ranging from 0.3 to 2.9; Duyck, Desmet, Verbeke, & Brysbaert, 2004).

Using this word pool, 240 words were randomly selected for each participant. These were divided into 12 groups (six blocks times two conditions: *old* vs. *new*, of 20 words each). All stimuli were displayed on a computer screen in white Arial font (12 mm height) against a black background. For pairs, the words were presented 25 mm to the left or to the right of the horizontal screen center, creating a gap of 50 mm between them. All stimuli were presented at the vertical screen center.

Procedure. After reading instructions, participants were seated approximately 60 cm in front of the computer screen. They were then presented with the learning phase of the first experimental block. Each block had a learning phase and a testing phase. In a learning phase, a participant viewed all 20 words that were to be classified old in the test phase one by one without having to respond, with each word presented for 3 s, divided by a 300 ms interstimulus interval. In the immediately following test phase, 40 word pairs were presented in a random sequence. Each pair was initiated by a fixation cross in the center of the screen (800 ms), followed by the word pair that could be of either of four types, “old-old” (5 pairs), “new-new” (5 pairs), “old-new” (5 pairs), or “new-old” (5 pairs), with the first word always appearing on the left side and the second word appearing simultaneously on the right side. This presentation opened an unlimited response window. After the participant’s response, a 600 ms intertrial interval followed until the next fixation cross appeared. The task was in each trial to indicate as quickly and accurately as possible, which of the four pair types one had just seen. As a response aid, a graphical arrangement of the four response alternatives was presented in the lower right corner of the screen, see Figure 1. In this display, the labels old-old, old-new, new-old, new-new were arranged in an order similar to the arrangement of the four directional arrow keys that were marked as the response keys for participants to use. The arrangement always showed the label old-new corresponding to the position of the left arrow key, the label new-old corresponding to the position of the right arrow key. The labels old-old and new-new were placed in the arrangement as corresponding to the positions of the up and down arrow keys, with the label positions and corresponding key assignments alternated between participants as a counterbalancing factor.

There were six blocks with the above structure. Interspersed between every two blocks was a short distraction block consisting of four easy arithmetic problems which participants responded to using marked keys. This was done to clear the words from the previous block from memory. Learning and test phases as described occurred six times, corresponding to the six experimental blocks. The experiment lasted about 25 min, including debriefing.

Results

The accuracy and latency data of the experiments in this article were each analyzed in two steps. In the first step, we estimated mixed linear models (for the accuracy data: generalized mixed linear models with logistic link function) with *participants* as random factor, to determine which random structure would best fit the data. Subsequently, a final model was defined using the above information, and fixed effects were statistically evaluated on the basis of that final model (see Jaeger, 2008; Judd, Westfall, & Kenny, 2012). The strategy for selecting a model with appropriate random-effects structure is described in the Appendix, along with information about the particular random-effect structure adopted for each model in each experiment.

Accuracy. The overall error level (i.e., pressing the incorrect key) was 27%, across all participants and all test pairs in all blocks. For the present analysis, only trials involving old-new and new-old trials were considered. The final model used for analysis contained fixed effects for the counterbalancing factor (between participants, counterbalancing response key positions: old-old corresponding to

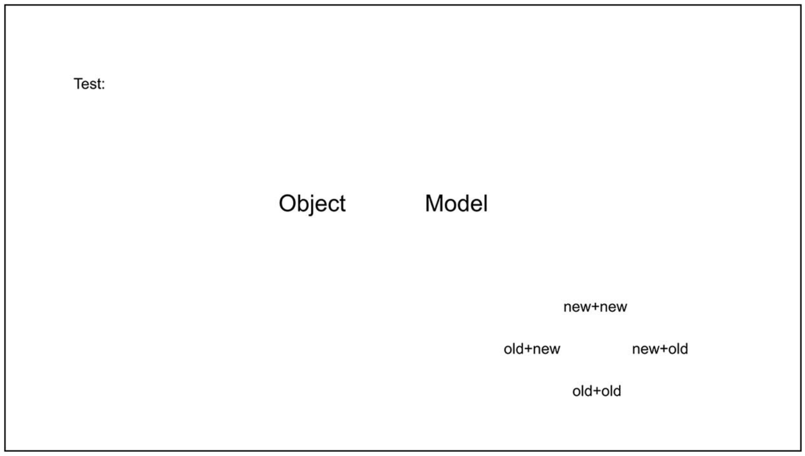

Figure 1. Experiment 1, experimental screen interface, for Test Phase. Example shown for condition with counterbalancing factor level "old-old" DOWN.

"arrow up" vs. corresponding to "arrow down"), block (1, . . . , 6), side of *old* word (left vs. right), as well as the interactions of these factors. There was a significant fixed factor effect for block ($\chi^2(5) = 11.33$; $p < .05$) indicating that responses were less correct in the last block ($M_{\text{Block 6}} = .66$) compared with previous blocks ($M_{\text{Blocks 1–5}} = .73$), as post hoc contrast analysis revealed ($p = .051$). There was also a significant interaction between the counterbalancing factor and side of old word ($\chi^2(1) = 8.95$; $p <$ .003). As Holm-corrected simple effects revealed, responses were more accurate in trials involving old words on the left side than old words on the right side, when old-old responses had to be made with the down-arrow key, ($M_{\text{OldOld down/Old left}} = .76$ vs. $M_{\text{OldOld down/Old right}} = .69$). When old-old responses had to be made with the up-arrow key, this difference was not significant ($M_{\text{OldOld up/Old left}} = .71$ vs. $M_{\text{OldOld up/Old right}} = .72$). A significant interaction between block and side of old word showed that the side of old word effect, which itself was at tendency level overall ($\chi^2(1) = 2.90$; $p = .09$; $M_{\text{Old left}} = .74$ vs. $M_{\text{Old right}} = .70$) varied in direction and size across the six blocks (see Table 1; $\chi^2(5) = 16.66$, $p < .005$). No other significant effects emerged.

Response latencies. For correct responses, latency data were trimmed according to the Tukey criterion based on excluding outliers with values larger (smaller) than the upper (lower) quartile by 1.5 times the interquartile range in the individual's distribution of latencies (see Clark-Carter, 2004, Chapter 9). In the linear mixed model, a significant main effect was found for the counterbalancing factor ($F(1, 38.00) = 9.82$; $p < .003$) showing that responses were faster when the old-old response option was associated with the up-arrow key than with the down-arrow key ($M_{\text{OldOld up}} = 1{,}897$ vs. $M_{\text{OldOld down}} = 2{,}644$ ms). A significant main effect for block indicated that across the duration of the experiment, participants tended to get faster in responding overall ($F(5, 31.66) = 4.42$, $p < .004$; $M_{\text{Block 1}} = 2{,}673$, $M_{\text{Block 2}} = 2{,}364$, $M_{\text{Block 3}} = 2{,}304$, $M_{\text{Block 4}} = 2{,}119$, $M_{\text{Block 5}} = 2{,}110$, $M_{\text{Block 6}} = 2{,}104$ ms), indicating a practice effect. Finally, and crucial to the hypothesis, there was a significant main effect of side of old word ($F(1, 37.37) = 6.75$; $p < .01$).[3] This effect indicated that, as predicted, correct responses to old words were faster when the old word in a pair was presented on the left, rather than the right, side of the screen ($M_{\text{Old left}} = 2{,}201$ vs. $M_{\text{Old right}} = 2{,}345$ ms, Cohen's $d = .37$).

Examining latencies for trials with incorrect responses, latencies were found to be at almost the same level for both sides of old word ($M_{\text{Old left}} = 2{,}355$ vs. $M_{\text{Old right}} = 2{,}336$ ms; $t(246.20) = 1.21$, $p = .22$) that makes it unlikely that quicker key pressing on the left side as such, as a general tendency, could explain the data.

Control via vertical response configuration. We also looked into the possibility that the vertical dimension, as likewise involved in reading direction (its being not only left-to-right but also top-to-bottom), could provide an equal scaffold for responding. It could be that the reading/writing habit (in horizontal *or* vertical terms) would generate faster responses irrespective of matching processes with a spatially represented category. Rather, in this case, a response could simply be made quicker if it was performed on top as compared with the bottom. Latencies associated with correct responses to old-old and new-new responses (for which the response keys were vertically arranged) were analyzed in a similar linear mixed model as above, with fixed effects for the counterbalancing factor (between participants, counterbalancing response key positions: old-old corresponding to arrow up vs. corresponding to arrow down), block (1, . . . , 6), type of pair (old-old vs. new-new), as well as the interactions of these factors. There was a significant effect for group ($F(1, 38.64) = 9.80$; $p = .003$) showing shorter response latencies for both vertical options in the group who had to press "up" for old-old ($M_{\text{OldOld up}} = 1{,}790$ ms) than in the group who had to use the same key for new-new ($M_{\text{OldOld down}} = 2{,}170$ ms). New-new stimuli ($M_{\text{NewNew}} = 1{,}924$ ms) were generally responded to faster than old-old stimuli, ($M_{\text{OldOld}} = 2{,}073$ ms; $F(1, 40.90) = 5.10$, $p = .03$) but crucially, the interaction between the counterbalancing factor and type of pair was insignificant (F(1,

[3] In the sample, three participants had extremely low accuracy levels, that is, they had produced more than 60% errors across all trials. We calculated the same type of statistical models for accuracy and latency under exclusion of these three, and found equivalent statistical results. The target effect of side of old word in the latency data was, in this case, at $F(1,76.61) = 7.11$; $p = .009$.

Table 1
Experiment 1, Accuracies by Experimental Block and Side of Old Word

Side of old word	Block 1	2	3	4	5	6
Left	.685 (.275)	.745 (.279)	.775 (.279)	.750 (.251)	.760 (.244)	.695 (.286)
Right	.795 (.245)	.725 (.258)	.675 (.266)	.670 (.295)	.730 (.254)	.630 (.298)

Note. Accuracies are given in proportion of correct responses. Standard deviations are presented in brackets.

40.90) = 1.23; p = .27). If the upper position of the response key was to generally engender faster responses than the bottom position, this interaction should have been significant. On the other hand, metaphorical blending could potentially take place in the vertical dimension, such that, in terms of representation, old-old stimuli were to be positioned on top of new-new stimuli. In this case again, we would have expected the above interaction to be significant, that is, to show that correct responses to old-old would be faster than responses to new-new, but only in the group in which the old-old-on-top response mapping had been administered.[4]

Discussion

In the first experiment, participants were faster to indicate the location of an old word on the screen when the old word was displayed as the left member of a pair of words. We interpret this finding as evidence for the representation of category assignment in mental space. Specifically, we argue that category membership is spatially represented as a binary dimension with member mentally positioned left, and nonmember positioned right. Faster responding simply based on a matching of reading/writing-primacy to response keys is unlikely insofar the vertical dimension, as likewise providing a primacy differential for responding, did not exhibit a systematic bias in the vertical control trials.

The original 2AFC-paradigm was modified. Instead of two response alternatives (old-new, new-old), we provided four alternatives, introducing the cases of old-old and new-new. In terms of internal validity, these changes had two consequences. First, by virtue of having to decide between four possibilities, a participant had to pay attention to both words in a pair, rather than being able to rely on individual-item recognition (Jou et al., 2016; Starns et al., 2017). This counteracted any potential left-bias because of only focusing on the first word when starting to read from the left. Such factors could not confound the target effect as reported here. Second, attention being given to both words means that a spatial labeling of each word in the pair as either "left" or "right" was possible. This seems crucial in a situation where one attempts to create a stimulus display that either matches or mismatches the hypothesized spatial mental representation of membership dimension as left (member) versus right (nonmember). It is argued here that as part of the process of generating a response, a participant mentally simulates, on the left side in mental space, the word which they think belongs to the target category, and on the right side the word which they think does not belong to the target category. If, in a trial, this mental orientation meets a perceptual display that is spatially in line with it, responding is quick. If the display runs counter that spatial representation, interference occurs and the response is slowed down.

