



Brief article

Re-evaluating evidence for linguistic relativity: Reply to Boroditsky (2001) [☆]

David January ^{a,*}, Edward Kako ^b

^a *University of Pennsylvania, USA*

^b *Swarthmore College, USA*

Received 1 February 2006; revised 30 May 2006; accepted 7 July 2006

Abstract

Six unsuccessful attempts at replicating a key finding in the linguistic relativity literature [Boroditsky, L. (2001). Does language shape thought?: Mandarin and English speakers' conceptions of time. *Cognitive Psychology*, 43, 1–22] are reported. In addition to these empirical issues in replicating the original finding, theoretical issues present in the original report are discussed. In sum, we conclude that Boroditsky (2001) provides no support for the Whorfian hypothesis.

© 2006 Elsevier B.V. All rights reserved.

Keywords: Linguistic relativity; Whorfian hypothesis; Temporal cognition; Language and thought

1. Introduction

Recently, the psychological community has seen renewed interest in the linguistic relativity hypothesis, the claim (Whorf, 1939/1956, 1941/1956) that the language one speaks structures how one thinks (for varying perspectives, see, e.g., Li & Gleitman, 2002; Levinson, Kita, Haun, & Rasch, 2002; Lowenstein & Gentner, 2005; Papafr-

[☆] This manuscript was accepted under the editorship of Jacques Mehler.

* Corresponding author. Tel.: +1 215 898 0360; fax: +1 215 573 9247.

E-mail address: djanuary@sas.upenn.edu (D. January).

agou, Massey, & Gleitman, 2002 and the volume edited by Gentner & Goldin-Meadow, 2003). Though most researchers currently reject the strongest form of this hypothesis, which claims that one can only think in accordance with the grammatical structure of one's language, (though see, e.g., Roberson, Davidoff, Davies, & Shapiro, 2005), several researchers have adopted "weaker" versions of the hypothesis, claiming language "influences" or "suggests" thought patterns or default modes of interpreting the world rather than determining cognition absolutely (e.g., Hunt & Agnoli, 1991). One influential study, receiving 35 citations in PsycInfo and ~55 citations in Google Scholar, is Boroditsky (2001), which serves as the point of departure for the work reported below.

In a study on the mental representation of time, Boroditsky (2001) noted that temporal relations are cross-linguistically expressed using spatial metaphors. For example, in English we can talk of "pushing a meeting back" in time or "looking forward to tomorrow". Though such metaphors are pervasive in the world's languages, there is at least one dimension on which they can differ: orientation. As the examples just given suggest, English typically uses a horizontally oriented spatial metaphor for temporal relations.¹ Events are described as points or expanses on a horizontal line and can be moved along it. In Mandarin Chinese, by contrast, spatial metaphors for temporal relations tend to be vertically oriented, as the example in (1), from Boroditsky (2001), shows.

- (1) Space
 tā xià le shān méi yǒu
 has she descended the mountain or not?
 Time
xià ge yuè
next (or following) month

Boroditsky found that these prevalent spatial metaphors had lasting effects on temporal cognition. Specifically, she reported that English speakers exhibit cross-domain priming from horizontal spatial relations to temporal relations whereas Mandarin speakers exhibit such priming from vertical spatial relations to temporal relations.

The current report raises some theoretical and empirical issues for Boroditsky (2001) that challenge the validity of its experimental outcome and by extension its potential larger implications. Specifically, we report six unsuccessful attempts to replicate the basic finding of Boroditsky (2001) that English speakers think of time as horizontal. Additionally, we point out some apparent empirical and theoretical inconsistencies in the original report that would pose a serious challenge to Boroditsky's interpretation of her results, even if they were replicable.

¹ This is indeed only a typicality difference, as Boroditsky (2001) notes, for English has such expressions as "traditions are handed *down* through generations" and "the midterm is coming *up*". Mandarin also makes use of both horizontal and vertical metaphors.

2. Boroditsky (2001): procedure and results

Subjects in the Boroditsky (2001) paradigm were presented with a series of pictures and sentences, with their task being to indicate whether the sentences were true or false. The stimuli were composed of targets and primes. The primes were composed of pictures depicting spatial relations (for example, two balls, one above the other) accompanied by sentences that described them (see Fig. 1). Targets were sentences describing the order of months in a year. Half of the targets (the spatial targets) used a spatial metaphor to describe the order (e.g., *June comes before August*). The other half (the temporal targets) used purely temporal terms to describe the order (e.g., *June comes earlier than August*).

