

Casasanto, D., Fotakopoulou, O., Lucero, C., Pita, R. & Boroditsky, L. (2025). The long and small of it: Language shapes duration estimation in speakers of English and Greek.

Language & Cognition. DOI: 10.1017/langcog.2025.10055

**The long and small of it:
Language shapes duration estimation in speakers of English and Greek**

Daniel Casasanto¹, Olga Fotakopoulou², Ché Lucero¹, Ria Pita³, Lera Boroditsky⁴

¹Cornell University
Department of Psychology
Martha Van Rensselaer Hall
Ithaca, NY 14853 USA

²Birmingham City University
Psychology Department
Curzon Building, 4 Cardigan Street
Birmingham, B4 7BD, UK

³Aristotle University of Thessaloniki
School of Psychology
Administration Building K. Karatheodori
54124 Thessaloniki, GR

⁴University of California at San Diego
Department of Cognitive Science
9500 Gilman Drive, CSB 166
Mail Code: 0515
La Jolla, CA, 92093, USA

Please address correspondence to:

Daniel Casasanto
Cornell University
Department of Psychology
Martha Van Rensselaer Hall
Ithaca, NY 14853 USA
casasanto@alum.mit.edu

Abstract

Does the way people talk about time affect how they think about it? Whereas English speakers describe the duration of events most often in terms of spatial length (e.g., a *long* night), Greek speakers tend to talk about duration in terms of multidimensional spatial size (e.g., *mia megali nychta*, tr. a *big* night) or amount (e.g., *poli ora*, tr. *much* time). After quantifying these linguistic patterns, we gave non-linguistic tests of duration estimation to English and Greek speakers. English speakers' estimates were influenced more strongly by irrelevant length information, and Greek speakers' by irrelevant amount information, consistent with verbal metaphors for duration in English and Greek. Next, we tested duration estimation with concurrent verbal interference, to confirm that the observed effects did not depend on participants verbally labelling the stimuli during the task. Finally, we trained English speakers to use Greek-like metaphors for duration, which resulted in Greek-like performance on a non-linguistic duration estimation task. Results show that (a) people who talk about time differently also think about it differently, (b) these effects are not due to participants' using verbal labels during the task, (c) language can play a causal role in shaping even basic non-linguistic mental representations of time.

Keywords: Linguistic relativity; Greek; Metaphor; Space; Time; Whorfian hypothesis

1. Introduction

When people talk about time, they often use spatial words. Vacations can be “long” or “short”; meetings can be “moved forward” or “pushed back”; our imaginations can transport us to the “remote” past or the “distant” future. Such spatial metaphors for time have been analyzed extensively, within and across languages (e.g., Alverson, 1994; Evans, 2004; Haspelmath, 1997; Lakoff & Johnson, 1980; 1999; Moore, 2006; Traugott, 1978).

English speakers tend to use 1-dimensional terms to talk about temporal concepts like succession and duration: a tendency thought to reflect the linear nature of time, itself (Clark, 1973). However, this tendency does not appear to be universal. In this paper we explore alternative ways of talking about temporal duration in terms of size or amount, and test whether prior experience using different linguistic metaphors for duration can cause speakers of different languages to conceptualize time differently, even when they are not using language in the moment.

1.1 Talking about duration in terms of 1-dimensional or multidimensional space

In many contexts, English speakers can hardly avoid using 1-dimensional spatial terms to talk about duration. Try replacing the word “long” in the sentence, “Let’s take a *long* vacation,” and communicating the same message without using any word related to spatial length. What are the possibilities: Lengthy? Extended? Protracted? Drawn out? All have linear spatial meanings, presently or historically. It is difficult to fill in the blank with a purely temporal term (e.g., we cannot say “a *duratious* vacation”). Length-related words are the norm, and avoiding words with 1-dimensional spatial meanings or roots requires circumlocution – in English.

Yet, not all languages describe the durations of events using length terms like “long” or “short” by default. Greek speakers, for example, often talk about duration using multidimensional spatial terms like “big” (*megalos*) [μεγάλος] and “small” (*mikros*) [μικρός]. Rather than saying “a *long* night,” in Greek it would be natural to say “a *big* night” (*mia megali nychta*) [μία μεγάλη νύχτα]. As in English, “a big night” could mean an exciting night or an important night, but in Greek this phrase also naturally describes the night’s duration. Greek has spatial words for “long” (*makris*) [μακρύς] and “short” (*kontos*) [κοντός] that function much like their English translation equivalents in spatial contexts: “ena *makry* skoini” [ένα μακρύ σκοινί] means “a *long* rope.” But these 1-dimensional spatial terms are not extended to express time metaphorically in Greek as commonly as they are in English. It would be unnatural to translate “a *long* meeting” literally as “*mia makria synantisi*” [μία μακριά συνάντηση]. Compare how English and Greek speakers typically describe the durations of the following events (Table 1).

Table 1. English (e) and Greek (g) expressions for event durations (literal translations in parentheses).

1e. <i>long</i> night	1g. <i>megali nychta</i> [μεγάλη νύχτα] (big night)
2e. long relationship	2g. <i>megali schesi</i> [μεγάλη σχέση] (big relationship)
3e. long party	3g. <i>parti pou kratise poli</i> [πάρτυ που κράτησε πολύ] (party that lasts much)
4e. long meeting	4g. <i>synantisi pou dierkese poli</i> [συνάντηση που διήρκησε πολύ] (meeting that lasts much)

In examples 1g. and 2g., the “source domain” (Lakoff & Johnson, 1980) of spatial size provides a metaphorical basis for the “target domain” of temporal duration. The same adjective “megalos” [μεγάλος] can be used to describe the size of a big building or a big

pile of sand, as well as a long-lasting event. Although size words can sometimes indicate 2-dimensional spatial magnitude (e.g., a *big* piece of paper), more prototypically they describe objects that are big or small in 3-dimensional space (e.g., a *big* dog; a *small* car).

Examples 3g. and 4g. use a related source-domain word, “*poli*” [πολύ], which means “much.” Thus, in 3g. and 4g. duration is described in terms of an *amount* of time. Whereas time is abstract (i.e., invisible, intangible), and amounts of time do not occupy any space at all, the word “*poli*” [πολύ] is also used to describe amounts of concrete substances that occupy 3-dimensional space, and occupy more space as they accumulate: “*poli nero*” [πολύ νερό] means “*much* water”; “*perissotero nero*” [περισσότερο νερό] means “*more* water.” Analogously, “*poli ora*” [πολλή ώρα] means “*much* time”; “*perissotero chrono*” [περισσότερο χρόνο] means “*more* time.”

Greek is not the only language that uses size or amount metaphors for duration. In many dialects of Spanish the most natural way to say “a long time” is “*mucho* tiempo” (tr. *much* time), not “*largo* tiempo” (tr. *long* time). But there may be no need to look any farther than English to find duration described in multidimensional spatial terms. It is difficult to produce a clear example of the size metaphor for duration in English: a “big” time does not mean a long time; a “large” meeting does not mean a long meeting. However, there are many common amount metaphors: “We’ll need *a lot* of time”; “We haven’t got *much* time”; “that’s the perfect *amount* of time.” If “amount” is mentally represented as an amount of a substance (as opposed to a purely abstract quantity), this would suggest that English speakers *can* talk about duration in terms of multidimensional space.

Conversely, Greek speakers *can* talk about duration in terms of 1-dimensional spatial extent, even though these constructions are not as compact or productive as those

using size or amount metaphors. For example, the closest Greek translation equivalent of “long time” is “*makry chroniko diastima*” [μακρύ χρονικό διάστημα], which literally means “long time period.” Cross-linguistic differences in speakers’ reliance on 1-dimensional versus multidimensional spatial metaphors for duration, therefore, are not absolute but rather a matter of degree: quantitative rather than qualitative differences.

1.1.1 Quantifying length vs. size/amount duration metaphors across languages

A preliminary study compared the use of length and amount metaphors for duration across languages (Casasanto, et al., 2004). The most natural phrases expressing the ideas “long time” and “much time” were elicited from native speakers of English, Greek, Spanish, and Indonesian, and their frequencies were compared in a very large multilingual text corpus: www.google.com. Google’s language tools were used to find exact matches for each expression, and to restrict the search to web pages written only in the appropriate languages. Results showed that in English and Indonesian, length metaphors were dramatically more frequent than amount metaphors, whereas the opposite pattern was found in Greek and Spanish. Although all four languages can use *both* length and amount metaphors for duration, their relative frequencies were reversed across languages.

1.2 Thinking about duration in terms of 1-dimensional or multidimensional space

Do these differences in duration metaphors across languages correspond to differences in how their speakers think about time? Many behavioral studies have demonstrated cross-linguistic or cross-cultural differences in spatial representations of time, which researchers predicted on the basis of differences in space-time metaphors in

language, or in the spatialization of time in culture-specific artifacts or practices (e.g., Boroditsky, 2001; Boroditsky, Fuhrman, & McCormick, 2011; Boroditsky & Gaby; 2010; Casasanto & Bottini, 2014; de la Fuente, Santiago, Román, Dumitrache, & Casasanto, 2014; Fedden & Boroditsky, 2012; Fuhrman & Boroditsky, 2010; Fuhrman, et al., 2011; Miles, Tan, Noble, Lumsden, & Macrae, 2011; Lai & Boroditsky, 2013; Nuñez & Sweetser, 2006; Núñez, Cooperrider, Doan, & Wassmann, 2012; Ouellet, Santiago, Israeli, & Gabay, 2010; Pitt et al., 2021; Tversky, Kugelmass, & Winter, 1991). These studies of cross-linguistic or cross-cultural differences in temporal thinking have focused on higher-level temporal concepts that may be uniquely human, and which are likely to be constructed via language and cultural practices (Sinha, Sinha, Zinken, & Sampaio, 2011): concepts like past and future (Miles, et al., 2011), breakfast, lunch, and dinner (Tversky et al., 1991), Nixon's presidency (Boroditsky, 2001), and Lady Gaga's rise to fame (Casasanto & Bottini, 2014).

Here we focus on an aspect of temporal thinking that is more basic inasmuch it develops in human infants and nonhuman animals with no knowledge of language or cultural conventions. The capacity to discriminate the durations of brief non-symbolic stimuli has been documented in rats (Meck & Church, 1983), monkeys (Merritt, Casasanto, & Brannon, 2010), and human infants as young as 0-3 days old (de Hevia, Izard, Coubart, Spelke, & Streri, 2014). Yet, even though infants and nonhumans can represent brief approximate durations independent of language, it is possible that experience with language could influence this pre-existing ability. Suppose each time people use a space-time metaphor in language they activate a corresponding association between non-linguistic representations of space and time in memory. By using expressions like "a long time" frequently, English speakers may strengthen an implicit

association between duration and length and, as a consequence, weaken the competing association between duration and size / amount (Casasanto, 2017; Song, 2000). By using expressions like “megali nychta” [μεγάλη νύχτα] (*tr.* big night) frequently, Greek speakers may strengthen an implicit association between duration and size / amount, weakening the association between duration and length. If so, people who talk about duration using different space-time metaphors should think about it differently, as a consequence (for further discussion of this proposed process see Section 6.1).

To test this proposal, Casasanto et al. (2004) compared duration representations across speakers of different languages, adapting a task that Casasanto and Boroditsky (2003; 2008) designed to test whether English speakers think about duration using representations of spatial length. In the canonical version of the task, lines of different spatial lengths and durations “grew” gradually across a computer screen and disappeared when they had reached their maximum extent in both space and time. Participants were asked to reproduce either the spatial or temporal extent of each stimulus by clicking the mouse to indicate the beginning and end points of either the spatial or temporal interval. Results showed that when participants were instructed to reproduce spatial length they could effectively ignore duration. By contrast, participants were unable to ignore the spatial dimension of the stimuli when reproducing the temporal dimension: For lines of the same average duration, those that extended longer in space were judged to take a longer time, and those that extended a shorter length in space were judged to take a shorter time. English speakers incorporated task-irrelevant spatial information into their temporal representations, more than vice versa, using mental representations of 1-dimensional spatial extent to think about duration just as they tend to use 1-dimensional spatial words to talk about duration.

Is thinking about duration in terms of spatial length a human universal, or do non-linguistic mental representations of duration depend in part on people's experience of talking about time in terms of length vs. size or amount? To find out, Casasanto et al. (2004) tested speakers of four languages on a pair of non-linguistic time reproduction tasks. In the "length interference" task participants saw lines growing across the screen and reproduced their durations while ignoring how far they extended in space, as in Casasanto and Boroditsky's (2003; 2008) studies described above. The "amount interference" task was constructed analogously. Participants saw schematically drawn containers gradually filling up with water, and reproduced the amount of time that each container remained on the screen while ignoring its fullness, responding with mouse clicks. English and Indonesian speakers' duration estimates were more strongly affected by irrelevant length information than by irrelevant amount information; conversely, Greek and Spanish speakers' duration estimates were more strongly affected by amount than by length, consistent with their preferred duration metaphors in language. This relationship between temporal language and temporal thinking was subsequently extended to an additional pair of languages, Swedish (a time-as-length language) and Spanish (a time-as-amount language; Bylund & Athanasopoulos, 2017).

1.3 The present study

The goal of the present study was to build on the results of Casasanto et al.'s (2004) results in English and Greek speakers, in several ways. First, although the corpus data from Casasanto et al. (2004) validated native speakers' intuitions about length and size/amount metaphors for time, this preliminary study did not capture the complexities of how duration is metaphorized in terms of length vs. size or amount, within or between

languages. Experiment 1 here provides a further test of cross-linguistic differences in the use of length, size, and amount metaphors for duration in English and Greek. Experiment 2 tested the effects of length and amount interference on duration estimation in larger samples of English and Greek speakers. Experiment 3 confirmed that the patterns of space-time interference found in the duration estimation tasks are not mediated by the use of language online during the task. Experiment 4 went beyond demonstrating correlations between temporal language and thought and tested whether using linguistic metaphors can play a causal role in shaping non-linguistic representations of duration.

2. Experiment 1: Describing event durations in English and Greek

The corpus search reported in Casasanto et al. (2004) measured how often English and Greek speakers use length vs. amount metaphors to talk about *time*, per se (e.g., long time, much time). Here we conducted a questionnaire study to investigate cross-linguistic differences in the use of length vs. size or amount metaphors to describe the durations of *events*. Participants were presented with pairs of stimuli (tones or dots) with unequal durations and asked to describe the difference in duration between the members of each pair.

2.1 Method

2.1.1 Participants A total of 85 participants (53 English speakers, 32 Greek speakers) completed the experiment online, in exchange for payment. Native English speakers were recruited via Amazon Mechanical Turk, and native Greek speakers were recruited through the Aristotle University of Thessaloniki. We intended to recruit 50 participants in each language group; the sample size of 53 English speakers reflects a slight overshoot in

recruiting. The sample size of 32 Greek speakers reflects the number of participants who volunteered during the semester in which the study was run. To avoid differences between different dialects of English spoken around the world, the sample of native English speakers was restricted to people who were born in the USA and were living in the USA at the time of testing. Native Greek speakers were recruited from the Aristotle University of Thessaloniki community. Participants were considered native speakers of a language if it was the only language they learned before age 5 and it was their strongest language at the time of testing, according to a language background questionnaire.

2.1.2 Materials and procedure

Each participant compared the durations of four pairs of stimuli: two pairs of serially-presented tones, followed by two pairs of serially-presented dots that appeared and disappeared in the center of the screen. The two tones/dots in each pair were identical to one another except for their durations. For one pair of tones/dots the first stimulus lasted 2000 ms. and the second lasted 4000 ms. For the other pair of tones/dots the first stimulus lasted 4000 ms. and the second lasted 2000 ms., with the order counterbalanced across participants.

The instructions told participants that the experimenters were “interested in how people use language to compare the everyday things that they hear and see,” and that “the questions may seem very simple.” Participants were asked to “try to answer them naturally, even if they seem easy,” and told that, “there's no need to over-think

your answers.” Participants were then led through a test of their computer’s speakers to make sure they would hear the tone stimuli clearly.

Before each pair of tone/dot stimuli was presented, participants were asked to “pay attention to the time each tone [or dot] takes.” After each pair, participants answered two questions about the relative durations of the stimuli. The first question asked them to type a free response to the prompt: “How was the first tone [or dot] different from the second?” The second question asked participants to: “Fill in the blank with one word that most naturally describes the difference between the tones [or dots]: The first tone [or dot] was _____ than the second.”