Experiment 2: Category Discrimination

The general assumption to be tested here is that because of metaphorical blending (Casasanto, 2009; Fauconnier & Turner, 1998; Pecher et al., 2010), membership in a target category is simulated as a positive magnitude (implying dominance), has higher primacy than nonmembership and is, for this reason, simulated on the left side. In Experiment 1 however, it could be argued that not membership per se, or exclusively, might have driven the effect, but that a stronger memory trace of old words as compared with new words in terms of recollection, or a greater familiarity of old words over new words, might have at least codetermined dominance and, consequently, the blending with primacy. Thus, the role of membership per se to the target category still needs further exploration as a determinant of primacy. Experiment 2, therefore, aims at using materials that do not imply differences in memory traces, or familiarity differences, between the sets of stimuli that belong, or do not belong, to the target category. Using taxonomic materials also addresses a possible confound in Experiment 1, as the relation between old and new may be simulated along the time line from left to right (Boroditsky & Ramscar, 2002; Fuhrman & Boroditsky, 2010); thereby, implying primacy for old, which would predict a left-bias in the experiment on the basis of this particular time-related connotation, without necessarily giving support to our basic hypothesis that claims a spatial representation of category membership as left by a blend of dominance and primacy based on reading/writing.

We chose basic categories (Leinster, 2016; Rosch, 1978) as being at a level of generality most preferred by humans in their reasoning and memorization (Rosch, 1978). Because the exemplars of such categories can be uniformly assumed as well accessible within typical populations (Rosch, 1978), we did not anticipate systematic differences in memory trace or familiarity between target and nontarget categories as used in the study, provided these were chosen randomly from an original pool of basic categories. Selecting words with high typicality for each category helped minimizing differences between the exemplars on that dimension (see below).

[4] Because no vertical arrangement on the screen is used, the described procedure in the vertical dimension does not test congruence between the spatial positioning of stimuli and the mental representation of primacy (as our horizontal procedures in Experiments 1, 2, and 4 do), but tests the congruence of an assumed spatial representation of primacy (UP-DOWN) with spatial features of the response keys. Insofar, this test is more germane to the rationale used in Experiment 3.

Method

Participants. Forty-three undergraduate students from the University of Freiburg (32 female, 11 male, mean age = 23.4 years) took part in the experiment, all with German-spoken backgrounds. They received course credits or €5.00 for their participation. One participant was excluded for extremely slow average response time, so the final $N = 42$.[5]

Materials. The word pool used for this experiment provided German norms for semantic typicality, age of acquisition, and concept familiarity for 824 exemplars of 11 semantic categories (Schröder, Gemballa, Ruppin, & Wartenburger, 2012). Using this word pool, the 20 most typical words (with typicality highly correlated with familiarity) were selected from 10 categories, *animal*, *bird*, *clothes*, *fruit*, *furniture*, *musical instrument*, *profession*, *sport*, *tool*, and *vegetable*. Five experimental blocks were used, and for each block, two categories were randomly selected from the 10, one of them to be used as target category, the other as distractor category in that particular block. For example, in a given block, "clothes" might serve as target category and "bird" as distractor category. This way, potential differences in frequency or familiarity between the categories that could still exist were counterbalanced across blocks and participants. In each presented word pair, either both, one, or none of the two words belonged to the target category. All presentation specifications were the same as in Experiment 1.

Procedure. After reading instructions, participants were seated approximately 60 cm in front of the computer screen. Different from Experiment 1, there was no learning phase, and each block started directly with a test phase. In each test phase, 20 word pairs were presented in a random sequence, with the target category name (e.g., "furniture") being constantly displayed in the upper left corner of the screen, as reminder. Participants were asked to indicate as quickly and accurately as possible, which of the four category pair types "target-target " (5 pairs), "distractor-distractor" (5 pairs), "target-distractor" (5 pairs), or "distractor-target" (5 pairs) one had just been presented with. The response labels used were "yes-yes," "no-no," "yes-no," and "no-yes," with *yes* meaning a target category instance, and *no* a nontarget category instance.

Presentation modalities and response options were the same as in Experiment 1. Between every two of the five blocks, a block of four easy arithmetic problems was presented as in Experiment 1. The experiment lasted between 20 and 25 min, including debriefing.

Results

Accuracy. The overall error level (i.e., pressing the incorrect key) was 5%, across all participants and all test pairs in all blocks, showing a lower level of difficulty as compared with Experiment 1. For the present analysis, only trials involving target-distractor and distractor-target trials were considered. The final model used for analysis contained fixed effects for the counterbalancing factor (between participants, counterbalancing response key positions: target-target corresponding to arrow up vs. corresponding to arrow down), block (1, . . . , 5), side of *target category* word (left vs. right), as well as the interactions of these factors. There was a significant fixed factor effect for side of *target category* word ($\chi^2(1) = 12.38$; $p = .0004$) indicating that responses were more accurate when the target category word was shown on the left side ($M_{\text{Target left}} = .98$) compared with the right side in a pair ($M_{\text{Target right}} = .95$). No other significant effects emerged.

Response latencies. For correct responses, latency data were trimmed in the same way as in Experiment 1. The final linear mixed model used for the analyses had the same fixed effect structure as the one just reported for accuracy. In this model, two significant main effects occurred. First, the effect of block ($F(4, 45.41) = 9.22$; $p < .0001$) showed that responses became faster from Block 1 through Block 5, again reflecting a practice effect ($M_{\text{Block 1}} = 1{,}503$, $M_{\text{Block 2}} = 1{,}316$, $M_{\text{Block 3}} = 1{,}287$, $M_{\text{Block 4}} = 1{,}276$, $M_{\text{Block 5}} = 1{,}218$ ms). Second, the effect of side of category word ($F(1, 44.93) = 4.19$; $p < .05$) revealed that test pairs in which the target category word was presented left were associated with faster correct responses than pairs in which that word was presented on the right ($M_{\text{Target left}} = 1{,}299$ vs. $M_{\text{Target right}} = 1{,}341$ ms, Cohen's $d = .25$). No further effects were significant. We did not analyze the latencies of trials with incorrect responses because of the overall infrequency of errors.

Control via vertical response configuration. Again, we looked into the possibility that the vertical dimension could provide an equal scaffold for responding. Latencies associated with correct responses to yes-yes and no-no responses (for which the response keys were vertically arranged) were, therefore, analyzed in a similar linear mixed model as above. There were no significant effects except for the block factor ($F(4, 48.36) = 9.71$; $p < .001$) indicating a practice effect in terms of responses becoming faster from the first block ($M_{\text{Block 1}} = 1{,}490$ ms) to the last block ($M_{\text{Block 5}} = 1{,}332$ ms). In particular, the interaction between the counterbalancing factor and type of pair was again insignificant ($F(1, 1641.99) = .81$; $p = .37$). This interaction would be expected under the assumption that either the upper position of the response key was to generate quicker responding in general, or that a metaphorical blending with respect to target and distractor categories, and with respect to the yes-yes and no-no stimuli, did happen in the vertical dimension.

Discussion

The present experiment did not use materials that would differ in memory strength, or familiarity, between stimuli in the target category and stimuli in the distractor category. Using basic categories also meant that these categories were well accessible throughout, such that the task became considerably easier (5% error against 27% error in Experiment 1). Under these conditions, the pattern obtained in Experiment 1 replicated. We found that participants were faster and also more accurate when the word from the target category was presented as the left rather than right member of a test pair of words. The explanation for this pattern can now be related even more stringently to membership as the critical stimulus characteristic to be simulated, because target and distractor categories in each block did not differ in memory strength or familiarity. Membership in the target category, therefore, appears to be simulated as a positive asset, as compared with nonmembership, which then affords to be blended with primacy.

[5] Mean latency for correct responses = 3,257 ms in a sample with overall mean = 1,330 ms, $SD = 427$ ms.

Experiment 3: Stimulus-Response Compatibility

In the modified forced-choice method used so far, exemplars of both target and distractor categories form the stimulus ensemble within a given trial, and the response represents the outcome of processing both simultaneously. It is, however, desirable to examine target and distractor words in isolation, as explained below. In the present experiment, we use a stimulus-response-compatibility methodology in which participants are asked to respond either with their left or right hand, as used in studies on the *spatial-numerical-associations-of-response-codes* (SNARC) effect (e.g., Chatterjee, 2001; Dehaene et al., 1993; Gevers et al., 2005; Shaki, Fischer, & Petrusic, 2009). In this paradigm the relevant effects, in line with our basic question, rest on congruence versus incongruence between the mental representation of a stimulus (in mental space) and the left-right response key locations, that is, the motor response that is asked for, and that is executed by the efferent system (in real space). Different from the previous methodology which used pairs of stimuli, the present procedure allows us to work with individual words as stimuli and, therefore, investigate asymmetries that might exist between target and distractor categories with respect to the spatial effects under study. According to the present research hypothesis, when instructed to decide whether or not a presented word belongs to the target category (*yes* or *no*), a correct *yes* decision is associated with mentally placing the word to the left side (member), or to the right side for a correct *no* decision (nonmember). If the response has to be made with the hand that, laterally, corresponds to the location in mental space where the stimulus was placed, a faster response should occur as compared with the opposite case. In other words, a left-bias is predicted for correct *yes*-responses to target category words, and a right-bias for correct *no*-responses to distractor category words. A priori, this prediction is symmetric; that is, it holds in the same way for individual stimuli from both target and distractor categories.

On the other hand, a general argument has been made about different processes being involved in correct responses to targets (hits) versus correct responses to distractors (correct rejections), during episodic retrieval (Yonelinas, 1994). Evidence shows a difference in response latency between correct responses to targets versus distractors in paradigms comparable to ours. For both episodic memory (old/new discrimination) as well as for category discrimination tasks (semantic classification), Bucur et al. (2008) found correct responses to target stimuli to be approximately 80 ms shorter than those to distractors. Similar differences between targets and distractors are reported by Kahn, Davachi, and Wagner (2004), as well as Kahana (2014). Small differences, but still showing a latency advantage of hits over correct rejections, had been found by Nobel and Shiffrin (2001). In addition, using search paradigms with visual materials (pictures), longer latencies have been reported for correct responses to distractors than targets (Czernochowski, 2005; Hockley, 1984; Kunar, Rich, & Wolfe, 2010), such that the phenomenon appears to have cross-modal relevance. From the present perspective, in explaining these differences one needs to consider the markedness dimension (Hamilton & Deese, 1971); that is, the fact that correct *yes*-responses are generated at the unmarked end of the dimension (member), whereas correct *no*-responses are generated at the marked end (nonmember). In general, making a judgment or a decision at the marked end of a dimension has been shown to be more difficult than at the unmarked end of the dimension (Schriefers, 1990; Schubert, 2005; van der Schoot, Bakker Arkema, Horsley, & van Lieshout, 2009). It has been argued that working at the marked end might involve more elaborate thinking with the potential to mask the influence of any spatial input. In the context of sentence comprehension, Sherman (1973, 1976) argued that processing at the marked end imposes more cognitive load than at the unmarked end because marked stimuli are often treated as a negation of the unmarked version which induces additional load. Also, negation is usually conceived as an operator that draws the reasoning focus away from the concept in question, and reduces its accessibility (Kaup & Zwaan, 2003; Lea & Mulligan, 2002; MacDonald & Just, 1989; Sanford, Moxey, & Paterson, 1996). Finally, as mentioned above, in our own research (von Hecker et al., 2016, Experiment 4b) we found a left-anchoring bias when rank judgments had to be made on pairs from a series of "taller" object (unmarked end, positive magnitude), but no spatial effect at all when judgments had to be made on "shorter" objects (marked end, lack of magnitude; see Schubert, 2005, for similar asymmetries).