On experimental trials, subjects saw 2 primes, followed by 1 target. On only these trials, the first prime was always false, the second prime was always true, and the target was always true. On filler trials, subjects saw 2 screens, which were randomly assigned true and false primes and targets. Examples of target trials are given in Fig. 1.





 <p>The white worm is ahead of the black worm.</p>	 <p>The black worm is ahead of the white worm.</p>	<p>December comes after November.</p>
 <p>The white ball floats below the black ball.</p>	 <p>The black ball floats above the white ball.</p>	<p>August comes earlier than October.</p>

Fig. 1. Target stimuli from Boroditsky (2001). Each row represents 1 target trial. The first prime was always false, the second prime always true, and the target always true. Prime orientation and target type were fully crossed within subjects.

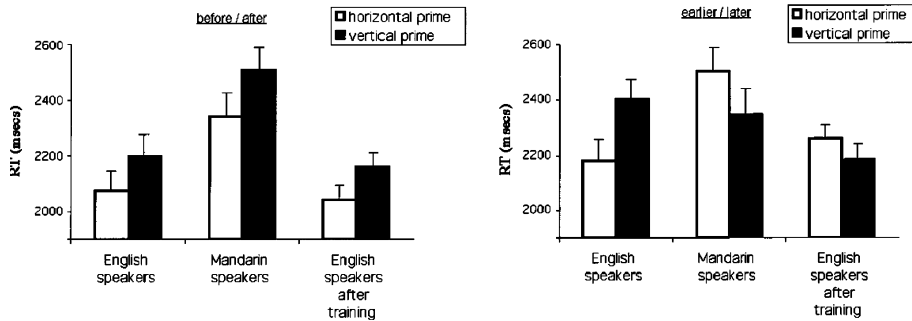


Fig. 2. Results from Boroditsky (2001). The left panel (before/after) corresponds to the spatial condition, and the right panel to the temporal condition.

Boroditsky (2001) predicted that both English speakers and Mandarin–English bilinguals would be faster to verify the spatial target sentences following horizontally oriented primes because the spatial targets make use of spatial terms, which are horizontally oriented. For the temporal targets, however, the prediction was that both language groups would rely on their “habitual” mental representations of temporal events. Thus, English speakers would make use of a horizontal representation because of the metaphors prevalent in their language, and Mandarin–English bilinguals would use a vertical representation because of the metaphors prevalent in their native language. These predictions were confirmed, as shown in Fig. 2. According to Boroditsky, the facilitation she measured was due to cross-domain priming. These results seem to comport with the idea of metaphoric structuring of abstract domains, which maintains that cognitive domains to which we do not have immediate sensory access, such as time, are structured according to our sensory experience via conceptual mapping (See, e.g., Boroditsky, 2000; Gentner, Imai, & Boroditsky, 2002; Lakoff & Johnson, 1999). Here, time would seem to be structured according to experience with spatial relations, to which we have sensory access via touch, locomotion, vision, and audition; language serves to determine any remaining characteristics of the abstract domain, such as orientation.

3. Attempted replications

This section reports on six different attempts to replicate Boroditsky’s finding that English speakers think of time horizontally, specifically that they exhibit cross-domain priming from spatial relations to temporal relations only when the spatial primes are horizontal. As in Boroditsky (2001), only trials on which the subjects correctly responded to the primes and responded within the time limit were analyzed. Overall, we observed target error rates very similar to, though typically slightly higher than, those in Boroditsky (2001) (Boroditsky: 7.1%, Attempt 1: 10%, Attempt 2: 11.6%, Attempt 3: 8.2%, Attempt 4: 11.4%, Attempt 5: 5.8%, Attempt 6: 7%). Additionally, as in Boroditsky (2001), only trials in which the subject responded correctly

to the target were included in reaction time analysis. Thus, only RTs on correct targets (following correct primes) were entered into a 2(prime orientation: horizontal or vertical) \times 2(target type: spatial or temporal) within-subjects, repeated-measures ANOVA.