Each participant made a total of 8 duration comparison responses on the 4 pairs of stimuli: 4 Free Responses and 4 Fill-in-the-Blank responses. English instructions and response prompts were translated into Greek by a native Greek speaker. All instructions, in both languages, avoided using any length, size, or amount metaphors for time. Stimuli were presented and responses collected using Millisecond software. Data and analysis scripts for all studies can be found in an Open Science Framework archive: <https://osf.io/3ECNP/>

2.2 Results and Discussion

English and Greek responses were coded by a native speaker of each language. Each response was coded as describing the relative durations of the stimuli in terms of either (a) length, (b) size, (c) amount, or (d) other. The “other” category included both non-

spatial descriptions of duration and erroneous responses that described non-temporal aspects of the stimuli.

2.2.1 Descriptive analyses

2.2.1.1 English responses Length metaphors (e.g., “longer”, “shorter”) were produced most frequently by English speakers, and accounted for 85% of their responses (Fig. 1, left). By contrast, size and amount metaphors, combined, accounted for only about 1% of English speakers’ responses. The amount responses used conventional metaphors (e.g., “less time”). Only one size response was given, and this appeared to be an error: the participant responded that the first dot was “smaller” than the second, presumably because the participant was mistakenly reporting that the dots differed in their spatial size – not their duration. To be maximally conservative, however, we coded this response as a size metaphor for duration, since we cannot be certain that it was *not* intended to describe duration, and since this coding worked against our predictions for English speakers. Given that size and amount responses were so infrequent in English speakers, overall (a total of 5 responses out of 424), these responses were combined into the category of “Size / Amount” responses in Figure 1, and in subsequent inferential analyses (section 2.2.2).

“Other” responses accounted for about 13% of total responses. The most frequent non-spatial responses were speed metaphors for duration (e.g., one stimulus was “faster” than the other), accounting for about 7% of the total responses. About 1% of total responses used purely temporal expressions (e.g., one stimulus was “briefer” than the other). The remaining responses (about 5%) were erroneous inasmuch as they were irrelevant to duration (e.g., indicating that one tone was “softer” or “mellower” than the other, even though objectively the stimuli only differed in duration).

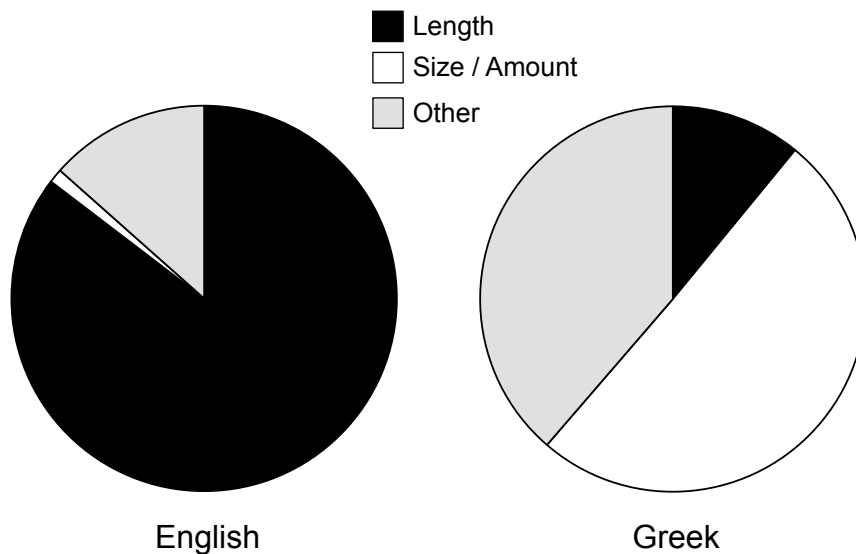


Figure 1. Results of Experiment 1. Proportion of responses using length metaphors (black), size / amount metaphors (white), or other duration expressions (grey), in English speakers (left) and Greek speakers (right).

2.2.1.2 Greek responses Size and amount metaphors, combined, were most the frequent in Greek speakers, accounting for 50% of their responses (Fig. 1, right). Size metaphors (e.g., “megaliteros” [μεγαλύτερος], tr. bigger; “mikróteros” [μικρότερος], tr. smaller) accounted for 29% of the total responses. Amount metaphors (e.g., “perissotero” [περισσότερο], tr. more; “ligotero” [λιγότερο], tr. less) accounted for about 22% of the total responses. By contrast, length metaphors (e.g., “makríteros” [μακρύτερος], tr. longer; “pio kontós” [πιο κοντός], tr. shorter) accounted for only 11% of the total responses.

“Other” responses accounted for 39% of total responses. The most frequent non-spatial responses used purely temporal words (e.g., “diarkésteros” [διαρκέστερος], tr. enduring) and accounted for 31% of the total responses – notably more purely temporal

responses than English speakers produced. This difference may reflect the relatively low frequency of purely temporal words in English (e.g., “enduring” and “brief” are far less frequent than their spatio-temporal alternatives like “long” and “short”). Speed metaphors (e.g., “pio aryí” [πιο αργή], tr. slower) accounted for only 1% of the total responses. The remaining 6% of responses were irrelevant to duration (e.g., “melodikós” [μελωδικός], tr. melodic).

2.2.2 Inferential analyses

The rates of length and size/amount responses were compared between the English and Greek groups using a repeated-measures binary logistic regression. Overall, language (English vs. Greek) was a significant predictor of participants’ use of length vs. size/amount words to compare the durations of the stimuli ($\chi^2=108.38$, $df=1$, $p=.00001$).¹ This difference was significant both in the Free Response condition, alone ($\chi^2=91.01$, $df=1$, $p=.00001$), and in the Fill-in-the-blank condition, alone ($\chi^2=41.54$, $df=1$, $p=.00001$).

2.2.3 Summary of linguistic elicitation data

These elicitation results show how English and Greek speakers tend to describe the relative durations of brief events. Participants were free to describe the durations of the stimuli any way they chose (especially during the first, unconstrained “free response” that they gave on each trial). The descriptions they chose included overwhelmingly more length metaphors (compared to size/amount metaphors) in English, and more size/amount

¹ Throughout the paper, “p=.00001” reflects a smaller p-value rounded to .00001.

metaphors (compared to length metaphors) in Greek. One caveat is that these data reflect language use in an experimental context, so they may not reflect the distribution of length vs. size/amount words across all contexts. These results on event descriptions corroborate the results of Casasanto et al.'s (2004) corpus search on descriptions of time, *per se*. Together, these convergent results establish a previously unexplored cross-linguistic difference in spatial metaphors for time: English speakers strongly prefer 1-dimensional length metaphors to describe duration, whereas Greek speakers prefer multidimensional size/amount metaphors.

Whereas size metaphors for duration accounted for almost one third of Greek speakers' responses, English speakers produced only a single (potential) size metaphor out of 424 responses, and this one response appears to have been given in error, referring to the literal spatial sizes of the stimuli rather than their durations. These results confirm native speakers' intuitions that size metaphors for duration are common in Greek but not in English.

Amount metaphors for duration were also strikingly more common in Greek speakers' responses than in English speakers'. Yet, they were not absent from English speakers' responses, and it would have been sensible for English speakers to use a greater number and variety of amount metaphors (e.g., compared to the first stimulus, the second lasted "*more* time"; "*less* time"; etc.) Amount metaphors for duration are sensible and well formed in English, even though length metaphors appear to be used by default. Conversely, length metaphors for duration are sensible and well formed in Greek, even though size/amount metaphors appear to be used by default. The double dissociation we find in the use of length vs. size/amount metaphors across languages provides a

quantitative basis for our predictions about English and Greek speakers' non-linguistic mental representations of duration.

3. Experiment 2: Do people think about duration like they talk about it?

Do differences in the use of length vs. size/amount metaphors correspond to differences in the way English and Greek speakers use space to *think* about time – even when they're not speaking or understanding language in the moment? To find out, in Experiment 2 we asked English and Greek speakers to estimate the duration of events that contained either distracting information about linear spatial extent (length interference) or about amount (amount interference). These tasks used non-linguistic stimuli and responses.

Furthermore, they were designed so that labeling the temporal content of the stimuli using spatial words would not produce the predicted results, because the task-relevant (temporal) and task-irrelevant (spatial) dimensions of the stimuli varied orthogonally (see Section 4).

If people's use of linguistic metaphors for time affects their subsequent temporal thinking, then task-irrelevant length and amount information should interfere with English and Greek participants' duration estimates differently; the prevalence of length vs. amount metaphors in these languages should correlate with the strength of length vs. amount interference on duration estimation. Alternatively, if non-linguistic mental representations of duration are universal, and are unaffected by cross-linguistic differences in space-time metaphors, then there should be no systematic differences in the effects of length and amount information on participants' duration estimates across language groups.

3.1 Method

3.1.1 Participants

A total of 145 adults participated in Experiment 2, for payment. 98 participants were native English speakers (tested at the Massachusetts Institute of Technology or Stanford University in the USA), and 47 were native Greek speakers (tested at the Aristotle University of Thessaloniki in Greece), according to their responses on a language background questionnaire. Forty-eight participants performed the length interference task (25 English, 23 Greek) and 97 performed the amount interference task (73 English, 24 Greek). We intended to recruit 20 participants for each task from each language group; sample sizes slightly greater than 20 reflect a small overshoot in recruiting. The initial sample included only 20 English speaking participants in the amount interference task. Participants were added ($n=53$) to this sample to address the concern that the relative weakness of the effect of amount on time estimation in this data set could be due to a lack of subject-wise power. The pattern of results was essentially unchanged by the addition of these participants, and the initial sample ($n=20$) and additional sample ($n=53$) were combined for all analyses reported here. Data from 72 of the 145 participants in Experiment 2 were reported previously in a conference proceedings paper (Casasanto et al., 2004) which was written while data collection was still in progress; we note that the effects reported here do not provide an independent replication of those findings. A total of 4 participants (2.8%) were removed from the analyses reported below for performing the task incorrectly: 1 participant (a Greek speaker) was removed from the length interference task, and 3 participants (English speakers) were removed from the amount interference task. For these participants, the slope of the relationship between the actual duration of a stimulus and the estimated duration was negative, indicating a complete failure to reproduce the stimulus durations as instructed. The measure by which

participants were excluded (i.e., the within-domain effect of actual time on estimated time) was orthogonal to our dependent measure of interest (i.e., the between-domain effect of actual space on estimated time).

3.1.2 Materials

For the length interference task, participants were shown lines growing across a computer screen (resolution: 1024 x 768 pixels). The growing line events varied in length and duration. Durations ranged from 1000 ms. to 5000 ms. in 500 ms. increments. Lengths ranged from 100 to 500 pixels in 50 pixel increments. Nine durations were fully crossed with nine lengths to produce 81 distinct line types. Lines “grew” horizontally across the screen one pixel at a time, from left to right, along the midline. Lines were situated in a square box (700 x 700 pixels), to minimize any influence that the asymmetric rectangular shape of the computer monitor might have on perception of horizontally growing lines vs. vertically filling containers. The starting point of each line was jittered with respect to the average starting point (+/- up to 50 pixels), so that the box would not provide a reliable spatial frame of reference. Each line remained on the screen until it reached its designated length, and then it disappeared.

The amount interference stimuli were constructed analogously. Participants watched schematically drawn containers filling on the computer screen. Containers were 600 pixels high x 500 pixels wide. Nine durations were fully crossed with nine fill levels to produce 81 distinct trial types. Durations ranged from 1000 ms. to 5000 ms. in 500 ms. increments. Fill levels ranged from 100 to 500 pixels, in 50 pixel increments. Empty containers filled gradually, one row of pixels at a time, for varying durations and fill

levels, and disappeared when they reached their designated fullness. The stimulus presentation and response collection script was programmed in REALBasic.

3.1.3 Procedure

Each participant performed either the length interference or amount interference task. Instructions were presented on the screen before each task in the participant's native language. No length or amount metaphors for time were used in the instructions. The tasks, themselves, were entirely non-linguistic, consisting of growing lines or filling containers (stimuli) and mouse clicks (responses).

For the length interference task, participants viewed all 162 line stimuli, one at a time in random order, from a viewing distance of about 50 cm. For each trial, participants reproduced either the length or the duration of the stimulus, never both. Length-reproduction and duration-reproduction trials were randomly intermixed. Before each stimulus, an icon appeared for 2 seconds in the center of the screen alerting participants that they would need to reproduce either the length that the line traveled (if an "X" icon appeared) or the duration for which it remained on the screen (if an "hourglass" icon appeared). Immediately after each stimulus was shown, the same icon appeared as a response prompt. Whereas stimuli grew from a jittered starting point on the vertical midline of the screen, responses were initiated at a fixed starting point in either the upper or lower left corner of the box (counterbalanced across participants). Thus, the response was translated both vertically and horizontally with respect to the stimulus; this spatial displacement ensured that remembering the final point to which a line grew would not be an effective strategy for reproducing its length.

To estimate length, participants clicked the mouse once on the center of the X icon, moved the mouse to the right in a straight line, and clicked the mouse a second time to indicate that they had moved a length equal to that of the stimulus. To estimate duration, participants clicked the mouse once on the center of the hourglass icon, waited the appropriate amount of time, and clicked again in the same spot.

The amount interference task was conducted analogously. Before each of the 162 stimuli, an icon appeared for 2 seconds in the center of the screen alerting participants that they would need to reproduce either the fullness of the container (if an “X” icon appeared) or the duration for which it remained on the screen (if an “hourglass” icon appeared).

Immediately after each stimulus was shown, the same icon appeared as a response prompt, level with the bottom of the container on either the left or right side (counterbalanced across participants). During the encoding phase the containers appeared in the center of the screen. During the reproduction phase the container was displaced to a jittered location at the bottom left of the screen. To estimate fullness, participants clicked the mouse once on the center of the X icon, moved the mouse up the side of the container in a straight line, and clicked the mouse a second time to indicate that they had moved a length equal to the fill level that had been reached. To estimate duration, participants clicked the mouse once on the center of the hourglass icon, waited the appropriate amount of time, and clicked again in the same spot, exactly as in the growing line task.

For both the length and amount tasks, all responses were self-paced. Response data were collected for both the trial-relevant and the trial-irrelevant stimulus dimensions, to ensure that participants were following instructions and only reproducing the relevant dimension of the stimulus for each trial. Each task lasted 30-40 minutes.

3.2 Results

To avoid experimental artifacts, the two highest and lowest durations and lengths/amounts were excluded from the analyses, leaving the middle five durations and lengths/amounts to be analyzed. Trimming the endpoints of perceptual continua is common practice in magnitude estimation tasks (Cordes, et al., 2007), as endpoints are susceptible to strategic responding. This decision was made prior to data analysis, on the basis of data from prior experiments using similar methods, and the same exclusion criteria were applied to all data sets from the length and amount interference tasks (i.e., 7 independent data sets constituting Experiments 2-4), and has also been used in subsequent experiments independently replicating and extending these experiments (Bylund & Athanasopoulos, 2017).

3.2.1 Time estimation results

3.2.1.1 Time estimation, cross-domain interference effects

To evaluate the effects of irrelevant length and amount information on time estimation, each participant's time estimates in milliseconds were plotted as a function of the actual change in line length measured in pixels (i.e., horizontal extent of change, for the length interference task), or the actual change in fill level in pixels (i.e., vertical extent of change, for the amount interference task). A line of best fit was computed, and the slope was used as a measure of cross-dimensional interference. The slopes of these lines were tabulated for each group and each task.

Of primary interest, a 2 x 2 ANOVA showed the predicted interaction of Language (English, Greek) by Task (Length Interference, Amount Interference) ($F(1,137)=5.47, p=.02$), consistent with the preferred use of length-duration metaphors in

English and amount-duration metaphors in Greek (Figure 2a-b; Figure 3). There were no significant main effects.

Further planned analyses tested whether performance differed from chance in each task and each group, and evaluated pairwise differences between groups and tasks. In English speakers, the spatial length of the stimuli influenced time estimates in the length interference task, according to a one-sample t-test comparing the mean of the slopes against zero, $\beta=1.56$, $t(24)=5.14$, $p=.00003$. Lines that grew a shorter length were estimated to take a shorter time, and lines that grew a longer length to take a longer time. Likewise, in the amount interference task fuller containers were estimated to remain on the screen for more time than emptier containers, $\beta=0.54$, $t(69)=2.52$, $p=.01$. Greek speakers showed a contrasting pattern. The fullness of containers had a significant effect on time estimates ($\beta=1.16$, $t(23)=2.82$, $p=.01$), but the spatial length of lines had only a marginally significant effect ($\beta=0.60$, $t(21)=1.65$, $p=.11$). The length interference effect was stronger than the amount interference effect in English speakers, according to an unpaired t-test ($t(93)=2.50$, $p=.01$). Likewise, the length interference effect was stronger in English speakers than in Greek speakers ($t(45)=2.03$, $p=.05$). No other pairwise differences were significant.

