LANGUAGE LEARNING
A Journal of Research in Language Studies
Language Learning ISSN 0023-8333

EMPIRICAL STUDY

Thinking Is Modulated by Recent Linguistic Experience: Second Language Priming Affects Perceived Event Similarity

Guillermo Montero-Melis,[a] T. Florian Jaeger,[b] and Emanuel Bylund[a,c]

[a]Stockholm University, [b]University of Rochester, and [c]Stellenbosch University

Can recent second language (L2) exposure affect what we judge to be similar events? Using a priming paradigm, we manipulated whether native Swedish adult learners of L2 Spanish were primed to use path or manner during L2 descriptions of scenes depicting caused motion events (encoding phase). Subsequently, participants engaged in a nonverbal task, arranging events on the screen according to similarity (test phase). Path versus manner priming affected how participants judged event similarity during the test phase. The effects we find support the hypotheses that (a) speakers create or select ad hoc conceptual categories that are based on linguistic knowledge to carry out nonverbal tasks, and that (b) short-term, recent L2 experience can affect this ad hoc process. These findings further suggest that cognition can flexibly draw on linguistic categories that have been implicitly highlighted during recent exposure.

Keywords motion events; linguistic relativity; event similarity; bilingual cognition; Spanish; Swedish

---

We thank Maya Hickmann and Henriette Hendriks for sharing their stimuli, and Maria Donoso for data collection. We are grateful for feedback on earlier presentations of these results at the Human Language Processing Lab (University of Rochester), the Experience and Cognition Lab (University of Chicago), and the Semantic Typology Lab (University at Buffalo), and to three anonymous reviewers who helped improve the manuscript.

This article has been awarded an Open Data badge. All data are publicly accessible via the Harvard Dataverse at http://dx.doi.org/10.7910/DVN/QOD8ZU. Learn more about the Open Practices badges from the Center for Open Science: https://osf.io/tvyxz/wiki.

Correspondence concerning this article should be addressed to Guillermo Montero-Melis, Centre for Research on Bilingualism, Stockholm University, SE-106 91, Stockholm, Sweden. E-mail: montero-melis@biling.su.se

DOI: 10.1111/lang.12172

Introduction

Is the way we think modified by second language (L2) learning? Research on bi- and multilingualism is increasingly focused on whether language learning affects cognition beyond language itself. That is, does learning a new language affect mental processes such as perception, categorization, similarity assessment, memory recognition, and reasoning? A growing body of evidence suggests that learning a new language indeed affects the way we think about reality in a variety of domains (see Bylund & Athanasopoulos, 2014b, for a recent overview). This research has focused on the effects of long-term L2 exposure on thought, typically comparing cognition in bilinguals against that of monolinguals. Together with other lines of research (discussed below), work on long-term L2 exposure suggests that linguistically based categories can be drawn on during nonverbal tasks (e.g., categorization, similarity assessment) that do not involve overt language use.[1] Here, we build on and extend this line of research. Whereas previous work has investigated whether nonverbal behavior reflects linguistic categories that differ crosslinguistically (e.g., Gennari, Sloman, Malt, & Fitch, 2002; Kersten et al., 2010; Papafragou, Massey, & Gleitman, 2002; Papafragou & Selimis, 2010; and others reviewed below), we ask whether nonverbal behavior can be modulated by recent linguistic experience.

Specifically, we ask whether L2 speakers will judge event similarity differently as a function of recent L2 exposure. To address this question we combine two paradigms: a priming paradigm allowing us to manipulate the semantic content of L2 primes and a subsequent nonverbal arrangement task that assesses perceived event similarity. This design directly compares performance on a nonverbal cognitive task as a function of manipulated L2 exposure. We hold language background and priming L2 constant, thus making it possible to compare bilinguals against each other (as proposed by Ortega, 2013), rather than with respect to monolingual baselines. The particular domain under investigation is caused motion, and the bilingual group being tested consists of native Swedish speakers of L2 Spanish.

Effects of Language on Cognition and ad hoc Category Formation

Do speakers of different languages think differently? This issue, known as the principle of linguistic relativity (Whorf, 1956), has generated vigorous debate in the cognitive sciences (e.g., Levinson, 2003; Pinker, 1994). Empirical research on linguistic relativity suggests that the answer to this question falls somewhere between the two extremes: Thought is neither determined by language nor is it entirely dissociated from it. On the one hand, there is by now ample evidence that the specific ways in which languages make us talk about reality can affect how we mentally represent colors (e.g., Roberson, Davidoff, Davies, & Shapiro, 2005), objects (e.g., Imai, Saalbach, & Stern, 2010), time (e.g., Boroditsky, Fuhrman, & McCormick, 2011), space (e.g., Haun, Rapold, Janzen, & Levinson, 2011), and motion (e.g., Athanasopoulos & Bylund, 2013a; Kersten et al., 2010; for recent overviews, see Gleitman & Papafragou, 2012; Wolff & Holmes, 2011). On the other hand, these effects are not always obtained (e.g., Chen, 2007; January & Kako, 2007; Li & Gleitman, 2002; