COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 43 (2019) 1–23

ISSN: 1551-6709 online
DOI: 10.1111/cogs.12709

The Space–Time Congruency Effect: A Meta-Analysis

Linda von Sobbe,[a] Edith Scheifele,[a] Claudia Maienborn,[b] Rolf Ulrich[c]

[a]*Collaborative Research Center 833, University of Tübingen*
[b]*German Department, University of Tübingen*
[c]*Department of Psychology, University of Tübingen*

Received 22 May 2018; received in revised form 28 November 2018; accepted 30 November 2018

Abstract

Several reaction time (RT) studies report faster responses when responses to temporal information are arranged in a spatially congruent manner than when this arrangement is incongruent. The resulting space–time congruency effect is commonly attributed to a culturally salient localization of temporal information along a mental timeline (e.g., a mental timeline that runs from left to right). The present study aims to provide a compilation of the published RT studies on this time–space association in order to estimate the size of its effect and the extent of potential publication bias in this field of research. In this meta-analysis, three types of task are distinguished due to hitherto existing empirical findings. These findings suggest that the extent to which time is made relevant to the experimental task has a systematic impact on whether or not the mental timeline is activated. The results of this meta-analysis corroborate these considerations: First, experiments that make time a task-relevant dimension have a mean effect size of $d = 0.46$. Second, in experiments in which time is task irrelevant, the effect size does not significantly deviate from zero. Third, temporal priming studies have a surprisingly high mean effect size of $d = 0.47$, which, however, should be adjusted to $d = 0.36$ due to publication bias.

Keywords: Mental timeline; Space–time congruency effect; Spatial metaphor of time; Deictic and sequential concepts of time; Effect size; Grounded cognition

1. Introduction

Time cannot be perceived because there is no adequate physical stimulus of time; thus, we seem to rely on our experience of space to frame the otherwise elusive concept of time (Burr, Tozzi, & Morrone, 2007; Gentner, Imai, & Boroditsky, 2002). Philosophers, linguists, and cognitive psychologists likewise assume that our conceptualization of time

---

Correspondence should be sent to Linda von Sobbe, Sonderforschungsbereich 833, Nauklerstr. 35, 72074 Tübingen, Germany. E-mail: linda.von-sobbe@uni-tuebingen.de

is based on our notion of space, since the latter can be traced back to sensorimotor experiences, whereas this is not the case for time (Eikmeier, Alex-Ruf, Maienborn, Schröter, & Ulrich, 2016; Núñez & Cooperrider, 2013). The linkage of time as an abstract concept and space as its concrete counterpart is conceived of as the *spatial metaphor of time* (Clark, 1973, p. 50), which becomes apparent in most languages across the world as spatial expressions, for example locative prepositions, are used to express temporal notions (Haspelmath, 1997). Moreover, co-speech gestures suggest that time is organized along a mental timeline (Casasanto & Jasmin, 2012; Núñez & Sweetser, 2006) with time moving from left to right, right to left, back to front, or top to bottom. While the lateral and vertical timelines presumably depend on a language's writing direction (Bergen & Chan Lau, 2012; Fuhrman & Boroditsky, 2007; Ouellet, Santiago, Israeli, & Gabay, 2010), the sagittal timeline is assumed to originate from our spatial orientation in the world, since foreseeable future events are ahead of us due to our experience of usually facing the objects that we are approaching (Lakoff & Johnson, 1980; Núñez & Sweetser, 2006).

To examine the psychological reality of the mental timeline, a considerable number of reaction time (RT) studies have been conducted within the last 15 years (e.g., Torralbo, Santiago, & Lupiáñez, 2006; Ulrich & Maienborn, 2010; Weger & Pratt, 2008; see also Eikmeier et al., 2016 for an overview). On the whole, these studies have revealed a space–time congruency effect; that is, the response to temporal stimuli is faster when the spatial response is consistent with the culturally salient direction of the mental timeline (congruent space–time mapping) than when it is reversed (incongruent condition). Thus, a person who is used to writing and reading from left to right will be faster at responding to past-related stimuli with the left hand as compared to responding with the right hand, whereas future-related stimuli will be responded to faster with the right than with the left hand. This congruency effect can be conceived of as a variation of the traditional spatial stimulus-response compatibility effect, which has been extensively studied in experimental psychology (see Proctor, Reeve, & Van Zandt, 1992). Moreover, the size of this effect varies in the range of milliseconds so that it seems likely to assume that these RT differences reflect an unconscious, but sound difference in the speed of mental processing (Eikmeier et al., 2016).

However, the space–time congruency effect is not corroborated uniformly by all studies. More precisely, in the majority of experiments, in which the concept of time is not focussed by the task, no facilitation of responses in the congruent space–time mapping has been observed (Maienborn, Alex-Ruf, Eikmeier, & Ulrich, 2015; Sell & Kaschak, 2011; Ulrich & Maienborn, 2010; Ulrich et al., 2012). This raises questions regarding the depth of the process that causes the space–time congruency effect. Instead of automatic sensorimotor activation, it might be a facilitated memory access that elicits the congruency effect. According to this memory access account, spatial locations that are associated with past and future work as cues for performing the RT task, when the task explicitly asks participants to respond to the temporal reference of an expression (Maienborn et al., 2015). Thus, the congruent mapping might simply be remembered more easily than the incongruent space–time mapping.

An alternative explanation for the congruency effect's dependency on the salience of the concept of time is that a coherent working model such as the mapping of time onto space happens only if its efforts are compensated by its benefits (Santiago, Román, & Ouellet, 2011). Thus, activating a mental timeline will only happen in those situations in which there is a gain to cope with that particular situation and its requirements, which is the case when the experimental task requires a temporal placement, but which is not the case when the experimental task draws the participant's attention away from the concept of time.

It is in our interest to assemble the diverging results of different types of task in order to frame their meaning and possible consequences. Another major motivation for providing a compilation of the so far published RT studies by means of a meta-analysis is to assess the size of the space–time congruency effect in those experiments, in which the concept of time is made salient, as there is further substantial variation concerning the design of the conducted studies beyond the salience of time. Under these circumstances a potential publication bias—that is, nonsignificant or negative results are put into the researcher's file drawer and do not get published—could imply that the actual space–time congruency effect is smaller than portrayed by the published studies and hence might not even significantly deviate from zero (Rosenthal, 1979; Ulrich, Miller, & Erdfelder, 2018). We will incorporate the potential publication bias into the estimation of the real effect size in order to examine whether it significantly deviates from zero and, thus, can be considered a sound effect.

For this purpose, a short overview of the designs of the studies that are included in the meta-analysis will be provided in Section 1.1. The studies have shown some variation concerning the temporal and spatial information (Sections 1.1.1 and 1.1.2, respectively) that are used as a cue for the activation of the mental timeline. In addition, different axes of the mental timeline have been investigated across the studies (Section 1.1.3). Type of temporal and spatial information, as well as direction of the mental timeline or language, however, will not be regarded as moderator variables in the meta-analysis. This is because even though they presumably have a small-sized impact on the size of the congruency effect, the resulting moderated effect sizes should be distributed narrowly around some common mean which will be accounted for adequately by the use of the random-effects model (Borenstein, Hedges, Higgins, & Rothstein, 2009, p. 61). Thus, the coding of corresponding moderator variables is not required. Apart from that, a too fine-grained subdivision of the gathered effect sizes would lead to smaller sample sizes, which would reduce the analysis's power. Such a fine-grained analysis would not provide an additional benefit for our purpose because we are interested in whether the space–time congruency effect is an empirically corroborated and thus acceptable fact.

In contrast, there is good reason to subdivide the studies into two steps. In a first step, the studies will be subdivided with regards to the type of task that is used. In a second step, the axis's origin will be considered. As mentioned above, the way time is made relevant by the task seems to have a major impact on whether or not the appearance of a congruency effect can be expected (e.g., Scheifele, Eikmeier, Alex-Ruf, Maienborn, & Ulrich, 2018; Ulrich & Maienborn, 2010). Thus, the different types of task will be introduced in Section 1.1.4 and taken into account as subcategories in the meta-analysis whose method will be outlined in Section 2. Consequently, not only an overall effect size for all gained studies will be calculated (Section 3), but also a separate analysis for each level of the factor *Task*, that is, *time is task relevant*, *time is task irrelevant*, and *temporal priming* (Sections 3.1–3.3). Additionally, an analysis of the moderator variable *Axis* with its two levels *lateral/vertical* and *sagittal* within each level of *Task* will be calculated, since the lateral and vertical axes' origin presumably lies in a culture's writing system, whereas the sagittal axis is based on the more profound experience of moving forward. Finally, after comparing the effect sizes of the level *Task* with one another (Section 3.4), the level *time is task irrelevant* will be addressed in more detail in Section 3.5.

1.1. Types of stimuli, responses, axes, and task

Three different axes of the mental timeline have been tested in the incorporated RT studies using different temporal and spatial cues in order to elicit its activation. In most studies, temporal information is incorporated by means of the stimulus material, whereas space usually gets activated via response mode. Thus, the stimulus-response mapping serves as the coding of the congruent and incongruent conditions of the mapping of time onto space. The extent to which both dimensions, space and time, are made salient, varies across studies. Especially the salience of the dimension of time has systematically been manipulated by means of different tasks.