3.2.1.2 Time estimation, within-domain effects

Control analyses were conducted to ensure that the predicted differences in cross-domain interference effects were not due to unexpected differences across groups or tasks in the accuracy with which participants estimated time, *per se*, collapsing over all levels of spatial interference. For each participant, we computed the slope of the effect of the

actual durations of the stimuli on their reproduced durations. These slopes were tabulated for each group and task.

According to a 2 x 2 ANOVA, Language did not interact with Task to predict the effect of actual duration on estimated duration ($F(1,137)=0.62, p=.43$; Figure 2c-d).

Duration estimation performance was compared against a slope of zero for each group and task, by one-sample t-tests (English length: $\beta=0.70, t(24)=15.00, p=.00001$; English amount: $\beta=0.65, t(69)=21.89, p=.00001$; Greek length: $\beta=0.67, t(21)=14.47, p=.00001$; Greek amount: $\beta=0.69, t(23)=17.05, p=.00001$). Duration estimation was significantly more accurate than chance for both groups and both tasks, but there were no significant pairwise differences, indicating that accuracy in time estimation did not differ significantly between any groups or tasks.

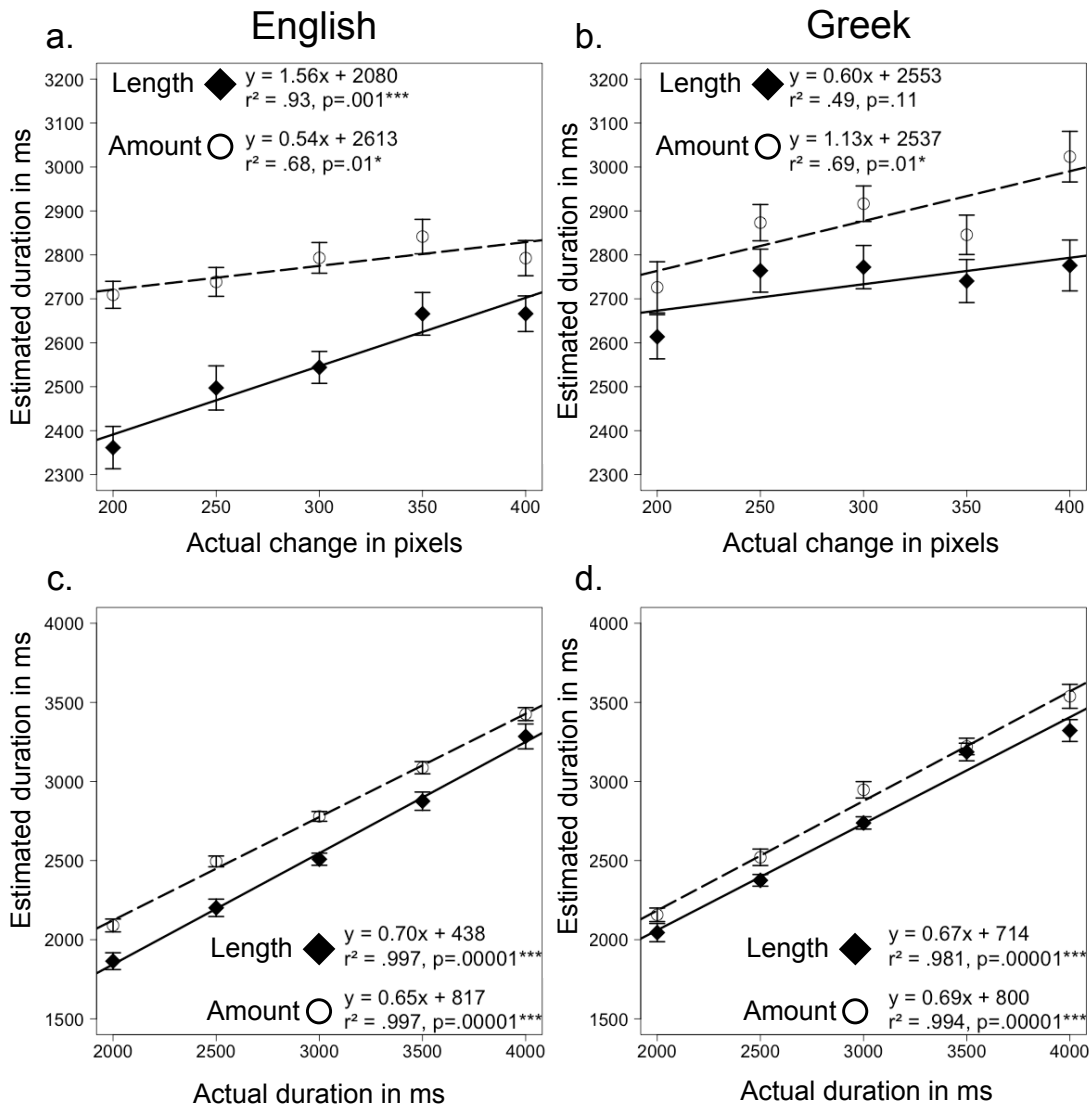


Figure 2. Results of Experiment 2, time estimation trials. Top: Cross-domain analyses. Duration estimates are plotted as a function of actual line length (diamonds) or actual container fullness (circles), in English speakers (left) and Greek speakers (right). Bottom: Within-domain analyses. Duration estimates are plotted as a function of the actual durations for which the lines (diamonds) or containers (circles) appeared, in English speakers (left) and Greek speakers (right). Error bars show s.e.m.

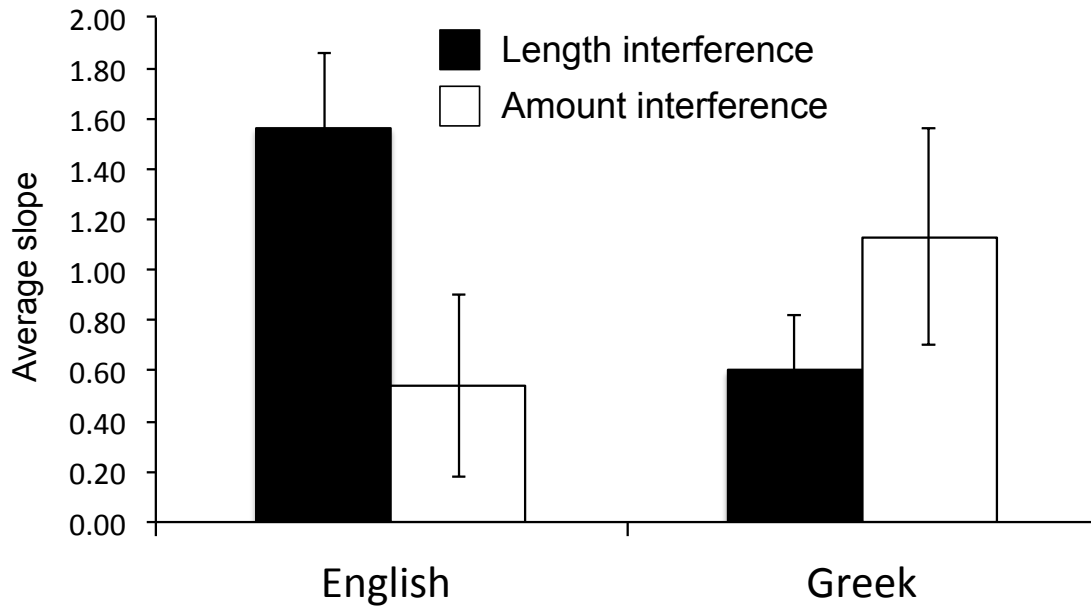


Figure 3. Results of Experiment 2, time estimation trials: Summary of the predicted cross-domain effects of length interference (black bars) and amount interference (white bars) on duration estimates in English speakers (left) and Greek speakers (right). Error bars show s.e.m.

3.2.1.3 Time estimation, comparison of cross-domain vs. within-domain effects

In the preceding two sections, we showed that there was a significant Language by Task interaction in the cross-domain interference data (i.e., effects of actual length or amount on duration estimation; section 3.2.1.1), but there was no significant Language by Task interaction in the within-domain time estimation data (i.e., effects of actual duration on estimated duration; section 3.2.1.2). In order to interpret the within-domain (non-) interaction effect as a control for the predicted cross-domain interaction effect, it is necessary to demonstrate a higher order interaction between them. We conducted a 2 x 2 x 2 ANOVA combining the cross-domain and within-domain analyses of time estimation. There was a 3-way interaction of Language (English, Greek), Task (length interference, amount interference) and Effect Type (within-domain effect, cross-domain effect; $F(1,274)=4.91, p=.03$), indicating that the predicted cross-domain interference effects

cannot be explained by the (non-significant) differences in within-domain performance across groups and tasks.

3.2.2 Space estimation results

Only the duration estimation trials (half of the total trials in each task) were of interest with respect to our experimental predictions. That is, on the basis of length and size/amount metaphors for duration in language, we predicted systematic differences in *duration estimates* across language groups. We have no reason to predict any systematic differences in spatial (length or amount) estimates, across groups or across tasks. Spatial responses were collected initially for historical reasons: our first pilot tests of cross-linguistic differences in duration estimation used one of the same length interference tasks used in Casasanto and Boroditsky's (2003; 2008) studies. Subsequent versions of the task preserved the structure of the original task to allow direct comparisons (across tasks, within English speakers).

We report the results of the length and amount estimation trials here for the sake of completeness, and also because they provide an additional control measure, to test the specificity of the predicted effects. The predicted interaction of Language (English, Greek) by Task (length interference, amount interference) should be found in the cross-domain interference analyses for time estimation trials, but not for space estimation trials.

3.2.2.1 Space estimation, cross-domain interference effects

To evaluate the effects of irrelevant duration information on length and amount estimation, each participant's spatial estimates (in pixels) were plotted as a function of the actual durations in milliseconds for which the growing lines remained on the screen

(for the length interference task), and the actual durations in milliseconds for which the filling containers remained on the screen (for the amount interference task). A line of best fit was computed, and the slope was used as a measure of cross-domain interference. The slopes of these lines were tabulated for both groups and both tasks.

Of primary interest, a 2 x 2 ANOVA showed no interaction of Language (English, Greek) by Task (length interference, amount interference; $F(1, 137) = 0.70, p = .41$; Figure 4a-b.) This null result contrasts with the significant interaction of Language by Task found in the time estimation condition.

There were no significant effects of duration on length estimation or on amount estimation, according to 1-sample t-tests comparing the slopes of these effects against a slope of zero (all p 's $> .09$; fig. 4a-b). That is, there were no significant effects of duration on space estimation. For the one effect that approached significance (amount estimation in English speakers; fig. 4a) the slope was negative. These null results contrast with the positive slopes found in the tests of length and amount interference on duration estimation, which were significantly greater than zero in three of the four data sets (English length, English amount, Greek amount), and marginally greater than zero in the fourth (Greek length).

3.2.2.2 Testing for a space-time asymmetry

The normalized data sets also allowed us to evaluate whether the cross-domain effects of space and time were symmetric or asymmetric; that is, they allowed us to test whether the cross-domain effect of space on time estimation was greater than the cross-domain effect of time on space estimation, in each group and task, separately.

Results showed that irrelevant spatial information influenced duration estimates significantly more than irrelevant temporal information influenced spatial estimates in three of the data sets (English length interference task: $t(48)=3.00, p=.004$; English amount interference task: $t(138)=3.14, p=.002$; Greek amount interference task: $t(46)=2.32, p=.02$); this difference was marginally significant in the fourth data set (Greek length interference task: $t(42)=1.77, p=.08$), where there was no significant effect of space on time or of time on space (compare fig. 2a-b to fig. 4a-b).

To summarize these analyses, in every data set where a significant effect of spatial information on time estimation was found, the cross-domain interference effects were asymmetric: the effect of space on time estimation was significantly greater than the effect of time on space estimation. This “space-time asymmetry” is a hallmark of metaphorical language about time (Lakoff & Johnson, 1980; 1999), and of metaphorical thinking about time (e.g., Boroditsky, 2000; Casasanto & Boroditsky, 2003; 2008; Casasanto, Fotakopoulou, & Boroditsky, 2010; Merritt et al., 2010; cf. Cai & Connell, 2015). Furthermore, this asymmetry underscores the specificity of the effects of spatial information on time estimation, predicted on the basis of spatial metaphors for time.

3.2.2.3 Space estimation, within-domain effects

In a final control analysis, we tested whether accuracy of space estimation, per se, differed across groups or across tasks. According to a 2 x 2 ANOVA, Language did not interact with Task to predict the effect of actual change in pixels on participants' estimated change in pixels ($F(1,137)= 0.23, p=.64$; Figure 4c-d). The slopes of the effects of actual change in pixels on estimated change in pixels were compared against a slope of zero for each group and task, by one-sample t-tests (English length: $\beta=0.71, t(24)=9.32$,

$p=.00001$; English amount: $\beta=0.85$, $t(69)=22.80$, $p=.00001$; Greek length: $\beta=0.73$, $t(21)=13.79$, $p=.00001$; Greek amount: $\beta=0.81$, $t(23)=20.87$, $p=.00001$). Spatial estimation was significantly more accurate than chance for both language groups and both tasks, and there were no significant pairwise differences between any of the data sets, indicating that accuracy in space estimation did not differ significantly between any groups or tasks.

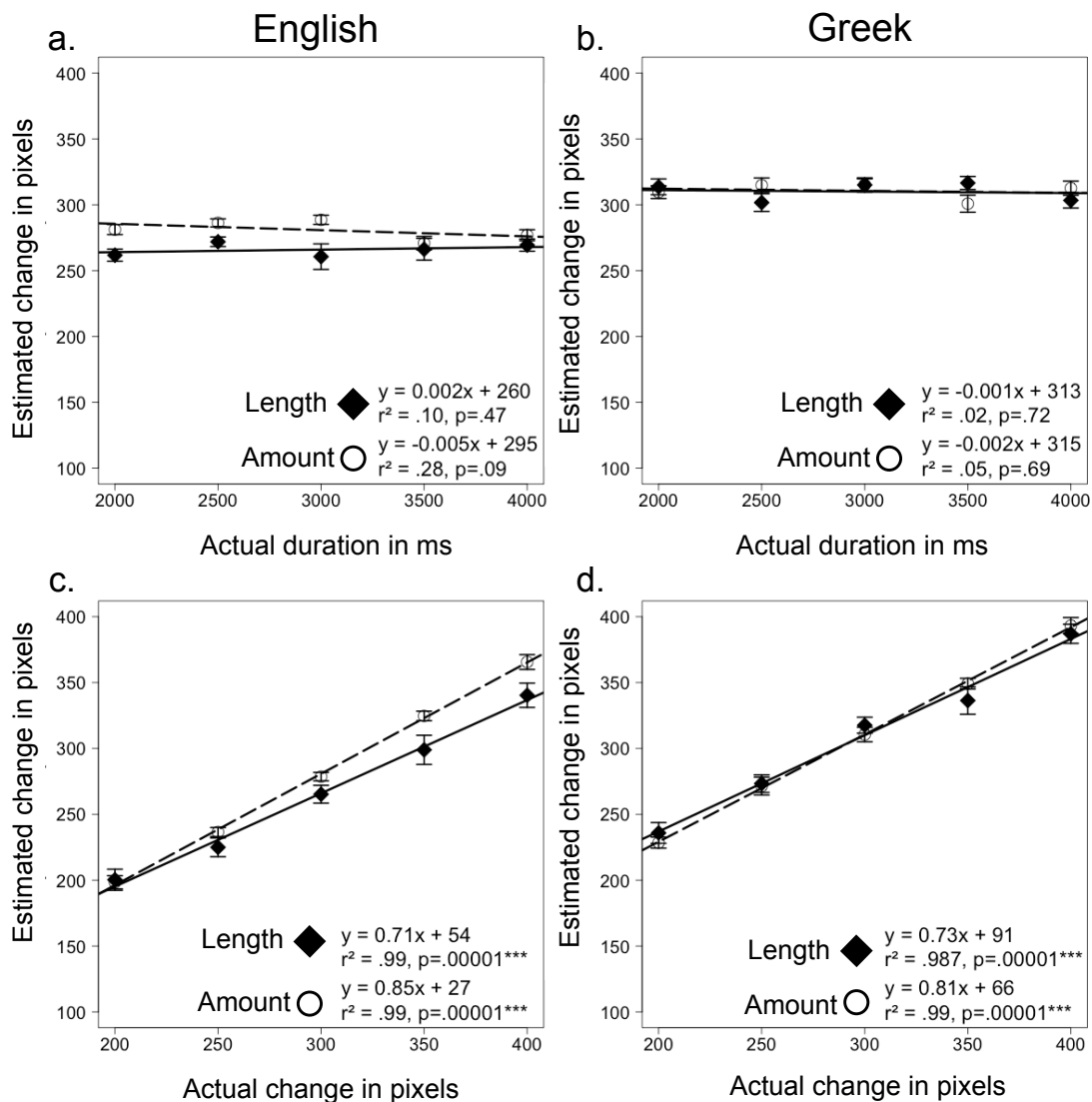


Figure 4. Results of Experiment 2, space estimation trials: Top: Cross-domain analyses. Spatial estimates are plotted as a function of the actual durations of the lines (diamonds) or the containers (circles), in English speakers (left) and Greek speakers (right). Bottom: Within-domain analyses. Spatial estimates are plotted as a function of the actual lengths of the lines (diamonds) or fill levels of the containers (circles), in English speakers (left) and Greek speakers (right). Error bars show s.e.m.

