Title:

The Whorfian time warp: Representing duration through the language hourglass

Running head:

The Whorfian time warp

Word count:

3489

Authors:

Emanuel Bylund [a, b] (corresponding author)

Panos Athanasopoulos [c]

Author affiliations:

[a)] Stellenbosch University, South Africa [b)] Stockholm University, Sweden

[c)] Lancaster University, UK

Author contact details:

[a, b] manne.bylund@biling.su.se; Department of General Linguistics, Private Bag X1, 7602 Matieland, Stellenbosch, South Africa

[c] p.athanasopoulos@lanc.ac.uk; Department of Linguistics and English Language, Bailrigg, Lancaster, UK, LA14YW

Author note: Parts of this work were presented by the second author as part of a public lecture at the Einstein Foundation, Berlin School of Mind and Brain, Germany, in 2015, and as part of a workshop on linguistic relativity at the Institute of Multilingualism, University of Fribourg, Switzerland, in 2016. We are grateful for the feedback those audiences provided. Athanasopoulos and Bylund contributed equally to the paper.

The Whorfian time warp:
Representing duration through the language hourglass

Human existence revolves around time, yet we cannot touch or see time. How do humans experience the passage of time, and how do they construct their mental representations of it? A key finding is that concepts from the more tangible domain of space are often used to represent the passage of time (for overviews; Bender & Beller, 2014; Núñez & Cooperrider, 2013). This tendency is particularly visible in language, where spatial concepts are often recruited to express both duration (e.g., "a <u>long</u> meeting") and succession ("Monday comes <u>before</u> Tuesday") (e.g., Clark, 1973; Traugott, 1978). A possibility then is that duration is primarily represented through spatial schemas, possibly due to an innate tendency to associate time with space (e.g., Lourenco & Longo, 2010; Srinivasan & Carey, 2010) and/or overlapping cortical processing of these domains (Walsh, 2003).

However, another possibility is that language not only reflects our inclination to associate time with space, but it actually shapes our thinking about it. In this view, the spatial schemas reflected in conceptual metaphors (e.g., "<u>long</u> time") provide the basis for our mental representation of duration (Lakoff & Johnson, 1980). This idea receives further support in the finding that our representations of abstract domains may be more susceptible to linguistic influence compared to concrete domains such as colour or objects (Borghi & Binkofski, 2014). Under this account, then, speakers of different languages would represent time differently (consistent with the linguistic relativity hypothesis, Whorf, 1956). For instance, speakers of English and Swedish, who primarily talk about time as "long" and "short", would represent duration differently from speakers of Spanish and Greek, who talk about time as "big" and "small" (Casasanto, 2005b; Lakoff & Johnson, 1980; Núñez & Cooperrider, 2013).

Here, we ask whether these different ways of talking about duration indeed lead speakers of different languages to represent the passing of time differently. We show that duration representation can change as a function of the presence of different linguistic cues in a simple duration estimation task. Surprisingly, switching the language context in the same bilingual individual also transforms the way they estimate duration. In three experiments, we implement a psychophysical task (Casasanto, 2005b) to investigate the effects of language on duration estimation in monolingual and bilingual speakers of Spanish and Swedish. These languages exhibit robust differences in the preferred patterns of duration expressions, with Swedish speakers preferring distance-based metaphors (e.g., "long time") and Spanish speakers preferring amount-based metaphors (e.g., "much time") (Figure 1).

In the task, participants either reproduced the duration of computer-generated animations showing either a container that filled gradually with liquid, or a line gradually growing on the screen. The fill level of the containers and the distance of the lines were irrelevant for duration reproduction. Instead, these displacements tested to what extent participants were able to disregard spatial information when estimating duration (Casasanto & Boroditsky, 2008). If language usage patterns correlate with humans' representation of duration, it can be predicted that because Spanish speakers talk about duration in terms of amount, the fill level of the containers should interfere with their duration estimations to a larger extent than among Swedish speakers, who do not talk about duration in the same way. Conversely, because Swedish speakers talk about duration in terms of distance, the distance of the growing lines should interfere with duration estimation to a larger extent than is the case for Spanish speakers. We expect these cross-linguistic differences to be more apparent in the presence (Experiment 1) rather than in the absence (Experiment 2) of linguistic cues. If language meddles with temporal cognition, then we also expect Swedish-Spanish bilinguals to behave differently in this task depending on the language of the cues (Experiment 3).

