Journal of Memory and Language 143 (2025) 104644

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

Can the lateral mental timeline be automatically activated in language comprehension?

Julio Santiago [a,*], Alessia Beracci [b], Andrea Flumini [a], Eva Sanjuan [a], Marc Ouellet [a], Pablo Solana [a]

[a] *Mind, Brain, and Behaviour Research Centre, University of Granada, Spain*
[b] *Department of Psychology "Renzo Canestrari", University of Bologna, Italy*

ARTICLE INFO

Keywords:
Time
Mental timeline
Automaticity
Conceptual metaphor
Embodied and grounded cognition

ABSTRACT

The mental representation of time recruits spatial representations, but is space an essential, inescapable feature of mental time? Supporting a positive answer to this question, recent research has reported that lateral (left-right) space is automatically activated in lexical decision tasks in which the temporal reference of the words is irrelevant for the goals of the task (implicit tasks). Here, using always the same set of Spanish verbs and pseudoverbs marked for past or future tense, we assess the space–time congruency effect in reaction time and mouse trajectories, both in an explicit time judgement task and an implicit lexical decision task. Moreover, we report the first confirmatory (preregistered) study in this field of research using long lateral movements in lexical decision. The congruency effect was always significant in time judgement, but non-significant in lexical decision. Moreover, in reaction time this effect was significantly smaller than a Smallest Effect Size Of Interest (SESOI) of 10 ms, and even smaller than a recently reported 9 ms effect. Therefore, it was considered negligible. We conclude that there is no convincing evidence for an automatic activation of the lateral mental timeline in lexical decision.

Introduction

According to Conceptual Metaphor Theory, abstract concepts are represented through conceptual metaphors: borrowing from more concrete concepts (Lakoff & Johnson, 1980, 1999). The space–time conceptual metaphor lets us think of time as space, leading us to conceive of time as the motion from one spatial location in the past to another location in the future. This conceptual metaphor is supported by how people speak about time (Clark, 1973; Haspelmath, 1997; Radden, 2004), and also by a variety of non-linguistic processing tasks (Boroditsky, 2000; see Núñez & Cooperrider, 2013, for a review). Subsequent psychological research has shown that such mental timeline can be mapped on all three spatial axes (as forward, lateral, or vertical movement; Beracci & Fabbri, 2022; Boroditsky et al., 2011; Dalmaso et al., 2023; Ding et al., 2020; Torralbo et al., 2006; Ulrich et al., 2012). An important piece of evidence in this line of research is the space–time congruency effect (for reviews, see Bonato et al., 2012; von Sobbe et al., 2019). Its lateral form consists in faster responses to words or sentences with past reference, past/earlier events, or short durations, when they are presented on the left or responded to with the left hand, and faster responses to words/sentences with future reference, future/later events, or long durations, when they are presented on the right or responded to with the right hand (Fuhrman & Boroditsky, 2010; Ishihara et al., 2008; Santiago et al., 2007, 2010). Such interaction is interpreted as the result of the congruency between the spatial aspects of the task and the spatial representation of time in the mind.

Yet, is space an essential, inescapable feature of the mental representation of time? In many domains of cognition, a central strategy to respond to this kind of question has been to test whether diagnostic effects (such as the space–time congruency effect) arise automatically, i. e., when the key dimension is irrelevant for the task (what we will call here implicit tasks). For example, in the domain of numerical cognition, Dehaene et al. (1993) reported a space-number congruency effect (faster responses to small numbers with the left hand and larger numbers with the right hand, widely known as the SNARC effect), which arose both in an explicit magnitude task (judge whether the presented number is smaller or larger than 5) and an implicit task (judge whether the number is odd or even). However, although unrelated to magnitude, parity is still a numerical feature, so when Fischer et al. (2003) reported that just looking at a number was able to move visual attention to the side

* Corresponding author.
E-mail address: santiago@ugr.es (J. Santiago).

https://doi.org/10.1016/j.jml.2025.104644
Received 11 June 2024; Received in revised form 6 April 2025; Accepted 8 April 2025
Available online 19 April 2025

congruent with its magnitude, the finding had a wide impact on the field and started a heated debate (leading even to a concerted effort of 17 labs to replicate the effect, see Colling et al., 2020). The question of automaticity has fueled debates across many different fields of research, from lexical access in reading (Besner et al., 1997) to conceptual activation (Lebois et al., 2015), attention (Ruz & Lupiáñez, 2002), imitation (Heyes, 2011), perception of social interactions (McMahon & Isik, 2023), and a variety of social-cognitive processes (Bargh et al., 2012).

The automaticity of the space–time congruency effect has recently been at the center of a strong debate (Abbondanza et al., 2024; Flumini & Santiago, 2013; Grasso, Ziegler, Mirault, et al., 2022; Ulrich & Maienborn, 2010). Since the discovery of the effect, the majority of studies used explicit tasks, i.e., tasks in which the temporal dimension is a relevant part of the task. For example, Santiago et al. (2007) presented words that referred to either past or future on the left or right sides of the screen and asked participants to judge their temporal reference by pressing a left or a right key. Participants were faster using the congruent past-left future-right response mapping than the incongruent mapping.[1] Such explicit space–time congruency effect is now well established, independently of the spatial axis (sagittal, lateral, or vertical), the type of materials (physical durations, words, sentences), or their modality (auditory, visual; Beracci et al., 2022; Bonato et al., 2012; Malyshevskaya et al., 2024; Mariconda et al., 2022; Ouellet et al., 2012; Scozia et al., 2023; Ulrich et al., 2012; Ulrich & Maienborn, 2010; von Sobbe et al., 2019). However, the picture is more complex regarding the implicit space–time congruency effect, that is, the effect that arises in tasks in which time is not a relevant dimension. In these tasks, the participant is asked to judge a dimension of the stimulus other than time (e.g., to perform a lexical decision), thus avoiding time becoming part of the definition of the task for the participant. We will focus here on the implicit space–time congruency effect in fast decisions to linguistic materials with a temporal reference, decisions that are not about time (often sensicality or lexicality decisions). In our target tasks, available evidence mostly supports the absence (or, at least, the undetectability) of the implicit congruency effect, for both sentences and single words. For example, using sentences, Ulrich and Maienborn (2010) for the lateral axis, Ulrich et al. (2012) for the front-back axis, and Maienborn et al. (2015) for both axes observed clear congruency effects in a temporal judgement task, but no effect in a sensicality task. Using verbs and pseudoverbs inflected in either future or past tense and left–right keypresses, Flumini and Santiago (2013) and Aguirre and Santiago (2017) found the congruency effect only in time judgement, but not in decisions about lexicality or potentiality, respectively (see von Sobbe et al., 2019, for a wider review; see Table 1). Yet, there have been recent reports of implicit space–time congruency effects, both with sentences and single words. We will here examine the reports using single word tasks, as they directly motivate the present experiment series, and leave a discussion of sentence processing studies and other methodologies to the General Discussion.

There have been four recent reports of space–time congruency effects in single-word implicit tasks, three of them in lexical decision and one using a Stroop-like task. As Flumini and Santiago (2013), Grasso, Ziegler, Mirault, et al. (2022) presented words and pseudowords inflected for tense (e.g., "*je marchais*": "I walked"; "*je rêverai*": "I will dream"). Participants judged lexical status by means of sideways movements on the trackpad or with the mouse, or pressing response keys. Word latencies (but not pseudowords) showed significant space–time congruency effects only in the lateral movement conditions of three experiments, ranging from 25 to 13 ms,[2] but there was a non-significant 5 ms effect in the keypress condition (see Table 1). This led the authors to conclude that long lateral movements are key for the automatic activation of the lateral mental timeline in implicit tasks. Long front-back movements had proven ineffective in whole-sentence sensicality tasks (Maienborn et al., 2015; Ulrich et al., 2012), but it might be different for lateral movements, as the lateral mental timeline is related to reading and writing experiences (Casasanto & Bottini, 2014; Ouellet, Santiago, Israeli, et al., 2010). In reading and writing, both the hands and the eyes perform long lateral movements as time passes and the sequence of events in the text unfolds, generating an experiential correlation between lateral movement and time. The authors suggested that "the very representation of a time word (like future- and past-tense verbs) might include the motor execution component of left-to-right movements" (Grasso, Ziegler, Mirault, et al., 2022, p. 11). Consistently, in a subsequent study, Grasso, Ziegler, Coull, et al. (2022) observed a 13 ms congruency effect for lateral saccade latencies in lexical decision. We will call this possibility the *lateral movement hypothesis* hereafter.

The third report (Abbondanza et al., 2024) also used tensed verbs and pseudoverbs in both time judgement (exp. 1) and lexical decision (exp. 2 and 3) with left or right keypress responses. As expected, there was an explicit space–time congruency effect, but, in stark contrast to very similar prior studies (Flumini & Santiago, 2013; Grasso, Ziegler, Mirault, et al., 2022, keypress condition), they also found a significant 9 ms effect in the implicit task (exp. 2; see Table 1). The authors speculated that this was due to the smaller variety and greater consistency of their inflectional endings. In contrast, they argued, Flumini and Santiago (2013) used several different inflections varying in person and number, while Grasso and colleagues (2022; 2022), who used only one inflection, used an inconsistent one: the first-person past ending "-ais", which is also often used as a nominal suffix to indicate place of origin (as in "Français"). This speculation, however, has an important problem: Grasso and colleagues (2022; 2022) used the same endings for both the words and pseudowords, and against the expectations from cue inconsistency, found the implicit effect in the words. Models of morphological processing agree that morphologically complex words are likely to be parsed into their constituent morphemes (see discussion in Abbondanza et al., 2024). Thus, cue inconsistency should have prevented the congruency effect also in the word condition. Finding the effect in words makes cue consistency an unlikely account of the differences between the two studies.

Even more recently, Carmona et al. (2024) tested the automaticity of the space–time congruency effect using only the word materials from Santiago et al. (2007). In their exp. 1 they took a very original approach: they had participants judging the temporal reference of the words (an explicit task), but half the words were presented below the level of consciousness (33 ms followed by a mask, what led to a chance level of recognition), while the other half could be clearly seen. This experiment does not meet the definition of our target tasks, as the decision is about time, so we classify it as explicit, but we will discuss the subliminal condition in detail in the General Discussion section together with other methodologies. Directly relevant is their exp. 2, in which they presented the words in either red or blue ink for a very short time (33 ms, unmasked) and participants discriminated their color by pressing a left or right key. Here they observed an astonishing 84 ms (Table 1) space–time congruency effect (partial eta-squared was 0.24, well above the 0.18 value considered conventionally to be a large effect; Cohen, 1988). The authors did not attempt to integrate their findings with the prior studies.

[1] A note on terminology: Hereafter, we use the term "effect" to refer to the difference between two conditions without regard to its statistical significance. That is, an observed effect may not be significant and may have any size (including zero). The space–time congruency effect is thus the difference between the congruent and incongruent conditions. When this effect is assessed in tasks where the temporal dimension is irrelevant (i.e., implicit tasks), we call it the "implicit effect", in contrast to the "explicit effect" that is assessed in time-relevant (explicit) tasks.

[2] Thinking of effect sizes in absolute values has the advantage of being more intuitive whenever the measured scale has intrinsic meaning, such as reaction time (see, e.g., Dienes, 2019).

Table 1
Relevant studies for the present experimental series, all using "fast decisions to linguistic materials with a temporal reference, decisions that are not about time", specifically to single inflected words and pseudowords. The relevant spatial axis is always the lateral axis (both keypresses and long movements of the hand and eyes are toward the left or right). "E" = Experiment. "EF" = congruency effect size. "EF-W" = congruency effect only in words. "EF-PW" = congruency effect only in pseudowords. "Interact. with lex. status" = significativity of the interaction of the congruency effect with lexical status. "time judg." = time judgement. "lex. dec" = lexical decision. "potentiality dec." = factual vs. potential decision. "primed" = targets preceded by temporal primes. "irrel. fillers" = irrelevant fillers. "50 % subl." = 50 % of the words presented subliminally. "trackpad" = long movements using a trackpad. "mouse" = long movements using a mouse. "RT" = reaction time. "MT" = movement time. "within subj." = within subjects. "bet. groups" = between groups. "?" = not reported. Effect are rounded to the nearest millisecond. Superscripts: "*" = significant at p < .05; "ns" = non-significant; "?" = not reported; "c" = the effect size is estimated from a chart. Blank = unapplicable.