3.3 Control experiment: Dimensionality or orientation?

The filling-container task differed from the growing-line task insofar as the lines showed changes in 1-dimensional spatial extent (following the convention that thin lines are idealized to be 1-dimensional), whereas the containers showed changes in multidimensional spatial magnitude (in reality the stimuli increased in 2-dimensional area on the screen, but by convention such 2-dimensional images often represent 3-dimensional space). Thus, the line and container stimuli differed in their spatial dimensionality, by design. However, they also differed in their spatial orientation: lines grew horizontally, whereas containers filled vertically. Could this difference in the spatial orientation of the stimuli be responsible for the differences in duration estimation we found across tasks in English and Greek speakers?

To test this alternative explanation, we conducted an additional control experiment in which English and Greek participants estimated the durations of vertical lines, while ignoring their vertical spatial extents (data are reported in Appendix E of the online Supplementary Materials). The results of the vertical growing line task were compared to the results of the horizontal growing line task from Experiment 2. We reasoned that if spatial orientation was responsible for the observed differences between groups in the effects of horizontal growing lines vs. vertically filling tanks, then a similar pattern should be found in a comparison of the effects of horizontal growing lines vs. vertical growing lines. Specifically, (a) the effect of vertical growing lines should be greater in Greek speakers than in English speakers, and (b) there should be a significant interaction of language group (English, Greek) by task (horizontal lines, vertical lines).

Neither of these effects was found. Vertical height had a significant influence on duration estimation in both English and Greek speakers, but there was no significant difference in the cross-domain effects of height on duration estimation between groups, and no significant group by task interaction (p 's > .22). The slope of the effect of actual height on estimated duration was numerically *shallower* in Greek speakers than in English speakers; although this difference did not approach significance, the trend goes in the opposite direction from the pattern that would support an orientation-based explanation of the critical result from Experiment 2.

The results of this control experiment rule out the possibility that the differential effects of growing lines and filling containers on English and Greek speakers' duration estimates were due to the spatial orientation of the stimuli. The preferred ways that Greek and English speakers talk and think about duration appear to vary in their spatial dimensionality.

3.4 Discussion

In summary, the pattern of cross-domain interference of length vs. amount on duration estimation found in English and Greek speakers was predicted by the prevalence of length vs. size/amount metaphors for duration in English and Greek. The predicted Language by Task interaction was highly specific: it was found in analyses of the cross-domain interference effects of space on time estimation, but was not found: (a) in cross-domain interference effects of time on space estimation; (b) in within-domain effects of actual time on estimated time; (c) in within-domain effects of actual space on estimated space. English and Greek speakers appear to mentally represent duration, in part, using

spatial schemas that differ in their dimensionality, as predicted by the differences we have documented in their length vs. size/amount space-time metaphors in language.

4. Experiment 3: Does space still affect time when verbal labels are suppressed?

The time estimation tasks in Experiment 2 used non-linguistic stimuli and responses. But were participants labeling the stimuli covertly? Attempting to covertly label the temporal dimension of the stimuli during time trials might seem sensible. A pre-trial cue informed participants when time was the relevant dimension, so a reasonable strategy might be to try to label this dimension using common terms for duration (e.g., long, short). Indeed, participants *may have tried* to label the durations of the temporal stimuli using spatial words, but we can be certain that this strategy did not give rise to the effects of space on time that we predicted and found. The experimental design rules out this possibility.

In the design, nine different line lengths or container fill-levels were fully crossed with nine different durations: there was no correlation between the spatial magnitudes of the stimuli and their durations. Therefore, verbally labeling the task-relevant dimension of the stimuli could not cause the predicted pattern of interference from the task-irrelevant dimension (i.e., interference of spatial information on time estimates), because the task-relevant and task-irrelevant dimensions varied orthogonally. On the contrary, covertly labeling each stimulus as lasting a “long” or “short” time, for example, could only *diminish* the effect of space on time that we predicted and found. Perfect application of such a labeling strategy would result in a perfectly null effect of the spatial extents of the stimuli on participants’ estimates of their durations (i.e., a slope of zero, rather than the predicted positive slope).

We posited, therefore, on the basis of the design of these psychophysical tasks, that the effects of space on time estimation we found in Experiment 2 were due to participants' previous (pre-experimental) use of length or size/amount metaphors, not to their covert use of length or size/amount words during the task. This proposal makes a prediction: If the effects of language on time estimation are not due to the use of language 'online', during the task, then the space-time interference effects observed in these tasks should persist in the presence of a secondary task that interferes with participants' ability to use language in the moment, while perceiving the stimuli and reproducing their durations.

Experiment 3 was conducted to test this prediction. We replicated the non-linguistic length interference task in English speakers, with the addition of a concurrent verbal suppression task. Participants rehearsed a novel five-letter string aloud during the entire time that they were perceiving and reproducing each of the growing line stimuli. If the effect of spatial length on time estimation was somehow driven by participants covertly labeling the spatial dimension of the temporal stimuli in Experiment 2, then this effect should be diminished or extinguished in Experiment 3. Alternatively, if the language-specific effects we observed in Experiment 2 were due to participants' prior, pre-experimental use of length or size/amount words to talk about time – not to the use of length or size/amount words (or any other words) during the task – then the effect of space on duration estimation should persist even with the secondary verbal suppression task added. Specifically, the effect of length on time estimation should be just as strong for English speakers in Experiment 3 as it was in Experiment 2.

4.1 Method

4.1.1 Participants

A new sample of native English speakers (N=20) from the University of Chicago community participated for payment or course credit. A sample size of 20 was selected on the basis of the results of Experiment 2.

4.1.2 Materials

For the primary task, the same materials were used as in the length interference task from Experiment 2 (English instructions), with 6 practice trials added at the beginning. For the secondary verbal suppression task (adapted from Winawer et al., 2007), 162 unpronounceable 5-letter target strings were constructed; 162 foils were also constructed, which differed from the target strings by 1 letter. The stimulus presentation and response collection script was programmed in Python.

4.1.3 Procedure

After 6 practice trials, participants were presented with the 162 growing line stimuli, one at a time, and they reproduced either the length or duration of each stimulus, as in Experiment 2. Before each line appeared, a 5-letter string appeared for two seconds. Participants were instructed to begin rehearsing the string aloud as soon as it appeared, and to continue rehearsing it aloud during the entire time they were viewing and reproducing the growing line stimuli. An experimenter remained in the room to ensure they complied. Recognition memory for the letter strings was tested using a 2-alternative forced choice test. Tests occurred unpredictably after 56 of the responses (about 35%), randomly distributed throughout the 162 growing line trials. The target and foil strings appeared side by side, with the left-right positions of targets and foils randomized. Participants responded by pressing the key corresponding to either the correct string or a

foil (“c” or “m” key). This forced-choice response was collected to ensure that participants encoded the target strings accurately enough to distinguish them from nearly-identical strings, indicating that their attention was occupied by the secondary task. The next trial began after a 1000 ms inter-trial interval. Testing lasted about 45 minutes.

4.2 Results

4.2.1 Verbal suppression (secondary task)

Participants’ recognition of the correct letter strings was much greater than chance (Mean percent correct=98%, SD=2.17%; $t(19)=99.47$, $p=.00001$), indicating that they were engaged in the verbal interference task.

4.2.2 Length interference (primary task)

The data from the time estimation trials were analyzed as in Experiment 2.

4.2.2.1 Time estimation, cross-domain interference effects

Of primary interest, we tested the cross-domain effect of irrelevant length information on time estimation. The mean slope of the effect of actual length on estimated duration was significantly greater than zero, by 1-sample t-test ($\beta=2.27$, $t(19)=4.23$, $p=.0005$), indicating that spatial length still influenced duration estimates, even under verbal suppression (Figure 5a). This slope did not differ from the slope of the effect of actual length on estimated duration found in the length interference task in English speakers from Experiment 2 ($t(43)=1.21$, $p=.23$).

4.2.2.2 Time estimation, within-domain effects

An analysis of within-domain performance showed that participants estimated duration with high accuracy in spite of the attention-demanding verbal suppression task, according to a 1-sample t-test comparing the mean slope of the effect of actual duration on estimated duration against a slope of zero ($\beta=0.75$, $t(19)=11.73$, $p=.00001$; Figure 5b). This slope did not differ from the slope of the effect of actual duration on estimated duration found in the length interference task in English speakers from Experiment 2 ($t(43)=0.60$, $p=.55$).

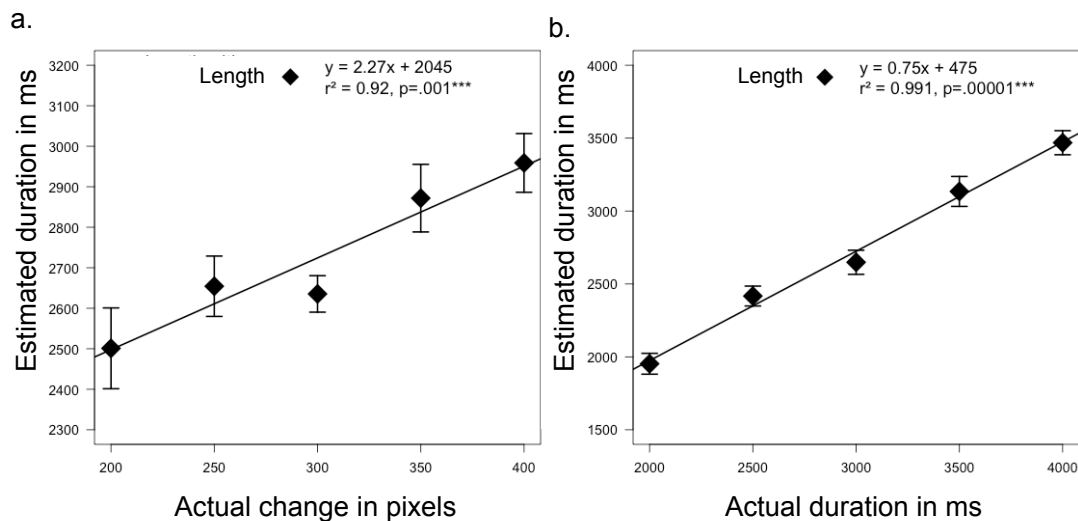


Figure 5. Results of Experiment 3. Length affected English speakers' duration estimates even when they were performing a concurrent verbal suppression task. Left: Cross-domain analysis. Effect of actual line length on estimated duration. Right: Within-domain analysis. Effect of actual line duration on estimated line duration. Error bars show s.e.m.

4.2.2.3 Space estimation

We had no hypothesis regarding the space estimation trials, the results of which closely resembled the results of the space estimation trials in the comparable data set from Experiment 2. To summarize these results, the within-domain effect of actual length on estimated length was highly significant, and there was no significant cross-domain

interference effect of actual duration on estimated length, replicating the space-time asymmetry discussed in section 3.2.2.2.1. (The space estimation data are reported in the Appendix B of the online Supplementary Materials.)

4.3 Discussion

Irrelevant length information influenced English speakers' duration estimates even while they were performing a verbal suppression task. The strength of the cross-domain interference effect, as measured by the slope of the effect of actual length on estimated duration, did not differ statistically between English speakers who performed the length interference task with concurrent verbal suppression (Experiment 3) and without verbal suppression (Experiment 2). The correlation coefficients for the effect of length on time estimation were nearly identical across these experiments (Expt. 2: $r=.96$; Expt. 3: $r=.96$), and the slope of this effect was numerically *greater* with concurrent verbal suppression than without (Expt. 2: $\beta=1.56$; Expt. 3: $\beta=2.27$), providing no support for the hypothesis that the language-specific effects of space on time estimation in Experiment 2 were due to (or enhanced by) participants covertly labeling the task-irrelevant spatial dimension of the stimuli.

These results support our interpretation of the data from these psychophysical tasks as showing effects of participants' previous experience using linguistic metaphors on their subsequent non-linguistic duration representations: effects that were not mediated by covert activation of verbal labels for the stimuli during the time estimation task. These results corroborate results from Starr and Brannon (2016) who showed that, verbal interference did not affect participants' performance in a length interference task adapted from Casasanto and Boroditsky (2008).

5. Experiment 4: A causal role for language in shaping temporal thinking?

Did English and Greek speakers' tendency to use different linguistic metaphors for duration give rise to the cross-linguistic differences in duration estimation we found? Or could there be non-linguistic cultural factors that caused the differences in time-estimation patterns, and perhaps also give rise to the observed differences between the languages? We are not aware of any potential alternative explanations for the between-group differences reported in Experiment 2. Yet, because these cross-linguistic comparisons were quasi-experimental (i.e., participants were not randomly assigned to be native speakers of English or Greek), it is not possible to conclude that linguistic differences *caused* the observed non-linguistic differences on the basis of these results.

In Experiment 4 we conducted a two-part training experiment to test whether linguistic metaphors can play a causal role in determining what kind of non-linguistic spatial schemas people activate when they estimate duration. During the first part, English speakers were randomly assigned to one of two linguistic training groups. One group completed fill-in-the-blank sentences comparing objects or events using the words *longer* or *shorter* (length training). The second group completed the same sentences comparing using the words *more* or *less* (amount training). The length-training group served as a control for the amount-training group. In effect, we assigned one group of participants to remain English speakers, and another group speakers to become more like Greek speakers in the relevant regard, talking about duration in terms of amount.

If using a linguistic metaphor activates a corresponding non-linguistic mapping between the source and target domains, then training English speakers to talk about duration using amount words should (at least transiently) strengthen the mapping

between amount and time, which is ordinarily weaker for English speakers than for speakers of a language like Greek.

After the linguistic training phase, all participants performed the non-linguistic amount interference task from Experiment 2. We predicted that participants in the amount training group would be more distracted by amount information during the non-linguistic duration estimation task and would show more cross-domain amount interference than participants in the length-training control group.

5.1 Method

5.1.1 Participants

A total of 90 Stanford University students participated for payment. Participants were randomly assigned to perform either the amount training task (n=47) or the length training control task (n=43). The sample size reflects the number of participants who volunteered during the semester in which the study was run. Two participants (2.2%) were removed, both from the length training condition, for performing the experiment incorrectly according to the criteria described in Experiment 2 (i.e., the slope of the effect of actual duration on estimated duration was negative).

5.1.2 Materials and Procedure

The experiment had two parts. For the first part, participants performed either the length-training or amount-training task. They completed 192 fill-in-the-blank sentences using the words “longer” or “shorter” for the length-training task, and “more” or “less” for the amount-training task. Half of the sentences compared the length or size of physical objects (e.g., An alley is *longer* / *shorter* than a clothesline; A teaspoon is *more* / *less* than

an ocean), and the other half compared the duration of events (e.g., A sneeze is *longer* / *shorter* than a vacation; A sneeze is *more* / *less* than a vacation; for a complete list of stimuli, see Appendix A in the online Supplementary Materials). Sentence types were randomly intermixed. Trials were self-paced, and the training tasks lasted about 20 minutes. Stimuli were presented and responses collected for the training task using PsyScope software. After training, all participants performed the same non-linguistic amount interference task used in Experiment 2.

5.2 Results and Discussion

5.2.1 Training Phase (fill-in-the-blank sentences)

Participants filled in the blanks with high accuracy for both the amount training (Mean percent correct=98.2%, SD=1.8%) and the length training tasks (Mean percent correct=97.5%, SD=1.9%).