1.1.1. Temporal information

To keep the effect sizes of the incorporated studies comparable, our meta-analysis focuses on the processing of deictic and sequential concepts of time. Thus, the analysis will not include studies in which the duration of a stimulus was manipulated for the assessment of the corresponding duration judgements (e.g., Di Bono et al., 2012). The latter line of research focuses on the mechanisms underlying the representation of time intervals and detaches too much from findings regarding the processing of abstract time concepts, which however is of major interest for the purpose of this study. The studies that will be included in the meta-analysis all examine the space–time congruency effect by measuring RT to a stimulus that requires a manual or vocal response that is either congruent or incongruent with the participant's culturally learned and salient mental timeline.

Incorporated are studies that use the following temporal cues as stimulus, prime, or response, presented either visually or auditorily:

1. Past- or future-related words or phrases (Aguirre & Santiago, 2017; Bottini, Crepaldi, Casasanto, Crollen, & Collignon, 2015; Casasanto & Bottini, 2014; De la Vega, Eikmeier, Ulrich, & Kaup, 2016; Ding, Feng, Cheng, Liu, & Fan, 2015; Eikmeier, Alex-Ruf, Maienborn, & Ulrich, 2015, experiment 2; Eikmeier, Hoppe, & Ulrich, 2015; Eikmeier, Schröter, Maienborn, Alex-Ruf, & Ulrich, 2013, experiment 2; Hartmann & Mast, 2012; Kong & You, 2012; Ouellet, Román, & Santiago, 2012; Ouellet, Santiago, Funes, & Lupiáñez, 2010; Ouellet, Santiago, Israeli, et al., 2010; Rolke et al., 2013; Rolke, Ruiz Fernández, Seibold, & Rahona, 2014; Santiago, Lupiáñez, Pérez, & Funes, 2007; Torralbo et al., 2006; Weger & Pratt, 2008, experiment 2);

2. Sentences containing temporal information (Eikmeier, Alex-Ruf, et al., 2015, experiment 1; Eikmeier et al., 2013, experiment 1; Maienborn et al., 2015; Scheifele et al., 2018; Sell & Kaschak, 2011; Ulrich & Maienborn, 2010; Ulrich et al., 2012); and
3. Triplets of pictures showing the progression of an event at which the middle stage represents the reference point for an earlier and a later stage (Boroditsky, Fuhrman, & McCormick, 2011; Fuhrman & Boroditsky, 2007; Fuhrman et al., 2011) and entities such as buildings, actors, or life events that can be categorized as earlier or later compared to some given reference point (Loeffler, Raab, & Cañal-Bruland, 2017; Miles, Tan, Noble, Lumsden, & Macrae, 2011; Walker, Bergen, & Núñez, 2014, 2017; Weger & Pratt, 2008, experiment 1).

There are two references known to us in which spatial instead of temporal cues were used as stimuli that were combined with temporal instead of spatial responses to assess the space–time congruency effect. First, Eikmeier et al. (2013, experiment 2) presented sounds that originated in front of or behind the participant to assess the sagittal axis. Second, in Eikmeier, Alex-Ruf, et al. (2015, experiment 2) the sounds were played on the participant's right or left side for examining the lateral axis. In both experiments the participant was asked to respond vocally to the location of the sound's source either with "back"/"front"/"left"/"right" in the control condition or with "past"/"future" in the experimental condition. All other studies mentioned above used temporal cues as stimulus or prime and spatial information as response but not as stimulus.

1.1.2. Spatial information

In most of the studies considered in our meta-analysis, spatial cues were encoded by requesting a manual response along the axis in question. For testing the lateral axis, participants were asked to respond with a left or right keypress either with their left and right index fingers (e.g., Santiago et al., 2007) or with one finger only in order to cause a movement (e.g., Miles et al., 2011). Similarly, the sagittal axis was tested by corresponding keypresses (further away or close to the body, e.g., Fuhrman et al., 2011; Sell & Kaschak, 2011) or by a slider that was moved forward and backward (e.g., Ulrich et al., 2012). Walker et al. (2017) used mouse presses instead of keypresses as response mode and there are two studies that asked for vocal responses, in which space was encoded semantically through spatial expressions like "back"/"front" or "left"/"right" as response to temporal information (Eikmeier, Alex-Ruf, et al., 2015, experiment 1; Eikmeier et al., 2013, experiment 1). For the vertical axis, up and down keypresses were used as response keys (e.g., Boroditsky et al., 2011).

There are a few studies that used rather subtle spatial cues. Torralbo et al. (2006) presented pictures of human silhouettes with a speech bubble either in front of or behind the silhouettes' face containing a time-related word in Spanish. Participants in their first experiment were asked to respond vocally with "pasado" and "futuro" to indicate whether the word was past- or future-related. In this experiment, space was only salient by means of the speech bubble's position relative to the silhouette. In a likewise subtle manner, Hartmann and Mast (2012) mapped time onto space by moving the participants forward and backward on a motion platform while they were reading words and evaluating their temporal reference. The participants' bodily movement was task- and response-irrelevant but was still evaluated in relation to the temporal reference of the presented words (Loeffler et al., 2017 used a similar set-up). Similarly, in Walker et al. (2014) spatial information was only implicitly salient. Here, space was not encoded through bodily movement but through the location of the target's auditory presentation: Participants had to respond vocally to categorize the temporal reference of the auditorily presented target irrespective of the sound's source that was arranged either along the lateral or the sagittal axis.

1.1.3. Axes

Three axes, the lateral, the sagittal, and the vertical, have been considered in the underlying studies. In most studies, the axis that has been examined is culturally salient for the corresponding group of participants. Only a few studies tested the space–time congruency effect for axes that were not common in that particular culture, usually to control the effect of the experimental condition (e.g., Hartmann & Mast, 2012) or to compare two different cultures (e.g., Boroditsky et al., 2011).[1] Since this applies only to a minority of studies, it is not reasonable to code an axis's cultural salience as a moderator variable. Instead, the issue will be dealt with by excluding those effect sizes from the meta-analysis that are gained by testing culturally irrelevant axes.

However, in order to account for the different origins of the axes as described in the introduction, the lateral and vertical axes will be treated as one level of the factor *Axis*, whereas the sagittal axis will be considered another level of the factor *Axis*. Thus, within each group of *Task*, the level *sagittal* will be compared to the level *lateral/vertical* in order to assess whether the more profound origin of the sagittal axis will also yield a larger mean effect size.

1.1.4. Task

The way time is made salient or relevant for the task has a major influence on the effect size (Scheifele et al., 2018; Ulrich & Maienborn, 2010). Thus, it is of eminent importance for the following meta-analysis to distinguish between the different types of task.

Most of the studies that present temporal information as stimulus material ask their participants to categorize the temporal reference of the stimulus. This procedure makes the temporal cues salient for the participants since they are task relevant. The appearance of a space–time congruency effect in these cases is highly predictable as the spatial responses that categorize the temporal reference can be remembered better, when the key-assignment is congruent to the known and culturally transmitted assignment of time onto space. Also making use of a mental timeline as a coherent model to cope with the task is very economical with respect to cognitive capacities. Subsequently, this type of task will be called *time is task relevant*.

However, it is also possible to obscure the object of investigation in order to examine whether the mental timeline gets activated automatically when temporal reference is task irrelevant. For instance, Ulrich and Maienborn (2010) asked their participants to evaluate whether the presented sentence was sensible or not by pressing a left or a right key, respectively. The alignment of the keys was reversed in a second block so that they recorded RTs of both the left and right hand to both past and future-related sensible stimuli. While RTs of the right hand were expected to be shorter for future-related stimuli than for past-related stimuli, the reverse pattern was expected for left hand responses. In this kind of design, the stimulus sentences still have to be processed thoroughly in order to execute the experimental task properly; however, their temporal reference is of no importance to accomplish the task. Most of the studies assessing the mental timeline's automatic activation by making time task irrelevant fail to find a congruency effect. Thus, this second type of task is referred to as *time is task irrelevant*.[2] Usually a sensicality judgment as exemplified above by means of Ulrich and Maienborn (2010) is used in this type of task. There is only one experiment in which a different technique has been implemented to distract the participants' attention from the concept of time: In their experiment 3, Aguirre and Santiago (2017) asked their participants to judge whether the stimulus expression referred to a real or a potential event. Thus, time was not a task-relevant dimension in their instructional design either.