The present experiment, allowing us to examine target and distractor words in isolation, can address the question of symmetry between stimulus categories. We assume that once a stimulus is presented, participants will mentally position it either on the left (target) or on the right (distractor). As such, when representing the stimulus at the unmarked end of the dimension, discrimination between congruent versus incongruent positioning of the response options will unfold more efficiently (faster), compared with a stimulus representation at the marked end where processing is more difficult (and slower). Therefore, in terms of an alternative, asymmetric prediction for target versus distractor stimuli, we expect an interaction between type of category and side of response: For target category words (working at the unmarked end of the dimension), participants should be quicker making a correct response with their left hand than with their right hand. In contrast, for distractor category words (working at the marked end), we expect any spatial effect to be minimized or nonexistent for the reasons explained above. Moreover, if the markedness argument is correct, then trials with distractor category words (marked end) should exhibit longer response times overall than trials with target category words (unmarked end).

Method

Participants. Forty-five undergraduate students from the University of Freiburg (32 female, 13 male, mean age = 24.6 years) took part in the experiment, all with German-spoken backgrounds. They received course credit or €5.00 for their participation. No participant was excluded.

Materials. The same word pool was used as in Experiment 2. Five experimental blocks were used, and in each block, the assignment of materials, selection of categories and so forth was identical to Experiment 2.

Procedure. After reading instructions, participants were seated approximately 60 cm in front of the computer screen. Each block started directly with a test phase. In each test phase, 20 target category words, and 20 distractor category words were presented individually, one in each trial, in a random sequence, with the category name (e.g., "furniture") being constantly displayed in the upper left corner of the screen, as reminder. Participants were asked to indicate as quickly and accurately as possible, whether the word on the screen was an exemplar of the target category, or not. Each trial started with the presentation of a fixation cross in the center of the screen, for 800 ms. After this, the stimulus word was presented centrally for an open response interval. Along with it, and 7 cm below the word at two lateral positions 3 cm to the left and 3 cm to the right of the screen center, the words "YES" and "NO," or, alternatively NO and YES, one of these arrangements randomly assigned to each trial, were displayed as response cues (see Figure 2). Participants had to use the marked horizontal arrow keys to make their response in accordance with the arrangement of the response cues. Care was taken that the forefingers of *both* hands (left hand for left, right hand for right) were always used to operate the two directional keys.

All other presentation modalities and response options were the same as in Experiment 2. Between blocks, four easy arithmetic problems were presented as in Experiment 1 and 2. The experiment lasted between 20 and 25 min, including debriefing.

Results

Accuracy. The overall error level (i.e., pressing the incorrect key) was 5%, across all participants and all test trials. The final model used for analysis contained fixed effects for block (1, . . . , 5), orientation of response options ("YES-NO" vs. "NO-YES"), category (target vs. distractor) as well as the interactions of these factors. No significant effects emerged.

Response latencies. For correct responses, latency data were trimmed in the same way as in Experiment 1 and 2. The linear mixed model used for the analysis had the same fixed effect structure as the one just reported for accuracy. There was a significant effect of block ($F(4, 55.58) = 8.04$; $p < .0001$) with responses becoming faster from Block 1 through Block 5, again reflecting a practice effect ($M_{\text{Block 1}} = 856$, $M_{\text{Block 2}} = 838$, $M_{\text{Block 3}} = 820$, $M_{\text{Block 4}} = 814$, $M_{\text{Block 5}} = 792$ ms). Second, a significant category effect ($F(1, 59.99) = 45.74$; $p < .0001$) revealed that words from the target category were associated with faster correct responses than words from the distractor category ($M_{\text{Target}} = 801$ vs. $M_{\text{Distractor}} = 848$ ms). Crucially, the interaction between orientation of response options and category was significant ($F(1, 7863.86) = 4.45$; $p = .03$) see Figure 3. According to a subsequent simple effects analysis, correct responses to target category words were given quicker with the left hand than with the right hand ($M_{\text{left}} = 792$ vs. $M_{\text{right}} = 808$ ms; $z = -2.734$, $p = .006$, Cohen's $d = .19$), whereas for distractor category words, there was no difference ($M_{\text{left}} = 849$ vs. $M_{\text{right}} = 846$ ms; $z = .199$, $p = .84$). No further effects were significant. We did not analyze the latencies of trials with incorrect responses because of the overall infrequency of errors.

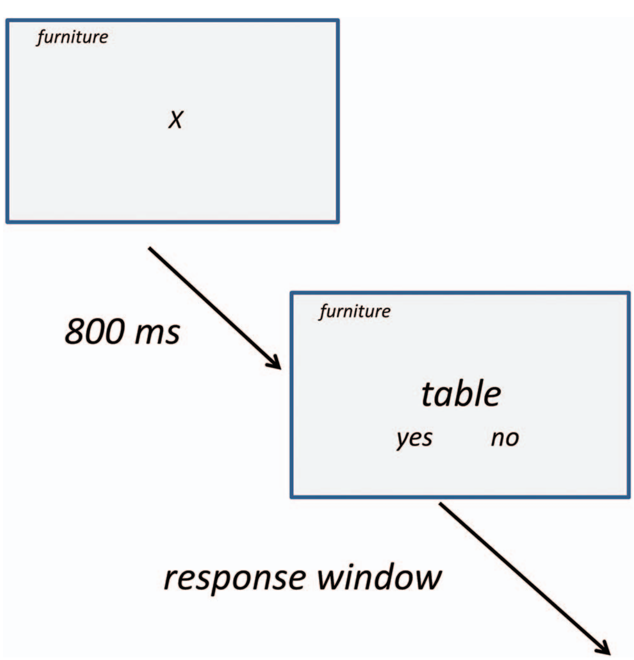

Figure 2. Experiment 3, experimental screen interface. Trial shown for a trial with "YES NO" response mapping.

Discussion

Using a stimulus-response compatibility paradigm, this experiment examined words from target and distractor categories in isolation, which enabled us to examine the unmarked and marked ends of the membership dimension separately. Response times overall were shorter than in the previous two experiments in which participants had to deal with four response options, as opposed to only two in the present case. Participants were quicker making a correct response with their left hand than with their right hand in target category trials. In contrast, they did not exhibit any difference in response speed between their hands when correctly responding to a distractor category trial. Overall, they were slower in distractor than in target category trials. This pattern supports the view that markedness does play a role in the processes under study. At the unmarked end of the dimension (member) a spatial simulation of the type "positive assignment to category means left" can be observed with isolated stimuli. At the marked end (nonmember) the same is not the case. For target category words, the simple effect (17 ms difference, $d = .19$) is small, and considerably smaller than the main effect associated with category type (48 ms difference, $d = .51$) such that one may assume that the factors affecting processing at the marked end (see above) may have obscured a potentially existing, symmetric effect of spatial congruence for distractor category words.

To the extent that one accepts the different paradigms (Experiments 1 and 2 vs. the present experiment) to be addressing, in essence, the same phenomenon, the present experiment can help clarify a further concern with respect to internal validity. In Experiments 1 and 2, one may argue that at least part of the reported left-anchoring effect might be because of a general motor tendency to respond faster when a stimulus is shown on the left side than on the right (1). Alternatively, an automatic tendency might exist to press left that facilitates left responses when they are appropriate and sometimes captures the response when a right response would have been appropriate (2). If (1) was true, only response speed as such could be affected but not accuracy, so the accuracy effects in Experiment 2 would remain unexplained, as would the interaction with category, as observed in the present experiment. The second possibility (2) can indeed explain the accuracy effect, but not the present interaction with stimulus category. Moreover, both alter-

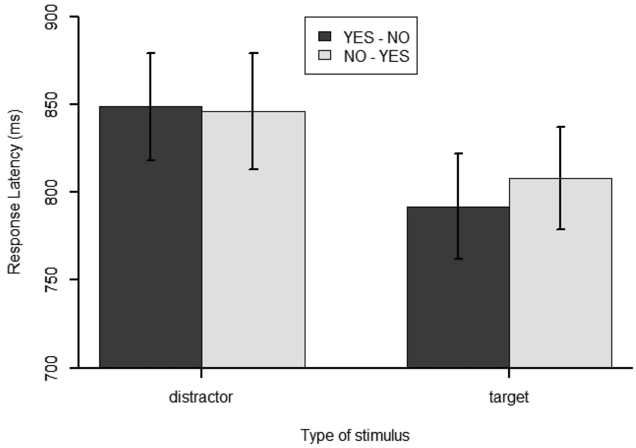

Figure 3. Experiment 3, mean response latency as a function of orientation of response options and category. Error bars show 1 *SE* above and below the mean.

native explanations predict a general advantage for pressing the left key that runs counter the evidence: The vast majority of all participants in Experiments 1–3 were right-dominant-handers, and with bilateral RTs to visual stimuli, responses by the dominant hand are usually made significantly faster than by the nondominant hand (Kerr, Mingay, & Elithorn, 1963, or, for a more recent null result, see Hiraoka et al., 2018). As to the present body of evidence, in Experiment 1 there was no significant difference in left versus right response latency for incorrect responses (19 ms), compared with a left-side advantage for correct responses (176 ms), which counts against a general "left = fast" explanation; however, this interaction is not significant. More convincingly, as hinted above, the present experiment provides evidence for left-anchoring only for target category words but not for distractor category words. This speaks against an overall explanation in terms of faster left-responses. The pattern overall lends support to an explanation based on congruence versus incongruence between an existing mental representation of membership and the required response mapping in each trial.

Experiment 4a: Category Discrimination (Replication With Response Key Labels Changed)

The aim of this experiment was a replication of Experiment 2, to rule out the possibility that the standard word ordering of yes-no (rather than no-yes) would be an alternative explanation for the results of Experiment 2. In that experiment, the correct answer to targets (category members) was always yes whereas the correct answer to nontargets (nonmembers) was always no. If participants imagined yes before, or left of, no, then the faster reactions to targets on the left side, with the yes-no response key positioned on the left, thereby making the assumed word order salient only for a left-key response but not for a right-key response, could potentially be explained by the canonical word order of yes and no. The rationale now used for Experiment 4a is that if response labels were changed into member and others, replacing yes and no, results should show convergence with those from Experiment 2 for our main hypothesis to be supported.

Another aim of Experiment 4a was to provide test power at an adequate level given the effect size obtained in Experiment 2, as $d = .25$. For one-tailed testing, with Type-I- and Type-II-errors fixed at $\alpha = .05$ and $\beta = .2$, this yielded a required sample size of $n = 101$ (GPower, see Faul, Erdfelder, Lang, & Buchner, 2007). This method can be seen as conservative because the linear mixed model analyses used here have more test power than the simple *t* test that underlies the GPower analysis.

Method

Participants. One hundred and one undergraduate students from the University of Freiburg (75 female, 26 male, mean age = 23.8 years) took part in the experiment, all with German-spoken backgrounds. They received course credits or €5.00 for their participation. One participant was excluded for data recording failure, so the final $N = 100$.

Materials. All methodology was identical to Experiment 2, except that the response labels used in this experiment were the German words "Mitglied-Mitglied," "Andere-Andere," "Mitglied-Andere," and "Andere-Mitglied," with *Mitglied (Member)* meaning a target category instance, and *Andere (Others)* a distractor category instance.[6] The experiment lasted between 20 and 25 min, including debriefing.

Results

Accuracy. The overall error level (i.e., pressing the incorrect key) was 7%, across all participants and all test pairs in all blocks. For the present analysis, only trials involving target-distractor and distractor-target trials were considered. The same effect structure as in Experiment 2 was used for the final model. There was a significant fixed factor effect for *side of target category word* ($\chi^2(1) = 4.07$; $p = .04$) indicating that responses were more accurate when the target category word was shown on the left side ($M_{\text{Target left}} = .95$) compared with the right side in a pair ($M_{\text{Target right}} = .93$). No other significant effects emerged.

Response latencies. For correct responses, latency data were trimmed in the same way as before, and the final linear mixed model used for the analyses had the same fixed effect structure as in Experiment 2. In this model, a significant main effect occurred for block ($F(4, 92.55) = 14.05$; $p < .001$) showing that responses became faster from Block 1 through Block 5, reflecting a practice effect ($M_{\text{Block 1}} = 1,530$, $M_{\text{Block 2}} = 1,415$, $M_{\text{Block 3}} = 1,376$, $M_{\text{Block 4}} = 1,324$, $M_{\text{Block 5}} = 1,292$ ms). Crucially, the effect of side of category word was also significant ($F(1, 100.22) = 5.17$, $p = .03$, Cohen's $d = .09$; $M_{\text{Target left}} = 1,367$ ms; $M_{\text{Target right}} = 1,406$ ms). No further effects were significant. Again, latencies of trials with incorrect responses were not analyzed because of the overall infrequency of errors.

[6] A minor methodological issue from Experiment 2 was removed. In Experiment 2, "birds" and "animals" were not excluded from being paired as target and distractor categories in one and the same block. This becomes an issue in a situation where "animals" is the target category. If "birds" is target category, this is only problematic if there were birds amongst the animal instances. Only the latter was found to be the case in two single trials from Experiment 2. These trials were excluded and the data reanalyzed with statistically identical results. In the present Experiment 4 the possibility that "animals" and "birds" could be paired as target and distractor in any block was removed.

Control via vertical response configuration. Again, we looked into the possibility that the vertical dimension could provide an equal scaffold for responding. Latencies associated with correct responses to target-target and distractor-distractor responses (for which the response keys were vertically arranged) were, therefore, analyzed in a similar linear mixed model as above. There were three significant effects. First, the block factor ($F(4, 91.48) = 15.06$; $p < .001$) indicated a practice effect in terms of responses becoming faster from the first block ($M_{\text{Block 1}} = 1{,}597$ ms) to the last block ($M_{\text{Block 5}} = 1{,}396$ ms). Second, a significant effect of the counterbalancing factor ($F(1, 97.61) = 9.48$; $p = .003$) revealed faster responses when the target-target response option was allocated to the UP arrow key ($M_{\text{up}} = 1{,}380$ ms), as compared with the DOWN arrow key ($M_{\text{down}} = 1{,}546$ ms). Third, the triple interaction between the counterbalancing factor, block, and type of pair was significant ($F(4, 3742.95) = 3.49$; $p = .008$) which was, upon inspection of the means, probably because of the practice effect appearing more pronounced for whatever type of pair was allocated to the UP arrow key. No further effects were significant. In particular, the interaction between the counterbalancing factor and type of pair was again insignificant ($F(1, 96.31) = .56$; $p = .46$). This interaction would be expected under the assumption that either the upper position of the response key was to generate quicker responding in general, or that a metaphorical blending with respect to target and distractor categories, and with respect to the target-target and distractor-distractor stimuli, did happen in the vertical dimension.

Discussion

By means of changing response labels from *yes—no* to *member—others*, Experiment 4a, now using a sample size designed to replicate the effect found in Experiment 2 with sufficient test power, demonstrated that this effect cannot be explained by the canonical word order of yes and no. Processing advantages for trials in which the target category member was presented left compared with right also appeared in a setting where response options were now labeled as *member* versus *others*. These two words are assumed to not stand in any culturally shaped order a priori and, therefore, do not imply any spatial configuration among each other per se, for participants used to read and write from left to right. The crucial effect in Experiment 4a emerged for accuracies, as well as for latencies. Our previous results on spatial biases (left anchoring of linear orders; von Hecker et al., 2016, see above) did show that the target effects indicative of the processing bias would mostly occur in latencies, but sometimes also in accuracies and not latencies, or in both types of DVs.

Experiment 4b: Category Discrimination (Replication of Experiment 4a With Target Category Label Centered and Changed Instructions)

In Experiments 2, 3, and 4a we used category labels to indicate the target category in each block. These labels were presented in the left upper corner of the screen in all three experiments, which could have constituted a spatial cue for response generation. The present experiment was conducted as an additional replication of the target effect under a condition where the spatial position of that indicative label (that was useful as response aid as the target category was different in each block) could not serve as a cue for responding. To this end, the target category label was now centered on the display during blocks, for the label to have no lateral implication by its position on the screen.

We also attempted to change instructions to a most conservative wording in terms of counteracting any primacy implications for the membership case (as opposed to nonmember) that could stem from the verbal delivery of instructions, in terms of the sequence of wording. As in Experiment 4a, participant recruitment aimed at a test power commensurate to the effect size observed in Experiment 2 ($d = .25$).

Method

Participants. One hundred and one undergraduate students from the University of Freiburg (53 female, 48 male, mean age = 24.8 years) took part in the experiment, all with German-spoken backgrounds. They received course credits or €5.00 for their participation.

Materials and procedure. All methodology was identical to Experiment 4a, except that in each block, the label indicating the target category was placed in the center of the screen. Also different from Experiment 4a, we used a new set of instructions which explicitly gave the nonmembership case spatial and temporal primacy within the sequence of terms used in the wording. The particular parts of the instructions as changed vis-à-vis Experiment 4a now read as follows (translated from German, and note the sequence of mentioning between others and member):

> In this experiment you will be presented with words that may be either nonmembers (others) or members (member) of a given target category (e.g., "buildings"). You will be shown 20 pairs of such words. One of the two words in a pair could be a nonmember, or both words could be nonmembers, or none of them. On the screen you will see four response options, corresponding to the four direction arrow keys. Within these options, others means that the word does not belong to the target category whereas member means that the word belongs to the target category. (The experiment lasted between 20 and 25 min, including debriefing.)

Results and Discussion

Accuracy. The overall error level (i.e., pressing the incorrect key) was 6%, across all participants and all test pairs in all blocks, after exclusion of three participants by boxplot analysis, for low overall accuracy. For the present analysis, only trials involving target-distractor and distractor-target trials were considered. The same effect structure as in Experiment 4a was used for the final model. There was a significant fixed factor effect for *side of target category word* ($\chi^2(1) = 5.89$; $p = .02$) indicating that responses were more accurate when the target category word was shown on the left side ($M_{\text{Target left}} = .96$) compared with the right side in a pair ($M_{\text{Target right}} = .95$). No other significant effects emerged.

Response latencies. For correct responses, latency data were trimmed in the same way as before, and the final linear mixed model used for the analyses had the same fixed effect structure as in Experiment 4a. In this model, a significant main effect occurred for block ($F(4, 95.39) = 19.31$; $p < .001$) showing that responses became faster from Block 1 through Block 5, reflecting a practice effect ($M_{\text{Block 1}} = 1{,}650$, $M_{\text{Block 2}} = 1{,}501$, $M_{\text{Block 3}} = 1{,}440$,

$M_{\text{Block 4}}$ = 1,399, $M_{\text{Block 5}}$ = 1,360 ms). Crucially, the effect of side of category word was also significant ($F(1, 97.76) = 4.34$, $p = .04$, Cohen's $d = .06$; $M_{\text{Target left}}$ = 1,456 ms; $M_{\text{Target right}}$ = 1,484 ms). Additionally, there was a significant three-way interaction showing that the practice effect in the counterbalancing group that responded to target-target with arrow up was more pronounced than in the group that responded to target-target with arrow down, but only for those trials that had the target category showing on the right ($F(4, 3886.82) = 3.02$; $p = .02$). Latencies of trials with incorrect responses were not analyzed because of the overall infrequency of errors.

Control via vertical response configuration. Again, we looked into the possibility that the vertical dimension could provide an equal scaffold for responding. Latencies associated with correct responses to target-target and distractor-distractor responses (for which the response keys were vertically arranged) were analyzed in a similar linear mixed model as above. There were three significant effects. First, the block factor ($F(4, 92.86) = 18.40$; $p < .001$) indicated a practice effect in terms of responses becoming faster from the first block ($M_{\text{Block 1}}$ = 1,710 ms) to the last block ($M_{\text{Block 5}}$ = 1,446 ms). Second, a significant effect of the counterbalancing factor ($F(1, 96.00) = 5.31$; $p = .02$) revealed faster responses when the target-target response option was allocated to the UP arrow key (M_{up} = 1,452 ms), as compared with the DOWN arrow key (M_{down} = 1,602 ms). Third, the triple interaction between the counterbalancing factor, block, and type of pair was significant, ($F(4, 3838.04) = 3.37$; $p = .009$) which was, upon inspection of the means, probably because of the practice effect appearing more pronounced for whatever type of pair was allocated to the DOWN arrow key. No further effects were significant. In particular, the interaction between the counterbalancing factor and type of pair was again insignificant ($F(1, 3839.72) = .88$; $p = .35$). This interaction would be expected under the assumption that either the upper position of the response key was to generate quicker responding in general, or that a metaphorical blending with respect to target and distractor categories, and with respect to the target-target and distractor-distractor stimuli, did happen in the vertical dimension.

By means of changing the position of the target category labels per block from the left upper corner of the screen (in previous experiments) to upper center screen (in the present experiment) we demonstrated that the left spatial position of the label was unlikely to have influenced the previous pattern of results, as the data from the present experiment very closely matched those from Experiment 4a. The crucial effect emerged for accuracies, as well as for latencies, replicating the pattern from Experiment 4a. Participants were more correct and responded faster in trials that presented the category member on the left side as compared with the right side, in a pair of test stimuli. Furthermore, these results were obtained under instructions that deliberately gave spatial and temporal primacy to the nonmembership case, in terms of the sequence of wording instructions, such that the paradigm variant used here constitutes a strongly conservative test of our assumptions.

Experiment 5a: Spontaneous Category Assignment

A simple sorting task was chosen to directly test the assumption that people envisage target categories (members) to the left of nontarget categories (nonmember). This was deemed an incremental step to our argument because it is desirable to show that a psychological mechanism, as theoretically postulated and as demonstrated in the context of small effect sizes and detected via response times in rigorous laboratory arrangements, would also have more straightforwardly observable, behavioral consequences. As in all previous experiments reported here, the basic hypothesis is that when it comes to assign a stimulus to a given category as either member or nonmember, Western participants would spatially simulate membership on the left side and nonmembership on the right side, based on a blending of primacy (member is unmarked) with dimensional dominance.

Method

This experiment was conducted at Cardiff, United Kingdom The procedures were approved by the Ethics Committee at the School of Psychology, Cardiff University. All participants gave written consent and were formally debriefed.

Participants. Fifty undergraduate students (all female, mean age = 19.4 years) took part in the experiment. They received course credit for their participation. We sampled participants for individual sessions from the student population at Cardiff University, exclusively with left-to-right reading/writing habit. No participant was excluded.

Materials and procedure. We prepared 80 cards of 5.5 cm (width) × 3 cm (height) to form two decks of 40 cards, A and B. In each deck there were 20 cards representing exemplars of the target category, and 2 × 10 cards representing exemplars of two distractor categories. In deck A *animal* was target with *musical instrument* and *furniture* as distractor categories; in deck B *sports discipline* was target with *clothes* and *fruit* being distractor categories. For each experimental session, the participant was presented with either deck A or B, according to an odd/even alternation with participant number. Each session started with all cards from the three categories being presented as an unsorted heap in front of the participant, at a distance of approx. 30 cm. The participant was then given a target category (e.g., "animal") and asked to sort the cards from the heap into two new, separate heaps closer to her/himself, as approx. 10 cm in front. The locations of the new heaps were suggested to be on both sides of a cardboard stripe that was laid directly in front as a central borderline, dividing a left and a right subspace on the table. The following instructions were given for deck A, avoiding any reference to "left" or "right": "This is a sorting task. I need you to make two heaps of cards out of those cards lying in front of you. Please put all cards with words describing animals on one side, and the remaining cards on the other side." Analogous instructions were provided for deck B, referring to "sports discipline" as target category. The duration of the experiment varied slightly between participants but was mostly within a few minutes, as procedures were terminated at the point in time at which it was clearly established by the participant, what sides each of the two heaps would occupy (see below).