Additionally, findings from a different perceptual domain (colour) suggest that language is more likely to affect perceptual judgements of stimuli that are more difficult to discriminate (colours that are closer together in the spectrum) than of stimuli that are more easily discriminated (colours that are further apart in the spectrum) (Winawer et al., 2007). The current psychophysical task (Casasanto, 2005b) includes stimuli with varying degrees of duration and physical growth. We compared spatial interference for 'extreme stimuli' (those with the longest and shortest durations and physical growth in the stimulus set) and 'medium stimuli' (those with medium durations and physical growth) in all experiments. Because the perceptual properties of extreme stimuli in a given stimulus set are more easily discerned (Winawer et al., 2007), we hypothesized that medium stimuli would be more difficult to process, thus yielding a stronger language effect on temporal cognition. A pre-experimental study independently confirmed that extreme stimuli elicited more precise duration reproductions than medium stimuli (Experiment 1).

Our findings show that duration estimation varied as a function of the presence of linguistic cues (Experiments 1 and 2), language context of operation (Experiment 3), and stimulus type (all experiments). We conclude that human temporal cognition is malleable, as part of a highly adaptive computational system, in which language flexibly functions as a source of information for duration estimation.

Experiment 1. Crosslinguistic differences in time reproduction in the presence of verbal prompts

Method

Participants

Forty native Spanish speakers and 40 native Swedish speakers were randomly assigned to either the container condition or the line condition. Participants were University students in Madrid and Stockholm, respectively. In the absence of previous studies, we assumed a medium-sized effect (η_p^2=.06) of the interaction between language group and stimulus type, alpha level of .05, along with default sample correlation and nonsphericity values (Faul et al., 2007), which yielded a recommended sample size of 36 participants (Experiments 1&2). We thus slightly over-sampled (40 participants), factoring in certain attrition (see below).

Materials

The Growing Lines and Filling Containers Experiments (Casasanto, 2005b) was used to assess duration reproduction.

Line condition: computer-generated black lines grew from left to right against a white background. Nine different line distances, ranging from 100 to 500 pixels (with 50 pixels' increments), and 9 different line durations, ranging from 1000 ms to 5000 ms (with 500 ms increments) were fully crossed to produce 81 unique line stimuli.

Container condition: a 600 pixels high x 500 pixels wide black frame against a white background represented an empty container. Containers were filled in black from the bottom and up. Nine fill levels (ranging from 100 to 500 pixels with 50 pixels' increments) were crossed with

9 durations (ranging from 1000 to 5000 ms with 500 ms increments) to produce 81 distinct container stimuli.

All animations were presented in a 700 x 700 pixels' field.

Procedure

Participants were tested individually by a native speaker of the relevant language, using a 15.6" laptop. Each stimulus (line or container) was presented twice, resulting in a total of 162 trials. Half of the times, the participants estimated stimulus duration, and the other half they estimated displacement (distracter task). Spatial reproduction trials and temporal reproduction trials were randomly intermingled. Instructions for duration reproduction were void of spatial expressions (e.g., "estimate the time it took…" instead of "estimate how <u>long</u> it took…").

A prompt preceded each stimulus, indicating whether duration or displacement was to be estimated. The prompt consisted of a symbol (an hourglass for duration and a cross for displacement estimation) and a verbal label. For duration reproduction, the labels were default expression of duration in Spanish (*duración*) and Swedish (*tid*). For displacement reproduction, the labels were either *avstånd* ('distance') or *mängd* ('amount') in Swedish, and *distancia* ('distance') or *cantidad* ('amount') in Spanish. Participants reproduced duration by clicking the computer mouse once, waiting the appropriate time, and clicking again. Displacement was reproduced by clicking the mouse once, moving it the appropriate distance/height, and clicking again. Presentation orders were fully randomised.