Additionally, a third category of task will be distinguished in the following analysis. The studies in this third category have used temporal information not as stimulus but as a prime that shows up shortly before the main task has to be carried out. This is the case in experiment 2A of Weger and Pratt (2008), who presented a prime word with either a prospective or a retrospective cue that was followed by a white circle, which appeared on the right or the left side of the computer screen. Participants were instructed to indicate as fast as possible that they had detected the target by pressing a key that corresponded to the side of the computer screen at which the circle appeared. The space–time congruency effect that arises in this kind of set-up (i.e., responses with the left hand to a target appearing on the left side are faster after seeing a past-related word as compared to a future-related word) shows that the processing of temporal information unconsciously shifts the visual attention along to the mental timeline which causes a facilitation of response velocity (Weger & Pratt, 2008). Since *temporal priming* involves a rather subtle activation of the mental timeline, it will be differentiated from tasks, in which temporal cues explicitly have to be evaluated with respect to their temporal reference.

2. Method

2.1. Sample of studies

We searched for articles using the databases PsycINFO and MEDLINE with the keyword "mental timeline" and scanned for articles containing terms such as "space-time congruency effect," "space-time mapping," "space time reaction time," "spatial metaphor of time," and "space time alignment." Furthermore, we looked through the reference lists of current articles to find out whether we had overlooked some relevant papers and publications.

We had to exclude all experiments that examined the mental timeline with methods other than RT recordings such as assessing gestures or laying out temporal sequences (Hendricks & Boroditsky, 2015). We only included studies with adult participants because the mental timeline is most probably caused by cultural imprint so needs time to evolve in individuals (Tillman, Tulagan, & Barner, 2015). Of the remaining studies we could only include those that reported statistics with which we could accurately estimate an effect size. This resulted in 30 references. Some of these, which had several sub-experiments, only reported *F*-values in experiments that actually showed a significant congruency effect, so that it was not possible to include all sub-experiments of each paper. Other sub-experiments had a design that tested more than one congruency effect such as experiment 1 of Fuhrman et al. (2011), in which the lateral, the vertical, and the sagittal axes were tested on English and Mandarin natives, resulting in five estimated effect sizes for one experiment. However, since independence of effect sizes is a precondition for the meta-analysis, the mean of the corresponding effect sizes is taken as an estimate in those cases, in which there is more than one effect size for the same group of participants.[3] Eventually this method led to 62 estimated effect sizes for the 30 incorporated studies on which this meta-analysis is based. We thank Roberto Aguirre, Roberto Bottini, Daniel Casasanto, Verena Eikmeier, Marc Ouellet, Bettina Rolke, and Andrea Sell for providing us with suitable data on their published experiments so that we could incorporate them into the meta-analysis, which we would not have been able to do otherwise. The Supporting Information contains an overview of all incorporated studies as well as specific information on how we determined the effect size from the statistics of each study.

2.2. *Coded factors*

As outlined above, we analyzed *Task* as a factor, of which there are three levels: *time is task relevant*, which applies to 41 out of the 62 estimated effect sizes; *time is task irrelevant*, for which we were only able to gain 10 effect sizes from the underlying references; and as a third level we coded *temporal priming*, which only holds for 11 estimated effect sizes. Within each subgroup of the factor *Task*, we also calculated an analysis of the factor *Axis* with the two levels *lateral/vertical* and *sagittal* to embrace the different axes' origin. We did not differentiate between the lateral and vertical axis, as they are assumed to share a common origin, as outlined above. Since all *temporal priming* studies examine the lateral axis, the factor *Axis* cannot be applied for this level of *Task*. For the level *time is task irrelevant,* a further subdivision into *low* and *high temporal complexity* derives from a theoretical and empirical point of view. This additional subdivision is only carried out because the considered effect sizes vary systematically and can be deduced from theoretical considerations.

We did not code language as a moderator variable since the corresponding differences in effect sizes are expected to be low and can, apart from that, not be derived appropriately from the theoretical background. Additionally, a further subdivision would lead to smaller sample sizes, thus reducing the analyses' power. The gained informative value would not have justified the loss of statistical explanatory power. This is why we also refrained from further clustering into, for example, type of stimulus material.

2.3. *Effect size analyses*

Each effect size was calculated as Cohen's *d*, that is, the standardized difference of mean RTs for the congruent as compared to the incongruent space–time mapping, using the Comprehensive Meta-Analysis (CMA, Version 3.3.070, 2014, Biostat, Englewood, NJ, USA) software. The effect size was defined as positive when the mean RT of the congruent condition was lower than the mean RT of the incongruent condition, which was the case for all of the effect sizes. *d* was calculated from *t*-values combined with the sample size. If only a *F*-value was reported, we calculated the corresponding *t*-value by $t = \sqrt{F}$. For the majority of studies, congruency was a within-subjects factor. Only four effect sizes are gained from experiments that treat congruency as a between-subjects factor, so *d* was estimated on the basis of independent groups in these cases, using the sample sizes of the two groups and the *t*-value for the calculation of *d*. If two or more effect sizes were gained from the same pool of participants, the mean of the corresponding *t*-values was taken to calculate *d*.

We chose the random-effects model for all analyses, because it assumes that the true effect sizes vary across studies but are distributed around some common grand mean, of which the given data represent a random sample (Borenstein et al., 2009). Since the studies included in this meta-analysis have examined different axes on different groups of population, this model is appropriate. Even when clustering the effect sizes into different levels of *Task* or *Axis*, the studies still vary considerably regarding type of language, stimulus material, and response mode. Despite this a-priori assumption, the *Q*-statistic that tests on heterogeneity and is implemented in CMA will be reported for all analyses since it indicates whether the assumption of a distribution of the true effect sizes around some common mean is justified by the variability of the data. For two analyses of *Task*, namely the estimation of the weighted mean effect size of *time is task irrelevant* and of *temporal priming* studies, the *Q*-statistic does not support the choice of the random effects model. However, while for the former the results are exactly the same compared to using a fixed-effect model, for the latter, the observed overall variance is largely due to heterogeneity of the true effect sizes as indicated by a high I^2-value. Thus, for both effect size estimations, the use of the random effects model seems appropriate despite a nonsignificant *Q*-statistic.[4] Details will be given in each corresponding section. For estimation and correction of publication bias, Duval and Tweedie's (2000) Trim and Fill that is implemented in CMA and that uses the linear (*L*) estimator has been employed (Borenstein, 2005, p. 203).

3. Results

We carried out a preliminary analysis for which the overall weighted mean effect size of the 62 underlying estimated effect sizes is significant ($d = 0.403$, 95% CI [0.335, 0.470], $p < .0005$). For the preliminary analysis, the Q-statistic supports our choice of the random-effects model since 54.1% of the observed overall variance is due to heterogeneity of the true effect sizes ($I^2 = 54.1$). Thus the data seem to be based on more than one true effect size ($Q(61) = 132.9$, $p < .0005$), which is in line with the random-effects model but not with the fixed-effect model, which assumes one common true effect size (Borenstein et al., 2009, p. 61).

We then calculated separate mean effect sizes for each level of the factor *Task* (see Fig. 1 for an overview of observed effect sizes and Table 1 for an overview of adjusted effect sizes). Furthermore, we decided to examine the publication bias for each level of *Task* separately by the use of distinct funnel plots, since heterogeneity can cause asymmetry of the funnel plots, even in the absence of publication bias (Terrin, Schmid, Lau, & Olkin, 2003). An analysis of the moderator variable *Axis* has also been carried out separately for each level of *Task*.

3.1. Time is task relevant

The choice of the random-effects model to calculate the weighted mean effect size for the level *time is task relevant* (Fig. 2), was again supported by the test for heterogeneity, which is significant ($Q(40) = 67.3$, $p = .004$) and 40.5% of the observed variation is due to heterogeneity of the true effect sizes ($I^2 = 40.5$). The weighted mean effect size significantly deviates from zero ($d = 0.463$, 95% CI [0.389, 0.536], $p < .0005$). Using the random-effects model to look for missing studies reveals that there is no publication bias (Fig. 3). Since no studies are trimmed and filled when applying Duval and Tweedie's trim and fill, the mean effect size stays the same ($d = 0.463$).

3.1.1. Time is task relevant—lateral/vertical versus sagittal axis

Nine out of the 41 effect sizes of *time is task relevant* are yielded by an assessment of the sagittal axis. Further, four effect sizes are gained by experiments that test the lateral as well as the sagittal axis (Walker et al., 2014, 2017). In order to guarantee for independence of subgroups, the effect sizes of the lateral axis are removed of these four effect

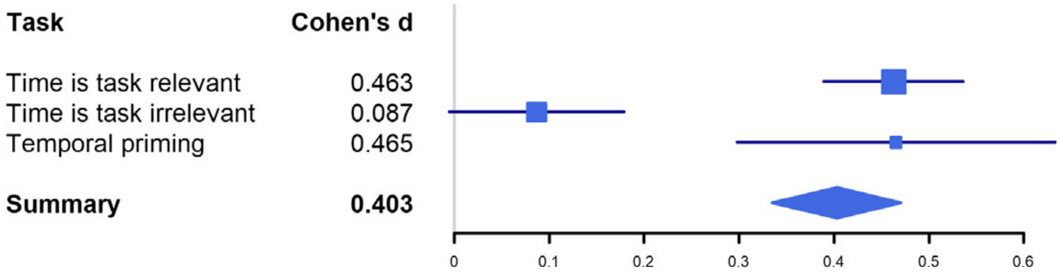

Fig. 1. Forest plot for the different levels of *Task*.