Results and Discussion

The rationale for sequential testing (Ghosh & Sen, 1991; Wald, 1947) was followed. According to this methodology, sample size is not prespecified. Rather, observations (here equivalent to an experimental session with one participant) are collected one at a time. An observation consisted in a participant arranging for the heap corresponding to the target category either to be on the left or on the right side of the cardboard divider. An individual session was terminated when the two new heaps had been clearly established, such that not all cards from the unsorted heap had to be exhaustively sorted. After each session, we calculated the Wald parameters with respect to acceptance and rejection boundaries for H_1 and H_0 (Bernoulli-distribution), yielding a decision between the three alternatives of a) reject H_0 and accept H_1, b) retain H_0, or c) run another session. The R package SPRT (Bottine, 2015) was used, and we specified, as null hypothesis, H_0 $p = .5$, and as alternative hypothesis H_1 $p = .7$, indicating a preference for placing the target category heap left. Type-I- and Type-II-errors were fixed at $\alpha = .05$ and $\beta = .2$, for a test power of $1\text{-}\beta = .8$. Procedures were terminated after the 50st participant, with all 49 previous sessions having had led to the Wald-decision of "run another session." At that point, 35 participants had placed the target category heap on the left side of the divider, and the statistical decision upon running the Wald test was to reject H_0 and accept H_1. The results of this experiment provided evidence for a spontaneous placement of category members to the left of nonmembers in our student population.

Experiment 5b: Spontaneous Category Assignment With Changed Instructions

Method

Eighty-two undergraduate students (71 female, 11 male, mean age = 19.6 years) took part in the experiment. They received course credit for their participation. We sampled participants for individual sessions from the student population at Cardiff University, exclusively with left-to-right reading/writing habit. No participant was excluded.

The above experiment was replicated with two instructional variants. We wanted to test whether spontaneous sorting behavior would replicate the left-bias in category assignment when spatial/temporal primacy of instructional wording was counterbalanced. Thus, two counterbalancing conditions were created that varied only in the instruction given to participants. In Condition 1, they were told: "You are to sort these cards into two heaps, one for *nonmembers* and one for *members* of a certain target category. The target category is 'animals' ('sports disciplines')." Note that the negative concept of nonmember has spatial/temporal primacy on the instruction sheet. Instructions for Condition 2 read as follows: "You are to sort these cards into two heaps, one for *members* and one for *nonmembers* of a certain target category. The target category is 'animals' ('sports disciplines')." In this case, the positive concept of member has spatial/temporal primacy in terms of reading the instructions. We retained from Experiment 5a decks A and B as representing two different kinds of noun materials, so that in following sequential testing sessions, the four combinations of conditions (1 or 2) and decks (A or B) were rotated in a counterbalanced way through increasing participant numbers. All other methods and procedures were the same as in Experiment 5a.

Results and Discussion

Procedures were terminated after the 82nd participant, with all 81 previous sessions having had led to the Wald-decision of "run another session." At that point, 52 participants had placed the target category heap on the left side of the divider, and the statistical decision upon running the Wald test was to reject H_0 and accept H_1. In terms of the instructional variants used in this experiment, this means that in Condition 1, administered to $N = 42$ participants, 25 of them created the heap for animals/sports disciplines on the left side, whereas in Condition 2, administered to $N = 40$ participants, 27 made the same lateral decision. Conclusively, spontaneous placement of category members to the left of nonmembers, as captured in this card-sorting paradigm, did not vary as a function of instructional wording between Experiments 5a and 5b. The spatial representations of category membership as left, as opposed to nonmembership as right, in Western participants, can express itself in straightforward behavioral observations. The apparent left-sorting tendency for members does not seem to be influenced by potential spatial or temporal primacy as derived from the reading of instructions.

General Discussion

For humans, sensorimotor experiences, that is, experiences involving action in physical space, play an important role in the acquisition of reasoning abilities. More complex forms of knowledge are based on, and linked to, previously encountered spatial concepts (Clark, 1973; Niedenthal et al., 2005; Tomasello, 2009; Williams, Huang, & Bargh, 2009). For example, abstract notions of animacy, sequential action, or containment are rooted in early perceptual experiences in space that are later generalized as analog representations, and made available for domain-unspecific use (Barsalou, 1999; Mandler, 1992). As a consequence, abstract thought, although it may not be specifically of spatial content, may still be associated with spatial content, or spatial simulations (Barsalou, 2003). This is reflected in the general notion of mental space (Huttenlocher, 1968), but also in findings, for example, concerning the representation of social power (Schubert, 2005), or motor implications and the time line (SNARC effect; Chatterjee, 2001). At the most general level, the question asked in the present article is to what extent such spatial representations are involved in the case of an even more abstract, but quite fundamental, cognitive operation, that is, the assignment of an entity (that can be a concrete, perceived object or itself an abstract concept under consideration) to a superordinate conceptual category.

In seven experiments, using different paradigms, we presented evidence in support of the idea that category membership is spatially represented as a binary dimension with member mentally positioned left, and nonmember positioned right; that is, people who have been raised with left-to-right reading/writing background. Using a modified 2AFC paradigm, Experiment 1 showed faster responding to indicate the location of an old word on the screen when the display of the old word was on the left within a pair of horizontally arranged words. Experiments 2 and 4 (a, b) replicated this pattern using materials that did not differ in memory strength, or familiarity, between stimuli with membership versus not, in the target category. Experiment 3 yielded convergent results using a stimulus-response compatibility paradigm. Participants were faster when responding with their left hand than with their right hand in target category trials. In contrast, there was no difference in response times between both hands when correctly responding to a distractor category trial. Finally, using a spontaneous sorting task in Experiment 5 (a, b) we demonstrated that a psychological mechanism, as theoretically postulated and as demonstrated in the context of relatively small effect sizes and detected via response times in rigorous laboratory arrangements, would also have more straightforwardly observable, behavioral consequences.

To explain this pattern, two assumptions are made. First, in analogy to current explanations of the time line (Gevers et al., 2003), the number line (Chatterjee, 2001; Zebian, 2005), the spatial agency bias (Suitner & Maass, 2016), and the construction of abstract orders (von Hecker et al., 2016), we submit that there is a tendency to use the learned reading/writing habit as a scaffold for determining the origin of an ordered dimension in mental space (Lourenco & Longo, 2010). Second, the dimension is assumed to have exactly two ordered levels, that is, member (with high primacy and located at the origin, that is, on the left side, discussed at the beginning of the article), and nonmember with low primacy, and located on the right side. The rationale for member to be placed to the left of nonmember is based on the notion of metaphoric *blending* (Fauconnier & Turner, 1998; Pecher et al., 2010). Blending means that conceptual integration takes place in the service of construction of new meaning and mental representations (see Casasanto, 2009; Coulson & Oakley, 2005). Assuming that primacy on the membership dimension is fixed to be high on the left and low on the right (by virtue of the sequential action schema derived from reading/writing proceeding from left to right), blending assumes in addition that the meaning of member, representing a positive entity/magnitude, is integrated with high primacy to yield a placement on the left side, in contrast to nonmember that is positioned on the right side.

Note that the present set of results is difficult to explain on the basis of an activated schema of sequential action alone (from left to right), that is, without making the blending assumption. According to a "sequential action alone" assumption, faster responding on the left side is generally expected, irrespective of stimulus condition, and irrespective of whether or not a response was correct. Looking at incorrect trials in Experiment 1 we do not find a difference in key pressing time between responses on the left and right side, meaning that participants' response speed is unlikely to be determined by a sequential action schema alone, that is, without considering the match/mismatch between the perceived membership values of the two stimuli on the screen and in mental space. Also, Experiment 3 showed that, in target category trials, participants responded faster with their left hand than with their right hand when making a correct response. However, their response speed did not differ between both hands when correctly responding to a distractor category trial. This pattern, as well, speaks against a "left = fast" tendency as overall explanation. Instead, the pattern gives support to the idea that at the unmarked end of the category membership dimension (see Hamilton & Deese, 1971), that is, in the positive case of member, the blend between the positive dimensional asset (see above) and primacy yields a congruence between the lateral position in mental space where the target is being placed onto (member = left), and the side of the responding hand (in trials where the YES-response had to be made with the left hand). Correspondingly, in trials requiring a YES-response to be made with the *right* hand, there is incongruence. In other words, for positive target category instances, the difference between congruence and incongruence is rooted in the blending of the dimensional asset (member) with dimensional primacy (left), thereby causing "positive membership" to be positioned left. In contrast, for stimuli with no membership in the target category, a correct response had to be made at the marked end of the dimension, which is associated with processing disadvantages (Schriefers, 1990; Schubert, 2005; van der Schoot et al., 2009). This meant longer response times overall as compared with responding at the unmarked end, such that a symmetric difference, that is, a possible right-hand advantage in responding as a result of spatial congruence between nonmember and right, was most likely obscured.

Another alternative explanation to be addressed is polarity correspondence. The literature on stimulus-response-compatibility has developed the concept of polarity correspondence (PC) to explain spatial mapping effects, mostly in binary classification tasks in which stimuli have to be classified into one of two response categories (for an overview see Proctor & Cho, 2006). The basic idea here is that the two conceptual alternatives in a given dimension (e.g., old-young) are coded as plus and minus, whereby the assignments of these codes typically follow the logic of linguistic markedness. For example, in old-young, the adjective old is plus-coded because it neutralizes to describe the dimension whereas young is minus-coded because it does not (Lakens, 2012). Spatial mapping effects are explained in the PC framework by assuming that not only stimuli, but also responses are plus/minus-coded; such that, for example an up-response key would be plus-coded as opposed to a down-key being minus-coded, as well as a right-response key being plus-coded as opposed to a left-response key being minus-coded (Lakens, 2012; Proctor & Cho, 2006). Responses should then meet interference and occur with longer latencies when stimulus-code and response-code are different as compared with when they are the same. Polarity correspondence, as we see it, has difficulties explaining the present results. With respect to Experiments 1 and 2 it is difficult to reconcile with previous research on which the polarity correspondence framework is erected, that left (as side of presentation) would have to be plus-coded (instead of right) to yield correspondence with the plus-coded member (vs. the minus-coded nonmember). Similarly, with respect to Experiment 3, it is not clear why responding with the left hand (instead of the right hand) should be plus-coded to yield correspondence with member. The easier explanation in these cases is that, according to the present framework, compatibility (or correspondence, or a lack thereof) exists not between two particular codes but between (a) the spatial mental representation of the membership dimension and (b) the perceptual spatial input on the screen (or, the spatial features of the motor response).