Study	E	Task	Type	Response	Measure	Sample size	Congruency design	EF	EF-W	EF-PW	Interact. with lex. status
Flumini & Santiago (2013)	E1	time judg.	explicit	keypress	RT	24	within subj.	64*	$78^{?}$	$53^{?}$	ns
	E2	lex. dec.	implicit	keypress	RT	24	within subj.	-2^{ns}	?	?	ns
Aguirre & Santiago (2017)	E1	time judg.	explicit	keypress	RT	28	within subj.	?	146*		
	E2	time judg.	explicit	keypress	RT	34	within subj.	?	59*		
	E3	potentiality dec.	implicit	keypress	RT	30	within subj.	?	2^{ns}		
Grasso, Ziegler, Mirault, et al. (2022)	E1	lex. dec. (primed)	explicit	trackpad	RT + MT	45	within subj.	?	25^{*c}	$?^{ns}$	?
	E2	lex. dec. (primed)	explicit	trackpad	RT + MT	294	bet. groups (online)	?	13^{*c}	$?^{ns}$	?
		lex. dec. (primed)	explicit	mouse	RT + MT	208	bet. groups (online)	?	13^{*c}	$?^{ns}$	?
		lex. dec. (primed)	explicit	keypress	RT	516	bet. groups (online)	?	$5^{ns\ c}$	$?^{ns}$	?
	E3	lex. dec.	implicit	trackpad	RT + MT	48	bet. groups	?	22^{*c}	$?^{ns}$	?
Grasso, Ziegler, Coull, et al. (2022)	E	lex. dec. (primed)	explicit	saccade	RT + MT	58	within subj.	?	13^{*c}	$2^{ns\ c}$	?
Abbondanza et al. (2024)	E1	time judg.	explicit	keypress	RT	39	bet. groups	13*	$18^{?}$	$9^{?}$	ns
	E2	lex. dec.	implicit	keypress	RT	37	within subj.	9*	$10^{?}$	$8^{?}$	ns
	E3	lex. dec. (irrel. fillers)	implicit	keypress	RT	55	within subj.	6^{ns}	$9^{?}$	$3^{?}$	ns
Carmona et al. (2024)	E1	time judg. (50 % subl)	explicit	keypress	RT	141	within subj.		164		
	E2	color dec.	implicit	keypress	RT	46	within subj.		84		

Summing up, evidence for the implicit space–time congruency effect, and therefore, the automatic activation of the lateral mental timeline, seems to be on the rise. Table 1 shows the relevant studies. Two reports (Grasso, Ziegler, Mirault, et al., 2022, and Grasso, Ziegler, Coull et al., 2022) suggest that the implicit effect arises only in words when long lateral movements are used (the lateral movement hypothesis) and two reports (Abbondanza et al., 2024, and Carmona et al., 2024) find the implicit effect in both words and pseudowords without using long lateral movements. However, upon close scrutiny, several of these findings occurred under conditions that render the time dimension relevant, leaving only a few experiments as clearly implicit. Grasso, Ziegler, Mirault, et al. (2022) obtained their finding in three experiments, but the first two presented a temporal word as prime before the lexical decision target, a procedure that may have rendered the temporal dimension highly salient and task relevant from the standpoint of the participant (see Santiago et al., 2011, 2012, for data and discussion of how saliency can affect the manifestation of an irrelevant dimension in a congruency task; see also de General Discussion section). Grasso, Ziegler, Coull, et al. (2022) also used a temporal prime in their only experiment using lateral saccades. In other words, these three experiments may be cases of explicit congruency effects (they are labeled as such in Table 1). This leaves Grasso, Ziegler, Mirault, et al. (2022, exp. 3), Abbondanza et al. (2024, exp. 2), and Carmona et al. (2024, exp. 2) as the only candidates for an implicit effect. Moreover, none of these studies have been replicated, and there are hints that they could be difficult to replicate. First, the lexical decision experiments provide inconsistent evidence in the pseudoword condition, as discussed above. Second, the effect found by Abbondanza et al. (2024, exp. 2) did not replicate when irrelevant fillers were added to the materials (exp. 3). Third, in a paradigm that is quite close to Carmona et al.'s (2024, exp. 2), Rolke et al. (2013) presented temporal words before a square whose color was to be discriminated using left–right keypresses. In stark contrast with the 84 ms effect observed by Carmona and colleagues, they observed a tiny (though significant) implicit congruency effect of 5 ms with visual presentation of the words, but then they could not replicate it in several subsequent experiments using auditory presentation unless attention was directed to the temporal meaning of the words (turning the temporal dimension explicit).[3] Finally, none of the published experiments in our target field of research has been preregistered and, therefore, it is impossible to assess the influence of the large number of experimenter's degrees of freedom (Nosek et al., 2019; Simmons et al., 2011). Thus, we will also consider here the possibility that the reports of implicit effects are false positives (the *false positive hypothesis* hereafter).

Last, but not least, we will consider an additional possibility: that the implicit congruency effect may be detected with a more sensitive measure. Reaction times index the end moment of a long processing chain. It is possible that implicit congruency effects are fleeting and vanish soon without leaving a trace in response time. Potentially useful measures include mouse tracking and evoked potentials (ERPs), but the available evidence is quite limited (see the General Discussion for studies using other measures). There are no ERP studies using our target tasks, but using more perceptual discriminations Li et al. (2019, 2023) found congruency effects in ERPs in conditions where reaction times failed to show them. Only two studies have examined space–time congruency using mouse tracking, and they both used explicit conditions. Miles et al. (2010) used a time judgement task and found a clear congruency effect in mouse trajectories. Malyshevskaya et al. (2023) presented temporal units (hours, weekdays, and months) after which participants moved the mouse upwards to bisect a horizontal line, an apparently implicit task. However, participants were asked to report which kind of time unit had been presented at the end of 40 % of the trials, thereby increasing the

[3] The reader may wonder why we do not further discuss here exp. 1 in Rolke et al. (2013) as another case of an implicit congruency effect. As with exp. 1 in Carmona et al. (2024), Rolke et al.'s (2013) experiments do not match the definition of our target tasks: participants see word primes, but respond to the color of a subsequent square. We will further discuss this study together with other methodologies in the General Discussion section.

relevance of time in the task. Even so, the evidence of lateral biases in mouse trajectories linked to the early vs. late temporal reference of units was mixed and inconsistent. The possibility that a more sensitive measure can capture the implicit congruency effect thus remains open. We will call this possibility the *measure sensitivity hypothesis* hereafter.

All in all, available evidence provides a mixed picture regarding the implicit lateral space–time congruency effect in single word processing tasks. In the present paper we report on a long-lasting quest to test the automatic activation of the left–right mental timeline using a single set of inflected verbs and pseudoverbs in time judgement vs. lexical decision, with as little procedural variation as possible. This quest started with Flumini and Santiago (2013). As Experiment 1, we here describe the methods of this study in detail (which was reported as a proceedings paper) because the following two experiments differed from it only in key respects, and we offer an up-to-date statistical re-analysis of those data that allows a close comparison to Experiments 2 and 3. Experiment 2 tested the measure sensitivity hypothesis by assessing mouse trajectories in the two tasks of Experiment 1. Finally, Experiment 3 provides a confirmatory test of the lateral movement hypothesis by conceptually replicating the final prime-less experiment of Grasso, Ziegler, Mirault, et al. (2022) with the present materials, a fully within-participant design, longer lateral movements of the hand, high power, and adequate methodological controls and open practices.

The first two experiments were carried out before the replication crisis (Bakker et al., 2012; Ioannidis, 2005; Open Science Collaboration, 2015) motivated the development of renovated guidelines for good experimental practice (see, e.g., Munafò et al., 2017), and thus, sample size was based on standard practice, there were not a priori power analyses, hypotheses and analytical pipelines were not preregistered, and so on. Therefore, they are explicitly reported here as exploratory attempts (see Wagenmakers et al., 2012, for the distinction between exploratory and confirmatory analyses). In contrast, the final experiment was the result of careful a priori planning regarding statistical power and sample size, and the hypotheses and analytical strategy were preregistered (https://aspredicted.org/6na98.pdf). In order to reduce the experimenter's degrees of freedom when analyzing the prior exploratory experiments, and to ease comparisons across studies, we preprocessed and analyzed all experiments using the preregistered criteria for the final experiment whenever possible. We did not update their sample sizes by collecting more participants, but we report (not preregistered) power analyses that allow us to assess their power to detect what we considered at the moment of designing the final experiment a smallest effect of theoretical interest (SESOI) in reaction time research (10 ms), as well as two additional benchmarks: the effect sizes detected by Grasso et al. (2022, exp. 3: 22 ms only in words) and Abbondanza et al. (2024, exp. 2: 9 ms). We established the SESOI for our final experiment based on extensive experience with the literature of single word processing, including priming and congruency effects, in which a 10 ms effect is usually considered to be a rather small effect. An even smaller effect may be important if there is a theory that specifically predicts an effect of that size, but this is not the case in the present context. To provide direct empirical grounding for our choice of SESOI, we examined the raw effect size of the explicit congruency effect between time and response side, whenever reported, of the single-word studies in the explicit condition in Von Sobbe et al.'s (2019) *meta*-analysis (Table 2). The average of 14 effect sizes was 60 ms (range 18–201 ms). The SESOI of 10 ms is thus 16.6 % of the average explicit congruency effect, and it is smaller than the smallest explicit effect size. After preregistering our final experiment (on 04/22/2022), Abbondanza et al. (2024, exp. 2) observed a significant even smaller effect (9 ms), so we also report not-preregistered analyses of this benchmark.

We also report equivalence analyses (Lakens, 2022; Lakens et al., 2018) to assess whether the effect sizes observed here were significantly smaller than those three benchmarks. We carried out additional (also not preregistered) analyses using variants of the preregistered pipeline aimed to show that the results do not depend on some specific decisions (i.e., latency cut-offs).

Table 2
Raw effect sizes of congruency between temporal reference and response side in explicit tasks in the studies included in the time-relevant condition of Von Sobbe et al.'s (2019) *meta*-analysis (when available). E = Experiment.

Study	Experiment	Effect size (in ms)
Aguirre and Santiago (2017)	E1	146
Aguirre and Santiago (2017)	E2	59
Bottini et al. (2015)	Sighted	18
Bottini et al. (2015)	Late blind	39
Bottini et al. (2015)	Early blind	28
de la Vega et al. (2016)	E	72
Ding et al. (2015)	E1: near time	63
Eikmeier et al. (2015)	Manual response	201
Kong and You (2011)	E1	45
Ouellet et al. (2012)	E	35
Ouellet et al. (2010)	Spanish natives	50
Ouellet et al. (2010)	Hebrew natives	19
Santiago et al. (2007)	E	42
Torralbo et al. (2006)	E2	29
	Mean	60

Data availability

We disclose and publicly share all materials, programs, raw data, and analysis scripts, which can be accessed at Open Science Framework (https://osf.io/ep5nb). The hypotheses and analytical strategy of Experiment 3 were preregistered (https://aspredicted.org/6na98.pdf).

Experiment 1 – Keypresses

Experiment 1 presented Spanish tensed verbs and pseudoverbs both for an explicit temporal judgment and an implicit lexical decision with bimanual left or right keypresses. Response latencies and accuracy were measured. The experiment is here described closely following the text in Flumini and Santiago (2013), but the data were re-analyzed following the preregistered strategy for Experiment 3, using up-to-date statistical techniques. Additional (not preregistered) analyses are also reported.

Method

Participants

Forty-eight psychology students from the University of Granada (6 males; age range 19–26 y.; 6 left-handed by self-report), all of them native Spanish speakers with normal or corrected vision, participated for course credit. All three experiments reported here are covered by resolution n° 763, 12/12/2012, of the Research Ethics Committee of the University of Granada. Participants were randomly assigned to two equal groups: temporal judgment and lexical decision.

Materials

We selected 148 Spanish intransitive verbs (or with at least one very common intransitive use). As Spanish allows subject dropping, single conjugated intransitive verbs can stand as full, grammatically correct sentences. The pseudoverbs were created by changing one letter of the morphological stem of each verb, with the only constraint of resulting in pronounceable sequences in Spanish (e.g., "*dormir*" – "to sleep" – was changed to "*dorpir*"). Both the 148 verbs and 148 pseudoverbs were conjugated in both the simple past perfect indicative and the simple future indicative, with all six possible combinations of grammatical person and number almost equally represented over the whole set (avoiding ambiguous forms such as "*amamos*" which means both "we love" and "we loved"). This resulted in 592 experimental stimuli of four types: past and future verbs, and past and future pseudoverbs. For example, from the item "*faltar/falbar*" the following versions were created: "*faltó*" ("he didn't show up" – past verb), "*faltará*" ("he will not show up" – future verb), "*falbó*" (past nonverb), "*falbará*" (future nonverb). Each of these versions was randomly assigned to one of four different lists and presented to different participants, thus avoiding lemma repetition. Each list contained 37 items from each stimulus condition (total 148 items). Direct control for factors such as frequency and length was not required because the theoretically relevant effect was the interaction between temporal reference and response hand when participants processed the very same list of stimuli using the two possible response/key mappings.

Procedure

Stimuli were presented at the center of a computer screen (Courier New font, 38 points, lower case), black printed on white background. Participants placed their left index finger on the Q key and their right index finger on the 9 key of the numerical keypad (located to the right of the typewriter keys) in a standard Spanish QWERTY keyboard. Each trial began with the presentation of a central fixation cross (500 ms) followed by the target verb or pseudoverb, which remained on the screen until a response was made. Incorrect trials were followed by a 500 ms red uppercase "X" at the same location of the stimulus plus a 1000 ms blank screen. Correct trials were followed by a 1500 ms blank screen. Participants in the explicit condition judged whether the target referred to either the past or the future. Participants in the implicit condition decided whether the target was a real Spanish verb or not.

In each task the same list of 148 items was presented twice in different blocks of trials (with a two minutes break) to be responded to using a different mapping of responses (past/future or word/pseudoword) to keys (left/right). The order of presentation of the two mappings was counterbalanced over participants. Each block was preceded by a four-trial training block with a different set of stimuli. The experiment was programmed and run using E-prime 2.0.