Table 1
Summary of the adjusted mean effect sizes (Cohen’s *d*) after applying Duval and Tweedie’s trim and fill

Task	Number of Estimated Effect Sizes	Number of Effect Sizes That Are Trimmed and Filled	Adjusted Mean Effect Size (*d*)	Adjusted 95% CI
Time is task relevant	41	0	0.463	0.389 to 0.536
Time is task irrelevant	10	3	0.051	−0.035 to 0.137
Temporal priming	11	3	0.359	0.179 to 0.540

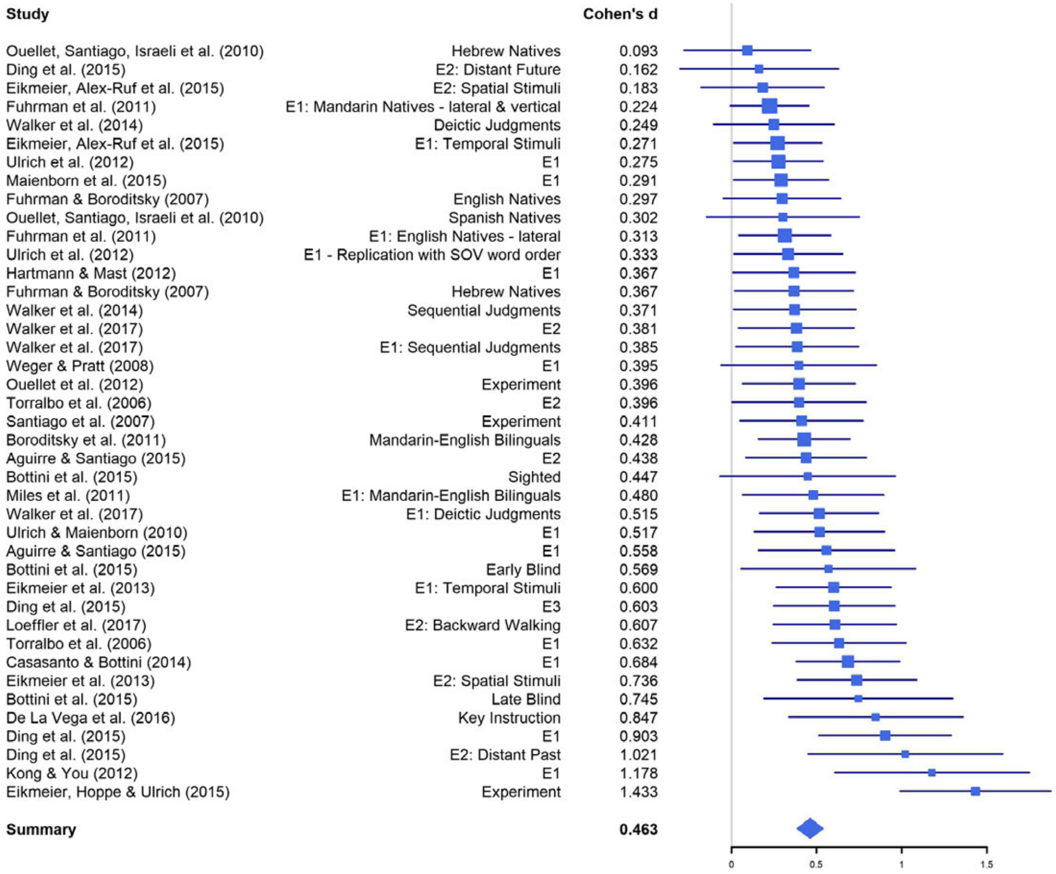

Fig. 2. Forest plot for *time is task relevant* (E = experiment).

sizes,[5] yielding a total of 13 effect sizes of the level *sagittal*. Three of the 41 effect sizes of *time is task relevant* are based on the testing of the lateral and the vertical axes. The remaining 25 of the 41 effect sizes are all based on the testing of the lateral axis only. Thus, in total there are 28 effect sizes of the level *lateral/vertical*. Both levels of *Axis* are

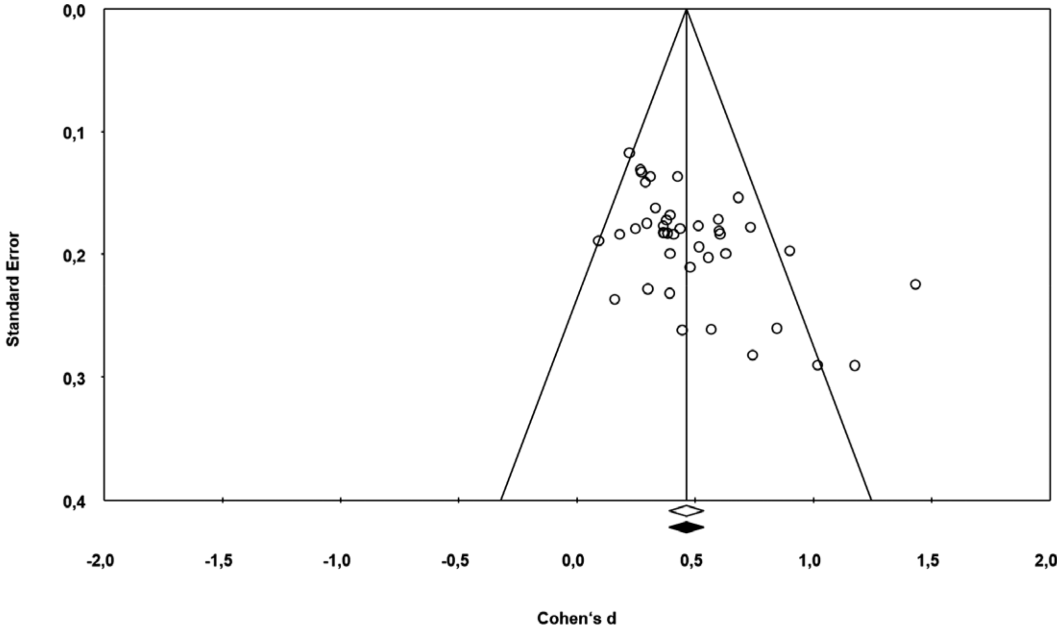

Fig. 3. Funnel plot for *time is task relevant*. The circles represent the observed effect sizes. No studies are trimmed and filled when using the random-effects model to look for missing studies.

significant. The weighted mean effect size of the level *sagittal* is numerically slightly higher ($d = 0.517$, 95% CI [0.362, 0.672], $p < .0005$, $Q(12) = 30.8$, $p = .002$, $I^2 = 61.1$) than the weighted mean effect size of the level *lateral/vertical* ($d = 0.437$, 95% CI [0.354, 0.520], $p < .0005$, $Q(27) = 38.3$, $p = .073$, $I^2 = 29.6$)[6]; however, this difference is not significant ($Q(1) = 0.8$, $p = .368$).

3.2. Time is task irrelevant

For the level *time is task irrelevant* (Fig. 4), the weighted mean effect size is not significant ($d = 0.087$, 95% CI [−0.005, 0.179], $p = .063$). We used the random-effects model for the analysis notwithstanding that the test for heterogeneity did not reach the level of significance ($Q(9) = 5.0$, $p = .832$, $I^2 = 0.0$).[7] The funnel plot (Fig. 5) reveals a slight publication bias, suggesting to trim and fill three studies. The estimation of the mean effect size after applying Duval and Tweedie's trim and fill differs only slightly from the observed mean effect size ($d = 0.051$, 95% CI [−0.035, 0.137]).

3.2.1. Time is task irrelevant—lateral/vertical versus sagittal axis

Six out of the 10 effect sizes that are included in the *time is task irrelevant* group are based on the examination of the sagittal axis. The remaining four effect sizes are gained on experiments testing the lateral axis. Interestingly, the weighted mean effect size of *sagittal time is task irrelevant* studies is significant ($d = 0.156$, 95% CI [0.019, 0.294],

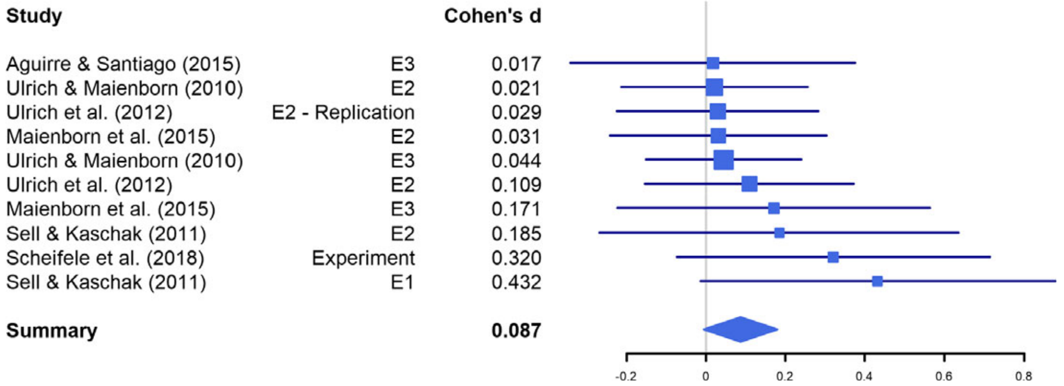

Fig. 4. Forest plot for *time is task irrelevant* (E = experiment).

p = .026, $Q(5) = 3.3$, p = .661, $I^2 = 0.0$), whereas the *lateral/vertical time is task irrelevant* studies do not yield a significant weighted mean effect size (d = 0.032, 95% CI [−0.092, 0.155], p = .617, $Q(3) = 0.0$, p = .999, $I^2 = 0.0$).[8] However, at the same time the Q-statistic does not suggest a significant difference between the weighted mean effect sizes of the two levels of *Axis* ($Q(1) = 1.7$, p = .186).