With respect to theories of embodiment, it is important to highlight that we are not advocating a *strong* embodiment formulation (see Boroditsky & Ramscar, 2002 for this terminology). According to such a view, activation is postulated to occur in relevant sensorimotor areas in the brain, upon activating a certain type of cognitive activity. To use an example by Casasanto and Gijssels (2015) whose paper focuses on this issue, for a strong view of embodiment to be empirically supported for a metaphor like "Knowing is Seeing," one would have to show activation of visual cortex areas in an experimental situation that activates an operationalized instance of "knowing." As Casasanto and Gijssels (2015) argue, so far there exists too little evidence of just this quality as to be in a position of making such a *strong embodiment* claim. In the present case, the *strong* claim would amount to be able to show that category assignment as a cognitive activity is directly associated with sensorimotor areas in the brain known to be crucially involved in spatial behavior in terms of movement or vision. Besides such evidence being outside the empirical scope of the present article, the *strong* claim does not correspond to our position in the first place. Instead, in light of the present data we argue for a *mild* embodiment view (Boroditsky & Ramscar, 2002). This is to say that the roots of the mental representation of dimensional primacy may originate in the physical experiences of some action dynamic (reading/writing). However, its application later on, that is, after consolidation upon early learning, proceeds largely independent of physical parameters, including presentational conditions. The milder view postulates mental space to be an abstraction from its physical origins, which turns space into a universally versatile cognitive tool, applicable in many ways and contexts. The embodiment literature sometimes asserts a version of just this argument by emphasizing that "image schemas" (Lakoff & Johnson, 2008) may have been acquired by perceptuo-motor experiences, but then "abstracted away from the particulars of those experiences" (Casasanto & Gijssels, 2015, p. 335). The point here is also that activation of such abstracted schemas seem to show correlated brain activity in modality-nonspecific areas, such as the intraparietal sulcus (Quadflieg et al., 2011) and other areas of the parietal lobe (Citron & Goldberg, 2014). These areas have been associated with spatial processing and binary decision making, as mentioned earlier (see Kiani & Shadlen, 2009; Vickery & Jiang, 2009). Our own research has shown that spatial effects in inferential reasoning processes are tightly connected to bold-response effects in parietal areas (Hinton, Dymond, von Hecker, & Evans, 2010). These results confirm the *mild* view of embodiment insofar as spatially interpretable behavioral effects (the "symbolic distance effect"; see Hinton et al., 2010) are associated with activation differences in a modality-nonspecific area of the brain, that is, the parietal lobe.

Implications

We mentioned above evidence for the parietal lobe contributions to brain activation during participants' performance on the 2AFC paradigm, in the context of recognition memory tasks (Kiani & Shadlen, 2009). Together with other findings on the functional overlap between decision making and spatial processing in the IPL (see Andersen et al., 1985; Vickery & Jiang, 2009) the question arises to what extent the observed spatial effect, that is, the representation of category assignment in mental space, is just an accidental byproduct, or alternatively, a necessary constituent of the cognitive process. Some authors have proposed that functions supported by the parietal lobe are related to a general simulation and magnitude comparison device (e.g., Barsalou, 2008; Fias et al., 2003; Walsh, 2003, see also Dehaene, Dehaene-Lambertz, & Cohen, 1998), and others have argued that such a common mechanism would still not necessarily imply the existence of a common mental representation of magnitude (Cohen Kadosh, Brodsky, Levin, & Henik, 2008). Without being able to provide a conclusive answer to this question yet, the present results at least support the notion that spatial simulation effects are indeed observable at the level of mental representation. Note that our model has much in common with the model of agency representation as proposed by Suitner and Maass (2016). Like them, we assume sequential action to be represented as flowing from left to right, as an abstract derivation from reading/writing. Different from their model, however, we additionally assume the blending mechanism as discussed above. Blending assures that not only the spatial position of the dimensional origin is fixed (= left), but that there is also a unique dimensional anchoring. Just because primacy (as derived from reading/writing) is blended with dimensional dominance (as derived from semantics: member dominates nonmember), an anchoring constraint arises for the model in terms of member being placed at the dimensional origin, that is, on the left.

Conclusion

We demonstrated spatial effects associated with category assignment in simple decision tasks from different domains and paradigms (recognition memory, category discrimination, stimulus-response compatibility, and spontaneous sorting). While it is not yet entirely clear what functional status these effects might have, and whether they are a necessary part of cognition in these tasks, their existence provokes ideas about possible explanations of functionality. As such, spatial processing might functionally overlap with abstract reasoning at a fundamental level, as future research will have to further examine. As well, we see potential in a unifying quality of the present assumptions. In the area of ordered relations of abstract entities, spatial mental modeling has been thoroughly investigated, and theoretical links between magnitude processing on an ordered dimension and spatial processing have been brought forward (Dehaene et al., 1993; DeSoto, London, & Handel, 1965; Holyoak & Patterson, 1981; Leth-Steensen & Marley, 2000; Moyer & Bayer, 1976; Ouellet et al., 2010; Smith & Foos, 1975; von Hecker, Klauer, & Sankaran, 2013, 2016). In all of these cases, the spatial simulation that is assumed to be constructed maps the preexisting order relations as perceived in reality. With category assignment, this is different because membership (or not) in a given category is per se information only at qualitative, not ordinal level. What the present research shows is that even in this case, the general mechanisms postulated for stimuli at the ordinal level (determination of dimensional dominance between stimuli, modeling of sequential action, via reading/writing habits, or blending) can be used to explain essential aspects of response generation. It is possible to treat the binary, qualitative case of category assignment with the same theoretical tools as used in the specification of ordinal mental models, which we see as progress in unification.

Context of the Research

The ideas leading to this research originated in earlier work on the left-anchoring bias. We had found that rank orders on abstract dimensions of comparison, such as *older* or *stronger*, had a tendency to be anchored with their maximum on the left side. This finding was based on responses to binary decisions, that is, determining which of two stimuli was the dominant one in a horizontal display. This led to the consideration that in essence, the left-anchoring bias could be understood as a bias of categorization. By interpreting category membership as a dimension in mental space we do not take a strong embodiment position, in the sense that the representation of membership might have directly to do with physical experiences of space via motion or body posture. Rather, we adopt a milder position (see Boroditsky & Ramscar, 2002), meaning that mental space, as it might have originated in, and solidified through, physical experience during early years of cognitive development, has become an abstract tool in the mind of adults. We believe that people use mental space for reasoning, that is, for representing abstract entities and the relations between them; and it is in this spirit that the research was undertaken.

References

Andersen, R. A., Essick, G. K., & Siegel, R. M. (1985). Encoding of spatial location by posterior parietal neurons. *Science, 230,* 456–458. http://dx.doi.org/10.1126/science.4048942

Barsalou, L. W. (1999). Perceptions of perceptual symbols. *Behavioral and Brain Sciences, 22,* 637–660. http://dx.doi.org/10.1017/S0140525X99532147

Barsalou, L. W. (2003). Situated simulation in the human conceptual system. *Language and Cognitive Processes, 18,* 513–562. http://dx.doi.org/10.1080/01690960344000026

Barsalou, L. W. (2008). Cognitive and neural contributions to understanding the conceptual system. *Current Directions in Psychological Science, 17,* 91–95. http://dx.doi.org/10.1111/j.1467-8721.2008.00555.x

Bates, D., Maechler, M., Bolker, B. M., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*. Retrieved from http://arxiv.org/abs/1406.5823

Boroditsky, L., & Ramscar, M. (2002). The roles of body and mind in abstract thought. *Psychological Science, 13,* 185–189. http://dx.doi.org/10.1111/1467-9280.00434

Bottine, S. M. (2015). SPRT: Wald's Sequential Probability Ratio Test (R package version 1.0) [Computer software]. Retrieved from https://CRAN.R-project.org/package=SPRT

Bröder, A., & Schütz, J. (2009). Recognition ROCs are curvilinear-or are they? On premature arguments against the two-high-threshold model of recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35,* 587–606. http://dx.doi.org/10.1037/a0015279

Bucur, B., Madden, D. J., Spaniol, J., Provenzale, J. M., Cabeza, R., White, L. E., & Huettel, S. A. (2008). Age-related slowing of memory retrieval: Contributions of perceptual speed and cerebral white matter integrity. *Neurobiology of Aging, 29,* 1070–1079. http://dx.doi.org/10.1016/j.neurobiolaging.2007.02.008

Casasanto, D. (2009). Embodiment of abstract concepts: Good and bad in right- and left-handers. *Journal of Experimental Psychology: General, 138,* 351–367. http://dx.doi.org/10.1037/a0015854

Casasanto, D., & Gijssels, T. (2015). What makes a metaphor an embodied metaphor? *Linguistics Vanguard: Multimodal Online Journal, 1,* 327–337. http://dx.doi.org/10.1515/lingvan-2014-1015

Chatterjee, A. (2001). Language and space: Some interactions. *Trends in Cognitive Sciences, 5,* 55–61. http://dx.doi.org/10.1016/S1364-6613(00)01598-9

Chatterjee, A., Southwood, M. H., & Basilico, D. (1999). Verbs, events and spatial representations. *Neuropsychologia, 37,* 395–402. http://dx.doi.org/10.1016/S0028-3932(98)00108-0

Citron, F. M., & Goldberg, A. E. (2014). Metaphorical sentences are more emotionally engaging than their literal counterparts. *Journal of Cognitive Neuroscience, 26,* 2585–2595. http://dx.doi.org/10.1162/jocn_a_00654

Clark, H. H. (1973). Space, time, semantics, and the child. In T. E. Moore (Ed.), *Cognitive development and the acquisition of language* (pp. 27–63). New York, NY: Academic Press. http://dx.doi.org/10.1016/B978-0-12-505850-6.50008-6

Clark-Carter, D. (2004). *Quantitative psychological research: A student's handbook*. New York, NY: Psychology Press.

Cohen Kadosh, R., Brodsky, W., Levin, M., & Henik, A. (2008). Mental representation: What can pitch tell us about the distance effect? *Cortex, 44,* 470–477. http://dx.doi.org/10.1016/j.cortex.2007.08.002

Coulson, S., & Oakley, T. (2005). Blending and coded meaning: Literal and figurative meaning in cognitive semantics. *Journal of Pragmatics, 37,* 1510–1536. http://dx.doi.org/10.1016/j.pragma.2004.09.010

Czernochowski, D. (2005). *Age differences in familiarity and recollection*. Irvine, CA: Universal-Publishers.

Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of parity and number magnitude. *Journal of Experimental Psychology: General, 122,* 371–396. http://dx.doi.org/10.1037/0096-3445.122.3.371

Dehaene, S., Dehaene-Lambertz, G., & Cohen, L. (1998). Abstract representations of numbers in the animal and human brain. *Trends in Neurosciences, 21,* 355–361. http://dx.doi.org/10.1016/S0166-2236(98)01263-6

DeSoto, C. B., London, M., & Handel, S. (1965). Social reasoning and spatial behavior. *Journal of Personality and Social Psychology, 2,* 513–521.

Dobel, C., Diesendruck, G., & Bölte, J. (2007). How writing system and age influence spatial representations of actions: A developmental, cross-linguistic study. *Psychological Science, 18,* 487–491. http://dx.doi.org/10.1111/j.1467-9280.2007.01926.x

Duyck, W., Desmet, T., Verbeke, L. P., & Brysbaert, M. (2004). WordGen: A tool for word selection and nonword generation in Dutch, English, German, and French. *Behavior Research Methods, Instruments, & Computers, 36,* 488–499. http://dx.doi.org/10.3758/BF03195595

Egan, J. P. (1975). *Signal detection theory and ROC analysis*. New York, NY: Academic Press.

Fauconnier, G., & Turner, M. (1998). Conceptual integration networks. *Cognitive Science, 22,* 133–187. http://dx.doi.org/10.1207/s15516709cog2202_1

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39,* 175–191. http://dx.doi.org/10.3758/BF03193146

Fias, W. (1996). The importance of magnitude information in numerical processing: Evidence from the SNARC effect. *Mathematical Cognition, 2,* 95–110. http://dx.doi.org/10.1080/135467996387552

Fuhrman, O., & Boroditsky, L. (2010). Cross-cultural differences in mental representations of time: Evidence from an implicit nonlinguistic task. *Cognitive Science, 34,* 1430–1451. http://dx.doi.org/10.1111/j.1551-6709.2010.01105.x

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. http://dx.doi.org/10.1080/09541440540000112

Gevers, W., Reynvoet, B., & Fias, W. (2003). The mental representation of ordinal sequences is spatially organized. *Cognition, 87,* B87–B95. http://dx.doi.org/10.1016/S0010-0277(02)00234-2

Ghosh, B. K., & Sen, P. K. (1991). *Handbook of sequential analysis*. New York, NY: Marcel Dekker.