Design and analysis

The main analyses followed the preregistered pipeline for Experiment 3. Participants with more than 20 % errors were excluded. For the analysis of latency, correct trials above 250 ms and below 1750 ms were log-10 transformed. Predictors were dummy coded as −0.5 and 0.5 and sum contrasts were used. Latency was analysed by means of linear mixed models, and accuracy by means of generalized linear mixed models, both using the R package *lme4* (Bates, Mächler, et al., 2015). ANOVA-tables and *p*-values were derived using the *Anova()* function of the package *lmerTest* (Kuznetsova et al., 2017), using Type III sums of squares. The mixed model analysis started with a maximal model including the main effects and all interactions of Time (past vs. future), Response (left vs. right), and Lexical Status (word vs. pseudoword), both as fixed factors as well as random slopes, plus random intercepts per participant and item. This maximal model was then simplified until finding the model with the simplest random term that did not produce neither convergence nor singular fit warnings, and did not lose a significant amount of fit (Bates, Kliegl, et al., 2015; Matuschek et al., 2017). Model comparison was based on the Akaike Information Criterion and Wald χ^2 tests using the function *anova()* of the R package *car* (Fox & Weisberg, 2019). The supplementary analysis script at the OSF repository details the process of model search and allows for a full replication of the analysis pipeline. The theoretically relevant effects were the two-way interaction between Time and Response, and the three-way interaction between Time, Response, and Lexical status. If the former is significant, and the latter is not, the space–time congruency effect is replicated across both words and pseudowords. If the latter is significant, the congruency effect is modulated by the lexical status of the target. When reporting the results, we will only describe in the text the theoretically relevant findings. All other results, including the full output of the models, are provided as supplementary tables at the OSF repository (see the document "Supplementary statistical results.doc").

Three additional sets of analyses are reported. First, as the cut-offs preregistered for Experiment 3 led to the exclusion of a slightly larger-than-usual percentage of latencies in the time judgement task (see below), we also analyzed latencies using a less conservative cut-off. Second, we carried out computer simulations to assess the power of the present design to detect three benchmark sizes of the space–time congruency effect, expressed in raw values: a) a Smallest Effect Size of Interest (SESOI) of 10 ms in both tasks; b) a 9 ms effect in the lexical decision task, as observed by Abbondanza et al. (2024, exp. 2); and c) a 22 ms effect only in the word condition of the lexical decision task, as observed by Grasso et al. (2022, exp. 3). The power for larger effect sizes (as in Carmona et al., 2024, exp. 2) can only be larger than these benchmarks. To this end, we followed the approach by Brysbaert and Stevens (2018): we added a fixed amount to all congruent latencies until reaching the effect size of interest, log-transformed them, analyzed them using the same model previously selected as the most adequate to describe the observed data set, and, using the *simr* package (Green & MacLeod, 2016), carried out 1000 simulations of the data generated by that model. We then computed the proportion of significant tests (the power) and the 95 % confidence interval around it.

It is important to note that this approach allows a researcher to assess statistical power after the completion of a study without falling in the circularities of the analytical computation of post hoc power. Each simulation generates the data set anew, randomly drawing numbers from the distributions assumed by the model using the estimated parameters, including the matrix of variances and covariances. The observed data thus provide the best guess about population values for the parameters of the design. The simulated datasets match exactly the design, including both the specific combinations at trial level of participants, items, and conditions, as well as the deviations that result from the actual running of the experiment (the missing trials due to errors or outliers). Manipulating the size of a specific fixed effect within the model (either by directly setting its beta value or by adding constants to the raw data), while keeping all other model parameters as observed, allows the use of simulations to estimate the power to detect that effect size in complex designs for which there are not yet analytical solutions, and whose parameters would be very difficult to estimate otherwise (Arend & Schäfer, 2019; Brysbaert & Stevens, 2018; DeBruine & Barr, 2021; Kumle et al., 2021; Westfall et al., 2014). Moreover, by manipulating the data structure, simulations can also assess the impact on power of increasing the number of participants and/or items, thereby allowing the estimation of a priori power for future studies (Brysbaert & Stevens, 2018; Green & MacLeod, 2016). This was the approach that we took to establish the required sample size of Experiment 3, based on the data obtained in Experiment 1.

Finally, we also report equivalence analyses (Lakens, 2022; Lakens et al., 2018) to assess whether the observed theoretically relevant effects in the lexical decision task are significantly smaller than the SESOI and benchmark effect sizes (and any other larger effect). Using Null Hypothesis Significance Testing it is impossible to assert the absence of an effect (see also Harms & Lakens, 2018). Instead, we can claim that an effect is statistically smaller than a Smallest Effect Size of Interest (SESOI), and therefore it should be considered negligible or practically zero. Equivalence tests are simple extensions of the logic of hypothesis testing, but the null hypothesis is that the observed effect is equal to the SESOI.[4] As with power, there are not well-established analytical procedures to test for equivalence within complex designs with random factors, so we again opted for computer simulations. We used the same models described above that generated the simulated data for the power analyses for each benchmark, and computed 90 % confidence intervals[5] around the estimates of the fixed effects of the model using bootstrapping procedures (specifically, we used the R function *confint.merMod()* of package *lme4*). If the closer-to-zero limit of the 90 % CI of the estimate of the relevant effect for a given benchmark does not include the observed estimate in the analysis of the original dataset, we can claim that the latter is significantly smaller than that benchmark.

Results

Analysis pipeline preregistered for Experiment 3

No participant made more than 20 % errors. There were errors in 6.4 % of trials in temporal judgment and 5.2 % of trials in lexical decision. The fixed cut-offs (250 ms and 1750 ms) left out 9.5 % of correct trials in temporal judgement, and 2.1 % in lexical decision. As the proportion of latencies removed in the explicit task seems slightly high, additional analyses using a less conservative cut-off (2500 ms) are reported below. Log-transformed latencies were independently analyzed for each task.

Time judgement. The space–time congruency effect (incongruent minus congruent conditions) amounted to 21 ms (see Fig. 1). The selected model for the analysis of latencies was (in R notation): logRT ~ Time*Response*Lex + (1 + Time + Lex|Subject) + (1 + Response|Item). The interaction between Time and Response was significant ($\chi^2(1) = 15.46$, $p < .001$), whereas the three-way interaction between Time, Response, and Lexical status was not ($\chi^2(1) = 0.21$, $p = .65$). The analysis of accuracy was based on the following model: ACC ~ Time*Response*Lex + (1 + Time|Subject) + (1|Item). Neither the two-way ($\chi^2(1) = 0.03$, $p = .87$) nor the three-way interactions of interest ($\chi^2(1) = 1.92$, $p = .17$) were significant. See Table S1 for full ANOVA tables, and Tables S2 and S3 for the full output of the models, including parameter estimations, standard errors, degrees of freedom, and p-values.

Lexical decision. The congruency effect amounted to 0 ms (see Fig. 1). The model for the latency analysis was: logRT ~ Time*Response*Lex + (1 + Lex|Subject) + (1|Item). The congruency effect was not significant (two-way interaction: $\chi^2(1) = 0.33$, $p = .56$) nor interacted with Lexical Status (three-way interaction: $\chi^2(1) = 0.44$, $p = .51$). The model for accuracy was: ACC ~ Time*Response*Lex + (1 + Lex|Subject) + (1|Item). No relevant interaction was significant (Time x Response: $\chi^2(1) = 2.77$, $p = .10$; Time x Response x Lexical Status: $\chi^2(1) = 0.05$, $p = .82$). See Table S4 for full ANOVA tables, and Tables S5 and S6 for the full output of the models.

Additional exploratory analyses

Latency analyses using an upper cut-off of 2500 ms. This cut-off led to the removal of 1.8 % of correct trials in the time judgement task, and 0.2 % in the lexical decision task. The analysis pipeline was the same (see the supplementary analysis script). The pattern of results did not change: there was a significant interaction between Response and Time in time judgement ($\chi^2(1) = 31.61$, $p < .001$) but not in lexical decision ($\chi^2(1) = 0.16$, $p = .69$). The Time x Response x Lexical Status interaction was not significant in any task (Time judgement: $\chi^2(1) = 0.98$, $p = .32$; Lexical decision: $\chi^2(1) = 0.75$, $p = .38$). Table S7 shows full ANOVA tables.

Power analysis. To compute the power of the present design to detect the SESOI of 10 ms, we first brought the latencies to the SESOI by adding a constant to all congruent latencies, then log-transformed and run 1000 simulations of the latencies using the model described above. In time judgement, we decreased the congruency effect by adding 11.19 ms to congruent latencies. The Response x Time interaction was significant in 53.8 % (*95 % CI:* [50.7, 56.9]) of the simulations. In lexical decision, the congruency effect was increased to 10 ms by subtracting 10.1 ms to congruent latencies. The simulations rendered a power of 50.1 % (*95 % CI:* [46.7, 53.2]). Thus, neither task was properly powered to detect an effect of 10 ms, and obviously, even less so for the smaller benchmark of 9 ms (the supplementary analysis script includes commands to carry out this analysis). To measure the power to detect the benchmark of 22 ms in words in the lexical decision task we followed the same procedure only in the word condition: the congruency effect in words was −5 ms, so we subtracted 27 ms from them. Simulations rendered a power of 80.0 % (*95 % CI:* [77.38, 82.44]) to detect the Time x Response x Lexical Status interaction, and a power of 82.5 % (*95 % CI:* [80.00, 84.81]) to detect the Time x Response interaction. Therefore, the present design was well powered to detect differing congruency effects in words and pseudowords as reported by Grasso, Ziegler, Mirault, et al. (2022) in their third experiment.

Equivalence analysis of the lexical decision task. We compared the beta value of the Time x Response interaction observed in lexical decision ($\beta = 0.002$) with the 90 % confidence interval around the beta value obtained in the analysis of the 9 ms benchmark ([-0.0145, −0.0002]). The observed estimate fell outside that CI. In other words, the observed effect was significantly smaller than 9 ms (it actually had the opposite sign), what implicates that it is also smaller than the SESOI of 10 ms. The observed effect was also significantly smaller than the 22 ms benchmark in words (the supplementary script provides code to compute this equivalence analysis).

Discussion

In Experiment 1 we found a standard left–right space–time congruency effect in time judgement: responses were faster in the congruent than in the incongruent condition. This effect was not detected in the implicit lexical decision task on the same materials. Neither in temporal judgment nor in lexical decision, the effect was modulated by the lexicality of the stimuli. Additional a posteriori analyses supported that the present design was adequately powered to detect implicit effects of the size reported by Grasso, Ziegler, Mirault, et al. (2022, exp. 3), and therefore, also the larger effect reported by Carmona et al. (2024). Moreover, we also showed that the observed implicit effect was significantly smaller than the 9 ms effect reported by Abbondanza et al. (2024, exp. 2). The present study was not preregistered and, therefore, its findings are exploratory (as all the prior relevant literature), but the additional analyses suggest that it provides quite healthy evidence. Thus, Experiment 1 joins a large number of studies (Aguirre & Santiago, 2017; Maienborn et al., 2015; Santiago et al., 2007; Ulrich & Maienborn, 2010, among others) that suggest that a left-to-right mental time line is activated in the explicit time judgement task, but not in the implicit lexical decision task. We now turn to another not-preregistered study using an on-line, potentially more sensitive measure: mouse tracking.

Experiment 2 – Mouse trajectories

Experiment 2 used the same materials as in Experiment 1, but changed the data collection technique. As well as latencies, we measured

[4] When the directionality of the effect is not known, a double test is used to assess whether the observed effect is both significantly smaller than a positive SESOI and larger than a negative SESOI. In the present case, a negative effect would mean that the latencies in the incongruent conditions are faster than in the congruent conditions. Such reversed congruency effect would be clearly diagnostic against the presence of a congruency effect, so we will focus only on the hypotheses of finding a congruency effect that is smaller than each benchmark.

[5] Note that the test is one-sided, thus the 90% CI.

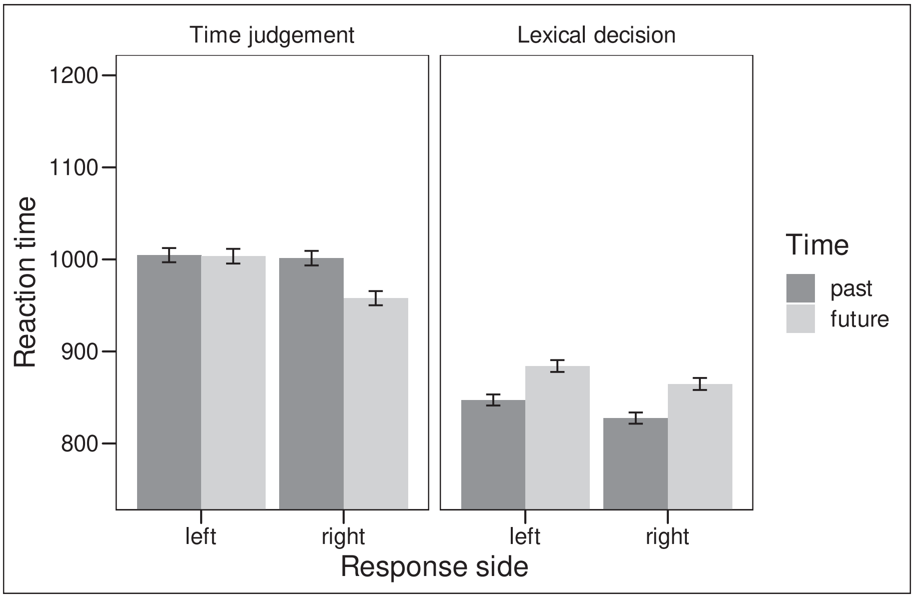

Fig. 1. Reaction times in Experiment 1 for time judgement and lexical decision tasks. Note. Reaction times are in ms. Error bars show ± 1 SEM.

mouse trajectories using the MouseTracker Software Package (Freeman & Ambady, 2010). Participants clicked on a square horizontally centered at the lower part of the screen, and immediately a word or pseudoword was presented at the center of the screen. The participant judged either its temporal reference or lexical status by clicking on one of two response boxes located at the upper left and right corners, moving the mouse diagonally. MouseTracker recorded the stream of x-y coordinates of participants' mouse trajectories, allowing a precise characterization of both the spatial and temporal dynamics of the responses. Thus, we were able to observe the unfolding competition between responses.