3.3. Temporal priming

For the *temporal priming* studies the use of the random-effects model again was not utterly supported by the Q-statistic ($Q(10) = 16.7$, p = .082); however, the proportion of observed variance that can be explained by heterogeneity is high ($I^2 = 40.0$). In addition to the above outlined assumption of a distribution of the true effect sizes around some common mean, the high I^2-value was taken as an indication to still use the random-effects model. The hereby resulting weighted mean effect size (Fig. 6) significantly deviates from zero (d = 0.465, 95% CI [0.298, 0.633], $p < .0005$).[9] According to the funnel plot (Fig. 7), there is a slight publication bias, recommending to trim and fill three studies which leads to an adjusted mean effect size of d = 0.359 (95% CI [0.179, 0.540]). All effect sizes that are incorporated in the *temporal priming* studies are based on an examination of the lateral axis. Thus, it is not possible to assess the impact of *Axis* on the weighted mean effect size of *temporal priming*.

3.4. *Comparing effect sizes of the different levels of* Task

The estimated mean effect size of *time is task irrelevant* studies significantly deviates from the estimated mean effect size of *time is task relevant* studies ($Q(1) = 39.2$, $p < .0005$) and of *temporal priming* studies ($Q(1) = 15.0$, $p < .0005$). However, the estimated mean effect sizes of *time is task relevant* studies and *temporal priming* studies do not differ significantly ($Q(1) = 0.0$, $p < .978$).

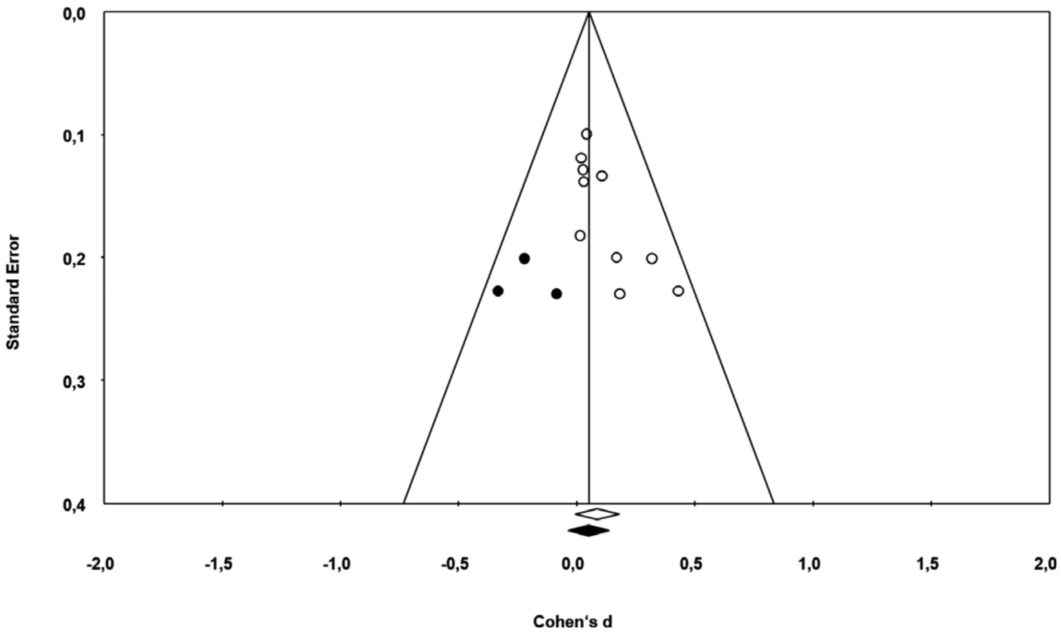

Fig. 5. Funnel plot for *time is task irrelevant*. Blank circles represent the observed effect sizes, and filled circles illustrate trimmed and filled studies.

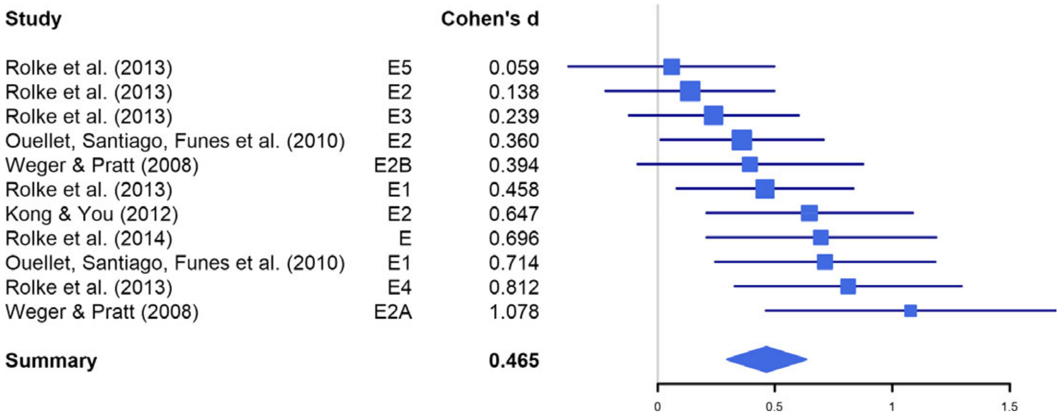

Fig. 6. Forest plot for *temporal priming* (E = experiment).

3.5. *Subdividing* time is task irrelevant

The results of the level *time is task irrelevant* suggest a systematic variation of the effect size depending on the type of temporal stimuli: Single sentences (e.g., The witness remembers the pistol-shot, Maienborn et al., 2015) and discourses with small time shifts (e.g., Jackie is taking a painting class; Tomorrow, she will learn about paintbrushes [. . .], Sell & Kaschak, 2011) show no effect, whereas discourses with large time shifts (e.g.,

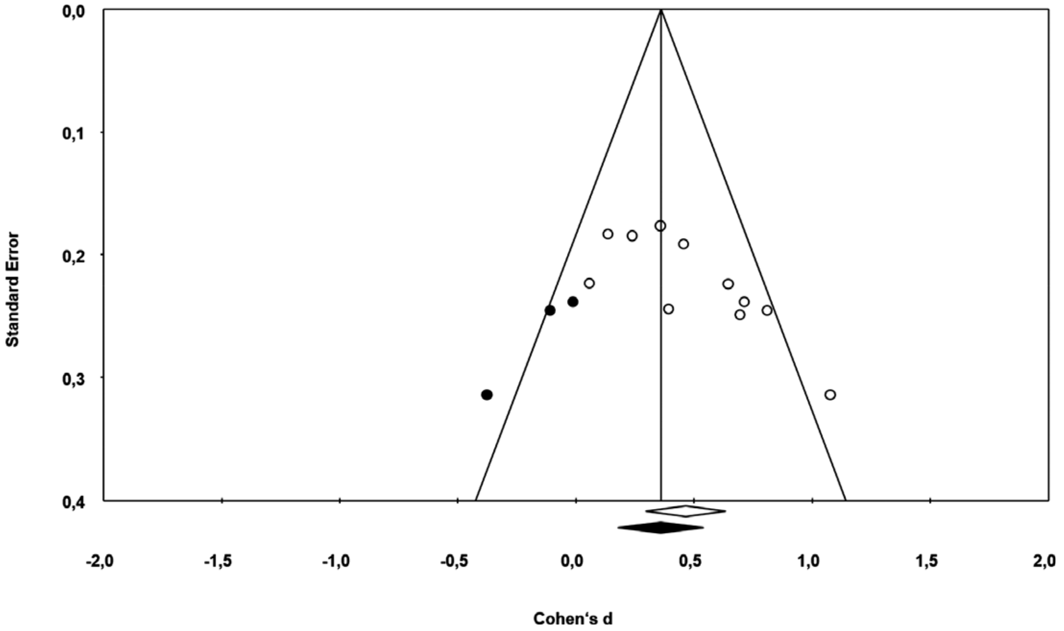

Fig. 7. Funnel plot for *temporal priming*. Blank circles represent the observed effect sizes, and filled circles illustrate trimmed and filled studies.