Since Boroditsky (2001) submitted the data from English speakers and Mandarin–English bilinguals to separate ANOVAs, we directly parallel the conditions under which the Whorfian hypothesis was tested for the English speaking population. It is true that we are only testing one language population, and so we cannot speak to whether Boroditsky’s effect is replicable with a Mandarin-speaking population. However, the failure to find a crucial facilitation effect in one language group renders moot any discussion of the potential differences *between* language groups. Given the experimental setup that Boroditsky herself has provided, the effects of language on thought ought to be detectable within a single language. If such effects are not detectable under these circumstances, then language cannot be the mechanism underlying them when they appear to be found in cross-linguistic contexts.

3.1. Attempts 1–4: procedure and results

In replication attempts (RAs) 1–4, Swarthmore college undergraduates were compensated \$7 for participation (RA1: $N = 22$, RA2 $N = 24$, RA3: $N = 24$, RA4: $N = 24$. Compare with Boroditsky’s (2001) $N = 26$). Subjects were presented with temporal and spatial primes as in Boroditsky (2001). Several rounds of personal communication with Boroditsky from summer 2002 to early winter 2003 served to correct any differences in procedure between our experiments and hers. These differences were as follows. In RAs 1–6, subjects saw each target only once, whereas in Boroditsky (2001) they saw each target twice (once after each orientation of prime). Since the analysis was not performed on difference scores, this difference should be negligible. The spatial primes for RA1 were composed of geometrical figures on a grey background. Communication with Boroditsky revealed that original stimuli were exclusively of “worms” or “balls” (of the type shown in Fig. 1) and all subsequent replications (including 5 and 6) used the very same image files used in Boroditsky (2001). In RAs 1–4, there was an enforced maximum RT of 5 s for every prime and target such that if the subject failed to respond within the time limit the next prime appeared automatically. In Boroditsky (2001), this time limit was enforced post-hoc such that only trials in which subjects responded within the time limit were included in the analysis. In RAs 1–3, the response keys for true and false were adjacent on the keyboard whereas communication with Boroditsky revealed that in the original study the response keys were on opposite sides of the keyboard. Additionally, in RAs 1–3 filler trials were composed of three screens each, whereas in Boroditsky (2001), they were only two screens each. Such a difference would not have been apparent to the subject at any rate, though, since the subject simply saw a successive presentation of screens without interruption (except for a break halfway through, present in all RAs and in the original). Additionally, an untimed practice session was added before the timed practice session in RA3 to assess whether lack of familiarity with the materials or procedure might be causing subjects

to perform counter to expectations. This alteration had no effect on the response pattern. Finally, in RAs 4–6, targets were accidentally segregated such that subjects saw only temporal targets in the first half of the experiment (before a break) and only spatial targets in the second half of the experiment. In Boroditsky (2001), and in RAs 1–3, the target types were randomly interspersed (an additional replication was begun to address this issue; see footnote 2).

The results for each of RAs 1–4 are presented in Fig. 3a–d. As is evident from the graphs, in no case was there a significant benefit for English speakers in the horizontal prime condition, contra Boroditsky (2001).

For RA1: There were no main effects of Prime Orientation ($F(1,22) = .477, p > .4$) or Target Type ($F(1,21) = .774, p > .3$). There was also no interaction ($F(1,22) = 1.968, p > .1$).