5.2.2 Test Phase (amount interference task)

5.2.2.1 Time estimation, cross-domain effects. The data were analyzed as in the amount interference task in Experiment 2. Of primary interest, we evaluated the cross-domain effect of irrelevant amount information on time estimation in each training group separately, and then compared them. The mean slope of the effect of actual amount on estimated duration was significantly greater than zero after amount training ($\beta=1.18$, $t(46)=4.77$, $p=.00002$), but not after length training ($\beta=0.04$, $t(40)=0.10$, $p=.92$). Of primary interest, the amount interference effect was significantly greater after amount training than after length training ($t(86)=2.64$, $p=.01$; Figure 6a), indicating a significant effect of language training, in the predicted direction.

In further planned analyses, we compared cross-domain interference effects in amount-trained participants from Experiment 4 with those of untrained native speakers of English and Greek from Experiment 2. The amount interference effect was marginally greater in amount-trained English speakers (Experiment 4) than in untrained English speakers (Experiment 2; $t(115)=1.92, p=.06$). By contrast, there was no difference in the magnitude of the amount interference effect between amount-trained English speakers (Experiment 4) and untrained Greek speakers (Experiment 2; $t(69)=0.06, p=.95$; Fig. 7). We note that these cross-experiment comparisons are supplementary to the main results of Experiment 4 and should be interpreted with caution. It is not possible to know what effect a training task might have on participants' subsequent time estimation, independent of the particular metaphors used. That is, simply asking participants to compare the durations of almost 200 pairs of events in Experiment 4 could affect how they think about duration relative to participants who were not asked to think about duration prior to the estimation task, making performance in the trained participants hard to compare to that of the untrained participants from Experiment 2. For these reasons, the main comparison of interest in Experiment 4 was between the two randomly assigned training groups: not between a trained group and an untrained group from an earlier experiment.

5.2.2.2 Time estimation, within-domain effects. The effect of language training on the magnitude of cross-domain amount interference cannot be attributed to differences between training groups in the accuracy of time estimation, *per se*. The mean slope of the effect of actual duration on estimated duration was significantly greater than zero for both training groups (amount training: $\beta=0.66, t(46)=13.52, p=.00001$; length training: $\beta=0.70,$

$t(40)=18.52, p=.00001$), and did not differ between groups ($t(86)=0.65, p=.51$; Figure 6b).

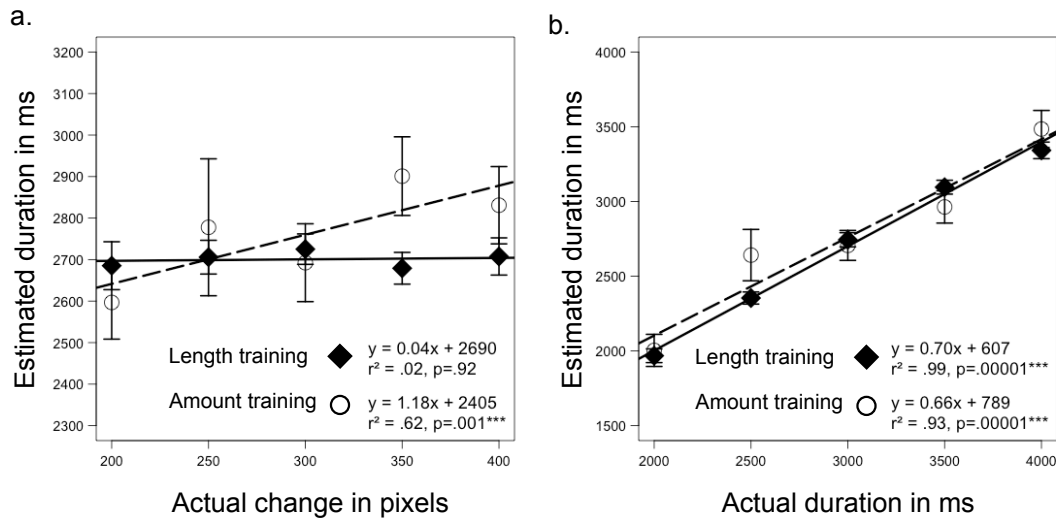


Figure 6. Results of Experiment 4, time estimation trials in length-trained English speakers (diamonds) and amount-trained English speakers (circles). Left: Cross-domain analyses. Duration estimates are plotted as a function of actual container fullness. Right: Within-domain analyses. Duration estimates are plotted as a function of the actual durations for which the containers appeared. Error bars show s.e.m.

5.2.2.3 Space estimation

We had no hypothesis regarding the amount estimation trials, the results of which closely resembled the results in the comparable data set from Experiment 2, for both training groups. To summarize these results, the within-domain effects of actual amount on estimated amount were highly significant, but there were no significant cross-domain interference effects of actual duration on estimated amount. (These data are reported in the Appendix C of the online Supplementary Materials).

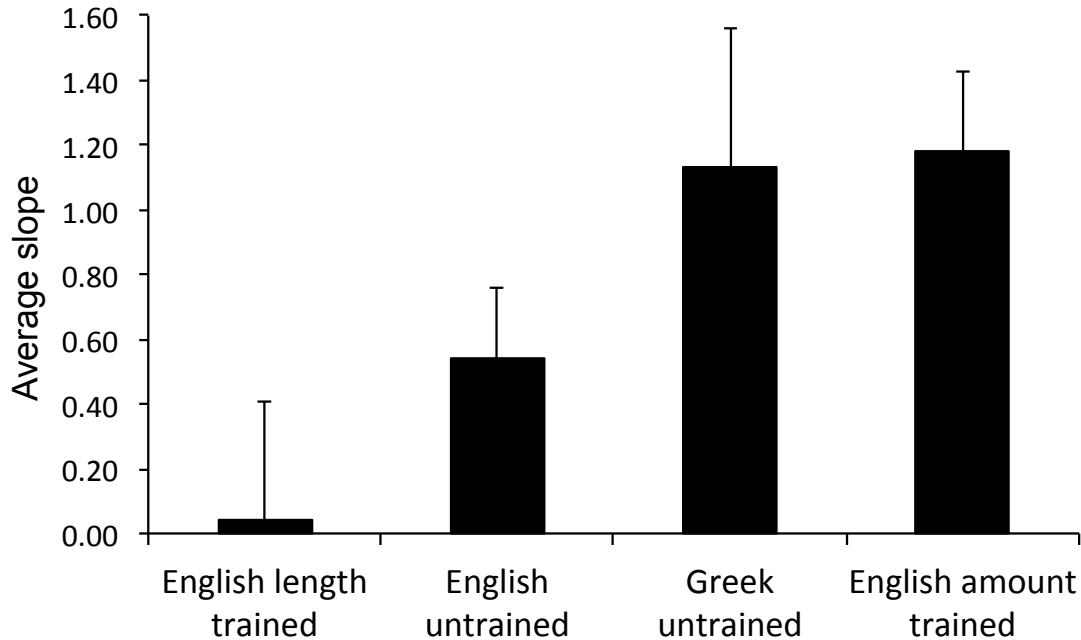


Figure 7. Comparison of the duration estimation results from the amount interference tasks in Experiments 2 (inner columns) and 4 (outer columns). Amount interference was significantly greater after amount training than after length training (compare outer columns). Amount interference was statistically indistinguishable between length-trained English speakers and untrained English speakers (left columns). By contrast, amount interference was statistically indistinguishable between amount-trained English speakers and untrained Greek speakers (right columns). Error bars show s.e.m.

5.3 Discussion

Experience using amount metaphors, but not length metaphors, caused native English speakers to perform the amount-interference task more like native speakers of Greek than like native speakers of English, demonstrating that language can have a causal influence on time representations. Using amount metaphors during the training phase influenced subsequent temporal thinking during the test phase. The increased effect of amount interference on time estimation found in the amount-training group cannot be due to participants covertly naming the task-relevant temporal dimension of the stimuli using amount words online, during the task, as doing so would only work *against* the observed effect (see also the verbal interference results from Experiment 3).

The results of Experiment 4 license several inferences. First, whereas Experiment 2 was quasi-experimental, Experiment 4 was a true experimental intervention in which participants were randomly assigned to language ‘treatment’ conditions. Therefore, unlike quasi-experimental studies, which can only demonstrate language-thought correlations, Experiment 4 showed a causal effect of linguistic experience on performance of a non-linguistic task.

Second, because language is a part of culture, it is often difficult to determine whether cognitive differences between speakers of different languages are caused by linguistic experience, *per se*, or by some other aspect of their cultures that merely correlates with language. By using a true experimental intervention, we were able to observe effects of linguistic experience while controlling for effects of culture, since participants in Experiment 4 were sampled from the same population and tested in the same cultural setting (i.e., the same US university psychology lab) – the only change in the amount-trained participants’ ‘cultural’ experience was the change in their use of linguistic metaphors for time.

Third, the training experiment complements a quasi-experiment we conducted in native Greek speakers living in different language environments. Prior to developing the tasks used in Experiments 2-4, we ran a pilot study using an earlier version of the length interference task (Casasanto & Boroditsky, 2008, Experiment 2). We compared length interference effects in two groups of native Greek speakers, one group living in Greece and the other living in the USA. Both groups were fluent in English as a second language, and both were tested with instructions and prompts written in English. The group tested in Greece showed no significant length interference effect, and their results did not differ statistically from those of the Greek length interference group reported in

Experiment 2. By contrast, the group of Greeks tested in the USA showed a significant length interference effect, which was statistically indistinguishable from the effect in English speakers tested in the US. Both this pilot study and Experiment 4 address a question that has been raised about the results of Experiment 2: Could the differences between English and Greek speakers' performance on the non-linguistic time estimation tasks be driven by the language in which the task instructions were written? In the pilot experiment, both groups of native Greek speakers received instructions in English, but only the Greeks living in the US (and presumably using English length-time metaphors routinely) showed a significant length interference effect. In Experiment 4, both groups of native English speakers received instructions in English, but only the English speakers trained to use Greek-like amount metaphors showed a significant amount interference effect. Together, these results show that the language used in the task instructions did not predict the results (cf., Bylund & Athanasopoulos, 2017). Rather, the results of the non-linguistic tasks were predicted by the type of space-time metaphors that participants had been using most frequently just prior to the experiment, either due to living in a Greek- or English-speaking environment or to participating in one metaphor training condition or the other. (Data from this pilot study are reported in Appendix D of the online Supplementary Materials).

The training task in Experiment 4 illustrates a process by which our everyday linguistic experience may shape our non-linguistic mental representations of duration. Presumably, instructing English speakers to use amount metaphors in language caused them to activate a pre-existing association between representations of amount and time in their long-term memories, strengthening this association (at least transiently). If giving English speakers a concentrated dose of amount metaphors can have this effect on

duration representations in the lab, perhaps ordinary doses of Greek have a similar effect on Greek speakers' amount-time associations, over a longer time course. This conjecture is supported by the pilot study suggesting that ordinary doses of English led Greek expatriates in the US to rely on length-time associations. On this view, using one's native language, or being immersed in a second language environment, is like participating in a natural 'training experiment'.

6. General Discussion

In four experiments, we showed that English and Greek speakers tend to talk about duration using different kinds of spatial metaphors, and tend to think about duration using correspondingly different spatial representations. In Experiment 1, English speakers described the relative durations of events most often using length metaphors (e.g., longer, shorter). Greek speakers, by contrast, used size or amount metaphors most often, reporting, for example, that one event's duration was "bigger" (tr. *Megalyteros* [μεγαλύτερος]) than another's, or "more" (tr. *Perissoteros* [περισσότερος]) than another's. In Experiment 2, Greek and English speakers performed non-linguistic duration reproduction tasks. The results showed that spatial information interfered with people's duration estimates in ways that reflected the preferred space-time metaphors in their native languages. English speakers' duration estimates were influenced more strongly by irrelevant length information, and Greek speakers' by irrelevant amount information. Experiment 3 showed that the observed effect of length on time estimation in English speakers was unaffected by a concurrent verbal suppression task, confirming that the language-specific effects of space on time we observed in Experiment 2 were not due to participants using spatial language covertly to describe the stimuli during the task.

Rather, it appears that people's previous experience using space-time metaphors in language affects their subsequent temporal thinking. In Experiment 4, after being randomly assigned to use about 200 amount expressions in language, native English speakers' duration estimates were strongly influenced by irrelevant amount information – like native Greek speakers'. Together, these experiments suggest that people's mental representations of duration differ according to the kinds of space-time metaphors they tend to use, and that experience using metaphors in language can play a causal role in creating these differences in people's non-linguistic representations of duration.

6.1 Influences of spatial language on temporal thinking: When and how?

What role do verbal space-time metaphors play in shaping temporal thinking? The answer to this question may differ for different metaphors, describing different aspects of time. Language may contribute to the construction of uniquely human, culture-specific temporal notions like “moving Wednesday's meeting forward” (Boroditsky, 2000). For some temporal concepts, exposure to conventional space-time metaphors may invite language learners to build analogical bridges between space and time that they might not have constructed otherwise (Boroditsky, 2001; Gentner, 2001).

In the case of duration, however, it appears that associations between space and time can be constructed, at least initially, independent of language. Non-human animals have been found to associate greater duration with longer spatial length (Merritt et al., 2010). Human infants have been trained to associate duration with length (Srinivasan & Carey, 2010) and with size (Lourenco & Longo, 2010), suggesting that both of the space-time associations that we explored in this study can be created pre-linguistically. It remains a subject of ongoing investigation whether these associations are part of infants'

innate cognitive endowment (de Hevia et al, 2014) or whether they are learned on the basis of observable correlations between space and time in the natural world (e.g., more time passes as objects travel farther, and as quantities accumulate; Lakoff & Johnson, 1999).

Whether innate or learned, non-linguistic duration-length and duration-size/amount associations could be present in the minds of language learners universally. Initially, the relative strengths of these associations could be similar across language groups, since presumably the laws of physics are the same in all language communities, and the parts of people's bodies and brains that enable them to perceive and remember spatio-temporal events are similar across groups. Yet, even if universal space-duration associations are established pre-linguistically, it appears that experience using language can change them.

How? Suppose each time people produce or understand an expression like “a long night” or “megali nyhta” [μεγάλη νύχτα] (tr. a big night) they activate the corresponding association between duration and length or size/amount. English speakers would activate their non-linguistic duration-length association most often, thus strengthening it relative to the duration-size/amount association. The opposite would be true for Greek speakers, whose language would cause them to activate their duration-size/amount association most often, strengthening it relative to the duration-length association. As a result, English speakers would tend to activate representations of spatial length more strongly or more automatically than representations of size/amount when they encode or reproduce duration, whereas Greek speakers would tend to activate representations of size/amount more strongly or more automatically, giving rise to the language-specific patterns of spatial interference we found on the non-linguistic duration reproduction tasks.

This proposal suggests that the same act of using a space-time metaphor in language influences people's non-linguistic representations of duration on two timescales. First, using a length or size/amount metaphor causes people to activate the corresponding space-time association *in the moment*, while they are producing or understanding the verbal metaphor. Some theorists have suggested that effects of language on thought may be limited to such "momentary" influences (e.g., Gennari, Sloman, Malt, & Fitch, 2002; Landau, Dessalegn, & Goldberg, 2010; Papafragou, Hulbert, & Trueswell, 2008; Slobin, 1996; Ünal & Papafragou, 2016). Yet, in the case of duration, it appears that by causing people to activate one space-time association or another in the moment, language also influences their *subsequent thinking*. Each time a verbal metaphor causes the momentary activation of a duration-length or duration-size/amount association, the strength of this association is changed relative to its competitors (Casasanto, 2017; Song, 2000). Thus, using a length or size/amount metaphor in language influences the kind of spatial representation that is likely to be activated subsequently, when the language user processes duration in the future – whether or not they are producing or understanding linguistic space-metaphors at that future time. The effect of previous language use on subsequent temporal thinking was demonstrated here via tasks in which the covert use of task-relevant language could not produce the predicted effect of space on time estimation, and the effect was unchanged by the addition of a concurrent verbal suppression task (see sections 4-4.3; see also Dolscheid, Shayan, Majid, & Casasanto, 2013).

Assuming that non-linguistic duration-length and duration-size/amount associations remain plastic throughout the lifetime, changes in the frequency with which people use verbal length or size/amount metaphors should change the strength or automaticity with which they activate length vs. size/amount representations when they

process the durations of events, subsequently. The data from Experiment 4 support this proposal. For amount-trained participants, the frequency of duration-amount metaphors was boosted far above the usual rate for English speakers (judging from corpus data (Casasanto et al., 2004) and from the event description data in Experiment 1). As a result, the amount-trained English speakers showed a pattern of amount interference that was statistically indistinguishable from native Greek speakers’.