We focused on the following measures. MouseTracker provides two temporal indexes: initiation time (from the click on the start button to the beginning of the mouse movement; note that the click is followed by the immediate presentation of the target string); and trajectory time (from the click on the start button to the click on the selected response box). Trajectory time thus corresponds to a standard reaction time (from stimulus presentation to response), and we will call it reaction time henceforth. Of special interest are the indexes of deviation of mouse trajectories. The Area Under the Curve is the most sensitive (AUC; see also Flumini et al., 2015): it is the area between the mouse trajectory and an ideal straight line from the start button to the correct response box. Any attraction toward the alternative response box will induce a deviation of the mouse trajectory from this ideal straight line, generating an area between them. The greater this area, the greater the competition between responses. We analyzed temporal measures following the preregistered pipeline for Experiment 3 as much as possible, and trajectory deviations following established practice, specifically the analysis decisions in Miles et al., (2010), the only available mouse tracking study of space–time congruency effects that is directly comparable to present methods. Miles et al. (2010) used only a time judgement task. Here we presented the same materials in both a time judgement and a lexical decision task, allowing a direct comparison between explicit and implicit procedures. If mouse tracking is more sensitive than reaction times, it should detect the space–time congruency effect in both tasks, thereby supporting the measure sensitivity hypothesis. As the mouse movements had a lateral component (they crossed the screen diagonally), the present experiment also provides a partial test of the lateral movement hypothesis, which makes the same predictions.

Method

Participants

Forty-eight Psychology students from the same population as in Experiment 1 (6 male; age range 20–29 y.; 4 left-handed by self-report) participated for course credit. Half of them were randomly assigned to each task.

Materials and procedure.

Everything was kept as in Experiment 1, with the exception that Mousetracker Software Package (Freeman & Ambady, 2010) was used to collect mouse trajectories of participants' responses. Individual trajectories were rescaled to a standard coordinate space, and then normalized into 101 time steps (Freeman & Ambady, 2010), using the utilities included in the MouseTracker software package.

Participants sat 60 cm from the computer screen, with their dominant hand placed over the mouse they found centrally positioned in front of them. Two grey squares (without any text) were always displayed at the top corners of the screen as response alternatives. Each trial began with the presentation of the start button (a grey rectangle at the bottom center of the screen with the word "start") which remained on the screen until clicked. Without delay, the target stimulus appeared at the center of the screen, and the participant moved the cursor to the chosen response box on the top left or right corner and clicked on it. All other details of the procedure and materials were as in Experiment 1.

Design and analysis

The design was the same as in Experiment 1, with four measures: initiation time, reaction time, accuracy, and AUC. The accuracy and latency measures were also analyzed following the same analysis strategy as in Experiment 1 (see the supplementary script). AUC was analyzed on the correct trials that had been selected to participate in the analysis of latency (following Miles et al., 2010), also using linear mixed models. We report the same set of additional analyses for reaction times as in Experiment 1 (less conservative upper cut-off, and power and equivalence for benchmark effects). As AUC showed a skewed distribution, we also report additional non-parametric analyses. Finally, a power analysis using simulations was also carried out for AUC.

Results

Analysis pipeline preregistered for Experiment 3 plus analysis of AUC

No participant was excluded for having more than 20 % errors. The exploration of initiation time (time from the click on the start button to the beginning of mouse movement) showed that 71.8 % of trials had latencies below 200 ms. This means that participants started moving the mouse immediately, even before the target stimulus could be processed. Thus, we did not analyze initiation time any further and focused instead on reaction time (the time between the initial click and the final click on one of the two response boxes). There were errors in 0.05 % of trials in the time judgement task and 0.02 % in the lexical decision task. The 250 ms and 1750 ms cut-offs were increased in 50 ms because MouseTracker is known to add that constant amount to latency measures (Freeman & Ambady, 2010). These cut-offs left out 22.3 % of correct trials in time judgement and 12.1 % in lexical decision. As these rejection percentages are larger than usual practice, we report additional analyses using less conservative cut-offs below. Finally, selected latencies were log-transformed and independently analyzed for each task.

Time judgement. There was a 35 ms space–time congruency effect in this task (Fig. 2). The selected model for reaction time was logRT ~ Time*Response*Lex + (1 + Time + Response|Subject) + (1|Item). There was a significant Time x Response interaction ($\chi^2(1) = 31.98$, $p < .001$), which did not interact with Lexical Status ($\chi^2(1) = 0.40$, $p = .53$). Accuracy was analyzed using the model ACC ~ Time*Response*Lex + (1 + Time + Response + Lex|Subject) + (1|Item). None of the relevant interactions was significant (Time x Response: $\chi^2(1) = 0.91$, $p = .34$; Time x Response x Lex: $\chi^2(1) = 0.41$, $p = .52$). The model for AUC (see Figs. 3 and 4) was AUC ~ Time*Response*Lex + (1 + Time + Response|Subject) + (1 + Response|Item). The Time x Response interaction was significant ($\chi^2(1) = 15.20$, $p < .001$) and it was not modified by Lexical Status ($\chi^2(1) = 1.47$, $p = .22$). See Table S8 for full ANOVA tables, and Tables S9-S11 for the full summary of the models.

Lexical decision. In this task, space–time congruency in reaction time amounted to −0.3 ms (see Fig. 2). The model for latencies was logRT ~ Time*Response*Lex + (1 + Response + Lex|Subject) + (1|Item). No relevant interaction reached significance (Time x Response: $\chi^2(1) = 0.32$, $p = .57$; Time x Response x Lex: $\chi^2(1) = 0.86$, $p = .36$). The model for accuracy was ACC ~ Time*Response*Lex + (1 + Lex|Subject) + (1 + Response|Item). Again, the relevant interactions were not significant (Time x Response: $\chi^2(1) = 0.13$, $p = .72$; Time x Response x Lex: $\chi^2(1) = 0.08$, $p = .77$). AUC (Figs. 3 and 4) was analyzed using the model AUC ~ Time*Response*Lex + (1 + Response + Lex|Subject) + (1 + Response|Item). It also failed to capture any of the relevant interactions (Time x Response: $\chi^2(1) = 3.27$, $p = .07$; Time x Response x Lex: $\chi^2(1) = 0.33$, $p = .57$). Table S12 shows full ANOVA tables, and Tables S13-15 the full output of the models.

Additional analyses for reaction time

Latency analyses using an upper cut-off of 2500 ms. The cut-offs were set at 300 ms and 2550 ms to compensate for the constant extra 50 ms added by MouseTracking (Freeman & Ambady, 2010). This led to filtering out 0.06 % of correct trials in the time judgement task and 0.03 % in the lexical decision task. There were no changes in the pattern of results, with a significant Time x Response interaction in the time judgement task ($\chi^2(1) = 42.60$, $p < .001$) and a non-significant interaction in lexical decision ($\chi^2(1) = 0.08$, $p = .77$). The three-way interaction Time x Response x Lex was not significant in any task (Time judgement: $\chi^2(1) = 1.52$, $p = .22$; Lexical decision: $\chi^2(1) = 1.59$, $p = .21$). See Table S16 for full ANOVA tables.

Power analysis. Using the same procedure as in Experiment 1, the design showed 24.50 % (*95 % CI:* [21.86, 27.29]) power to detect a 10 ms Time x Response interaction in the time judgement task. In lexical decision, power was 36.90 % (*95 % CI:* [33.90, 39.98]). Power was inadequate to detect an interaction of this effect size in both tasks. Therefore, the power of the lexical decision task was even smaller to detect the 9 ms benchmark (see supplementary script). The power to detect a significant Time x Response x Lexical Status interaction after increasing to 22 ms the congruency effect only in words was 46.30 % (*95 % CI:* [43.17, 49.45]), and for the Time x Response interaction was 58.40 % (*95 % CI:* [55.27, 61.48]). Thus, the present design was also not well powered to detect this benchmark.

Equivalence analysis of the lexical decision task. The observed beta value ($\beta = 0.0020$) of the Time x Response interaction was not included in the 90 % CI around the beta of the 9 ms effect ([-0.0105, 0.0010]). This means that the observed congruency effect in lexical decision was significantly smaller than the smallest benchmark, and thus also smaller than the 10 ms SESOI. The observed effect was also significantly smaller than the 22 ms benchmark in words (the supplementary script also provides code to compute this equivalence analysis).

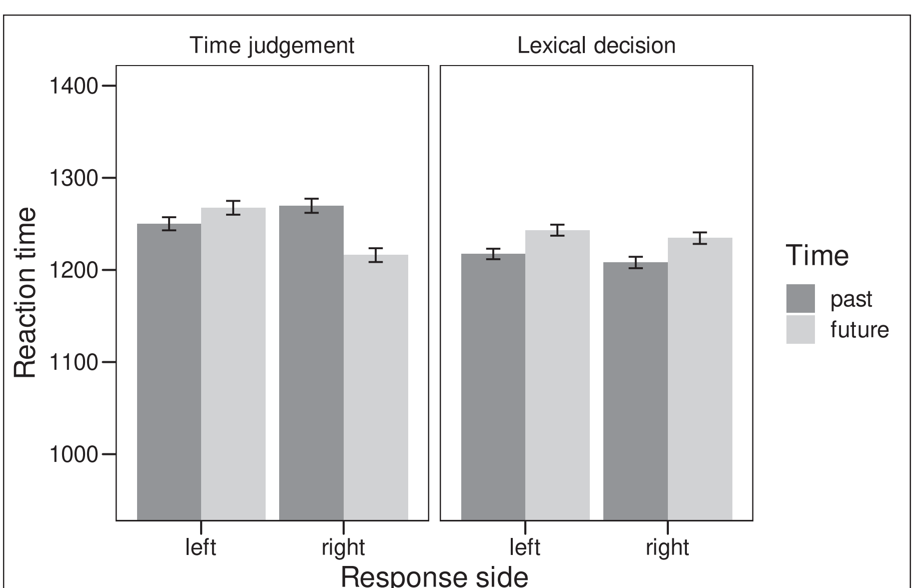

Fig. 2. Reaction time results in Experiment 2 in the time judgement and lexical decision tasks. Note. Reaction times are in ms. Error bars show ± 1 SEM.

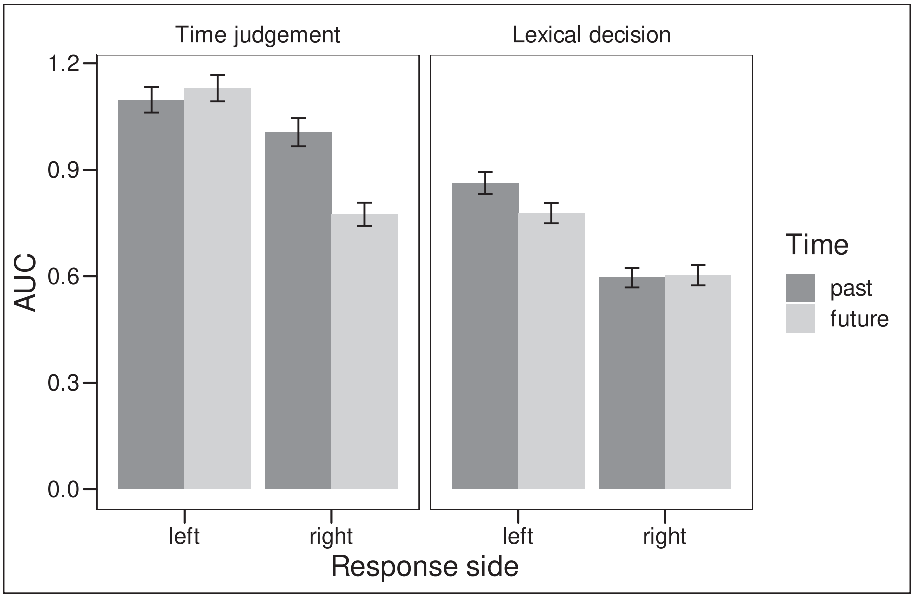

Fig. 3. Area Under the Curve (AUC) results of Experiment 2 in the time judgement and lexical decision tasks. Note. Error bars show ± 1 SEM.