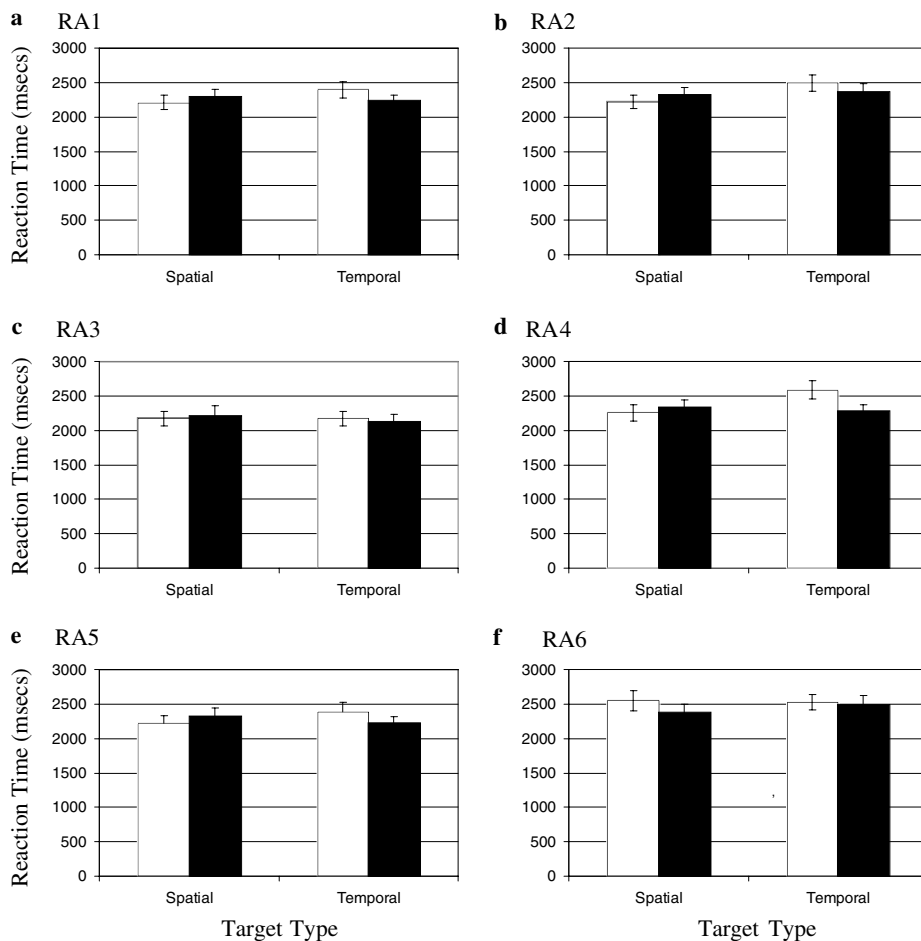


Fig. 3. Results of the non-replications. White bars indicate horizontal primes, black bars indicate vertical primes. Error bars indicate standard errors of the mean.

For RA2: There was no main effect of Prime Orientation ($F(1, 23) = .052, p > .8$). There was, however, a significant effect of Target Type ($F(1, 23) = 5.621, p < .05$), such that temporal targets took longer than spatial targets. Boroditsky (2001) also found such a result and attributed it to the greater number of syllables to be read in the temporal condition. That we also found this result in the current non-replication suggests that our finding of no difference between the horizontal prime condition and the vertical prime condition is not due simply to a lack of power or sensitivity but is caused by a genuine non-difference in the two conditions. Additionally, there was no interaction ($F(1, 23) = 1.085, p > .3$).

For RA3: There were no main effects of Prime Orientation ($F(1, 23) = .001, p > .9$) or Target Type ($F(1, 23) = .684, p > .4$), and there was no interaction ($F(1, 23) = .684, p > .4$).

For RA4: There was no main effect of Prime Orientation ($F(1, 23) = 2.769, p > .1$), but there was a marginal main effect of Target Type ($F(1, 23) = 2.949, p < .1$) and additionally a significant interaction such that targets preceded by horizontal primes had *longer* RTs in the temporal condition ($F(1, 23) = 10.373, p < .01$). This finding is clearly contrary to the prediction from a relativistic viewpoint. Again, the finding of a significant result here suggests that the non-difference between the horizontal prime condition and the vertical prime condition is not simply the result of insufficient power or sensitivity of the measure.

It is potentially worrying that we found this “counter-Whorfian” effect. If language patterns do in fact determine mental conceptions, why does our result run counter to the expected facilitation effect? If, on the other hand, language patterns do not determine mental conceptions, why is there any difference here at all? Whatever the source of this effect, however, our main point still stands: Boroditsky (2001) cannot be used to support the Whorfian position. The paradigm produces results too unreliable to base such a strong conclusion on.

3.2. Attempts 5 and 6

For RAs 5–6, there are no known differences in the implementation of the experimental procedure from Boroditsky (2001) except for the segregation of targets and the one-time versus two-time presentation of targets.² RA5 used 25 Swarthmore College undergraduates; RA6 used 25 University of Pennsylvania undergraduates compensated \$8. As in RAs 1–4, in no case is there a benefit to English speakers in the horizontal prime condition (ssee Fig. 3e–f).