Following previous studies on time estimation (Casasanto, 2005b; Casasanto & Boroditsky, 2008), participants were removed if they estimated distance instead of time, or their overall duration estimations were markedly inaccurate (if the slope of the correlation between actual and estimated duration was <.5). Five Swedish and 5 Spanish participants were consequently removed.[1]

Stimulus norming

The precision of duration reproductions for medium stimuli (2000-4000 ms and 200-400 pixels) and extreme stimuli (1000, 1500, 4500, 5000 ms and 100, 150, 450, 500 pixels) was measured by

[1] Extended, repetitive psychophysical tasks like the present one inevitably yield certain participant exclusion (e.g., 22% exclusion rate in Casasanto & Boroditsky, 2008; 30% in Casasanto, 2005b), presumably due to impatience and/or fatigue.

calculating the discrepancy between actual duration and reproduced duration. As the current design rests on a robust effect of stimulus type, we assumed a large effect size ($\eta_p^2 = .14$) and thus sampled twenty-four Spanish and 24 Swedish speakers (not part of the main experiments), who were randomly allocated to reproduce duration in either the lines or the container condition. Two participants were excluded due to poor performance (see above).

For the line condition, a 2 (group: Spanish vs. Swedish) x 2 (stimulus type: medium vs. extreme) mixed Anova yielded a significant main effect of stimulus type, F (22, 1) = 12.639, $p < .01$, $\eta_p^2 = .367$, showing that reproductions were significantly more precise for extreme than for medium stimuli. A similar result obtained for the container condition, F (22, 1) = 12.155, $p < .01$, $\eta_p^2 = .390$ (see Table 1). No significant interactions or main effects of group were found.

Design

Following Boroditsky and Casasanto (2008), the degree to which stimulus displacement interfered with duration reproduction was computed by calculating the duration estimates for each fill level/distance, and then entering these two variables into a regression to obtain the slope. Higher slopes indicated proneness to estimate larger displacements as having longer duration, and were thus indices of greater spatial interference. Slopes were calculated separately for medium and extreme stimuli, and served as the dependent variable in the analyses.

Participants' overall accuracy for reproducing duration and displacement (distracter task) was controlled for (Supplemental Material).

Results

In the container condition, a 2 (language: Spanish vs. Swedish) x (stimulus type: extreme vs. medium) mixed Anova showed a significant interaction, F (1, 33) = 10.59, $p = 0.003$, $\eta_p^2 = 0.24$. Spanish speakers showed more interference than Swedish speakers, but only for medium stimuli ($p = 0.009$). In the lines condition, a mixed Anova with the same variables revealed a significant interaction, F (1, 33) = 7.62, $p = 0.009$, $\eta_p^2 = 0.19$. Swedish speakers showed more interference than Spanish speakers only for medium stimuli ($p = 0.001$) (Figure 2A). Such crosslinguistic differences conform to previous evidence showing similar patterns for English and Greek speakers reproducing duration in the presence of linguistic cues (Casasanto, 2005b), and are compatible with the linguistic relativity hypothesis (Whorf, 1956).

Experiment 2: Do crosslinguistic differences in time reproduction persist in the absence of verbal prompts?

To test the persistence of the effects found in Experiment 1 in a strictly non-verbal context, we removed the verbal labels in the prompts appearing before each trial.

Method

Forty different native Swedish speakers (students in Stockholm) and 40 different native Spanish speakers (students in Madrid) were randomly assigned to either the container condition or the line condition. Three Spanish and 5 Swedish participants were removed due to poor performance.

Materials and procedures were identical to experiment 1, with the crucial exception that the label in the prompt was removed, leaving only the symbol (hourglass or cross).

Results

Contrary to what linguistic relativity would predict, we found no interaction between language and stimulus type, in either the line condition (F (1, 33) = 0.99, p = 0.328, η_p^2 = .029) or the container condition (F (1, 33) = 0.03, p = 0.871, η_p^2 = .001). Instead, interference patterns of Spanish and Swedish speakers were strikingly similar (Figure 2B). Moreover, both language groups seemed to display slightly greater spatial interference in the lines condition than in the containers condition. There were no significant main effects.