Jackie is taking a painting class; Next month, she will learn about paintbrushes [. . .], Sell & Kaschak, 2011) yield an ample effect. Hence, whether or not the mental timeline is activated automatically seems to depend on whether or not the temporal order information can be processed without using the mental timeline as part of the build-up of a situation model (Scheifele et al., 2018). Only when discourses with larger time shifts are processed, temporal complexity might “get the upper hand” (Scheifele et al., 2018, p. 10), thus eliciting the build-up of a mental situation model to manage the temporal order information: While single sentences usually only refer to one event, multi-sentences discourses with large time shifts refer to several clearly distinguishable and separate events that necessitate a sequencing by means of a localization on the mental timeline.

Accordingly, the subdivision of *time is task irrelevant* seems to be necessary for stimulus material, in which temporal complexity gets the upper hand (*high temporal complexity*) and stimuli that allow the processing of temporal information without the activation of the mental timeline (*low temporal complexity*). However, subdividing the effect sizes of *time is task irrelevant* into *low* and *high temporal complexity* requires gaining more than one effect size from one experiment, which violates the assumption of independence. Therefore, the following subdivision is only illustrative and the discovered trend has to be corroborated by further future experiments before strong conclusions can be drawn.

Since *high temporal complexity* as stimulus material is only part of three experiments (Scheifele et al., 2018; Sell & Kaschak, 2011, experiments 1 and 2), the effect sizes of these three experiments are subdivided into two distinct effect sizes each (one for *low*

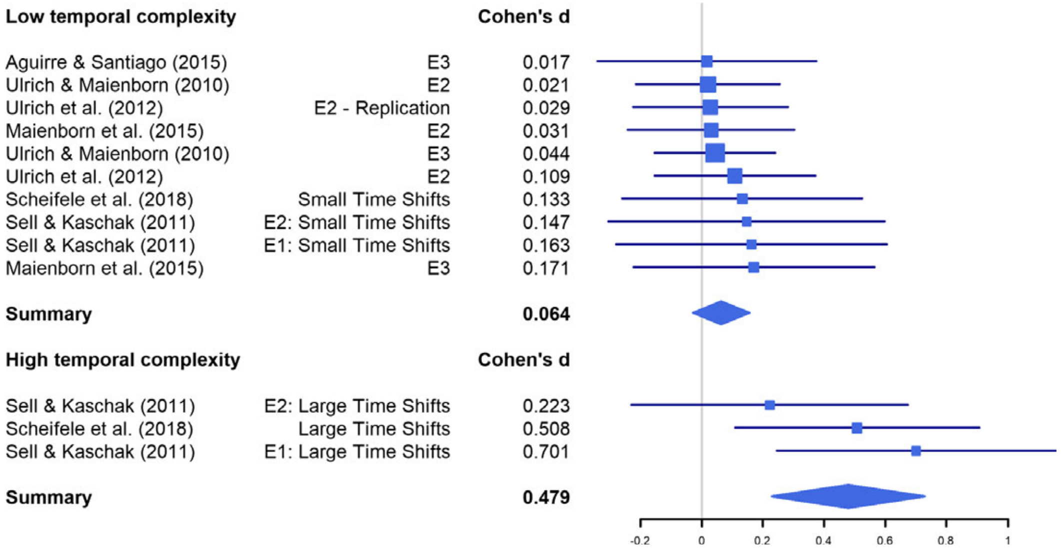

Fig. 8. Forest plot for *time is task irrelevant*, subdivided into *low* and *high temporal complexity* (E = experiment).

temporal complexity, i.e., small time shifts in this particular case; one for *high temporal complexity*, i.e., large time shifts). The resulting forest plot (Fig. 8) for *low temporal complexity* now shows more homogeneity of effect sizes than the overall forest plot (Fig. 4), while the three effect sizes of *high temporal complexity* are not any longer accounted for as being outliers but seem to depict an activation of the mental timeline that is comparable to the one of *time is task relevant*. The resulting mean effect size deviates from the mean effect size of *low temporal complexity* considerably.

4. Discussion

Of major interest for the purpose of this meta-analysis is whether the mean effect size of the space–time congruency deviates from zero, even when potential publication bias is accounted for, and what magnitude can be expected. For the default case, when time is a task-relevant dimension, which applies to the majority of effect sizes, our meta-analysis provides a clear answer for these questions. The estimated mean effect size of $d = 0.46$ implies that the space–time congruency effect is a sound effect when time is task relevant, which means that the mental timeline gets activated when temporal reasoning takes place, thereby facilitating responses congruent with the mental timeline's orientation. Beyond that, the mean effect size can be used to ensure adequate power for potential future experiments. More specifically, for a statistical power of 0.90, at least 41 participants are needed when manipulating the space–time congruency effect within subjects, and at least 81 participants when using a between-subjects design.

Since the aforementioned group of studies (i.e., *time is task relevant*) unifies effect sizes gained from different language populations for different axes of the mental timeline with diverse temporal cues used as stimulus, the random-effects model was used to estimate the publication bias. Using Duval and Tweedie's trim and fill suggests that the obtained overall effect size is not inflated by publication bias. The picture changes slightly, if the fixed-effect model is used for the estimation of the publication bias (suggested adjusted effect size of $d = 0.36$ after the trimming of 11 effect sizes). Indeed, the funnel plot displays some asymmetry that can be detected visually. Hence, assuming that there is no publication bias at all might be a little too optimistic, whereas it is also important to remark that according to the random-effects model a considerable proportion (if not all) of the found asymmetry is due to the remaining variation in experimental set-ups (see also Terrin et al., 2003 for the loose link between asymmetry and publication bias).

One could of course try to further differentiate between the designs of the studies. However, it was not the purpose of this analysis to distinguish between the effect sizes of different language populations or stimulus material, and, in addition, subdividing the level *time is task relevant* would lead to a fragmentation that is no longer statistically meaningful. The only further subdivision that has been carried out is an analysis of the influence of an axis's origin. Even though the sagittal axis with its origin in the fundamental human experience of moving forward in the world (Lakoff & Johnson, 1980; Núñez & Sweetser, 2006) yields a numerically slightly larger weighted mean effect size than the lateral and vertical axes that can presumably be traced back to the cross-culturally more flexible writing and reading direction (Bergen & Chan Lau, 2012; Fuhrman & Boroditsky, 2007; Ouellet, Santiago, Israeli, & Gabay 2010), the difference between the sagittal and lateral/vertical axes is not significant.

For the second type of task (i.e., *time is task irrelevant*), which, overall, does not significantly deviate from zero, a similar pattern occurs when looking at the vertical and lateral as opposed to the sagittal axis: Even though the weighted mean effect size of the vertical and lateral axes is slightly smaller than the one of the sagittal axis, this difference is not statistically significant. However, while the former does not significantly deviate from zero, the latter does. This suggests a tendency that will have to be investigated by future studies. Specifically, the sagittal axis might be strongly rooted in human experience so that it gets activated as a mental timeline even when time is task irrelevant. Since, however, the sample size is small, no clear conclusions can be drawn so far.

There is yet an alternative explanation for the dispersal of the level *time is task irrelevant*: The discrepant results of its additional subdivision reflect the current discussion about the automatic activation of the mental timeline (Scheifele et al., 2018). Most studies fail to find a space–time congruency effect when time is not a relevant response dimension (Maienborn et al., 2015; Ulrich & Maienborn, 2010; Ulrich et al., 2012). However, Sell and Kaschak (2011) reported an automatically activated mental timeline for discourses containing large time shifts. Since Scheifele et al. (2018) basically replicated Sell and Kaschak's findings, they do not seem to be a false-positive result but indicate that automatic activation of the mental timeline is dependent on the level of temporal complexity (Scheifele et al., 2018). When time shifts are small or only single sentences have to be processed, there is no need for the use of a mental timeline in order to comprehend the temporal information of the linguistic material. However, as the ordering of temporal information gets more demanding, a mental situation model including a mental timeline could be built up to facilitate this process (Scheifele et al., 2018; Zwaan, Madden, & Stanfield, 2001).

This conclusion is also supported by the mean effect sizes gained through the additional subdivision of *time is task irrelevant*: There seems to be no automatic activation of the mental timeline for a low level of temporal complexity (d = 0.06), whereas the effect size of the congruency effect of stimuli with higher temporal complexity (i.e., discourses with large time shifts, d = 0.48) seems to be comparable to the effect size that arises when time is task relevant (d = 0.46). It is important to note, however, that there are not yet enough measured effect sizes to give a reliable evaluation. Future research will have to further identify the exact conditions under which automatic activation takes place and whether the incorporated effect size really is comparable to the one that emerges when time is task relevant.