For RA5: There were no main effects of Prime Orientation ($F(1, 24) = .512, p > .4$) or Target Type ($F(1, 24) = .276, p > .6$) and no interaction between these two ($F(1, 24) = 2.751, p > .1$).

² We started yet another replication attempt to try to eliminate these remaining differences. With eight subjects, the pattern still did not conform to that in Boroditsky (2001), so the attempt was halted. RTs were as follows (standard errors in parentheses): spatial targets, horizontal primes – 2539.369 (198.880); spatial targets, vertical primes – 2523.553 (217.946); temporal targets, horizontal primes – 2684.489 (272.598); temporal targets, vertical primes – 2496.375 (262.089).

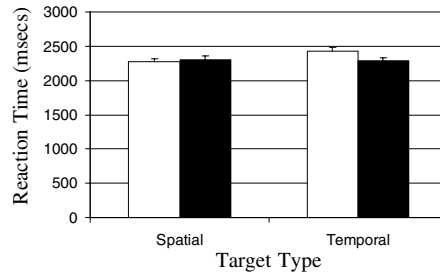


Fig. 4. Results, collapsing across all six non-replications.

For RA6: There was a marginal main effect of Prime Orientation ($F(1, 24) = 3.898, p = .06$) such that RTs to targets following horizontal primes were *longer* than RTs to targets following vertical primes (as in RA2). There was no main effect of Target Type ($F(1, 24) = .625, p > .4$) and no interaction ($F(1, 24) = 1.725, p > .2$).

3.3. Collapsing across replication attempts

The failure to replicate Boroditsky's (2001) original finding is made more striking when collapsing across all 6 failed replications. This move is not unwarranted since none of the differences between the various RAs has any theoretical significance, given the question at hand. Again, as shown in Fig. 4, there is no benefit to English speakers in the horizontal prime condition. Indeed, English speakers were overall *slower* following horizontal primes ($F(1, 143) = 5.825, p < .05$), exactly contrary to Boroditsky's (2001) prediction. Moreover, there was a significant interaction between Prime Orientation and Target Type such that targets following horizontal primes only had significantly longer RTs in the temporal condition ($F(1, 143) = 7.560, p < .01$).

4. Interpretive issues

Setting these empirical issues aside for the moment, Boroditsky (2001) also reports a result that vastly complicates, if not outright contradicts, the claim that native language spatial metaphors for temporal relations structure the time domain. In Experiment 3 of this series, English speakers were trained in a “new way to talk about time” by receiving 90 examples of sentences making use of the vertical metaphor for time found in Mandarin (e.g., *Bill Clinton was president below Ronald Reagan*). After these 90 training sentences, English-speaking subjects were run through the original procedure reported above, and there was a reversal of response pattern such that English speakers seem to have taken on the “Mandarin” way of thinking about time (see Fig. 2 (Boroditsky (2001)'s Fig. 4), English speakers after training). Boroditsky (2001) takes this as evidence for language influencing thought because it

seems that exposure to a new metaphor for temporal relations can cause a (temporary?) change in mental representation of time. However, this claim seems entirely incompatible with the claim from Experiment 1 in the same paper that native language metaphors have lasting effects on temporal representation. If 90 exposures to a new temporal metaphor can change English speakers' representation of time, why cannot years of a new metaphor change Mandarin speakers' representation? Or, conversely, if Mandarin speakers persist in a default mental representation of temporal relations as vertically oriented after years of speaking English, why don't English speakers persist in a default mental representation of time as horizontally oriented after only 90 examples of a different metaphor? Boroditsky (2001) is silent on these questions. (See Gleitman & Papafragou, 2005, for related discussion of this apparent interpretive contradiction.)

5. Discussion and conclusions

The present work has demonstrated that the Whorfian finding that native language structures mental representation of temporal relations is distinctly unreliable. Boroditsky (personal communication) does report being able to replicate the effect with English speakers. However, we maintain that such results must be treated with extreme caution in light of our repeated inability to do so. Given this questionable nature of the empirical basis for the claim of such relativistic effects on mental representation of time and the apparently contradictory evidence provided within the initial paper, it is clear that the results in Boroditsky (2001) cannot be cited as evidence for the Linguistic Relativity Hypothesis.