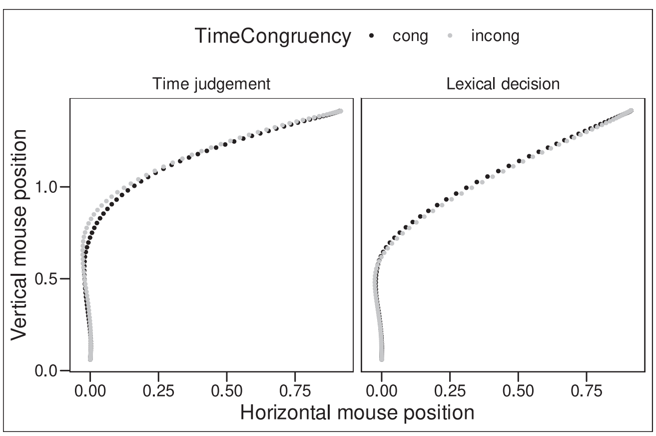

Fig. 4. Average mouse trajectories in Experiment 2 in the time judgement and lexical decision tasks. Note. Trajectories are normalized both in space and time, and rightward remapped.

Additional analyses for AUC.

AUC non-parametric analyses. As AUC had a skewed distribution, we also carried out non-parametric analyses using the Wilcoxon test. AUCs in congruent and incongruent trials were different in time judgement ($V = 70$, $p = 0.02$), but not in lexical decision ($V = 157$, $p = 0.86$).

Power analysis in the lexical decision task. Given the lack of prior studies using mouse tracking in implicit space–time congruency tasks, we set the target effect size of the interaction Time x Response equal to the effect observed in the time judgement task (*mean AUC* = 0.13), a value that closely replicated Miles et al. (2010). The obtained value in the lexical decision task was −0.05, so we subtracted 0.18 from AUC in all congruent trials. Simulations showed that the design had 99.90 % (*95 % CI:* [99.44, 100.00]) power to detect a Time x Response interaction of 0.13. If the target effect size is halved, the power comes down to 66.60 % (*95 % CI:* [63.58, 69.52]).

Discussion

Experiment 2 used the materials and tasks of Experiment 1 in the context of a mouse tracking procedure. The main measure of interest was the deviation of the mouse trajectories toward the alternative response box under conditions of response-time congruency versus incongruency. If mouse trajectories are a sensitive index of the on-line process of response decision (Flumini et al., 2015; Freeman & Ambady, 2009, 2010), they should reveal an implicit congruency effect that escapes reaction times. In contrast, Experiment 2 found the effect only in time judgement, with the same size reported by Miles et al. (2010). In lexical decision there were no traces of such interaction, even though the present design was very well-powered to detect an interaction of that size, and even had 66.6 % power to detect an interaction half as large. Reaction times showed the same pattern: the congruency effect was only present in time judgement, but not in lexical decision, consistently with Experiment 1. However, latency measures were not as well powered as in the prior experiment, probably because mouse click-move-click responses increased random noise in the data. Even so, equivalence analysis showed that the observed congruency effect in lexical decision was significantly smaller than the smallest benchmark (9 ms), suggesting that the effect was negligible.

Together, Experiments 1 and 2 provide consistent evidence, with two measures of different sensitivity and using long movements with a lateral component in the latter, against the existence of space–time congruency effects in lexical decision of the sizes observed by Grasso, Ziegler, Mirault, et al. (2022, exp. 3) and Abbondanza et al. (2024, exp. 2), and also of the larger effect observed by Carmona et al. (2024, exp. 2) using a Stroop-like task. Thus, the prior two experiments fail to support the predictions of the measure sensitivity hypothesis, and provide at least a partial refutation of the lateral movement hypothesis. However, as writing movements are horizontal, for a full test of the latter hypothesis we need to use perfectly lateral movements. Importantly, none of the two prior experiments were preregistered, and therefore, they cannot be properly used to disconfirm any hypothesis at the 5 % alpha level. Therefore, we carefully designed and preregistered Experiment 3.

Experiment 3 – Sideways Movement

Experiment 3 is a replication of the lexical decision task of Experiment 1 with the only difference being the use of a lateral response. To provide a stringent test of the lateral movement hypothesis, participants performed a long lateral movement of a handle, either leftwards or rightwards, until reaching the maximum length of the response apparatus (see Fig. 5). A power analysis based on Experiment 1 ensured over 80 % power to detect the SESOI of 10 ms, and sample size, together with

Fig. 5. Setup used in Experiment 3.

design, hypotheses, criteria for outlier exclusion, analysis pipeline, stopping points, and basically, all experimenter degrees of freedom, were preregistered before the beginning of data collection (see https://aspredicted.org/6na98.pdf). This allows the present experiment to be used for confirmatory testing of the hypothesis that, when lateral movements are used, a space–time congruency effect arises in lexical decision. This is the first confirmatory study reported on the automaticity of the activation of the mental timeline in our target tasks.

Method

Participants

Sample size was estimated based on the power analysis of the Response x Time interaction of Experiment 1 for the 10 ms SESOI (see the preregistration). As described above, the power to detect that SESOI was 50.10 %. We then doubled the sample size to 48 participants and run other 1000 simulations. In this case, 81.20 % (*95 % CI:* [78.64, 83.58])[6] of them rendered significant results. We finally increased the sample an additional 24 participants. New simulations showed that, with $N = 72$, the power to detect a 10 ms interaction was 93.10 % (*95 % CI:* [91.35, 94.59]). To maximize efficiency in data collection we preregistered a sequential analysis strategy (Pocock, 1977; Wald, 1947; see Lakens, 2014, for a tutorial): we divided alpha across two stopping points. After 48 participants, we would test the effect of interest at $\alpha = .03$. If not significant, we would collect 24 more participants, and test the effect again at $\alpha = .02$. This strategy allows incremental data collection without corresponding increases in Type I error. However, this strategy also comes with a slight reduction in power, so we run new simulations at those sample sizes and alpha levels. At the first stopping point, power reached 72.00 % (*95 % CI:* [69.11, 74.76]). At the second stopping point, it was 86.40 % (*95 % CI:* [84.12, 88.46]). Power thus remained at acceptable levels. Upon data collection, the interaction did not reach significance at the first stopping point and we proceeded to the second. Actually, we collected 76 participants (14 males; age range 18–29 y:, 7 left-handed by self-report), due to participant availability, with the same characteristics as in prior experiments. We here report analyses with the full sample. The supplementary script allows for analysis at the first stopping point.

Materials and procedure

Everything was kept identical to Experiment 1, with the only exception of the response apparatus. Participants moved the slider leftwards or rightwards until reaching the end point to indicate whether the stimulus was a word or a pseudoword. The handle returned automatically to the center after each response.

Design and analysis

We preregistered independent analyses of starting time (from stimulus presentation to the beginning of the movement of the slider) and travelling time (from the beginning to the end of the lateral movement, which we will call "end time" henceforth). Grasso, Ziegler, Mirault, et al. (2022) used lateral movements of the mouse or a pen on a trackpad and measured the reaction time from stimulus presentation until the response reached a boundary located on the far left or right sides of the screen. The analogous measure in the present study would be the addition of starting time and end time. Thus, we also analyzed this total reaction time measure in the additional exploratory analyses. Moreover, Grasso, Ziegler, Mirault, et al. (2022) observed the implicit congruency effect only in the word condition, so we report additional independent analyses of words and pseudowords. The additional analyses also include power and equivalence analyses of the smaller benchmarks: the

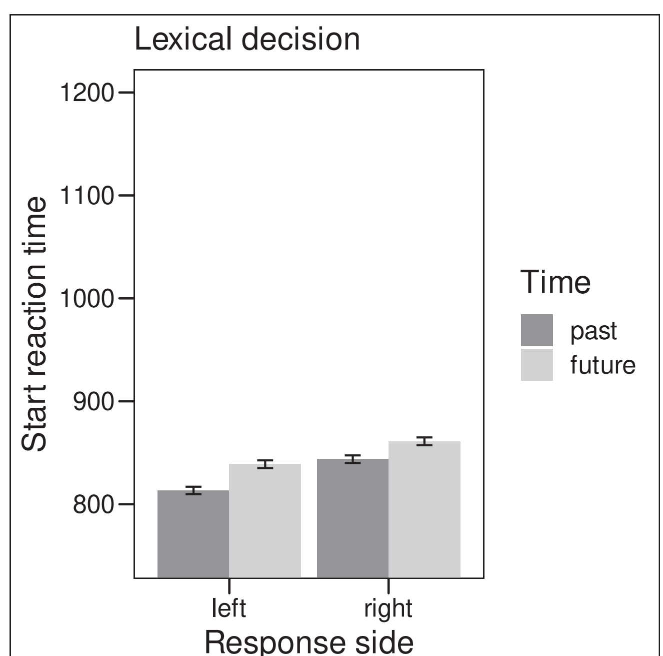

Fig. 6. Results of Experiment 3 in the start reaction time measure of the lexical decision task. Note. Error bars show ± 1 SEM.

[6] The slight differences with the preregistered numbers are due to having used a different random seed. We here report the results from the seed used in the supplementary script, which should be perfectly reproducible.

10 ms SESOI and the 9 ms congruency effect found by Abbondanza et al. (2024).

Results

Preregistered analysis pipeline

One participant was filtered out because of having more than 20 % errors. Thus, final sample size was 75. There were errors on 7.1 % of trials. Correct trials with start time below 250 ms and above 1750 ms (3.7 %) were removed. As this proportion is within normal practice, we did not carry out additional analyses with less conservative cut-offs. Remaining latencies were log-transformed. Latency and accuracy were analyzed by means of mixed models as detailed in Experiment 1.

The Time x Response interaction in starting time amounted to 4 ms (see Fig. 6). The selected model was logRTStart ~ Time*Response*Lex + (1 + Response + Lex|Subject) + (1|Item). The effect approached, but did not reach, the alpha level preregistered for the second stopping point ($\chi^2(1) = 5.16$, $p = .023$). The three-way interaction with Lexical Status was not significant ($\chi^2(1) = 0.27$, $p = .60$). The congruency effect on end time was 0 ms. The model was logRTEnd ~ Time*Response*Lex + (1 + Response|Subject) + (1|Item). None of the relevant interactions were significant (Time x Response: $\chi^2(1) = 0.85$, $p = .36$; Time x Response x Lexical Status: $\chi^2(1) = 0.58$, $p = .44$). Accuracy was analyzed with the model ACC ~ Time*Response*Lex + (1 + Lex|Subject) + (1|Item). Again, no relevant interaction reached significance (Time x Response: $\chi^2(1) = 1.32$, $p = .25$; Time x Response x Lexical Status: $\chi^2(1) = 1.93$, $p = .16$). See Table S17 for ANOVA tables, and Tables S18-20 for the full output of the models.

Additional exploratory analyses

Total reaction time. For better comparison with Grasso, Ziegler, Mirault, et al. (2022), we analyzed total reaction time (starting time plus end time) in both the full dataset and separately for words and pseudowords (although this analysis was not warranted due to the lack of a significant three-way interaction). The size of the Time x Response interaction in this measure was 4 ms (5 ms in words and 3 ms in pseudowords). The full dataset was analyzed using the model logRT ~ Time*Response*Lex + (1 + Response + Lex|Subject) + (1|Item). The congruency effect was clearly above the preregistered level ($\alpha = 0.02$) for the second stopping point ($\chi^2(1) = 3.79$, $p = .052$). Its interaction with Lexical Status was non-significant ($\chi^2(1) = 0.002$, $p = .97$). The subset of words was analyzed with the model logRT ~ Time*Response + (1 + Response|Subject) + (1|Item). The Time x Response interaction was not significant ($\chi^2(1) = 1.93$, $p = .17$). Pseudowords were analyzed using the model logRT ~ Time*Response + (1 + Response|Subject) + (1|Item). Again, the interaction was not significant ($\chi^2(1) = 1.64$, $p = .20$). See Table S21 for full ANOVA tables, and Tables S22-24 for the full output of the models.

Starting reaction time separately for words and pseudowords. For completeness, starting reaction time was also analyzed separately for words and pseudowords. The congruency effect in words was 3 ms, and in pseudowords was 5 ms and did not reach significance in any subset (Words: $\chi^2(1) = 1.43$, $p = .23$; Pseudowords: $\chi^2(1) = 3.71$, $p = .054$). See Tables S25-27.

Power analyses for benchmark effects. The power of the present design to detect a 9 ms congruency effect in starting time using alpha = .02 was 98.00 % (*95 % CI:* [96.93, 98.77]). Its power in total reaction time was 90.50 % (*5 % CI:* [88.51, 92.25]). Thus, Experiment 3 was very well powered to detect a 9 ms congruency effect and, therefore, also larger effects. The supplementary script provides code to carry out power simulations for the 10 ms SESOI in both starting time and total reaction time using the whole set of stimuli as well as only the subset of words. Halving the dataset reduced power, but only in the total reaction time measure power was below 80 % (66.70 %, *95 % CI:* [63.68, 69.62]). This is probably because end time is much less sensitive to processes that occur during response planning (as suggested by the independent analyses of starting and end time reported above), and adding it to starting reaction time only contributes some random noise. However, the design was very well powered to detect the 22 ms effect observed by Grasso, Ziegler, Mirault, et al. (2024) in words in total reaction time (100.0 %, *95 % CI* = [99.63, 100.0]).

Power analysis for the observed 4 ms congruency effect in starting time using the whole dataset at a standard significance level of.05. At both stopping points, the observed space–time congruency effect in the full item set approached the preregistered alpha levels, but did not reach them, thereby not qualifying to be claimed as significant. However, in both cases, the effect was below the standard 0.05 significance level. As an exploratory analysis, being aware that we are using here an alpha level that is unduly relaxed, we wanted to know what would be the power of the present design with the 75 participants sample to detect the observed 4 ms effect in starting time (the most sensitive measure) at a standard alpha level of 0.05. Simulations showed that power was 62.00 % (*95 % CI:* [58.91, 65.02]), well below the conventionally accepted 80 % level.