The weighted mean effect size of the subgroup *temporal priming* is high. This is a surprising result since temporal priming in advance of the execution of a spatial task is meant to make time a task–irrelevant dimension and thus is expected to be rather comparable to the effect size of *time is task irrelevant* (with low temporal complexity), which is close to zero. However, carefully examining the designs of the temporal priming studies reveals that the temporal reference of the primes is in some cases brought into focus by the instruction. For example, Ouellet, Santiago, Funes, et al. (2010) asked their participants to remember the temporal reference of the prime. After executing the spatial task (e.g., indicating the appearance of a circle in one of two laterally aligned boxes with a left or right keypress), participants had to answer whether the temporal prime had referred to the past or the future. This final probe question that was carried out in all three experiments of the study makes temporal reference much more salient as compared to a sensicality judgment that is used by most studies that examine automatic activation. Detecting a space–time congruency effect under these circumstances certainly is more likely compared to sensicality judgments where temporal information is irrelevant for performing the task. At least 3 out of the 11 effect sizes of *temporal priming* arise in settings in which the temporal reference of the primes is of essential relevance for the execution of the task (see also Rolke et al., 2013, experiment 4). This could explain why the mean effect size of *temporal priming* is as large compared to the other two levels of *Task*.

In some of the temporal priming experiments, the priming is more subliminal and can nevertheless positively be said to trigger an automatic activation of the mental timeline. In experiment 1 of Rolke et al. (2013), for example, a considerable effect size emerged (d = 0.46), even though temporal complexity was very low since single words were used as temporal primes. Here, the need for distinguishing temporal priming studies from studies in which the temporal dimension is task irrelevant becomes apparent: While the priming studies reveal a pronounced space–time congruency effect (adjusted mean effect size: d = 0.36), the *time is task irrelevant* studies do not (adjusted mean effect size: d = 0.05), although both are based on automatic activation. This may be due to the temporal interval given between prime presentation and target task, thus allowing for a build-up of the mental timeline. In the case of sensicality judgments, the responses have to be given much sooner after the processing of the temporal information. This, however, remains in the sphere of speculation and requires further investigation.

Overall, the underlying publication bias is surprisingly low, keeping in mind that the incorporated studies still vary with respect to their specific designs. Thus, our results can only give a coarse estimation. However, it seems relatively safe to expect an effect size somewhere between 0.39 and 0.54 when time is task relevant, an effect size between 0.18 and 0.54 when conducting a temporal priming study, and no effect when time is task irrelevant and temporal complexity of the stimuli is low. The different origins of the axes that are the mental timeline's basis on the contrary do not seem to modulate these effect sizes, although further studies are required to substantiate this conclusion—especially with regard to the *time is task irrelevant* studies. How large exactly the size of the congruency effect is when time is task irrelevant and temporal complexity is high remains an open issue until more studies have been undertaken. For now, at least, the corresponding effect size seems to be comparable to the effect size that appears when time is task relevant. The relatively large confidence interval for the effect size of temporal priming studies indicates that again more studies are needed to refine the picture.

The psychological reality of the mental timeline seems apparent after the outcome of this meta-analysis: Not only is space used to talk about time (Haspelmath, 1997), but space is also cognitively exerted to order events and to reason about the temporal reference of an entity. This corroborates the cognitive existence of the space–time metaphor. Even though the space–time congruency effect of experiments, in which time is a task-relevant dimension, could still be explained by means of the memory account (Eikmeier, Hoppe, et al., 2015; Ulrich & Maienborn, 2010), there still seems to be evidence for an automatic activation of the mental timeline under certain circumstances. Specifically, the mean effect size of temporal priming studies and the trend of studies using high temporal complexity for experiments in which time is a task-irrelevant dimension provide evidence for an automatic activation of the mental timeline. Future studies are required to examine the specific circumstances that facilitate automatic activation in more depth.

Acknowledgments

This research was supported by the Deutsche Forschungsgemeinschaft (SFB 833, B7: Language Processing and Temporal Cognition).

Notes

1. Note that there is one study that explicitly examines whether axes that are not culturally salient can become activated by manipulating the reading direction of the experimental stimuli (i.e., by using mirror reversed orthography, vertical downward

orthography, and vertical upward orthography). Since the premise of the emergence of the space–time congruency effect in this study by Casasanto and Bottini (2014) is very specific, the results cannot be integrated into the present meta-analysis. Only their standard orthography control group has been incorporated.

2. Please note that congruency thus is randomized within blocks: For *time is task irrelevant* studies there are congruent and incongruent trials in both blocks, whereas for *time is task relevant* studies congruency is manipulated between blocks: One block contains only congruent responses, the other block only incongruent responses. This methodological difference is a systematic confound when comparing the two levels of *Task*. However, it is not assumed to have a significant influence on the effect size.
3. For detailed information on how we treat the effect sizes of particular experiments concerning the independence of subgroups, see Supporting Information.
4. A similar pattern applies to three analyses of the factor *Axis*.
5. See Supporting Information for details. We decided to keep the sagittal effect size and eliminate the lateral effect size of these four effect sizes, since the majority of effect sizes of *time is task relevant* is based on the lateral axis. Thus, for the purpose of statistical evaluation increasing the amount of sagittal effect sizes is preferable.
6. Please note that even though the test for heterogeneity is not significant, I^2 is considerable. Using a fixed-effect model would lead to only slightly different results: $d = 0.421$, 95% CI [0.353, 0.488], $p < .0005$.
7. Please note that using the fixed-effect model would yield exactly the same results. The reason for the choice of the random-effects model besides the a-priori assumption outlined above is that a nonsignificant *p*-value cannot readily be taken as evidence for homogeneity (Borenstein et al., 2009, p. 113).
8. Again, using the fixed-effect model would yield exactly the same results for both the *lateral/vertical* and the *sagittal* axes.
9. Using the fixed-effect model would yield only slightly different results: $d = 0.442$, 95% CI [0.314, 0.570], $p < .0005$.

References

References marked with an asterisk (*) indicate studies included in the meta-analysis.

*Aguirre, R., & Santiago, J. (2015). Do potential past and future events activate the left-right mental timeline? *Psicológica*, *38*, 231–255.

Bergen, B. K., & Chan Lau, T. T. (2012). Writing direction affects how people map space onto time. *Frontiers in Psychology*, *3*, 1–5. https://doi.org/10.3389/fpsyg.2012.00109.

Borenstein, M. (2005). Software for publication bias. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), *Publication bias in meta-analysis: Prevention, assessment and adjustments* (pp. 193–220). Chichester, UK: John Wiley & Sons.

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

*Boroditsky, L., Fuhrman, O., & McCormick, K. (2011). Do English and Mandarin speakers think about time differently? *Cognition*, *118*, 123–129. https://doi.org/10.1016/j.cognition.2010.09.010.

*Bottini, R., Crepaldi, D., Casasanto, D., Crollen, V., & Collignon, O. (2015). Space and time in the sighted and blind. *Cognition*, *141*, 67–72. https://doi.org/10.1016/j.cognition.2015.04.004.

Burr, D., Tozzi, A., & Morrone, M. C. (2007). Neural mechanisms for timing visual events are spatially selective in real-world coordinates. *Nature Neuroscience*, *10*, 423–425. https://doi.org/10.1038/nn1874.

*Casasanto, D., & Bottini, R. (2014). Mirror reading can reverse the flow of time. *Journal of Experimental Psychology: General*, *143*, 473–479. https://doi.org/10.1037/a0033297.

Casasanto, D., & Jasmin, K. (2012). The hands of time: Temporal gesture in English speakers. *Cognitive Linguistics*, *23*(4), 643–674. https://doi.org/10.1515/cog-2012-0020.

Clark, H. H. (1973). Space, time, semantics, and the child. In T. Moore (Ed.), *Cognitive development and the acquisition of language* (pp. 27–63). New York: Academic Press. https://doi.org/10.1016/b978-0-12-505850-6.50008-6

*De la Vega, I., Eikmeier, V., Ulrich, R., & Kaup, B. (2016). The mental timeline in a crossed-hands paradigm: A matter of instruction. *Experimental Psychology*, *63*, 326–332. https://doi.org/10.1027/1618-3169/a000342.

Di Bono, M. G., Casarotti, M., Priftis, K., Gava, L., Umiltà, C., & Zorzi, M. (2012). Priming the mental time line. *Journal of Experimental Psychology: Human Perception and Performance*, *38*, 838–842. https://doi.org/10.1037/a0028346.

*Ding, X., Feng, N., Cheng, X., Liu, H., & Fan, Z. (2015). Are past and future symmetric in mental time line? *Frontiers in Psychology*, *6*, 1–8. https://doi.org/10.3389/fpsyg.2015.00208.

Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, *56*, 455–463. https://doi.org/10.1111/j.0006-341x.2000.00455.x.

Eikmeier, V., Alex-Ruf, S., Maienborn, C., Schröter, H., & Ulrich, R. (2016). The mental timeline during the processing of linguistic information. In B. Lewandowska-Tomaszczyk (Ed.), *Conceptualizations of time* (pp. 103–122). Amsterdam, The Netherlands: Benjamins. https://doi.org/10.1075/hcp.52.06eik

*Eikmeier, V., Alex-Ruf, S., Maienborn, C., & Ulrich, R. (2015). How strongly linked are mental time and space along the left-right axis? *Journal of Experimental Psychology: Learning Memory and Cognition*, *41* (6), 1878–1883. https://doi.org/10.1037/xlm0000129.