One question not addressed by this training study is how long the effects of transiently boosting the frequency of a linguistic metaphor might last. It seems likely that, as with many other types of learning, linguistic patterns that are learned early, repeated frequently, and have high availability in recent experience have the most weight in guiding behavior. That is, primacy, frequency and recency all should have some predictive power. In the case of the laboratory training, the trained instances have the benefit of recency and local frequency. How the system settles as the recency fades, and what kind of experience might be needed to maintain the newly learned pattern are topics for future investigation. The dynamics of the effects of language on subsequent temporal thinking remain to be more fully explored.

6.1.1 Linguistic relativity and Ad Hoc Cognition

Building on theorists from James (1890) to Barsalou (1983), the Ad Hoc Cognition proposal (Casasanto & Lupyan, 2015) maintains that thinking does not rely on a store of permanently existing concepts. Instead, all mental representations are constructed ad hoc, and vary systematically between groups of people, between individuals within a group, and between instantiations of ‘the same’ representation within individuals. This variation is posited to occur on three overlapping timescales: *activation dynamics* (milliseconds to

seconds), *local context effects* (minutes to hours), and *experiential relativity* (days to lifetimes).

The studies presented here illustrate variability in mental representations on two of these timescales and point to variability on the third. At the longest timescale, the differences in English and Greek speakers' temporal thinking predicted by their different preferred space-time metaphors (Experiments 1-2) are an example of *linguistic relativity* (Whorf, 1956): People's thinking about time varies relative to the language they use for talking about it. However, as the training experiment showed (Experiment 4), entrenched habits of thinking can be changed at least transiently by changes in the *local context* – in this case, by the local linguistic context in the laboratory. Language-driven changes at both of the longer timescales are presumably mediated by changes in how associations in long-term memory become activated on the timescale of *activation dynamics*. Using a length or amount metaphor in language presumably strengthens the corresponding mnemonic association between space and time gradually, one momentary activation at a time, and simultaneously weakens the alternative space-time association gradually via a process of competitive associative learning (Casasanto, 2017; Song et al., 2000).

From the Ad Hoc Cognition perspective, effects of language on thought appear not only explicable, but also inevitable. The natural language that we use most often – to talk, to read, and even to dream – is a nearly ever-present part of the context in which we use our minds; since thinking depends on context, it depends on language. Considering linguistic relativity effects at multiple timescales can help explain how language affects our thinking in ways that are long-lasting but not immutable.

6.2 Is ‘amount of time’ spatial?

Size metaphors such as *megali nyghta* [μεγάλη νύχτα] (big night) are unambiguously spatial: The duration of a night is metaphorized as 3-dimensional volume (i.e., bigness) rather than 1-dimensional length. Presumably, amount metaphors such as *poli ora* [πολλή ώρα] (much time) are also volume metaphors: Time is metaphorized as a substance that accumulates in 3-dimensional space, like water in a container. Ultimately, whether representations of an “amount of time” are spatial is an empirical question: In principle, “amount” could reduce more fundamentally to some notion of non-spatial, abstract quantity. However, three lines of reasoning suggest that, like size metaphors, amount metaphors for duration may also invoke multi-dimensional space rather than abstract quantity. First, Greek speakers appear to use size metaphors and amount metaphors in the same contexts: Whether a size or amount metaphor gets used is determined by the event whose duration is being described (e.g., a night vs. a party, see Table 1). Second, analyses of many linguistic metaphors suggest that source domains are typically more concrete (i.e., more available to perception) than target domains (Lakoff & Johnson, 1980), and that space is one of the most productive source domains of all. Gentner and colleagues (2001) suggest that space is the “universal donor” of concrete structure to relatively abstract domains, and in particular to the domain of time. Finally, it is unclear to what extent people actually understand abstract quantities *abstractly*, as opposed to understanding them at least partly in terms of space. Pure numbers like 3 or 387 are, according to their mathematical definitions, not associated with any physical quantities; thus, numbers would seem to provide a paradigm case of abstract magnitude. Yet, the psychology of number representation belies numbers’ pure abstraction. According to dozens of behavioral studies, people conceptualize numbers, in part,

spatially (see Wood et al, 2008, for a review). For these reasons, we posit that amount-time metaphors are one variety of space-time mappings, and that their source domain is 3-dimensional space, as it is in size-time metaphors.

An aspect of these 3-dimensional space-time metaphors that calls for further exploration concerns the distinction between objects and substances. In an expression like ‘a big night’, the source domain appears to be an object. By contrast, in an expression like ‘a party that lasts much’ the source domain appears to be a substance. If these assumptions are correct, it remains an open question whether there are systematic differences in people’s conceptions of duration for events that are metaphorized as objects of a particular size (like nights) vs. events that are metaphorized as particular amounts of a substance (like parties).

7. Conclusions

Time is often described as a line, and the durations of events as “long” or “short.” But this is not the only way to talk about duration in terms of space, and apparently it is not even the preferred way in some languages. Whereas English speakers most naturally describe duration using 1-dimensional spatial words, Greek speakers tend to use words denoting multidimensional spatial size or amount.

This cross-linguistic difference has consequences for people’s non-linguistic mental representations of duration. When encoding and reproducing the durations of brief events, English speakers’ time estimates were more strongly influenced by irrelevant length information, and Greek speakers’ by irrelevant amount information. Unlike many linguistic relativity effects, these results cannot be explained as effects of people using task-relevant language in the moment, during the duration reproduction tasks, to covertly

describe the stimuli: this possibility is ruled out both by the design of these experiments and by the results of a verbal suppression task. Rather, it appears that people's previous experience using space-time metaphors in language affects their subsequent temporal thinking.

Duration-length and duration-size/amount associations may be formed initially pre-linguistically, perhaps in infancy, but this does not make them impervious to the effects of language. Presumably, when people produce or understand either a length - duration or amount / size-duration metaphor in language, they activate the corresponding space-time association in memory, strengthening this association relative to its competitors. Consistent with this proposal, after native English speakers were exposed to a concentrated 'dose' of amount-duration metaphors in the laboratory, their duration estimates were influenced just as strongly by irrelevant amount information as native Greek speakers'. Together, these results show that spatial metaphors for duration differ across languages, and that these verbal metaphors can play a causal role in shaping people's basic non-linguistic mental representations of time.

Acknowledgements

We thank Webb Phillips, Shima Goswami, Jesse Green, and Kyle Jasmin for help with programming and data collection, Pieter A. M. Seuren for consultation on Greek metaphors, and Laura Casasanto, Herb Clark, Teenie Matlock, and Tyler Marghetis for helpful feedback. A preliminary report on Experiment 2 appeared in the *Proceedings of the 26th Annual Conference Cognitive Science Society*, and preliminary versions of Experiments 2 and 4 were included in D.C.'s doctoral dissertation (2005, MIT). This research was supported in part by: an NSF Graduate Research Fellowship, NSF

Dissertation Improvement Award, and NSF grant (#1257101) to D.C., NSF Grant (#1547901) to L.B., and by James S. McDonnell Foundation Scholar Awards to D.C. and to L.B.

References

- Alverson, H. (1994). *Semantics and experience: Universal metaphors of time in English, Mandarin, Hindi, and Sesotho*. Baltimore: Johns Hopkins University Press.
- Barsalou, L. W. (1983). Ad hoc categories. *Memory & Cognition*, *11*, 211-227.
- Boroditsky, L. (2000). Metaphoric structuring: Understanding time through spatial metaphors. *Cognition*, *75*(1), 1-28.
- Boroditsky, L. (2001). Does language shape thought? Mandarin and English speakers' conceptions of time. *Cognitive Psychology*, *43*(1), 1-22.
- Boroditsky, L., Fuhrman, O., & McCormick, K. (2011). Do English and Mandarin speakers think differently about time? *Cognition*, *118*(1), 123-129.
- Boroditsky, L. & Gaby, A. (2010). Remembrances of Times East: Absolute Spatial Representations of Time in an Australian Aboriginal Community. *Psychological Science*, *21*(11), 1635-1639.
- Bylund, E., & Athanasopoulos, P. (2017). The Whorfian time warp: Representing duration through the language hourglass. *Journal of Experimental Psychology: General*, *146*(7), 911.
- Cai, Z. G., & Connell, L. (2015). Space–time interdependence: Evidence against asymmetric mapping between time and space. *Cognition*, *136*, 268-281.

- Casasanto, D. (2017). The hierarchical structure of mental metaphors. In B. Hampe (Ed.), *Metaphor: Embodied cognition and discourse* (pp. 46-61). Cambridge: Cambridge University Press.
- Casasanto, D., & Boroditsky, L. (2003). Do we think about time in terms of space? In R. Alterman & D. Kirsh (Eds.), *Proceedings of 25th Annual Conference of the Cognitive Science Society* (pp. 216–221). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Casasanto, D., & Boroditsky, L. (2008). Time in the mind: Using space to think about time. *Cognition*, 106(2), 579–593.
- Casasanto, D., Boroditsky, L., Phillips, W., Greene, J., Goswami, S., Bocanegra-Thiel, et al. (2004). How deep are effects of language on thought? Time estimation in speakers of English, Indonesian, Greek, and Spanish. In K. Forbus, D. Gentner, & T. Regier (Eds.), *Proceedings of the 26th Annual Conference Cognitive Science Society*. (pp. 575–580). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Casasanto, D. & Bottini, R. (2014). Mirror-reading can reverse the flow of time. *Journal of Experimental Psychology: General*. 143(2), 473-9.
- Casasanto, D., Fotakopoulou, O., & Boroditsky, L. (2010). Space and Time in the Child's Mind: Evidence for a Cross-Dimensional Asymmetry. *Cognitive Science*, 34(3), 387-405.
- Casasanto, D., & Lupyan, G. (2015). All concepts are ad hoc concepts. In E. Margolis, & S. Laurence (Eds.), *The conceptual mind: New directions in the study of concepts* (pp. 543-566). Cambridge: MIT Press.

- Clark, H. H. (1973). Space, time, semantics and the child. In T. E. Moore (Ed.), *Cognitive development and the acquisition of language* (pp. 27-63). New York: Academic Press.
- Cordes, S., Gallistel, C. R., Gelman, R., & Latham, P. (2007). Nonverbal arithmetic in humans: Light from noise. *Perception & Psychophysics*, 69(7), 1185-1203.
- de Hevia, M. D., Izard, V., Coubart, A., Spelke, E. S., & Streri, A. (2014). Representations of space, time, and number in neonates. *Proceedings of the National Academy of Sciences*, 111(13), 4809-4813.
- de la Fuente, J., Santiago, J., Román, A., Dumitrache, C., & Casasanto, D. (2014). When you think about it, the past is in front of you: How culture shapes spatial conceptions of time. *Psychological Science*, 25(9), 1682-1690.
- Dolscheid, S., Shayan, S., Majid, A., & Casasanto, D. (2013). The thickness of musical pitch: Psychophysical evidence for linguistic relativity. *Psychological Science*, 24(5), 613–621.
- Evans, V. (2004). *The structure of time: Language, meaning and temporal cognition*. Amsterdam: John Benjamins.
- Fedden, S. and Boroditsky, L. (2012). Spatialization of time in Mian. *Frontiers in Psychology* 3:485. doi: 10.3389/fpsyg.2012.00485
- Fuhrman, O. & Boroditsky, L. (2010). Cross-Cultural Differences in Mental Representations of Time: Evidence from an Implicit Non-Linguistic Task. *Cognitive Science*. 34(8), 1430-1451.

- Fuhrman, O., McCormick, K., Chen, E., Jiang, H., Shu, D., Mao, S., Boroditsky, L. (2011). How linguistic and cultural forces shape conceptions of time: English and Mandarin time in 3D. *Cognitive Science*, 35(7), 1305-28.
- Gennari, S. P., Sloman, S. A., Malt, B. C., & Fitch, W. (2002). Motion events in language and cognition. *Cognition*, 83(1), 49-79.
- Gentner, D. (2001). Spatial metaphors in temporal reasoning. In M. Gattis (Ed.), *Spatial Schemas and Abstract Thought* (pp. 203-222). Cambridge: MIT Press.
- Gentner, D., Bowdle, B., Wolff, P., & Boronat, C. (2001). Metaphor is like analogy. In D. Gentner, K. J. Holyoak, & B. N. Kokinov (Eds.), *The analogical mind: Perspectives from cognitive science* (pp. 199–253). Cambridge, MA: MIT Press.
- Haspelmath, M. (1997). *From space to time: Temporal adverbials in the world's languages*. Munich & Newcastle: Lincom Europa.
- James, W. (1890). *The Principles of Psychology*. Henry Holt and Company.
- Lai, V.T. and Boroditsky, L. (2013) The immediate and chronic influence of spatio-temporal metaphors on the mental representations of time in English, Mandarin, and Mandarin-English speakers. *Frontiers in Psychology* 4:142.
- Lakoff, G. & Johnson, M. (1980). *Metaphors We Live By*. Chicago: University of Chicago Press.
- Lakoff, G., & Johnson, M. (1999). *Philosophy in the flesh: The embodied mind and its challenge to western thought*. Chicago: University of Chicago Press.

- Landau, B., Dessalegn, B., & Goldberg, A. (2010). Language and space: Momentary interactions. *Language, cognition and space: The state of the art and new directions. Advances in Cognitive Linguistics Series*, 51-78.
- Lourenco, S. & Longo, M. (2010). General magnitude representation in human infants. *Psychological Science*, 21:873-881.
- Moore, K. (2006). Space-to-time mappings and temporal concepts. *Cognitive Linguistics*, 17, 199–244.
- Meck, W. H., & Church, R. M. (1983). A mode control model of counting and timing processes. *Journal of Experimental Psychology – Animal Behavior Processes*, 9(3), 320–334.
- Merritt, D. J., Casasanto, D., & Brannon, E. M. (2010). Do monkeys think in metaphors? Representations of space and time in monkeys and humans. *Cognition*, 117(2), 191-202.
- Miles, L. K., Tan, L., Noble, G. D., Lumsden, J., & Macrae, C. N. (2011). Can a mind have two time lines? Exploring space–time mapping in Mandarin and English speakers. *Psychonomic Bulletin & Review*, 18(3), 598-604.
- Núñez, R., & Sweetser, E. (2006). Looking ahead to the past: Convergent evidence from Aymara language and gesture in the crosslinguistic comparison of spatial construals of time. *Cognitive Science*, 30(3), 401-450.
- Núñez, R., Cooperrider, K., Doan, D., & Wassmann, J. (2012). Contours of time: Topographic construals of past, present, and future in the Yupno valley of Papua New Guinea. *Cognition*, 124(1), 25-35.

- Ouellet, M., Santiago, J., Israeli, Z. & Gabay, S. (2010). Is the future the right time? *Experimental Psychology*, 57(4), 308–314.
- Papafragou, A., Hulbert, J., & Trueswell, J. (2008). Does language guide event perception? Evidence from eye movements. *Cognition*, 108(1), 155-184.
- Sinha, C., Sinha, V. D. S., Zinken, J., & Sampaio, W. (2011). When time is not space: the social and linguistic construction of time intervals and temporal event relations in an Amazonian culture. *Language and Cognition*, 3(1), 137-169.
- Slobin, D. I. (1996). From “thought and language” to “thinking for speaking.” In J. J. Gumperz & S. C. Levinson (Eds.), *Rethinking Linguistic Relativity* (pp. 70–96). Cambridge: Cambridge University Press.
- Song, S., Miller, K. D., & Abbott, L. F. (2000). Competitive Hebbian learning through spike-timing-dependent synaptic plasticity. *Nature Neuroscience*, 3(9), 919-926.
- Srinivasan, M., & Carey, S. (2010). The long and the short of it: on the nature and origin of functional overlap between representations of space and time. *Cognition*, 116(2), 217-241.
- Starr, A., & Brannon, E. M. (2016). Visuospatial working memory influences the interaction between space and time. *Psychonomic Bulletin & Review*, 23, 1839-1845.
- Traugott, E. (1978). On the expression of spatio-temporal relations in language. In J. H. Greenberg (ed.), *Universals of Human Language: Vol. 3. Word Structure*, 369–400. Stanford: Stanford University Press.
- Tversky, B., Kugelmass, S., & Winter, A. (1991). Cross-cultural and developmental trends in graphic productions. *Cognitive Psychology*, 23, 515-557.

Ünal, E. & Papafragou, A. (in press). Interactions Between Language and Visuo-spatial Representations. *Language Learning*.

Winawer, J. , Witthoft, N., Frank, M.C., Wu , L., & Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination . *Proceedings of the National Academy of Science*, 104 (19), 7780–7785.