Glanzer, M., & Adams, J. K. (1990). The mirror effect in recognition memory: Data and theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 5–16. http://dx.doi.org/10.1037/0278-7393.16.1.5

Glanzer, M., & Bowles, N. (1976). Analysis of the word-frequency effect in recognition memory. *Journal of Experimental Psychology: Human Learning and Memory, 2,* 21–31. http://dx.doi.org/10.1037/0278-7393.2.1.21

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York, NY: Wiley.

Hamilton, H. W., & Deese, J. (1971). Does linguistic marking have a psychological correlate? *Journal of Verbal Learning and Verbal Behavior, 10,* 707–714. http://dx.doi.org/10.1016/S0022-5371(71)80079-8

Hegarty, M., & Just, M. A. (1993). Constructing mental models of machines from text and diagrams. *Journal of Memory and Language, 32,* 717–742. http://dx.doi.org/10.1006/jmla.1993.1036

Hicks, J. L., & Marsh, R. L. (1998). A decrement-to-familiarity interpretation of the revelation effect from forced-choice tests of recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 1105–1120. http://dx.doi.org/10.1037/0278-7393.24.5.1105

Hinton, E. C., Dymond, S., von Hecker, U., & Evans, C. J. (2010). Neural correlates of relational reasoning and the symbolic distance effect: Involvement of parietal cortex. *Neuroscience, 168,* 138–148. http://dx.doi.org/10.1016/j.neuroscience.2010.03.052

Hiraoka, K., Igawa, K., Kashiwagi, M., Nakahara, C., Oshima, Y., & Takakura, Y. (2018). The laterality of stop and go processes of the motor response in left-handed and right-handed individuals. *Laterality, 23,* 51–66. http://dx.doi.org/10.1080/1357650X.2017.1311906

Hockley, W. E. (1984). Analysis of response time distributions in the study of cognitive processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 598–615. http://dx.doi.org/10.1037/0278-7393.10.4.598

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. http://dx.doi.org/10.1037/0096-1523.7.6.1283

Huttenlocher, J. (1968). Constructing spatial images: A strategy in reasoning. *Psychological Review, 75,* 550–560.

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. http://dx.doi.org/10.1016/j.jml.2007.11.007

Jou, J., Flores, S., Cortes, H. M., & Leka, B. G. (2016). The effects of weak versus strong relational judgments on response bias in Two-Alternative-Forced-Choice recognition: Is the test criterion-free? *Acta Psychologica, 167,* 30–44. http://dx.doi.org/10.1016/j.actpsy.2016.03.014

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 problem. *Journal of Personality and Social Psychology, 103,* 54–69. http://dx.doi.org/10.1037/a0028347

Kahana, M. J. (2014). *Foundations of human memory*. New York, NY: Oxford University Press.

Kahn, I., Davachi, L., & Wagner, A. D. (2004). Functional-neuroanatomic correlates of recollection: Implications for models of recognition memory. *The Journal of Neuroscience, 24,* 4172–4180. http://dx.doi.org/10.1523/JNEUROSCI.0624-04.2004

Kaup, B., & Zwaan, R. A. (2003). Effects of negation and situational presence on the accessibility of text information. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29,* 439–446. http://dx.doi.org/10.1037/0278-7393.29.3.439

Kellen, D., & Klauer, K. C. (2014). Discrete-state and continuous models of recognition memory: Testing core properties under minimal assumptions. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40,* 1795–1804. http://dx.doi.org/10.1037/xlm0000016

Kerr, M., Mingay, R., & Elithorn, A. (1963). Cerebral dominance in reaction time responses. *British Journal of Psychology, 54,* 325–336. http://dx.doi.org/10.1111/j.2044-8295.1963.tb00887.x

Kiani, R., & Shadlen, M. N. (2009). Representation of confidence associated with a decision by neurons in the parietal cortex. *Science, 324,* 759–764.

Knauff, M. (2013). *Space to reason: A spatial theory of human thought*. Cambridge, MA: MIT Press.

Kosslyn, S. M. (1994). *Image and brain: The resolution of the imagery debate*. Cambridge, MA: MIT Press.

Kunar, M. A., Rich, A. N., & Wolfe, J. M. (2010). Spatial and temporal separation fails to counteract the effects of low prevalence in visual search. *Visual Cognition, 18,* 881–897. http://dx.doi.org/10.1080/13506280903361988

Lahl, O., Göritz, A. S., Pietrowsky, R., & Rosenberg, J. (2009). Using the World-Wide Web to obtain large-scale word norms: 190,212 ratings on a set of 2,654 German nouns. *Behavior Research Methods, 41,* 13–19. http://dx.doi.org/10.3758/BRM.41.1.13

Lakens, D. (2012). Polarity correspondence in metaphor congruency effects: Structural overlap predicts categorization times for bipolar concepts presented in vertical space. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38,* 726–736. http://dx.doi.org/10.1037/a0024955

Lakoff, G., & Johnson, M. (2008). *Metaphors we live by*. Chicago, IL: University of Chicago Press.

Lea, R. B., & Mulligan, E. J. (2002). The effect of negation on deductive inferences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 303–317. http://dx.doi.org/10.1037/0278-7393.28.2.303

Leinster, T. (2016). Basic category theory. *arXiv preprint*. Retrieved from https://arxiv.org/pdf/1612.09375.pdf

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

Lourenco, S. F., Ayzenberg, V., & Lyu, J. (2016). A general magnitude system in human adults: Evidence from a subliminal priming paradigm. *Cortex, 81,* 93–103. http://dx.doi.org/10.1016/j.cortex.2016.04.013

Lourenco, S. F., & Longo, M. R. (2010). General magnitude representation in human infants. *Psychological Science, 21,* 873–881. http://dx.doi.org/10.1177/0956797610370158

Maass, A., & Russo, A. (2003). Directional bias in the mental representation of spatial events: Nature or culture? *Psychological Science, 14,* 296–301. http://dx.doi.org/10.1111/1467-9280.14421

Maass, A., Suitner, C., Favaretto, X., & Cignacchi, M. (2009). Groups in space: Stereotypes and the spatial agency bias. *Journal of Experimental Social Psychology, 45,* 496–504. http://dx.doi.org/10.1016/j.jesp.2009.01.004

Maass, A., Suitner, C., & Nadhmi, F. (2014). What drives the spatial agency bias? An Italian-Malagasy-Arabic comparison study. *Journal of Experimental Psychology: General, 143,* 991–996. http://dx.doi.org/10.1037/a0034989

MacDonald, M. C., & Just, M. A. (1989). Changes in activation levels with negation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 633–642. http://dx.doi.org/10.1037/0278-7393.15.4.633

Macmillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide*. New York, NY: Cambridge University Press.

Malmberg, K. J. (2008). Recognition memory: A review of the critical findings and an integrated theory for relating them. *Cognitive Psychology, 57,* 335–384. http://dx.doi.org/10.1016/j.cogpsych.2008.02.004

Mandler, J. M. (1992). How to build a baby: II. Conceptual primitives. *Psychological Review, 99,* 587–604. http://dx.doi.org/10.1037/0033-295X.99.4.587

Moyer, R. S., & Bayer, R. H. (1976). Mental comparison and the symbolic distance effect. *Cognitive Psychology, 8,* 228–246. http://dx.doi.org/10.1016/0010-0285(76)90025-6

Niedenthal, P. M., Barsalou, L. W., Winkielman, P., Krauth-Gruber, S., & Ric, F. (2005). Embodiment in attitudes, social perception, and emotion. *Personality and Social Psychology Review, 9,* 184–211. http://dx.doi.org/10.1207/s15327957pspr0903_1

Nobel, P. A., & Shiffrin, R. M. (2001). Retrieval processes in recognition and cued recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 384–413. http://dx.doi.org/10.1037/0278-7393.27.2.384

Ouellet, M., Santiago, J., Funes, M. J., & Lupiáñez, J. (2010). Thinking about the future moves attention to the right. *Journal of Experimental Psychology: Human Perception and Performance, 36,* 17–24. http://dx.doi.org/10.1037/a0017176

Patro, K., Nuerk, H. C., & Cress, U. (2015). Does your body count? Embodied influences on the preferred counting direction of preschoolers. *Journal of Cognitive Psychology, 27,* 413–425. http://dx.doi.org/10.1080/20445911.2015.1008005

Pecher, D., Van Dantzig, S., Boot, I., Zanolie, K., & Huber, D. E. (2010). Congruency between word position and meaning is caused by task-induced spatial attention. *Frontiers in Psychology, 1,* 30. http://dx.doi.org/10.3389/fpsyg.2010.00030

Proctor, R. W., & Cho, Y. S. (2006). Polarity correspondence: A general principle for performance of speeded binary classification tasks. *Psychological Bulletin, 132,* 416–442. http://dx.doi.org/10.1037/0033-2909.132.3.416

Quadflieg, S., Etzel, J. A., Gazzola, V., Keysers, C., Schubert, T. W., Waiter, G. D., & Macrae, C. N. (2011). Puddles, parties, and professors: Linking word categorization to neural patterns of visuospatial coding. *Journal of Cognitive Neuroscience, 23,* 2636–2649. http://dx.doi.org/10.1162/jocn.2011.21628

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

Rosch, E. (1978). Principles of categorization. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 27–48). Hillsdale, NJ: Erlbaum.

Sanford, A. J., Moxey, L. M., & Paterson, K. B. (1996). Attentional focusing with quantifiers in production and comprehension. *Memory & Cognition, 24,* 144–155. http://dx.doi.org/10.3758/BF03200877

Santiago, J., Lupiáñez, J., Pérez, E., & Funes, M. J. (2007). Time (also) flies from left to right. *Psychonomic Bulletin & Review, 14,* 512–516. http://dx.doi.org/10.3758/BF03194099

Schriefers, H. (1990). Lexical and conceptual factors in the naming of relations. *Cognitive Psychology, 22,* 111–142. http://dx.doi.org/10.1016/0010-0285(90)90005-O

Schröder, A., Gemballa, T., Ruppin, S., & Wartenburger, I. (2012). German norms for semantic typicality, age of acquisition, and concept familiarity. *Behavior Research Methods, 44,* 380–394. http://dx.doi.org/10.3758/s13428-011-0164-y

Schubert, T. W. (2005). Your highness: Vertical positions as perceptual symbols of power. *Journal of Personality and Social Psychology, 89,* 1–21. http://dx.doi.org/10.1037/0022-3514.89.1.1

Shaki, S., Fischer, M. H., & Petrusic, W. M. (2009). Reading habits for both words and numbers contribute to the SNARC effect. *Psychonomic Bulletin & Review, 16,* 328–331. http://dx.doi.org/10.3758/PBR.16.2.328

Sherman, M. (1973). Bound to be easier? The negative prefix and sentence comprehension. *Journal of Verbal Learning and Verbal Behavior, 12,* 76–84. http://dx.doi.org/10.1016/S0022-5371(73)80062-3

Sherman, M. (1976). Adjectival negation and the comprehension of multiply negated sentences. *Journal of Verbal Learning and Verbal Behavior, 15,* 143–157. http://dx.doi.org/10.1016/0022-5371(76)90015-3

Singmann, H., Bolker, B., Westfall, J., & Aust, F. (2018). *afex: Analysis of factorial experiments* (R package version 0.20–2) [Computer software]. Retrieved from https://CRAN.R-project.org/package=afex

Smith, D. G., & Duncan, M. J. J. (2004). Testing theories of recognition memory by predicting performance across paradigms. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 615–625. http://dx.doi.org/10.1037/0278-7393.30.3.615

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

Starns, J. J., Chen, T., & Staub, A. (2017). Eye movements in forced-choice recognition: Absolute judgments can preclude relative judgments. *Journal of Memory and Language, 93,* 55–66. http://dx.doi.org/10.1016/j.jml.2016.09.001

Suitner, C., & Maass, A. (2016). Spatial agency bias: Representing people in space. In J. M. Olson & M. P. Zanna (Eds.), *Advances in experimental social psychology* (Vol. 53, pp. 245–301). Cambridge, MA: Academic Press.