Experiment 3: Does switching the prompt language trigger different interference patterns within the same individual?

The results of Experiments 1 and 2 give rise to the possibility that human temporal cognition flexibly adapts to environmental constraints such as the presence or absence of verbal cues. Comparing the performance of bilingual speakers operating in different language contexts allows for this critical test. If temporal cognition is indeed adaptive as a function of language, then the same bilingual individual should exhibit different interference patterns depending on the language context.

Method

Seventy-four adult Swedish–Spanish bilinguals performed either the container condition or line condition. The sample size was based on the parameters outlined in Experiment 1 and adapted for within-subjects analyses. Materials and procedures were identical to Experiment 1 with the exception that the distracter task was removed to minimise fatigue. Nine participants were removed due to poor performance. Participants took the experiment twice, once with Spanish and once with Swedish prompt labels (order counterbalanced).

Results

For the bilinguals in the containers condition, a 2 (language context: Spanish vs. Swedish) x 2 (stimulus type: extreme vs. medium) repeated measures Anova showed a significant interaction, F (1, 34) = 9.38, p = 0.004, η_p^2 = 0.22. Specifically, for medium stimuli, interference levels changed significantly in the direction predicted by the prompt language (p = 0.018) (Figure 2C).

Likewise, for bilinguals in the lines condition, a 2 x 2 repeated measures Anova yielded a significant interaction between language context and stimulus type, F (1, 29) = 5.18, p = 0.028, η_p^2 = 0.15. Again, interference for medium stimuli changed significantly as the prompt language changed (p = 0.04) (Figure 2C). No significant main effects of language were found for extreme stimuli in either condition.

Discussion and Conclusions

Using a simple psychophysical task in different language populations we have shown that language can influence humans' representation of the passage of time. Linguistic cues yielded language-specific spatial interference in time estimation, while in the absence of such cues language-specific interference disappeared. Further, language-specific interference was confined to difficult discriminations. Our approach to manipulate different language prompts in the same population of bilinguals revealed context-induced adaptive behaviour: prompts in language A induced language A-congruent spatial interference. When the prompt switched to language B, interference became language B-congruent instead. To our knowledge, this study provides the first psychophysical demonstration of shifting duration representations within the same individual as a function of language context.

A central question then concerns the likely mechanism that underlies the reported effects. According to the label-feedback hypothesis (Lupyan, 2012), the perception of a given stimulus may be warped through co-activation of its corresponding verbal label. The degree of top-down modulation of language can then be regulated by manipulating the presence of the corresponding verbal label (Lupyan & Ward, 2013). In the current study, the label-feedback hypothesis would predict greater linguistic modulation in the presence of language prompts rather than in their absence, which was found in Experiments 1 and 2. However, the prompts used here do not contain the labels that would have to be involved in such modulation (i.e., spatio-temporal metaphors), and the fine temporal and spatial increments in the stimuli do not readily trigger labelling in the same way that colours or objects would (Casasanto & Bottini, 2014; Dolscheid, Shayan, Majid, & Casasanto, 2013). More importantly though, if the stimuli would readily lend themselves to labelling, the label-feedback hypothesis would predict greater language-specific spatial interference for the extreme stimuli, as these would be more likely to activate labels such as ‘long’ and ‘short’. However, it was precisely for the extreme stimuli that no language-specific effects could be detected.

The current results may still be interpreted as an online warping of temporal judgments, but one that goes beyond a one-to-one mapping between label and percept. Under the predictive processing account (A. Clark, 2013; Hohwy, 2013; Kanai, Komura, Shipp, & Friston, 2015), a percept is co-constructed in a continuous interplay between downward flowing predictions and upward flowing sensory signals. The downward predictions are the system’s expectations given its knowledge about the world and the context, but are also continuously updated by perceptual input. Language constitutes one such source of prior knowledge, but also a potentially powerful bottom-up contextual cue (Lupyan & Clark, 2015).