Equivalence analyses. In the analysis of starting time using the full dataset (words and pseudowords), the beta value observed for the Time x Response interaction was β = -0.0058. In the same analysis after bringing the latencies to the 9 ms benchmark, beta was −0.0111, with a *90 % CI* = [-0.0154, −0.0069]. The observed beta is outside the 9 ms beta CI. Same results were found in the analysis of total reaction time (observed β = -0.0042, 9 ms β = -0.0079, *90 % CI* = [-0.0114, −0.0044]). Thus, the observed effect was significantly closer to zero than the smallest benchmark, the 9 ms effect found by Abbondanza et al. (2024), even smaller that our preestablished SESOI of 10 ms. When using only the subset of words, the observed 5 ms effect (β = -0.0040) was not outside the 90 % CI around the 10 ms effect (β = -0.0080, *90 % CI* = [-0.0127, −0.0032]), but it fell outside the 90 % CI around the 22 ms effect (β for the 21.5 effect in words = -0.0167; *90 % CI* = [-0.0215, −0.0120]) observed by Grasso, Ziegler, Mirault, et al. (2022, exp.3). The supplementary script provides code for computing additional equivalence analyses.

Discussion

Experiment 3 used an implicit task, lexical decision, to test whether long sideways movements are able to produce the automatic activation of the lateral mental timeline. The experiment was preregistered and well-powered, and therefore, it provided a confirmatory test of the lateral movement hypothesis. The results were clear: there was a 4 ms space–time congruency effect in the latency between stimulus and movement start, and a virtually zero effect in the travelling time from movement start to end. The former effect approached, but did not reach significance, at any of the two preregistered sequential stopping points. Additional exploratory analyses also failed to observe the effect in total reaction time (the sum of start and end time, which is analogous to the measure used by Grasso, Ziegler, Mirault, et al., 2022). In no measure there was a significant interaction of the congruency effect with lexical status, but for the sake of comparison with Grasso, Ziegler, Mirault, et al. (2022, exp. 3), who observed the effect only in words, additional exploratory analyses showed that the effect was absent both in words and pseudowords. Power analyses (using the preregistered 2 % alpha for the second stopping point) showed that the design was well powered to detect the smallest benchmark (the 9 ms effect observed by Abbondanza et al., 2024), even smaller than our preestablished SESOI of 10 ms, both in starting and in total reaction time when using the full set of materials. When only words were included, the design was well-powered to detect the 10 ms SESOI in starting reaction time, but not in total reaction time. However, it was very well-powered to detect the 22 ms effect in words in total reaction time observed by Grasso, Ziegler, Mirault, et al. (2022, exp. 3).

Moreover, the observed (non-significant) congruency effects were significantly smaller than the smallest 9 ms benchmark in both starting and total reaction time when using the full set of materials, and also in starting time when using only words. In total reaction time, it was significantly smaller than the 22 ms effect observed by Grasso, Ziegler, Mirault, et al. (2022, exp. 3). Finally, an exploratory power analysis with the whole 75 participants sample and assuming the standard 5 % alpha level showed that power was only 62.00 % (*95 % CI:* [58.91, 65.02]). That is, if repeated identically with only one stopping point at $N = 75$, the present study would find a significant effect only 62 % of the time, in spite of the increase in Type I error that this power analysis implies.

All in all, we can conclude that the present experiment provides strong evidence against the existence of a 22 ms space–time congruency effect in word stimuli in a lexical decision task (as observed by Grasso, Ziegler, Mirault, et al. (2022, exp. 3), against the existence of a SESOI of 10 ms when pooling both words and pseudowords, and even against the existence of a 9 ms effect in the full set of materials (as observed by Abbondanza et al. 2024, exp. 2). This also casts strong doubts about the existence of an 84 ms effect in Stroop-like tasks as reported by Carmona et al. (2024). In other words, the implicit congruency effect should be considered negligible.

General Discussion

Can the left–right mental time-line be activated automatically? We examined this question focusing on the space–time congruency effect in tasks that require fast decisions to single words about dimensions of the stimulus other than their temporal reference, thus making time an implicit dimension of the task. Against a background of reports failing to find such effects (Aguirre & Santiago, 2017; Flumini & Santiago, 2013; Maienborn et al., 2015; Ulrich et al., 2012; Ulrich & Maienborn, 2010, see Introduction), three recent experiments have observed the implicit congruency effect, two of them in lexical decision (Abbondanza et al., 2024, exp. 2; Grasso, Ziegler, Mirault, et al., 2022, exp. 3) and one in a Stroop-like task (Carmona et al., 2024, exp. 2). The latter joins evidence by Rolke et al. (2013, exp. 1) in a similar task without direct responding to the linguistic stimulus, but to the color of a subsequent square. In this context, it would seem that the evidence is accruing on the side of the existence of an automatic space–time congruency effect in word processing tasks.

In the present study we tested the automaticity of the space–time congruency effect by means of lexical decision tasks in which both words and pseudowords carried temporal inflections. We also tested two potential moderators: the sensitivity of the measure and the use of long sideways movements as responses. We aimed to assess not only the presence of the effect, but also whether its size significantly differs from a Smallest Effect Size of Interest (SESOI), which we established at 10 ms based on prior reports and extensive experience in this field of research. Moreover, we also compared the observed effect with two benchmarks: the effect sizes reported by Grasso, Ziegler, Mirault, et al. (2022, exp. 3: 22 ms in the subset of words) and Abbondanza et al. (2024, exp. 2: 9 ms pooling words and pseudowords). We also assessed whether the present findings were adequately powered. We hereby reported two not-preregistered (thus, exploratory) studies using reaction time and mouse tracking, and the first preregistered (https://aspredicted.org/6na98.pdf), confirmatory study in this field of research, which secured high power and a 5 % Type I error. Importantly, we offer all materials, programs, raw data, analysis scripts, and detailed output tables as publicly available supplementary materials at OSF (https://osf.io/ep5nb).

The first two studies were exploratory, with sample sizes established following the standards that were usual before the advent of the replication crisis in psychology. We set the stage by reporting in detail and with up-to-date statistical analyses the study by Flumini and Santiago (2013), which was a proceedings paper. This study compared an explicit (time judgement) and an implicit (lexical decision) task. The congruency effect arose in the explicit, but not in the implicit task. Moreover, the observed (non-significant) effect in lexical decision was significantly smaller than the smallest benchmark. Computer simulations indicated that the design was not adequately powered to detect the 10 ms SESOI (and less the 9 ms benchmark), but it was well-powered to detect the 22 ms effect observed by Grasso, Ziegler, Mirault, et al. (2022) in words.

Experiment 2, also exploratory, assessed both reaction times and mouse trajectories to test the possibility that a more sensitive measure might detect the implicit effect. Mouse tracking is able to reveal the online accumulation of evidence for the finally selected response in a way that reaction time is not (see, e.g., Freeman & Ambady, 2009). The pattern of results was identical to Experiment 1, in both measures: a clear congruency effect in time judgement and a non-significant effect in lexical decision. There are no prior reports to serve as benchmarks for the implicit effect size in mouse deviations, so we compared it to the explicit effect size: the (non-significant) implicit effect was significantly smaller. The design was highly powered to detect an effect of the size of the explicit effect in mouse trajectories, and even had 66.6 % power to detect an effect half that size. The (non-significant) implicit effect in reaction time was also significantly smaller than the 9 ms benchmark, although the design was not adequately powered to detect even the 22 ms benchmark, probably because of the added noise due to mouse responses. Thus, mouse tracking also failed to detect the implicit congruency effect. It is an open question whether even more sensitive measures, such as ERPs, might suceed.

Although the first two experiments provided quite healthy evidence for a negligible implicit effect in reaction time (Experiments 1 and 2) and mouse trajectories (Experiment 2), their main problem is their exploratory nature. Lacking a detailed preregistration, it is impossible to know what role played the flexibility in experimenter choices in obtaining the results (Nosek et al., 2019; Wagenmakers et al., 2012). We tried to minimize this problem by following the preregistered plan for Experiment 3 as much as possible when analyzing Experiments 1 and 2, but this is not enough to secure that the findings meet the conventional 5 % Type I error level. This problem plagues psychology, and biomedical sciences in general, and together with publication bias, it is considered a chief cause of the lack of reproducibility of most published studies (Ioannidis, 2005; Open Science Collaboration, 2015; Simmons et al., 2011).

Therefore, Experiment 3 was designed strictly adhering to current guidelines for good scientific practice (Munafò et al., 2017) to provide a confirmatory test of the implicit space–time congruency effect in lexical decision, under conditions that have been suggested to increase the possibility of finding it: using long sideways movements as responses (Grasso, Ziegler, Mirault, et al., 2022). We carried out an a priori power analysis based on Experiment 1 (which only differed in the kind of response) to secure high power to detect a SESOI of 10 ms and established a sequential analysis strategy with two stopping points, dividing the alpha between them to maintain Type I error at 5 % (we also checked that a priori power remained high at these two stopping points, and a posteriori analyses corroborated this point). The study was preregistered and we reported here the results making a clear distinction between confirmatory and exploratory analyses. Preregistered analyses revealed a small congruency effect (4 ms) in the time to start the response, that was non-significant at either of the stopping points at the adjusted alpha levels. Moreover, additional exploratory analyses confirmed that the study was highly powered to detect even the smallest benchmark (9 ms). Additional exploratory analyses also assessed the significance, power and equivalence of the measure of total reaction time, which corresponds more closely to the measure used by Grasso, Ziegler, Mirault, et al. (2022), and provided independent analyses of the subsets of words and pseudowords (as Grasso and colleagues found the effect only in words). Power was high and the effect was significantly smaller than 9 ms in the full item set, and smaller than the effect observed by Grasso, Ziegler, Mirault, et al. (2022) in the subset of words. Finally, we checked the power of the present design to detect the observed effect size (4 ms) with the full sample at a conventional 5 % alpha, even though this implies that Type I error is probably higher than 5 %. The probability of finding a significant effect in identical repetitions of this study was only 62 %, well below the conventionally accepted level (80 %).

All in all, using Null Hypothesis Testing nobody can rule out the possibility that there truly is a very small automatic space–time congruency effect in lexical decision. But we can claim with a 5 % Type I error rate that such effect, if it exists, is smaller than 9 ms (even smaller that our preregistered SESOI of 10 ms), and we thus suggest that it can be considered negligible.

Is it possible to find implicit space–time congruency effects in language processing tasks?

There are two other lines of inquiry that have shown what appear implicit effects in the processing of linguistic materials with a semantic connection with time. It is our contention that these effects occur because the task favours the simultaneous activation of space and time by either requiring them or increasing their saliency. As proposed by the Coherent Working Models theory (Santiago et al., 2011, 2012; Torralbo et al., 2006), when participants carry out any task, they set up a mental model of the task in working memory, and the elements of this mental model enter into coherence interactions among them. These interactions are responsible for the observed congruency effects, as an incongruent model is less coherent than a congruent model. Frequently used models tend to be stored in long term memory and tend to be re-used depending on prior practice. Reading and writing experiences lead to the creation of models in which the contents of the story are placed in accordance with the direction of the script (Román et al., 2013, 2015, 2018). As earlier events tend to be mentioned before later events, the lateral dimension gets associated with sequences of events and the passage of time, although the association is far from perfect (Román et al., 2018; see also Maass et al., 2014, for a similar reasoning). This association is especially clear in the case of conventional sequences, such as numbers, weekdays, months, or the alphabet. This leads to the expectation that sentences and words can activate time to the degree that they bring time to mind. In other words, our suggestion is that a space–time congruency effect will not arise unless both dimensions are simultaneously activated in the working model of the task.

This account is consistent with the fact that some studies that assess the processing of units of highly practiced ordered sequences have found implicit effects (see Gevers et al., 2003, 2004, for weekday, months, and the alphabet; see also Dodd et al., 2008, for additional evidence). The extreme case is the counting sequence. Since first reported by Dehaene et al. (1993), it is now very well established that numbers are able to generate interactions with left and right keypresses in implicit tasks such as parity judgement (i.e., discriminating odd and even using left and right responses; see Fischer & Shaki, 2014, for a review). However, even in the case of numbers there is strong debate regarding whether the implicit effect vanishes when space is not relevant to the task in any way (Cleland et al., 2019; Colling et al., 2020; Miklashevsky et al., 2022; Pellegrino et al., 2019; Pinto et al., 2019, 2021; Salvaggio et al., 2022). Another clear case is provided by tasks in which participants are required to retrieve a just learned sequence of items in the same order and respond with left and right keypresses (Previtali et al., 2010; see Abrahamse et al., 2017, for a review). As long as both space and order (which is a central element of time) are kept simultaneously in the working model of the task, they can influence each other.