*Eikmeier, V., Hoppe, D., & Ulrich, R. (2015). Response mode does not modulate the space-time congruency effect: Evidence for a space-time mapping at a conceptual level. *Acta Psychologica*, *156*, 162–167. https://doi.org/10.1016/j.actpsy.2014.10.008.

*Eikmeier, V., Schröter, H., Maienborn, C., Alex-Ruf, S., & Ulrich, R. (2013). Dimensional overlap between time and space. *Psychonomic Bulletin and Review*, *20*, 1120–1125. https://doi.org/10.3758/s13423-013-0431-2.

*Fuhrman, O., & Boroditsky, L. (2007). Mental Time-Lines Follow Writing Direction: Comparing English and Hebrew Speakers. In D. S. McNamara & J. G. Trafton (Eds.), *Proceedings of the 29th Annual Cognitive Science Society* (pp. 1007–1011). Austin, TX: Cognitive Science Society.

*Fuhrman, O., McCormick, K., Chen, E., Jiang, H., Shu, D., Mao, S., & Boroditsky, L. (2011). How linguistic and cultural forces shape conceptions of time: English and mandarin time in 3D. *Cognitive Science*, *35*, 1305–1328. https://doi.org/10.1111/j.1551-6709.2011.01193.x.

Gentner, D., Imai, M., & Boroditsky, L. (2002). As time goes by: Evidence for two systems in processing space → time metaphors. *Language and Cognitive Processes*, *17*, 537–565. https://doi.org/10.1080/01690960143000317.

*Hartmann, M., & Mast, F. W. (2012). Moving along the mental time line influences the processing of future related words. *Consciousness and Cognition*, *21*, 1558–1562. https://doi.org/10.1016/j.concog.2012.06.015.

Haspelmath, M. (1997). *From space to time: Temporal adverbials in the world's languages*. München, Germany: Lincom Europa.

Hendricks, R. K., & Boroditsky, L. (2015). Constructing mental time without visual experience. *Trends in Cognitive Sciences*, *19*, 429–430. https://doi.org/10.1016/j.tics.2015.06.011.

*Kong, F., & You, X. (2012). Space-time compatibility effects in the auditory modality. *Experimental Psychology*, *59*, 82–87. https://doi.org/10.1027/1618-3169/a000129.

Lakoff, G., & Johnson, M. (1980). Conceptual metaphor in everyday language. *The Journal of Philosophy*, *77*, 453–486.

*Loeffler, J., Raab, M., & Cañal-Bruland, R. (2017). Walking back to the future: The impact of walking backward and forward on spatial and temporal concepts. *Experimental Psychology*, *64*, 346–358. https://doi.org/10.1027/1618-3169/a000377.

*Maienborn, C., Alex-Ruf, S., Eikmeier, V., & Ulrich, R. (2015). Do we map remembrances to the left/back and expectations to the right/front of a mental timeline? Space-time congruency effects with retrospective and prospective verbs. *Acta Psychologica*, *156*, 168–178. https://doi.org/10.1016/j.actpsy.2014.11.006.

*Miles, L. K., Tan, L., Noble, G. D., Lumsden, J., & Macrae, C. N. (2011). Can a mind have two time lines? Exploring space-time mapping in Mandarin and English speakers. *Psychonomic Bulletin and Review*, *18*, 598–604. https://doi.org/10.3758/s13423-011-0068-y.

Núñez, R., & Cooperrider, K. (2013). The tangle of space and time in human cognition. *Trends in Cognitive Sciences*, *17*, 220–229.

Núñez, R., & Sweetser, E. (2006). With the future behind them: Convergent evidence from Aymara language and gesture in the crosslinguistic comparison of spatial construals of time. *Cognitive Science*, *30*, 401–450.

*Ouellet, M., Román, A., & Santiago, J. (2012). A multisensory interaction effect in the conceptual realm of time. *Experimental Psychology*, *59*, 236–242. https://doi.org/10.1027/1618-3169/a000148.

*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. https://doi.org/10.1037/a0017176.

*Ouellet, M., Santiago, J., Israeli, Z., & Gabay, S. (2010). Is the future the right time? *Experimental Psychology*, *57*, 308–314. https://doi.org/10.1027/1618-3169/a000036.

Proctor, R. W., Reeve, T. G., & Van Zandt, T. (1992). Salient-features coding in response selection. In G. E. Stelmach & J. Requin (Eds.), *Tutorials in motor behavior II* (pp. 727–741). Amsterdam, The Netherlands: Elsevier.

*Rolke, B., Fernández, S. R., Schmid, M., Walker, M., Lachmair, M., López, J. J. R., Hervás, G., & Vázquez, C. (2013). Priming the mental time-line: Effects of modality and processing mode. *Cognitive Processing*, *14*, 231–244. https://doi.org/10.1007/s10339-013-0537-5.

*Rolke, B., Ruiz Fernández, S., Seibold, V., & Rahona, J. J. (2014). Crossed hands stay on the time-line. *Cognitive Processing*, *15*, 134–135. https://doi.org/10.1007/s10339-014-0632-2.

Rosenthal, R. (1979). The “file drawer problem” and tolerance for null results. *Psychological Bulletin*, *86*, 638–641.

*Santiago, J., Lupiáñez, J., Pérez, E., & Funes, M. J. (2007). Time (also) flies from left to right. *Psychonomic Bulletin & Review*, *14*(3), 512–516. https://doi.org/10.3758/BF03194099.

Santiago, J., Román, A., & Ouellet, M. (2011). Flexible foundations of abstract thought: A review and a theory. In A. Maas & T. Schubert (Eds.), *Spatial dimensions of social thought* (pp. 39–108). Berlin, Germany: Mouton de Gruyter.

*Scheifele, E., Eikmeier, V., Alex-Ruf, S., Maienborn, C., & Ulrich, R. (2018). A replication of “processing time shifts affects the execution of motor responses (Sell & Kaschak, 2011; experiment 1).” *The Quantiative Methods for Psychology*, *14*, 8–11. https://doi.org/10.20982/tqmp.14.1.r008

*Sell, A. J., & Kaschak, M. P. (2011). Processing time shifts affects the execution of motor responses. *Brain and Language*, *117*, 39–44. https://doi.org/10.1016/j.bandl.2010.07.003.

Terrin, N., Schmid, C. H., Lau, J., & Olkin, I. (2003). Adjusting for publication bias in the presence of heterogeneity. *Statistics in Medicine*, *22*, 2113–2126. https://doi.org/10.1002/sim.1461.

Tillman, K. A., Tulagan, N., & Barner, D. (2015). Building the mental timeline: Spatial representations of time in preschoolers. In D. C. Noelle, R. Dale, A. Warlaumont, J. Yoshimi, T. Matlock, C. Jennings & P. P. Maglio (Eds.), *37th Annual Meeting of the Cognitive Science Society (Cognitive Science 2015)* (pp. 2398–2403). Austin, TX: Cognitive Science Society.

*Torralbo, A., Santiago, J., & Lupiáñez, J. (2006). Flexible conceptual projection of time onto spatial frames of reference. *Cognitive Science*, *30*, 745–757.

*Ulrich, R., Eikmeier, V., de la Vega, I., Fernández, S. R., Alex-Ruf, S., & Maienborn, C. (2012). With the past behind and the future ahead: Back-to-front representation of past and future sentences. *Memory and Cognition*, *40*, 483–495. https://doi.org/10.3758/s13421-011-0162-4.

*Ulrich, R., & Maienborn, C. (2010). Left-right coding of past and future in language: The mental timeline during sentence processing. *Cognition*, *117*, 126–138. https://doi.org/10.1016/j.cognition.2010.08.001.

Ulrich, R., Miller, J., & Erdfelder, E. (2018). Effect size estimation from t-statistics in the presence of publication bias: A brief review of existing approaches with some extensions. *Zeitschrift für Psychologie*, *226*, 56–80.

*Walker, E. J., Bergen, B. K., & Núñez, R. (2014). Disentangling spatial metaphors for time using non-spatial responses and auditory stimuli. *Metaphor and Symbol*, *29*, 316–327. https://doi.org/10.1080/10926488.2014.948801.

*Walker, E. J., Bergen, B. K., & Núñez, R. (2017). The spatial alignment of time: Differences in alignment of deictic and sequence time along the sagittal and lateral axes. *Acta Psychologica*, *175*, 13–20. https://doi.org/10.1016/j.actpsy.2017.02.001.

*Weger, U. W., & Pratt, J. (2008). Time flies like an arrow: Space-time compatibility effects suggest the use of a mental timeline. *Psychonomic Bulletin and Review*, *15*, 426–430. https://doi.org/10.3758/PBR.15.2.426.

Zwaan, R., Madden, C., & Stanfield, R. (2001). Time in narrative comprehension: A cognitive perspective. In D. H. Schram & G. J. Steen (Eds.), *The psychology and sociology of literature: In honor of Elrud Ibsch* (pp. 71–86). Amsterdam, The Netherlands: Benjamins. https://doi.org/10.1075/upal.35.06zwa

Supporting Information

Additional Supporting Information may be found online in the section at the end of the article:

Supplemental Materials.