Swets, J., Tanner, W. P., Jr., & Birdsall, T. G. (1961). Decision processes in perception. *Psychological Review, 68,* 301–340. http://dx.doi.org/10.1037/h0040547

Tomasello, M. (2009). *The cultural origins of human cognition*. Cambridge, MA: Harvard University Press. http://dx.doi.org/10.2307/j.ctvjsf4jc

Tversky, B. (1993). Cognitive maps, cognitive collages, and spatial mental models. In A. U. Frank, & I. Campari (Eds.), *Spatial information theory: A theoretical basis for GIS, Proceedings COSIT ’93. Lecture notes in Computer Science, 716* (pp. 14–24). Berlin, Germany: Springer.

Tversky, B., Kugelmass, S., & Winter, A. (1991). Cross-cultural and developmental trends in graphic productions. *Cognitive Psychology, 23,* 515–557. http://dx.doi.org/10.1016/0010-0285(91)90005-9

van der Schoot, M., Bakker Arkema, A. H., Horsley, T. M., & van Lieshout, E. C. D. M. (2009). The consistency effect depends on markedness in less successful but not successful problem solvers: An eye movement study in primary school children. *Contemporary Educational Psychology, 34,* 58–66. http://dx.doi.org/10.1016/j.cedpsych.2008.07.002

Vickery, T. J., & Jiang, Y. V. (2009). Inferior parietal lobule supports decision making under uncertainty in humans. *Cerebral Cortex, 19,* 916–925. http://dx.doi.org/10.1093/cercor/bhn140

von Hecker, U., Klauer, K. C., & Sankaran, S. (2013). Embodiment of social status: Verticality effects in multilevel rank-orders. *Social Cognition, 31,* 374–389.

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. http://dx.doi.org/10.1037/xlm0000220

Wald, A. (1947). *Sequential analysis*. New York, NY: Wiley.

Walsh, V. (2003). A theory of magnitude: Common cortical metrics of time, space and quantity. *Trends in Cognitive Sciences, 7,* 483–488. http://dx.doi.org/10.1016/j.tics.2003.09.002

Williams, L. E., Huang, J. Y., & Bargh, J. A. (2009). The Scaffolded Mind: Higher mental processes are grounded in early experience of the physical world. *European Journal of Social Psychology, 39,* 1257–1267. http://dx.doi.org/10.1002/ejsp.665

Yonelinas, A. P. (1994). Receiver-operating characteristics in recognition memory: Evidence for a dual-process model. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 1341–1354. http://dx.doi.org/10.1037/0278-7393.20.6.1341

Yonelinas, A. P., & Parks, C. M. (2007). Receiver operating characteristics (ROCs) in recognition memory: A review. *Psychological Bulletin, 133,* 800–832. http://dx.doi.org/10.1037/0033-2909.133.5.800

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. http://dx.doi.org/10.1163/1568537054068660

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin, 123,* 162–185. http://dx.doi.org/10.1037/0033-2909.123.2.162

(*Appendix follows*)

Appendix

Modeling of Effects

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.

Experiments 1, 2 and 4

In the first step, we fitted three models for each data type (a1, a2 and a3 for accuracy data, as well as t1, t2 and t3 for latency data). All of these models had the same fixed effect structure, that is, group (counterbalancing variable for key positions assigned to "up" versus "down," see text) presentation side of the target element and block, as well as their interactions. All models had a random intercept for participants. Models a3 and t3 had only this intercept, so these models are minimal. Models a1/t1 also had a random slope for block as function of participant, whereas a2/tm2 had a random slope for side of target element instead. These models were then compared using the Chi square difference statistic τ^2. Models of a given Type 1 or 2 were compared with the corresponding model of Type 3, the minimal model. If there was a significant difference in fit, the particular type of random slope as specified in the nonminimal model under comparison was then retained for the final model, afinal, resp., tfinal. In a second step, these final models were assembled and run to evaluate the respective fixed effect structure from those models (see Jaeger, 2008). This strategy 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 used 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).

Experiment 3

The same general strategy as above was followed. Different from the above, the fixed effect structure for the tested models was presentation side of the target element, category (target vs. distractor, see text), and block, as well as their interactions. Thus, as compared with the minimal models a4/t4 (only with random intercept for participants), models a1/t1 had a random slope for block as function of participant, a2/t2 had a random slope for side of target element, and a3/t3 had a random slope for category type (target versus distractor).

Experiment 1

Accuracies

Model	*df*	AIC	BIC	Loglik	Deviance	$\Delta\chi^2$	Δdf	*p*
a3	25	2546.1	2690.7	−1248.1	2496.1			
a1	45	2560.9	2821.1	−1235.4	2470.9	25.243	20	.1923
a2	27	2549.7	2705.9	−1247.9	2495.7	0.3851	2	.8248

Note. AIC = Akaike's information criterion; BIC = Bayesian information criterion. afinal: No random slopes as a function of participants were kept.

Latencies

Model	*df*	AIC	BIC	Loglik	Deviance	$\Delta\chi^2$	Δdf	*p*
t3	26	3559.8	3699.9	−1753.9	3507.8			
t1	46	3515.3	3763.2	−1711.7	3423.3	84.472	20	6.751e-10***
t2	28	3525.5	3676.3	−1734.7	3469.5	38.356	2	4.688e-09***

Note. AIC = Akaike's information criterion; BIC = Bayesian information criterion. tfinal: Random slopes for block and side of target element, as a function of participants, are kept.
*** $p < .001$.

Experiment 2

Accuracies

Model	*df*	AIC	BIC	Loglik	Deviance	$\Delta\chi^2$	Δdf	*p*
a3	21	639.68	758.33	−298.84	597.68			
a1	35	654.71	852.45	−292.36	584.71	12.971	14	.5288
a2	23	642.37	772.31	−298.19	596.37	1.311	2	.5191

Note. AIC = Akaike's information criterion; BIC = Bayesian information criterion. afinal: No random slopes as a function of participants were kept.

(*Appendix continues*)

Latencies

Model	*df*	AIC	BIC	Loglik	Deviance	$\Delta\chi^2$	Δdf	*p*
t3	22	1325.8	1448.0	−640.88	1281.8			
t1	36	1228.0	1428.1	−577.98	1156.0	125.79	14	<2.2e-16***
t2	24	1321.7	1455.2	−636.87	1273.7	8.0137	2	.01819*

Note. AIC = Akaike's information criterion; BIC = Bayesian information criterion. tfinal: Random slopes for block and side of target element, as a function of participants, are kept.
$^{*}p < .05$. $^{***}p < .001$.

Experiment 2a + 2b, Combined Analysis

Accuracies

Model	*df*	AIC	BIC	Loglik	Deviance	$\Delta\chi^2$	Δdf	*p*
a3	21	1564.8	1698.7	−761.38	1522.8			
a1	35	1572.8	1796.0	−751.39	1502.8	19.973	14	.131
a2	23	1559.9	1706.5	−756.94	1513.9	8.8851	2	.01177*

Note. AIC = Akaike's information criterion; BIC = Bayesian information criterion. afinal: Random slopes as a function of participants were kept for side of target element.
$^{*}p < .05$.

Latencies

Model	*df*	AIC	BIC	Loglik	Deviance	$\Delta\chi^2$	Δdf	*p*
t3	22	3643.6	3781.5	−1799.8	3599.6			
t1	36	3407.7	3633.5	−1667.9	3335.7	263.84***	14	<2.2e-16
t2	24	3601.6	3752.1	−1776.8	3553.6	45.996	2	1.028e-10***

Note. AIC = Akaike's information criterion; BIC = Bayesian information criterion. tfinal: Random slopes for block and side of target element, as a function of participants, are kept.
$^{***}p < .001$.

Experiment 3

Accuracies

Model	*df*	AIC	BIC	Loglik	Deviance	$\Delta\chi^2$	Δdf	*p*
a4	21	3269.0	3418.2	−1613.5	3227.0			
a1	35	3229.5	3478.2	−1579.8	3159.5	67.431	14	5.612e-09***
a2	23	3271.3	3434.7	−1612.6	3225.3	1.6969	2	.4281
a3	23	3260.5	3423.9	−1607.3	3214.5	12.445	2	.001985**

Note. AIC = Akaike's information criterion; BIC = Bayesian information criterion. afinal: Random slopes for block and category, as a function of participants, are kept.
$^{**}p < .01$. $^{***}p < .001$.

Latencies

Model	*df*	AIC	BIC	Loglik	Deviance	$\Delta\chi^2$	Δdf	*p*
t4	22	−3406.8	−3252.7	1725.4	−3450.8			
t1	36	−3521.2	−3269.1	1796.6	−3593.2	142.47	14	<2.2e-16***
t2	24	−3403.3	−3235.2	1725.6	−3451.3	0.4853	2	.7845
t3	24	−3440.2	−3272.1	1744.1	−3488.2	37.418	2	7.495e-09***

Note. AIC = Akaike's information criterion; BIC = Bayesian information criterion. tfinal: Random slopes for block and category, as a function of participants, are kept.
$^{***}p < .001$.

Experiment 4a

Accuracies

Model	*df*	AIC	BIC	Loglik	Deviance	$\Delta\chi^2$	Δdf	*p*
a3	21	1890.7	2027.6	−924.35	1848.7			
a1	35	1893.2	2121.3	−911.62	1823.2	25.462	14	.03027*
a2	23	1890.7	2040.6	−922.34	1844.7	4.0121	2	.1345

Note. AIC = Akaike's information criterion; BIC = Bayesian information criterion. afinal: Random slopes for block as a function of participants are kept.
$^{*}p < .05$.

Latencies

Model	*df*	AIC	BIC	Loglik	Deviance	$\Delta\chi^2$	Δdf	*p*
t3	22	4907.5	5048.2	−2431.8	4863.5			
t1	36	4697.1	4927.4	−2312.6	4625.1	238.36	14	<2.2e-16***
t2	24	4844.3	4997.8	−2398.1	4796.3	67.207	2	2.548e-15***

Note. AIC = Akaike's information criterion; BIC = Bayesian information criterion. tfinal: Random slopes for block and side of target element, as a function of participants, are kept.
$^{***}p < .001$.

Experiment 4b

Accuracies

Model	*df*	AIC	BIC	Loglik	Deviance	$\Delta\chi^2$	Δdf	*p*
a3	21	1755.1	1891.6	−856.56	1713.1			
a1	35	1755.8	1983.2	−842.93	1685.8	27.276	14	.01772*
a2	23	1758.7	1908.2	−856.37	1712.7	0.3854	2	.8247

Note. AIC = Akaike's information criterion; BIC = Bayesian information criterion. afinal: Random slopes for block as a function of participants are kept.
$^{*}p < .05$.

Latencies

Model	*df*	AIC	BIC	Loglik	Deviance	$\Delta\chi^2$	Δdf	*p*
t3	22	5224.7	5365.5	−2590.4	5180.7			
t1	36	5038.3	5268.6	−2483.2	4966.3	14.42	14	<2.2e-16***
t2	24	5207.2	5360.7	−2579.6	5159.2	21.554	2	2.088e-05***

Note. AIC = Akaike's information criterion; BIC = Bayesian information criterion. tfinal = Random slopes for block and side of target element, as a function of participants, are kept.
$^{***}p < .001$.

Received June 29, 2018
Revision received April 1, 2020
Accepted May 17, 2020 ■