Under this account, the verbal prompts used in the current study constitute contextual cues (rather than ready stimulus labels), triggering spreading activation of semantically related linguistic knowledge (Çukur, Nishimoto, Huth, & Gallant, 2013), in this case language-specific duration metaphors, which in turn transitorily warp temporal processing. The powerful nature of this warping is most poignantly seen in the bilingual speakers: varying the language of the prompts changes what prior knowledge is recruited, thus yielding different language-specific modulations of duration estimation, within the same individual. Our findings also support the idea that the more ambiguous the input is, the less weight it has on perception, thus increasing the system's reliance on prior knowledge (Lupyan & Clark, 2015). This explains the reliance on linguistic knowledge in the harder-to-process medium stimuli, as opposed to the easier-to-process extreme stimuli, where no top-down linguistic modulation was found.

Predictive processing is Bayesian optimality-driven, striving for a complete match between the bottom-up signal and the top-down expectations. The strong inclination humans have to associate time and distance (Srinivasan & Carey, 2010) could explain both language groups' tendency to reproduce long line distances as longer in time in the absence of linguistic cues (Experiment 2). This raises a question regarding the precise impact of linguistic cues on a system that may already be optimised to predict correspondence between distance and duration. In Spanish speakers, the inclination seems to be overridden by the presence of linguistic cues triggering language-derived expectations, as indexed by the minimal spatial interference in the line condition (Experiment 1). In Swedish speakers, however, this inclination matches the linguistically derived knowledge, which could increase the expectation of distance-duration correspondence. Their difference in interference slopes between Experiment 1 and Experiment 2 is however within the error margin, indicating a potential ceiling effect of the influence of top-down information on duration estimation.

By showing that language, under certain circumstances, can transform the basic psychophysical experience of the passing of time, the current findings align with important advances in linguistic relativity research highlighting that the effects of language on cognitive processing is not an either-or phenomenon, but are instead highly dynamic and context-bound (e.g., Athanasopoulos et al., 2015; Athanasopoulos & Bylund, 2013; Kersten et al., 2010; Montero-Melis, Jaeger, & Bylund, 2016). Specifically, the findings show that the strong inclination among humans to represent time through spatial schemas may be modulated by the specific ways these schemas are instantiated in different languages. The attested modulations conform to behavioural and neural evidence of the role of words as targeted manipulations that selectively enhance or mute the influence of any other aspect of prior knowledge, including which specific language the system may rely upon to inform top-down modulations (Athanasopoulos et al., 2015; Lupyan & Clark, 2015). Our approach to vary the language context within the same bilingual individual reveals that the observed patterns of behaviour are language-induced (rather than the artifact of some between-subjects, extra-linguistic cultural factor (Casasanto, 2005a; Levinson & Majid, 2013). This resonates with the emerging view of a highly adaptive human computational system, in which language can serve as a critical source of information for processing experience.

References

Athanasopoulos, P., & Bylund, E. (2013). Does grammatical aspect affect motion event cognition? A cross-linguistic comparison of English and Swedish speakers. *Cognitive Science*, *37*(2), 286–309. https://doi.org/10.1111/cogs.12006

Athanasopoulos, P., Bylund, E., Montero-Melis, G., Damjanovic, L., Schartner, A., Kibbe, A., … Thierry, G. (2015). Two Languages, Two Minds Flexible Cognitive Processing Driven by Language of Operation. *Psychological Science*, 0956797614567509. https://doi.org/10.1177/0956797614567509

Bender, A., & Beller, S. (2014). Mapping spatial frames of reference onto time: A review of theoretical accounts and empirical findings. *Cognition*, *132*(3), 342–382. https://doi.org/10.1016/j.cognition.2014.03.016

Borghi, A. M., & Binkofski, F. (2014). *Words as Social Tools: An Embodied View on Abstract Concepts*. New York, NY: Springer New York. https://doi.org/10.1007/978-1-4614-9539-0

Casasanto, D. (2005a). Crying 'Whorf'. *Science (New York, N.Y.)*, *307*(5716), 1721-1722-1722.

Casasanto, D. (2005b). *Perceptual foundations of abstract thought* (Doctoral dissertation). MIT, Massachussets.