Wood, G., Willmes, K., Nuerk, H. C., & Fischer, M. H. (2008). On the cognitive link between space and number: A meta-analysis of the SNARC effect. *Psychology Science Quarterly*, 50(4), 489.

Supplementary Online Materials for:
The Long and Small of It:
Language shapes duration estimation in speakers of English and Greek.

Appendix A. Stimuli used in Experiment 4.

Participants filled in the blanks with “longer/shorter” in the length-training task and “more/less” in the amount-training task.

Temporal Sentences

- | | |
|-------------------------------------|-------------------------------------|
| A blink is ____ than a concert. | A lunch is ____ than a hiccup. |
| A blink is ____ than a lunch. | A lunch is ____ than a pregnancy. |
| A blink is ____ than a movie. | A lunch is ____ than a semester. |
| A blink is ____ than a pregnancy. | A lunch is ____ than a sneeze. |
| A blink is ____ than a semester. | A lunch is ____ than a summer. |
| A blink is ____ than a summer. | A lunch is ____ than a winter. |
| A blink is ____ than a wedding. | A movie is ____ than a blink. |
| A blink is ____ than a winter. | A movie is ____ than a breath. |
| A breath is ____ than a concert. | A movie is ____ than a hiccup. |
| A breath is ____ than a lunch. | A movie is ____ than a pregnancy. |
| A breath is ____ than a movie. | A movie is ____ than a semester. |
| A breath is ____ than a pregnancy. | A movie is ____ than a sneeze. |
| A breath is ____ than a semester. | A movie is ____ than a summer. |
| A breath is ____ than a summer. | A movie is ____ than a winter. |
| A breath is ____ than a wedding. | A pregnancy is ____ than a blink. |
| A breath is ____ than a winter. | A pregnancy is ____ than a breath. |
| A concert is ____ than a blink. | A pregnancy is ____ than a concert. |
| A concert is ____ than a breath. | A pregnancy is ____ than a hiccup. |
| A concert is ____ than a hiccup. | A pregnancy is ____ than a lunch. |
| A concert is ____ than a pregnancy. | A pregnancy is ____ than a movie. |
| A concert is ____ than a semester. | A pregnancy is ____ than a sneeze. |
| A concert is ____ than a sneeze. | A pregnancy is ____ than a wedding. |
| A concert is ____ than a summer. | A semester is ____ than a blink. |
| A concert is ____ than a winter. | A semester is ____ than a breath. |
| A hiccup is ____ than a concert. | A semester is ____ than a concert. |
| A hiccup is ____ than a lunch. | A semester is ____ than a hiccup. |
| A hiccup is ____ than a movie. | A semester is ____ than a lunch. |
| A hiccup is ____ than a pregnancy. | A semester is ____ than a movie. |
| A hiccup is ____ than a semester. | A semester is ____ than a sneeze. |
| A hiccup is ____ than a summer. | A semester is ____ than a wedding. |
| A hiccup is ____ than a wedding. | A sneeze is ____ than a concert. |
| A hiccup is ____ than a winter. | A sneeze is ____ than a lunch. |
| A lunch is ____ than a blink. | A sneeze is ____ than a movie. |
| A lunch is ____ than a breath. | A sneeze is ____ than a pregnancy. |

A sneeze is ____ than a semester.
 A sneeze is ____ than a summer.
 A sneeze is ____ than a wedding.
 A sneeze is ____ than a winter.
 A summer is ____ than a blink.
 A summer is ____ than a breath.
 A summer is ____ than a concert.
 A summer is ____ than a hiccup.
 A summer is ____ than a lunch.
 A summer is ____ than a movie.
 A summer is ____ than a sneeze.
 A summer is ____ than a wedding.
 A wedding is ____ than a blink.
 A wedding is ____ than a breath.
 A wedding is ____ than a hiccup.

A wedding is ____ than a pregnancy.
 A wedding is ____ than a semester.
 A wedding is ____ than a sneeze.
 A wedding is ____ than a summer.
 A wedding is ____ than a winter.
 A winter is ____ than a blink.
 A winter is ____ than a breath.

 A winter is ____ than a concert.
 A winter is ____ than a hiccup.
 A winter is ____ than a lunch.
 A winter is ____ than a movie.
 A winter is ____ than a sneeze.
 A winter is ____ than a wedding.

Spatial Sentences

A bathtub is ____ than a keg.
 A bathtub is ____ than a kettle.
 A bathtub is ____ than a mug.
 A bathtub is ____ than a pitcher.
 A bathtub is ____ than a shot glass.
 A bathtub is ____ than a teacup.
 A bathtub is ____ than a teapot.
 A bathtub is ____ than a thimble.
 A keg is ____ than a bathtub.
 A keg is ____ than a lake.
 A keg is ____ than a mug.
 A keg is ____ than a shot glass.
 A keg is ____ than a swimming pool.
 A keg is ____ than a teacup.
 A keg is ____ than a thimble.
 A keg is ____ than an ocean.
 A kettle is ____ than a bathtub.
 A kettle is ____ than a lake.
 A kettle is ____ than a mug.
 A kettle is ____ than a shot glass.
 A kettle is ____ than a swimming pool.
 A kettle is ____ than a teacup.
 A kettle is ____ than a thimble.
 A kettle is ____ than an ocean.
 A lake is ____ than a keg.
 A lake is ____ than a kettle.
 A lake is ____ than a mug.
 A lake is ____ than a pitcher.
 A lake is ____ than a shot glass.
 A lake is ____ than a teacup.

A lake is ____ than a teapot.
 A lake is ____ than a thimble.
 A mug is ____ than a bathtub.
 A mug is ____ than a keg.
 A mug is ____ than a kettle.
 A mug is ____ than a lake.
 A mug is ____ than a pitcher.
 A mug is ____ than a swimming pool.
 A mug is ____ than a teapot.
 A mug is ____ than an ocean.
 A pitcher is ____ than a bathtub.
 A pitcher is ____ than a lake.
 A pitcher is ____ than a mug.
 A pitcher is ____ than a shot glass.
 A pitcher is ____ than a swimming pool.
 A pitcher is ____ than a teacup.
 A pitcher is ____ than a thimble.
 A pitcher is ____ than an ocean.
 A shot glass is ____ than a bathtub.
 A shot glass is ____ than a keg.
 A shot glass is ____ than a kettle.
 A shot glass is ____ than a lake.
 A shot glass is ____ than a pitcher.
 A shot glass is ____ than a swimming pool.
 A shot glass is ____ than a teapot.
 A shot glass is ____ than an ocean.
 A swimming pool is ____ than a keg.
 A swimming pool is ____ than a kettle.
 A swimming pool is ____ than a mug.
 A swimming pool is ____ than a pitcher.

- A swimming pool is ____ than a shot glass.
 A swimming pool is ____ than a teacup.
 A swimming pool is ____ than a teapot.
 A swimming pool is ____ than a thimble.
 A teacup is ____ than a bathtub.
 A teacup is ____ than a keg.
 A teacup is ____ than a kettle.
 A teacup is ____ than a lake.
 A teacup is ____ than a pitcher.
 A teacup is ____ than a swimming pool.
 A teacup is ____ than a teapot.
 A teacup is ____ than an ocean.
 A teapot is ____ than a bathtub.
 A teapot is ____ than a lake.
 A teapot is ____ than a mug.
 A teapot is ____ than a shot glass.
 A teapot is ____ than a swimming pool.
 A teapot is ____ than a teacup.
 A teapot is ____ than a thimble.
 A teapot is ____ than an ocean.
 A thimble is ____ than a bathtub.
 A thimble is ____ than a keg.
 A thimble is ____ than a kettle.
 A thimble is ____ than a lake.
 A thimble is ____ than a pitcher.
 A thimble is ____ than a swimming pool.
- A thimble is ____ than a teapot.
 A thimble is ____ than an ocean.
 An ocean is ____ than a keg.
 An ocean is ____ than a kettle.
 An ocean is ____ than a mug.
 An ocean is ____ than a pitcher.
 An ocean is ____ than a shot glass.
 An ocean is ____ than a teacup.
 An ocean is ____ than a teapot.
 An ocean is ____ than a thimble.
-

Appendix B. Space estimation results for Experiment 3

B1.0 Only the duration estimation trials were of interest with respect to our experimental predictions; these results are reported in the main text. We report the results of the length estimation trials here for the sake of completeness, and also because they provide an additional control measure, to demonstrate the specificity of the predicted effects.

B1.1 Space estimation, cross-domain interference effect

To evaluate the effect of irrelevant duration information on length estimation, each participant's spatial estimates (in pixels) were plotted as a function of the actual durations in milliseconds for which the growing lines remained on the screen. A line of best fit was computed, and the slope was used as a measure of cross-domain interference.

There was no significant effect of actual duration on length estimation, according to a 1-sample t-test comparing the slope of this effect against a slope of zero ($t(19)=1.08, p=.30$; fig S1a). That is, there was no significant effect of duration on space estimation.

B1.1.1 Testing for a space-time asymmetry

We tested whether the cross-domain effects of space and time were symmetric or asymmetric; that is, whether the cross-domain effect of space on time estimation was greater than the cross-domain effect of time on space estimation. As in Experiment 2, the results showed that irrelevant spatial information influenced duration estimates significantly more than irrelevant temporal information influenced spatial estimates ($t(38)=2.51, p=.02$; see fig. 5a in the main text for the relevant time estimation results). This “space-time asymmetry” underscores the specificity of the effects of

spatial information on time estimation, and was predicted on the basis of spatial metaphors for time (for discussion see Casasanto & Boroditsky, 2008).

B1.2 Space estimation, within-domain effect

The slope of the effect of actual change in pixels on estimated change in pixels was compared against a slope of zero, by one-sample t-test ($t(19)=13.29$, $p=.00001$; fig. S1b), thus spatial estimation was significantly more accurate than chance. The effect of actual change in pixels on estimated change in pixels did not differ from the effect of actual duration on estimated duration, according to a 2-sample t-test comparing the slopes of these effects ($t(38)=1.17$, $p=.25$; see fig. 5b in the main text for the relevant time estimation results).

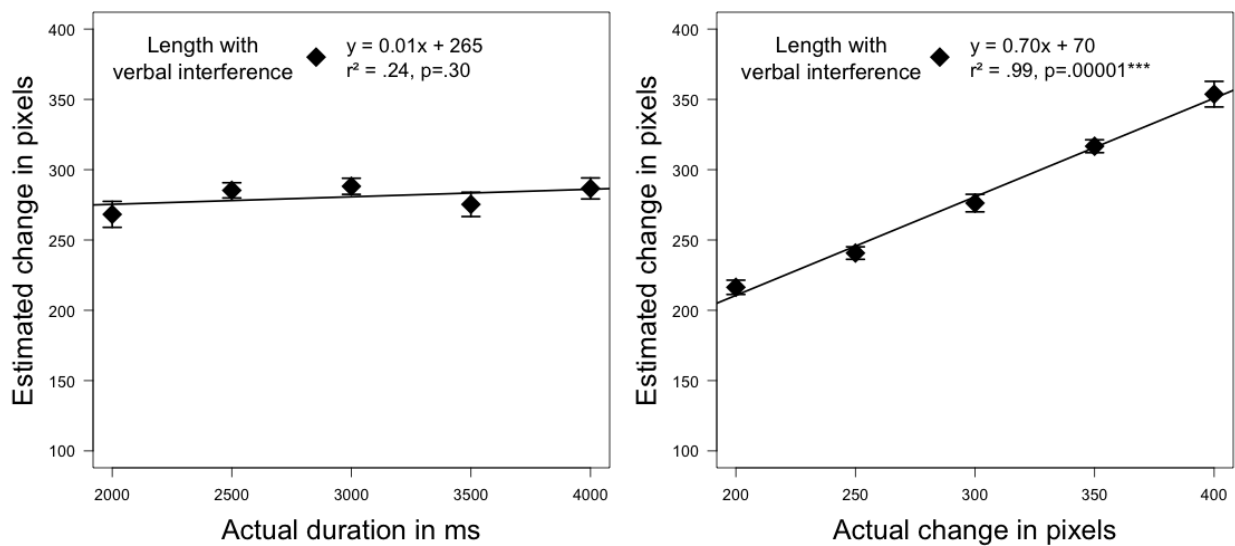


Figure B1. Space estimation results for Experiment 3. a (left): Cross-domain analysis. Effect of actual duration on estimated line length. b (right): Within-domain analysis. Effect of actual line length on estimated line length. Error bars show s.e.m.

Appendix C. Space estimation results for Experiment 4

C1.0 Only the duration estimation trials were of interest with respect to our experimental predictions; these results are reported in the main text. We report the results of the amount estimation trials here for the sake of completeness, and also because they provide an additional control measure, to demonstrate the specificity of the predicted effects.

C1.1 Space estimation, cross-domain interference effect

To evaluate the effect of irrelevant duration information on amount estimation, each participant's amount estimates (in pixels) were plotted as a function of the actual durations in milliseconds for which the containers remained on the screen. A line of best fit was computed, and the slope was used as a measure of cross-domain interference. Mean slopes were compared across the Length Trained and Amount Trained groups.

There was no significant effect of actual duration on amount estimation, in either training group, according to 1-sample t-tests comparing the slopes of these effects against a slope of zero (Length trained: $t(40)=0.38$, $p=.71$; Amount trained $t(46)=1.54$, $p=.13$; fig S2a). That is, there was no significant effect of duration on amount estimation.

C1.2 Space estimation, within-domain effect

For each training group the slope of the effect of actual change in pixels on estimated change in pixels was compared against a slope of zero, by one-sample t-tests (Length trained: $t(40)=22.66$, $p=.00001$; Amount trained $t(46)=16.84$, $p=.00001$, fig. S2b), thus spatial estimation was

significantly more accurate than chance, in both groups. The effect of actual change in pixels on estimated change in pixels did not differ between training groups, according to a 2-sample t-test comparing the slopes of these effects ($t(86)=0.15, p=.88$), thus differences in amount estimation, per se, cannot account for the difference in the cross-domain effect of amount on time estimation between the training groups (reported in the main text).

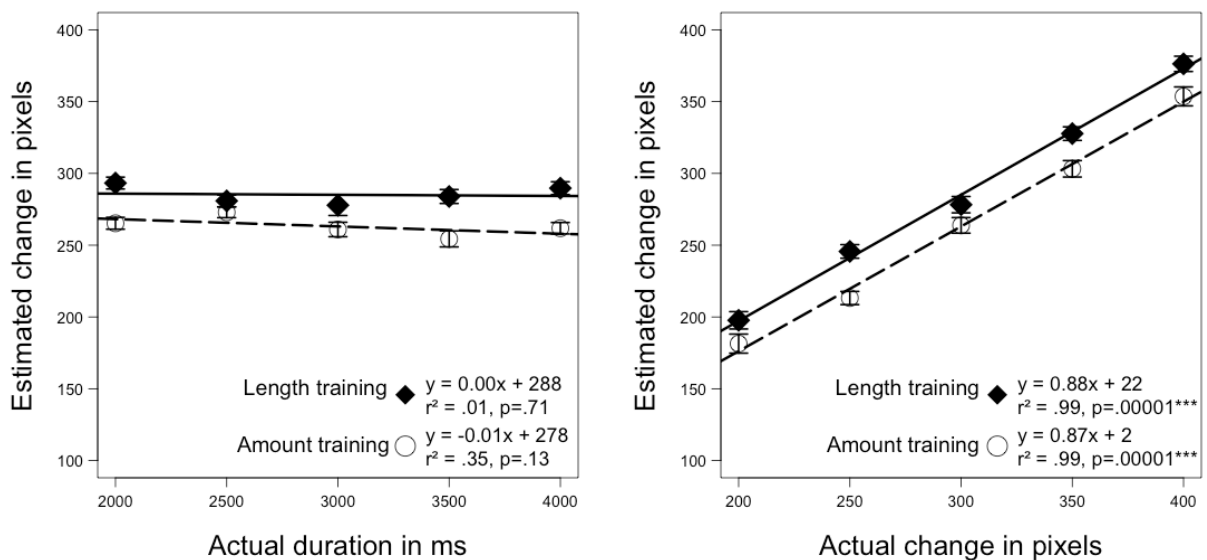


Figure C1. Space estimation results for Experiment 4, in Length-trained English speakers (diamonds) and amount-trained English speakers (circles). a (left): Cross-domain analyses. Amount estimates are plotted as a function of actual durations for which the containers appeared. b (right): Within-domain analyses. Amount estimates are plotted as a function of the actual change in pixels. Error bars show s.e.m.