A remaining objection to our results needs addressing, namely, that effects of a horizontal timeline have been documented in Western cultures using other experimental paradigms, suggesting that, indeed, English speakers think of time as horizontally oriented (e.g., Gevers, Reynvoet, & Fias, 2003). Such an objection is too broad, however, as our main argument is that the results in Boroditsky (2001) are unreliable and hence should not be used as evidence of the relativity effect; we are not claiming that all documentations of linguistic relativity should be disregarded based on one set of studies. Additionally, it is unclear what the mechanism for the effects is in such studies, and the claim here is crucially that language is the mechanism. Westerners may indeed typically order events in time horizontally from left to right when forced to do so (Tversky, Kugelmass, & Winter, 1991), but this does not necessarily reveal that their *internal* representation of time is so structured, nor that *language*, and not simply cultural convention, has driven the effect.

Acknowledgements

The work reported in this paper was accomplished predominantly in preparation for the first author's undergraduate thesis at Swarthmore College and was supported in large part by a grant from the Howard Hughes Medical Institute. Thanks go to

Lila Gleitman and John Trueswell for much advice, discussion, and encouragement, and to Samantha Crane for data collection at the University of Pennsylvania. Particular thanks to Lera Boroditsky for helping us to minimize procedural differences and for providing her stimuli for our replication attempts.

References

- Boroditsky, L. (2000). Metaphoric structuring: understanding time through spatial metaphors. *Cognition*, 75, 1–18.
- Boroditsky, L. (2001). Does language shape thought?: Mandarin and English speakers' conceptions of time. *Cognitive Psychology*, 43, 1–22.
- Gentner, D., & Goldin-Meadow, S. (2003). *Language in mind: Advances in the study of language and thought*. Cambridge: MIT Press.
- Gentner, D., Imai, M., & Boroditsky, L. (2002). As time goes by: evidence for two systems in processing space time metaphors. *Language and Cognitive Processes*, 17, 537–565.
- Gevers, W., Reynvoet, B., & Fias, W. (2003). The mental representation of ordinal sequences is spatially organized. *Cognition*, 87, B87–B95.
- Gleitman, L., & Papafragou, A. (2005). Language and thought. In K. Holyoak & R. Moore (Eds.), *The Cambridge Handbook of Thinking and Reasoning* (pp. 633–662). NY: Cambridge University Press.
- Hunt, E., & Agnoli, F. (1991). The Whorfian hypothesis: a cognitive psychology perspective. *Psychological Review*, 98(3), 377–389.
- Lakoff, G. & Johnson, M. (1999). *Philosophy in the flesh*. NY: Basic Books
- Li, P., & Gleitman, L. (2002). Turning the tables: Language and spatial reasoning. *Cognition*, 83, 265–294.
- Levinson, S., Kita, S., Haun, D., & Rasch, B. (2002). Returning the tables: language affects spatial reasoning. *Cognition*, 84, 155–188.
- Loewenstein, J., & Gentner, D. (2005). Relational language and the development of relational mapping. *Cognitive Psychology*, 50, 315–353.
- Papafragou, A., Massey, C., & Gleitman, L. (2002). Shake, rattle, 'n' roll: The representation of motion in language and cognition. *Cognition*, 84, 189–219.
- Roberson, D., Davidoff, J., Davies, I., & Shapiro, L. (2005). Color categories: evidence for the cultural relativity hypothesis. *Cognitive Psychology*, 50, 378–411.
- Tversky, B., Kugelmass, S., & Winter, A. (1991). Cross-cultural and developmental trends in graphic productions. *Cognitive Psychology*, 23, 515–557.
- Whorf, B. (1939/1956). The relation of habitual thought and behavior to language. In J. B. Carroll (Ed.), *Language, thought, and reality: selected writings of Benjamin Lee Whorf* (pp. 134–159). Cambridge, MA: The MIT Press.
- Whorf, B. (1941/1956). Language, mind, and reality. In J. B. Carroll (Ed.), *Language, thought, and reality: selected writings of Benjamin Lee Whorf* (pp. 134–159). Cambridge, MA: The MIT Press.