The study by Sell and Kaschak (2011) is revealing in the present context. They presented sentences describing two events in a sensicality task using front-back responses, and observed a significant congruency effect only when a) the response was a long movement of the hand, and b) the interval between the first and second events was one month, but not one day (see Scheifele et al., 2018, for a replication). That long movements are, all by themselves, key to observe the effect has been ruled out for sentences (Maienborn et al., 2015; Ulrich et al., 2012) and by present results for single words (Experiment 3). Maienborn et al. (2015) also ruled out the need to impose stronger language processing demands. In our view, this set of studies suggests that time is not easily activated from linguistic descriptions of events, unless a) there are two or more events to be ordered, b) their temporal sequence is emphasized (e.g., by having a long interval between them), and c) the spatial dimension is also emphasized (e.g., by using long movements). This is consistent with the eye tracking study by Stocker et al. (2016), who found evidence of spatial activation when participants listened to sequences of two events. Several aspects of the findings suggest that the effect was weak, plus there was no effect on lateral saccades. As also suggested by Stocker et al. (2016), it is possible that the paradigm did not get to activate time strongly enough. When Hartmann et al. (2014) asked participants to time travel to a year in the past or in the future, they found clear effects on both sagittal and lateral saccades.

Regarding single words, several studies have shown that temporal primes can influence subsequent spatial tasks when their temporal meaning must be held in mind during the trial (for different procedures to activate the temporal meaning of the primes see Ouellet et al., 2010, Rolke et al., 2013, exp. 4, Ding et al., 2020, exp. 2, or Weger and Pratt 2008, exp. 2). We believe that something similar may happen when prime temporal words are presented before targets without instructions to focus on their temporal meaning, as in Grasso, Ziegler, Mirault, et al. (2022, exps. 1 and 2), Grasso, Ziegler, Coull, et al. (2022), or Rolke et al. (2013, exp. 1). As the primes are fully visible, the participants may wonder what is their utility in the task, realize that they differ in temporal reference, and strategically decide to activate time. Participants may also strategically use any available cue (including temporal reference) when they are asked to "just choose any response" to temporal words (Ding et al., 2020; Kranjec & McDonough, 2011) or to carry out difficult perceptual judgements on the temporal words (such as word loudness; Lakens et al., 2011). If this is so, it is to be expected that these congruency effects should be difficult to replicate, but the lack of replications and the bias to publish only significant results may generate a distorted picture. Note that the use of this strategy is less likely when a single stimulus is presented to carry out an ostensively clear task with it (e.g., lexical decision), and this is why we have focused in the present study on such experimental paradigm.

In this context, it is relevant to mention the congruency effect that was observed by Carmona et al. (2024, exp. 1) when they asked participants to judge the temporal reference of words presented subliminally. Although the task is clearly time relevant, the subconcious processing of the words makes it qualify as a time irrelevant condition. Unfortunately, several concerns arise about the interpretation of the subliminal condition in Carmona et al. (2024, exp. 1). First, the congruency effect failed to interact with the visibility of the words (i.e., visible words produced a congruency effect of the same size). Second, it is difficult to understand on what basis the participants were taking their decision if they could not see the words. Subliminal words were responded to only 43 ms slower than unmasked words, what is not consistent with a guessing strategy. Accuracy data could be revealing, but they were not reported. Their raw data (which were asserted to be openly available at OSF) required permission for access. After several months, we have not yet received response to our request. Pending more information, we prefer to set this study momentarily aside in the debate.

It follows from our argument that there is no space–time congruency effect in fully implicit language processing tasks such as lexical decision or single sentence sensicality judgements because temporal reference is not automatically accessed from these materials, unless it is boosted somehow. Temporal reference is definitely not necessary to discriminate a word from a pseudoword (much less when both are inflected for tense), nor to judge whether a sentence is plausible or not.

To conclude, we here report two exploratory studies (using reaction time and mouse tracking) and one confirmatory study (to our knowledge, the first in this field) that show that space and time do not significantly interact in lexical decision, even when long sideways movements are used as response. There remains the possibility that there is a very small implicit effect in reaction times, but we can claim with 95 % confidence that such effect is smaller than 9 ms, a size we consider to be negligible. The implicit congruency effect might be found using measures even more sensitive than reaction time and mouse trajectories, but the evidence accrued so far is consistent with the claim that time is not activated automatically in implicit tasks with linguistic materials unless it is somehow boosted, and that space–time congruency effects arise only when both space and time are included in the working model of the task.

Transparency and openness

We disclose and publicly share all materials, programs, raw data, and analysis scripts, which can be accessed at Open Science Framework (https://osf.io/ep5nb). The hypotheses and analytical strategy of Experiment 3 were preregistered (https://aspredicted.org/6na98.pdf).

CRediT authorship contribution statement

Julio Santiago: Writing – review & editing, Writing – original draft, Supervision, Project administration, Methodology, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Alessia Beracci:** Writing – review & editing, Methodology, Investigation, Conceptualization. **Andrea Flumini:** Writing – review & editing, Methodology, Investigation, Conceptualization. **Eva Sanjuan:** Writing – review & editing, Investigation. **Marc Ouellet:** Writing – review & editing, Supervision, Funding acquisition, Conceptualization. **Pablo Solana:** Writing – review & editing, Supervision, Methodology, Investigation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was funded by the Consejería de Innovación, Ciencia y Empresa, Andalusian Government (ref. P09-SEJ-4772), and by the Ministerio de Economía y Competitividad, Spanish Government (ref. PSI2012-32464).

Data availability

All materials, programs, raw data, and analysis scripts can be accessed at Open Science Framework (https://osf.io/ep5nb).

References

Abbondanza, M., Amenta, S., Rinaldi, L., & Marelli, M. (2024). Time is -ending: Sublexical information activates the horizontal mental time line in word processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition.*. https://doi.org/10.1037/xlm0001331

Abrahamse, E. L., van Dijck, J.-P., & Fias, W. (2017). Grounding verbal working memory: The case of serial order. *Current Directions in Psychological Science, 26*(5), 429–433. https://doi.org/10.1177/0963721417704404

Aguirre, R., & Santiago, J. (2017). Do potential past and future events activate the Left-Right Mental Timeline? *Psicológica, 38*, 231–255.

Arend, M. G., & Schäfer, T. (2019). Statistical power in two-level models: A tutorial based on Monte Carlo simulation. *Psychological Methods, 24*(1), 1–19. https://doi.org/10.1037/met0000195

Bakker, M., van Dijk, A., & Wicherts, J. M. (2012). The rules of the game called psychological science. *Perspectives on Psychological Science, 7*(6), 543–554. https://doi.org/10.1177/1745691612459060

Bargh, J. a, Schwader, K. L., Hailey, S. E., Dyer, R. L., & Boothby, E. J. (2012). Automaticity in social-cognitive processes. *Trends in Cognitive Sciences*, *16*(12), 593-605. doi: 10.1016/j.tics.2012.10.002.

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious mixed models. *arXiv*, *arXiv:1506*. doi: arXiv:1506.04967.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Beracci, A., & Fabbri, M. (2022). Past on the ground floor and future in the attic: The vertical mental timeline. *Journal of Experimental Psychology: Human Perception and Performance, 48*(4), 380–399. https://doi.org/10.1037/xhp0000988

Beracci, A., Rescott, M. L., Natale, V., & Fabbri, M. (2022). The vertical space–time association. *Quarterly Journal of Experimental Psychology, 75*(9), 1674–1693. https://doi.org/10.1177/17470218211057031

Besner, D., Stolz, J. A., & Boutilier, C. (1997). The Stroop effect and the myth of automaticity. *Psychonomic Bulletin & Review, 4*(2), 221–225.

Bonato, M., Zorzi, M., & Umiltà, C. (2012). When time is space: Evidence for a mental time line. *Neuroscience and Biobehavioral Reviews, 36*(10), 2257–2273. https://doi.org/10.1016/j.neubiorev.2012.08.007

Boroditsky, L. (2000). Metaphoric structuring: Understanding time through spatial metaphors. *Cognition, 75*(1), 1–28.

Boroditsky, L., Fuhrman, O., & McCormick, K. (2011). Do English and Mandarin speakers think about time differently? *Cognition, 118*(1), 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.

Brysbaert, M., & Stevens, M. (2018). Power Analysis and Effect Size in Mixed Effects Models: A Tutorial. *Journal of Cognition, 1*(1), 9. https://doi.org/10.5334/joc.10

Carmona, I., Rodriguez-Rodriguez, J., Alvarez, D., & Noguera, C. (2024). Inhibition and working memory capacity modulate the mental space-time association. *Psychonomic Bulletin & Review, 31*, 2634–2645. https://doi.org/10.3758/s13423-024-02497-1

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

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). Academic Press.

Cleland, A. A., nueva, E. a sitio externo E. enlace se abrirá en una ventana, Bull, R., & nueva, E. a sitio externuno E. enlace se abrirá en una ventana. (2019). Automaticity of access to numerical magnitude and its spatial associations: The role of task and number representation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(2), 333-348. doi: 10.1037/xlm0000590.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* ((2nd ed.). L. Erlbaum Associates.

Colling, L. J., Szűcs, D., De Marco, D., Cipora, K., Ulrich, R., Nuerk, H.-C., Soltanlou, M., Bryce, D., Chen, S.-C., Schroeder, P. A., Henare, D. T., Chrystall, C. K., Corballis, P. M., Ansari, D., Goffin, C., Sokolowski, H. M., Hancock, P. J. B., Millen, A. E., Langton, S. R. H., & McShane, B. B. (2020). Registered Replication Report on Fischer, Castel, Dodd, and Pratt (2003). *Advances in Methods and Practices in Psychological Science, 3*(2), 143–162. https://doi.org/10.1177/2515245920903079

Dalmaso, M., Pileggi, S., & Vicovaro, M. (2023). Face Age is Mapped Into Three-Dimensional Space. *Cognitive Science, 47*(11), Article e13374. https://doi.org/10.1111/cogs.13374

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*(6), 326–332. https://doi.org/10.1027/1618-3169/a000342.

DeBruine, L. M., & Barr, D. J. (2021). Understanding Mixed-Effects Models Through Data Simulation. *Advances in Methods and Practices in Psychological Science, 4*(1), Article 2515245920965119. https://doi.org/10.1177/2515245920965119

Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of parity and number magnitude. *Journal of Experimental Psychology: General, 122*(3), 371–396. https://doi.org/10.1037/0096-3445.122.3.371

Dienes, Z. (2019). How Do I Know What My Theory Predicts? *Advances in Methods and Practices in Psychological Science, 2*(4), 364–377. https://doi.org/10.1177/2515245919876960

Ding, X., Feng, N., Cheng, X., Liu, H., & Fan, Z. (2015). Are past and future symmetric in mental time line? *Frontiers in Psychology, 6*(208). https://doi.org/10.3389/fpsyg.2015.00208.

Ding, X., Feng, N., He, T., Cheng, X., & Fan, Z. (2020). Can mental time lines co-exist in 3D space? *Acta Psychologica, 207*, Article 103084. https://doi.org/10.1016/j.actpsy.2020.103084

Dodd, M. D., Van der Stigchel, S., Adil Leghari, M., Fung, G., & Kingstone, A. (2008). Attentional SNARC: There's something special about numbers (let us count the ways). *Cognition, 108*(3), 810–818. https://doi.org/10.1016/j.cognition.2008.04.006

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

Fischer, M. H., Castel, A. D., Dodd, M. D., & Pratt, J. (2003). Perceiving numbers causes spatial shifts of attention. *Nature Neuroscience, 6*(6), 555–556. https://doi.org/10.1038/nn1066

Fischer, M. H., & Shaki, S. (2014). Spatial associations in numerical cognition—From single digits to arithmetic. *Quarterly journal of experimental psychology (2006)*, *67*(8), 1461-1483. doi: 10.1080/17470218.2014.927515.

Flumini, A., Barca, L., Borghi, A. M., & Pezzulo, G. (2015). How do you hold your mouse? Tracking the compatibility effect between hand posture and stimulus size. *Psychological Research, 79*, 928–938. https://doi.org/10.1007/s00426-014-0622-0

Flumini, A., & Santiago, J. (2013). Time (also) flies from left to right... If it is needed! En M. Knauff, M. Pauen, N. Sebanz, & I. Wachmuz (Eds.), *Proceedings of the 36th Annual Conference of the Cognitive Science Society* (pp. 2315-2320). Cognitive Science Society.