Appendix D. Length interference in Greek speakers tested in Greece vs. USA

D1.0 Two groups of native Greek speakers estimated the durations of horizontal lines, while ignoring their spatial extents (length interference task); about half of the subjects were tested in Greece (in a predominantly Greek language environment), and the other half in the USA (in a predominantly English language environment). All subjects were given instructions in English.

D1.1 Method

D1.1.1 Participants

All participants (N=21) were Native Greek speakers who were fluent in English as a second language. About half of the participants were tested in Greece (n=11), at the International Neuropsychological Society's Summer institute in Xylokastro. The others (n=10) were tested at MIT in Cambridge, Massachusetts (USA).

D1.1.2 Materials and procedure

The materials and procedures were identical to those used in the length interference experiment (Experiment 2 in the main text), with the following exceptions. First, the line lengths ranged from 200-800 pixels in 75-pixel increments, as in Casasanto & Boroditsky (2008, Expt 2), whereas in the length interference experiment reported in the main text the lines ranged from 100-500 pixels in 50-pixel increments. Also, in the present version of the length interference experiment there was no box surrounding the area of the screen where the lines grew. Finally, here the language of the instructions was English for all subjects, whereas in Experiment 2 instructions were given in the participants' native language.

D2.0 Results

Data were analyzed as described in Experiment 2 of the main text.

D2.1 Time estimation, cross-domain interference effects

Of primary interest, we compared the cross-domain effect of irrelevant length information on time estimation in Greek speakers who were tested either in Greece or in the USA. The mean slope of the effect of actual length on estimated duration was significantly greater than zero, by a 1-sample *t*-test, in the Greeks who were tested in the USA ($t(9)=2.62, p=.03$), but not in the Greeks who were tested in Greece; ($t(10)=0.11, p=.91$; fig. S3a). The slope of the length interference effect was (marginally) greater in the subjects tested in the USA than in the subjects tested in Greece (difference of slopes=1.10; $t(19)=2.02, p=.058$).

In further analyses, we compared the slope of the length-time interference effect in the Greek speakers tested in Greece from the present experiment to the slope of the length-time interference effect in the Greek speakers from Experiment 2 in the main text (after normalizing the slopes to correct for differences in the stimulus parameters between experiments). Both groups comprised native Greek speakers tested in Greece; however, instructions were presented in Greek for the group reported in Expt. 2, but in English for the group reported here. The difference between groups did not approach significance (difference in normalized slopes=0.049; $t(31)=0.691, p=.495$), indicating that there was no measurable effect of the language in which the instructions were given.

We also compared the slope of the length-time interference effect in the Greek speakers tested in the USA in the present experiment to the slope of the length-time interference effect in the English speakers from Experiment 2 in the main text. Although one group comprised native Greek

speakers and the other native English speakers, both groups were living in Cambridge, MA when they were tested. The difference between groups did not approach significance (difference in normalized slopes=.046; $t(33)=0.693$, $p=.493$), indicating that habitual exposure to English was sufficient to produce native-English-like behavior on the length interference task, regardless of one's native language.

D2.2 Time estimation, within-domain effects

An analysis of within-domain performance showed that participants in both groups estimated duration with high accuracy, according to 1-sample t-tests comparing the mean slopes of the effect of actual duration on estimated duration against a slope of zero (Tested in Greece: $t(10)=8.44$, $p=.00001$; Tested in USA: $t(9)=8.36$, $p=.00002$; Figure S3c). The slopes did not differ between groups ($t(19)=0.21$, $p=.84$).

D2.3 Space estimation, cross-domain interference effect

We had no hypothesis about the space estimation trials; these analyses are included for the sake of completeness. To evaluate the effect of irrelevant duration information on length estimation, each participant's length estimates (in pixels) were plotted as a function of the actual durations in milliseconds for which the growing lines remained on the screen. There was no significant effect of actual duration on length estimation, in either language group, according to 1-sample t-tests comparing the slopes of these effects against a slope of zero (Tested in Greece: $t(10)=0.66$, $p=.52$; Tested in USA: $t(9)=1.39$, $p=.20$; fig S3b). That is, there were no significant effects of duration on length estimation.

D2.4 Space estimation, within-domain effect

For each language group, the slope of the effect of actual change in pixels on estimated change in pixels was compared against a slope of zero, by one-sample t-tests (Tested in Greece: $t(10)=11.68$, $p=.00001$; Tested in USA: $t(9)=4.54$, $p=.001$, fig. S3d); thus spatial estimation was significantly more accurate than chance, in both groups.

D3.0 Discussion

These results show a significant length-time interference effect in native Greek speakers tested in Greece, but not in native Greek speakers tested in the USA. The strength of this effect did not differ statistically between the Greeks tested in Greece in the present experiment versus the Greeks (also tested in Greece) reported in Expt. 2 of the main text. Likewise, the strength of the effect in did not differ statistically between the Greeks tested the USA in the present experiment versus the English speakers (also tested in the USA) reported in Expt. 2.

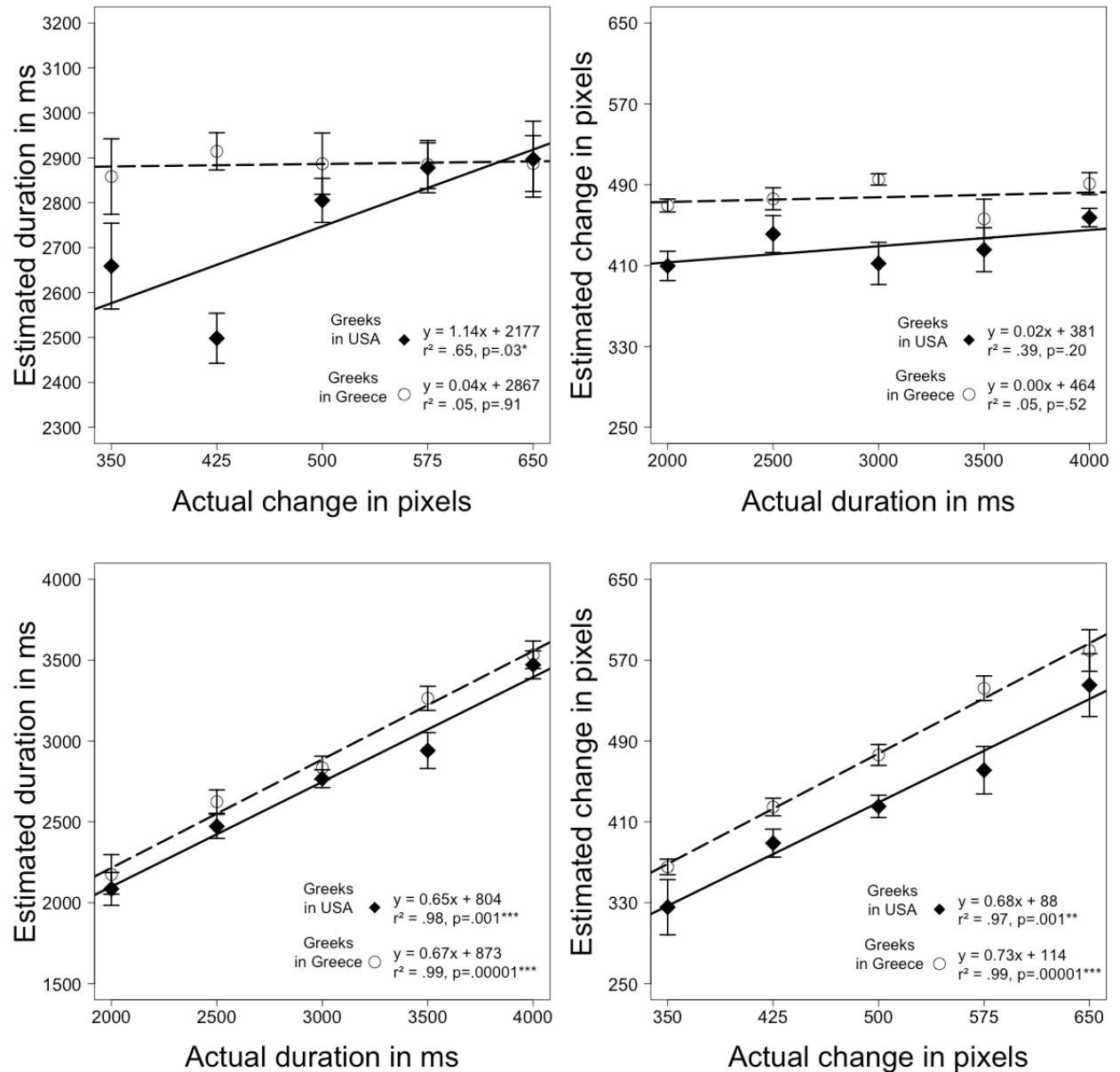


Figure D1. Results of the height interference experiment. a (top left): Cross-domain effect of actual length on duration estimation. b (top right): Cross-domain effect of actual duration on length estimation. c (bottom left): Within-domain effect of actual duration on estimated duration. d (bottom right): Within-domain effect of actual length on estimated length. Error bars show s.e.m.

Appendix E. Vertical growing lines experiment (height interference)

E1.0 English and Greek speakers estimated the durations of vertical lines, while ignoring their vertical spatial extents (height interference task).

E1.1 Method

E1.1.1 Participants

Native English speakers (N=32), tested at MIT, and Native Greek speakers (N=19), tested at the Aristotle University of Thessaloniki, participated for payment. Three English speakers and one Greek speaker were excluded for performing the task incorrectly, according to the criteria described in Experiment 2 of the main text. Thus, data from 29 English speakers and 18 Greek speakers are included in the analyses below.

E1.1.2 Materials and procedure

The materials and procedures were identical to those used in the length interference experiment (Experiment 2 in the main text), with the following exception. Rather than growing horizontally, each line grew vertically, from bottom to top. For temporal responses participants clicked twice on an hourglass icon, as in Experiment 2, but for the present experiment the icon appeared in either the bottom left or bottom right corner the 700 x 700-pixel box in which the stimuli appeared (location counterbalanced across participants). Each pair of mouse clicks indicated the beginning and ending of a remembered temporal interval. For spatial responses, participants clicked on an “X” icon in either the bottom left or bottom right corner the box (counterbalanced across participants), moved

the mouse upward in a straight line, and clicked a second time. Each pair of mouse clicks indicated the beginning and ending of a remembered spatial interval.

E2.0 Results

Data were analyzed as described in Experiment 2 of the main text.

E2.1 Time estimation, cross-domain interference effects

Of primary interest, we tested the cross-domain effect of irrelevant height information on time estimation in English and Greek speakers. The mean slope of the effect of actual height on estimated duration was significantly greater than zero, by 1-sample t-tests, in both language groups: (English: $t(28)=3.86$, $p=.0006$; Greek: $t(17)=3.47$, $p=.003$; fig. S3a).

Crucially, these slopes did not differ between groups $t(45)=0.17$, $p=.86$, and there was no significant interaction of language group (English, Greek) by task (horizontal lines, vertical lines; $F(1,90)=1.56$, $p=.22$).

E2.2 Time estimation, within-domain effects

An analysis of within-domain performance showed that participants in both groups estimated duration with high accuracy, according to 1-sample t-tests comparing the mean slopes of the effect of actual duration on estimated duration against a slope of zero (English: $t(28)=15.41$, $p=.00001$; Greek: $t(17)=13.47$, $p=.00001$; Figure S3c). The slopes did not differ between groups ($t(45)=0.42$, $p=.68$).

E2.3 Space estimation, cross-domain interference effect

We had no hypothesis about the space estimation trials; these analyses are included for the sake of completeness. To evaluate the effect of irrelevant duration information on height estimation, each participant's height estimates (in pixels) were plotted as a function of the actual durations in milliseconds for which the growing lines remained on the screen. There was no significant effect of actual duration on height estimation, in either language group, according to 1-sample t-tests comparing the slopes of these effects against a slope of zero (English: $t(28)=0.64$, $p=.53$; Greek: $t(17)=1.14$, $p=.27$; fig S3b). That is, there were no significant effects of duration on height estimation.

E2.4 Space estimation, within-domain effect

For each language group the slope of the effect of actual change in pixels on estimated change in pixels was compared against a slope of zero, by one-sample t-tests (English: $t(28)=6.02$, $p=.00001$; Greek: $t(17)=15.32$, $p=.00001$, fig. S3d), thus spatial estimation was significantly more accurate than chance, in both groups.

E3.0 Discussion

These results rule out the possibility that the differential effects of growing lines and filling containers on English and Greek speakers' duration estimates were due to the spatial orientation of the stimuli, as opposed to their dimensionality.

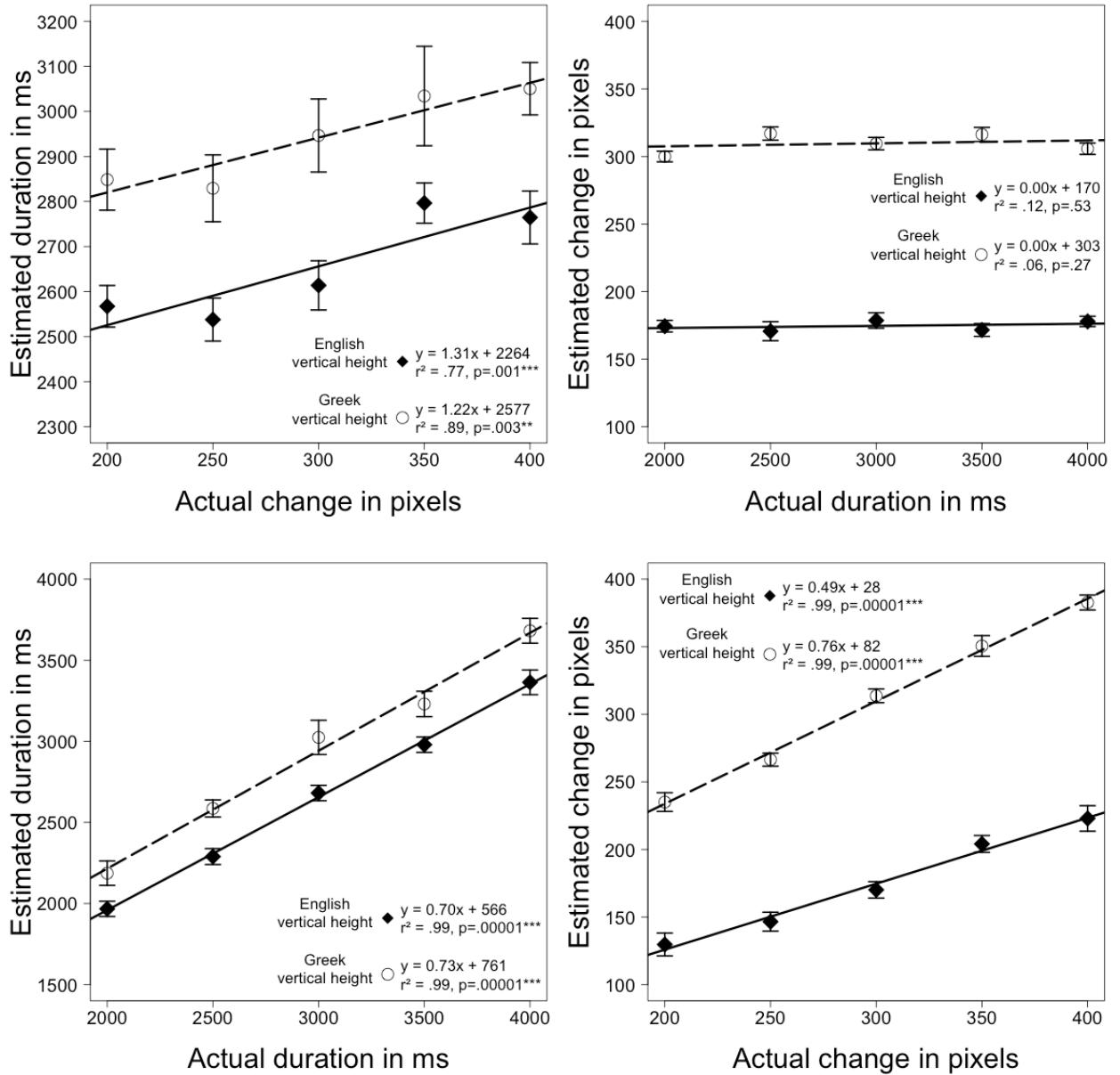


Figure E1. Results of the height interference experiment. a (top left): Cross-domain effect of actual height on duration estimation. b (top right): Cross-domain effect of actual duration on height estimation. c (bottom left): Within-domain effect of actual duration on estimated duration. d (bottom right): Within-domain effect of actual height on estimated height. Error bars show s.e.m.