Casasanto, D., & Boroditsky, L. (2008). Time in the mind: Using space to think about time. *Cognition*, *106*(2), 579–593. https://doi.org/10.1016/j.cognition.2007.03.004

Casasanto, D., & Bottini, R. (2014). Spatial language and abstract concepts. *Wiley Interdisciplinary Reviews: Cognitive Science*, *5*(2), 139–149. https://doi.org/10.1002/wcs.1271

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(03), 181–204. https://doi.org/10.1017/S0140525X12000477

Clark, H. H. (1973). Space, time, semantics, and the child. In T. Moore (Ed.), *Cognitive development and the acquisition of language* (pp. 27–63). New York: Academic Press.

Çukur, T., Nishimoto, S., Huth, A. G., & Gallant, J. L. (2013). Attention during natural vision warps semantic representation across the human brain. *Nature Neuroscience*, *16*(6), 763–770. https://doi.org/10.1038/nn.3381

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. https://doi.org/10.1177/0956797612457374

Hohwy, J. (2013). *The Predictive Mind*. OUP Oxford.

Kanai, R., Komura, Y., Shipp, S., & Friston, K. (2015). Cerebral hierarchies: predictive processing, precision and the pulvinar. *Phil. Trans. R. Soc. B*, *370*(1668), 20140169. https://doi.org/10.1098/rstb.2014.0169

Kersten, A. W., Meissner, C. A., Lechuga, J., Schwartz, B. L., Albrechtsen, J. S., & Iglesias, A. (2010). English speakers attend more strongly than Spanish speakers to manner of motion when classifying novel objects and events. *Journal of Experimental Psychology: General*, *139*(4), 638–653. https://doi.org/10.1037/a0020507

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. University of Chicago Press.

Levinson, S. C., & Majid, A. (2013). The island of time: Yélî Dnye, the language of Rossel Island. *Cultural Psychology*, *4*, 61. https://doi.org/10.3389/fpsyg.2013.00061

Lourenco, S. F., & Longo, M. R. (2010). General Magnitude Representation in Human Infants. *Psychological Science*, *21*(6), 873–881. https://doi.org/10.1177/0956797610370158

Lupyan, G. (2012). Linguistically modulated perception and cognition: the label-feedback hypothesis. *Frontiers in Psychology*, *3*, 114–128. https://doi.org/10.3389/fpsyg.2012.00054

Lupyan, G., & Clark, A. (2015). Words and the World Predictive Coding and the Language-Perception-Cognition Interface. *Current Directions in Psychological Science*, *24*(4), 279–284. https://doi.org/10.1177/0963721415570732

Lupyan, G., & Ward, E. (2013). Language can boost otherwise unseen objects into visual awareness. *Proceedings of the National Academy of Sciences*, *110*(35), 14196–14201. https://doi.org/10.1073/pnas.1303312110

Montero-Melis, G., Jaeger, T. F., & Bylund, E. (2016). Thinking Is Modulated by Recent Linguistic Experience: Second Language Priming Affects Perceived Event Similarity. *Language Learning*, *66*(3), 636–665. https://doi.org/10.1111/lang.12172

Núñez, R., & Cooperrider, K. (2013). The tangle of space and time in human cognition. *Trends in Cognitive Sciences*, *17*(5), 220–229. https://doi.org/10.1016/j.tics.2013.03.008

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. https://doi.org/10.1016/j.cognition.2010.05.005

Traugott, E. C. (1978). On the expression of spatio-temporal relations in language. *Universals of Human Language*, *3*, 369–400.

Walsh, V. (2003). A theory of magnitude: Common cortical metrics of time, space and quantity. *Trends in Cognitive Sciences*, *7*(11), 483–488. https://doi.org/10.1016/j.tics.2003.09.002

Whorf, B. L. (1956). *Language, thought, and reality: Selected writings (edited by J. B. Carroll)*. Cambridge, Mass.: MIT Press.

Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., & Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Sciences*, *104*(19), 7780–7785. https://doi.org/10.1073/pnas.0701644104