Fox, J., & Weisberg, S. (2019). *An R Companion to Applied Regression* ((3rd ed.).). SAGE Publications.
Freeman, J. B., & Ambady, N. (2009). Motions of the hand expose the partial and parallel activation of stereotypes. *Psychological Science, 20*(10), 1183–1188. https://doi.org/10.1111/j.1467-9280.2009.02422.x
Freeman, J. B., & Ambady, N. (2010). MouseTracker: Software for studying real-time mental processing using a computer mouse-tracking method. *Behavior Research Methods, 42*(1), 226–241. https://doi.org/10.3758/BRM.42.1.226
Fuhrman, O., & Boroditsky, L. (2010). Cross-cultural differences in mental representations of time: Evidence from an implicit nonlinguistic task. *Cognitive Science, 34*(8), 1430–1451. https://doi.org/10.1111/j.1551-6709.2010.01105.x
Gevers, W., Reynvoet, B., & Fias, W. (2003). The mental representation of ordinal sequences is spatially organized. *Cognition, 87*, 87–95. https://doi.org/10.1016/S0010-0277(02)00234-2
Gevers, W., Reynvoet, B., & Fias, W. (2004). The Mental Representation of Ordinal Sequences is Spatially Organised: Evidence from Days of the Week. *Cortex, 40*(1), 171–172. https://doi.org/10.1016/S0010-9452(08)70938-9
Grasso, C. L., Ziegler, J. C., Coull, J. T., & Montant, M. (2022). Space–time congruency effects using eye movements during processing of past- and future-related words. *Experimental Psychology, 69*(4), 210–217. https://doi.org/10.1027/1618-3169/a000559
Grasso, C. L., Ziegler, J. C., Mirault, J., Coull, J. T., & Montant, M. (2022). As time goes by: Space-time compatibility effects in word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 48*(2), 304–319. https://doi.org/10.1037/xlm0001007
Green, P., & MacLeod, C. J. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution, 7* (4), 493–498. https://doi.org/10.1111/2041-210X.12504
Harms, C., & Lakens, D. (2018). Making «null effects» informative: Statistical techniques and inferential frameworks. *Journal of Clinical and Translational Research, 3*(Suppl 2), 382–393.
Hartmann, M., Martarelli, C. S., Mast, F. W., & Stocker, K. (2014). Eye movements during mental time travel follow a diagonal line. *Consciousness and Cognition, 30*, 201–209. https://doi.org/10.1016/j.concog.2014.09.007
Haspelmath, M. (1997). *From space to time: Temporal adverbials in the world's languages*. Lincom Europa.
Heyes, C. (2011). Automatic imitation. *Psychological Bulletin, 137*(3), 463–483. https://doi.org/10.1037/a0022288
Ioannidis, J. P. A. (2005). Why Most Published Research Findings Are False. *PLOS Medicine, 2*(8), e124.
Ishihara, M., Keller, P. E., Rossetti, Y., & Prinz, W. (2008). Horizontal spatial representations of time: Evidence for the STEARC effect. *Cortex, 44*(4), 454–461. https://doi.org/10.1016/j.cortex.2007.08.010
Kong, F., & You, X. (2011). Space-time compatibility effects in the auditory modality. *Experimental Psychology, 59*(2), 82–87. https://doi.org/10.1027/1618-3169/a000129
Kranjec, A., & McDonough, L. (2011). The implicit and explicit embodiment of time. *Journal of Pragmatics, 43*(3), 735–748. https://doi.org/10.1016/j.pragma.2010.07.004
Kumle, L., Võ, M.-L.-H., & Draschkow, D. (2021). Estimating power in (generalized) linear mixed models: An open introduction and tutorial in R. *Behavior Research Methods, 53*(6), 2528–2543. https://doi.org/10.3758/s13428-021-01546-0
Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software, 82*(13). https://doi.org/10.18637/jss.v082.i13
Lakens, D. (2014). Performing high-powered studies efficiently with sequential analyses. *European Journal of Social Psychology, 44*, 701–710.
Lakens, D. (2022). *Improving Your Statistical Inferences*. *Zenodo*. https://doi.org/10.5281/ZENODO.6409077
Lakens, D., Scheel, A. M., & Isager, P. M. (2018). Equivalence Testing for Psychological Research: A Tutorial. *Advances in Methods and Practices in Psychological Science, 1*(2), 259–269. https://doi.org/10.1177/2515245918770963
Lakens, D., Semin, G. R., & Garrido, M. V. (2011). The sound of time: Cross-modal convergence in the spatial structuring of time. *Consciousness and Cognition, 20*(2), 437–443. https://doi.org/10.1016/j.concog.2010.09.020
Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. University of Chicago Press.
Lakoff, G., & Johnson, M. (1999). *Philosophy in the Flesh: The Embodied Mind and Its Challenge to Western Thought*. Basic Books.
Lebois, L. a M., Wilson-Mendenhall, C. D., & Barsalou, L. W. (2015). Are automatic conceptual cores the gold standard of semantic processing? The context-dependence of spatial meaning in grounded congruency effects. *Cognitive Science*, *39*, 1764-1801. doi: 10.1111/cogs.12174.
Li, Y., Casaponsa, A., Wu, Y. J., & Thierry, G. (2019). Back to the future? How Chinese-English bilinguals switch between front and back orientation for time. *NeuroImage, 203*, Article 116180. https://doi.org/10.1016/j.neuroimage.2019.116180
Li, Y., Oppenheim, G., & Thierry, G. (2023). Time flows vertically in Chinese. *Brain and Cognition, 170*, Article 106057. https://doi.org/10.1016/j.bandc.2023.106057
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*(3), 991–996. https://doi.org/10.1037/a0034989
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
Malyshevskaya, A., Fischer, M. H., Shtyrov, Y., & Myachykov, A. (2024). Horizontal mapping of time-related words in first and second language. *Scientific Reports, 14*(1), 9675. https://doi.org/10.1038/s41598-024-60062-1
Malyshevskaya, A., Miklashevsky, A., Fischer, M. H., Scheepers, C., Shtyrov, Y., & Myachykov, A. (2023). Keeping track of time: Horizontal spatial biases for hours, days, and months. *Memory & Cognition, 52*(894–908). https://doi.org/10.3758/s13421-023-01508-1
Mariconda, A., Prpic, V., Mingolo, S., Sors, F., Agostini, T., & Murgia, M. (2022). A systematic investigation reveals that Ishihara et al.'s (2008) STEARC effect only emerges when time is directly assessed. *Scientific Reports*, *12*(1), Article 1. doi: 10.1038/s41598-022-23411-6.
Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language, 94*, 305–315. https://doi.org/10.1016/j.jml.2017.01.001
McMahon, E., & Isik, L. (2023). Seeing social interactions. *Trends in Cognitive Sciences, 27* (12), 1165–1179. https://doi.org/10.1016/j.tics.2023.09.001
Miklashevsky, A., Fischer, M. H., & Lindemann, O. (2022). Spatial-numerical associations without a motor response? Grip force says 'Yes'. *Acta Psychologica, 231*, Article 103791. https://doi.org/10.1016/j.actpsy.2022.103791
Miles, L. K., Betka, E., Pendry, L. F., & Macrae, C. N. (2010). Mapping temporal constructs: Actions reveal that time is a place. *Quarterly Journal of Experimental Psychology, 63*(11), 2113–2119. https://doi.org/10.1080/17470218.2010.524932
Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., Percie, N., Simonsohn, U., & Wagenmakers, E. (2017). A manifesto for reproducible science. *Nature Human Behaviour, 1*(0021). https://doi.org/10.1038/s41562-016-0021
Nosek, B. A., Beck, E. D., Campbell, L., Flake, J. K., Hardwicke, T. E., Mellor, D. T., van Veer, A. E., & 't, & Vazire, S.. (2019). Preregistration Is Hard. *And Worthwhile. Trends in Cognitive Sciences, 23*(10), 815–818. https://doi.org/10.1016/j.tics.2019.07.009
Núñez, R., & Cooperrider, K. (2013). The tangle of space and time in human cognition. *Trends in Cognitive Sciences, 17*(5), 220–229. https://doi.org/10.1016/j.tics.2013.03.008
Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), aac4716-1-8. doi: 10.1126/science.aac4716.
Ouellet, M., Román, A., & Santiago, J. (2012). A multisensory interaction effect in the conceptual realm of time. *Experimental Psychology, 59*(4), 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*(1), 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*(4), 308–314. https://doi.org/10.1027/1618-3169/a000036
Pellegrino, M., Pinto, M., Marson, F., Lasaponara, S., Rossi-Arnaud, C., Cestari, V., & Doricchi, F. (2019). The Attentional-SNARC effect 16 years later: No automatic space–number association (taking into account finger counting style, imagery vividness, and learning style in 174 participants). *Experimental Brain Research, 237* (10), 2633–2643. https://doi.org/10.1007/s00221-019-05617-9
Pinto, M., Pellegrino, M., Marson, F., Lasaponara, S., Cestari, V., D'Onofrio, M., & Doricchi, F. (2021). How to trigger and keep stable directional Space–Number Associations (SNAs). *Cortex, 134*, 253–264. https://doi.org/10.1016/j.cortex.2020.10.020
Pinto, M., Pellegrino, M., Marson, F., Lasaponara, S., & Doricchi, F. (2019). Reconstructing the origins of the space-number association: Spatial and number-magnitude codes must be used jointly to elicit spatially organised mental number lines. *Cognition, 190*, 143–156. https://doi.org/10.1016/j.cognition.2019.04.032
Pocock, S. J. (1977). Group sequential methods in the design and analysis of clinical trials. *Biometrika, 64*(2), 191–199. https://doi.org/10.1093/biomet/64.2.191
Previtali, P., de Hevia, M. D., & Girelli, L. (2010). Placing order in space: The SNARC effect in serial learning. *Experimental Brain Research, 201*(3), 599–605. https://doi.org/10.1007/s00221-009-2063-3
Radden, G. (2004). The metaphor TIME AS SPACE across languages. En *Uebersetzen, Interkulturelle Kommunikation, Spracherwerb und Sprachvermittlung—Das Leben mit mehreren Sprachen: Festschrift fuer Juliane House zum 60. Geburtstag.* (pp. 225-238). AKS-Verlag.
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*(3), 231–244. https://doi.org/10.1007/s10339-013-0537-5
Román, A., El Fathi, A., & Santiago, J. (2013). Spatial biases in understanding descriptions of static scenes: The role of reading and writing direction. *Memory & Cognition, 41*, 588–599. https://doi.org/10.3758/s13421-012-0285-2
Román, A., Flumini, A., Lizano, P., Escobar, M., & Santiago, J. (2015). Reading and writing direction causes spatial biases in mental model construction in language understanding. *Scientific Reports, 5*(18248). https://doi.org/10.1038/srep18248
Román, A., Flumini, A., & Santiago, J. (2018). Scanning of speechless comics changes spatial biases in mental model construction. *Philosophical Transactions of the Royal Society B, Biological. Sciences, 373*(20170130.). https://doi.org/10.1098/rstb.2017.0130
Ruz, M., & Lupiáñez, J. (2002). A review of attentional capture: On its automaticity and sensitivity to endogenous control. *Psicológica, 23*, 283–309.
Salvaggio, S., Andres, M., Zénon, A., & Masson, N. (2022). Pupil size variations reveal covert shifts of attention induced by numbers. *Psychonomic Bulletin & Review, 29*(5), 1844–1853. https://doi.org/10.3758/s13423-022-02094-0
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., Ouellet, M., Román, A., & Valenzuela, J. (2012). Attentional factors in conceptual congruency. *Cognitive Science, 36*(6), 1051–1077. https://doi.org/10.1111/j.1551-6709.2012.01240.x

Santiago, J., Román, A., & Ouellet, M. (2011). Flexible foundations of abstract thought: A review and a theory. En A. Maass & T. W. Schubert (Eds.), *Spatial dimensions of social thought* (pp. 41-110). Mouton de Gruyter.

Santiago, J., Román, A., Ouellet, M., Rodríguez, N., & Pérez-Azor, P. (2010). In hindsight, life flows from left to right. *Psychological Research, 74*(1), 59–70. https://doi.org/10.1007/s00426-008-0220-0

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 Quantitative Methods for Psychology, 14*(1), r8–r11. https://doi.org/10.20982/tqmp.14.1.r008

Scozia, G., Pinto, M., Pellegrino, M., Lozito, S., Pia, L., Lasaponara, S., & Doricchi, F. (2023). How time gets spatial: Factors determining the stability and instability of the mental time line. *Attention, Perception, & Psychophysics.*. https://doi.org/10.3758/s13414-023-02746-w

Sell, A. J., & Kaschak, M. P. (2011). Processing time shifts affects the execution of motor responses. *Brain and Language, 117*(1), 39–44. https://doi.org/10.1016/j.bandl.2010.07.003

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science, 22*(11), 1359–1366. https://doi.org/10.1177/0956797611417632

Stocker, K., Hartmann, M., Martarelli, C. S., & Mast, F. W. (2016). Eye movements reveal mental looking through time. *Cognitive Science, 40*, 1648–1670. https://doi.org/10.1111/cogs.12301

Torralbo, A., Santiago, J., & Lupiáñez, J. (2006). Flexible conceptual projection of time onto spatial frames of reference. *Cognitive Science, 30*(4), 745–757. https://doi.org/10.1207/s15516709cog0000_67

Ulrich, R., Eikmeier, V., de la Vega, I., Ruiz Fernández, S., Alex-Ruf, S., & Maienborn, C. (2012). With the past behind and the future ahead: Back-to-front representation of past and future sentences. *Memory & Cognition, 40*(3), 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*(2), 126–138. https://doi.org/10.1016/j.cognition.2010.08.001

von Sobbe, L., Scheifele, E., Maienborn, C., & Ulrich, R. (2019). The Space–Time Congruency Effect: A Meta-Analysis. *Cognitive Science, 43*(1), Article e12709. https://doi.org/10.1111/cogs.12709

Wagenmakers, E.-J., Wetzels, R., Borsboom, D., van der Maas, H. L. J., & Kievit, R. A. (2012). An Agenda for Purely Confirmatory Research. *Perspectives on Psychological Science, 7*(6), 632–638. https://doi.org/10.1177/1745691612463078

Wald, A. (1947). *Sequential analysis. Courier Corporation.*

Weger, U. W., & Pratt, J. (2008). Time flies like an arrow: Space-time compatibility effects suggest the use of a mental timeline. *Psychonomic Bulletin & Review, 15*(2), 426–430.

Westfall, J., Kenny, D. a, & Judd, C. M. (2014). Statistical power and optimal design in experiments in which samples of participants respond to samples of stimuli. *Journal of Experimental Psychology: General*, *143*(5), 2020-2045. doi: 10.1037/xge0000014.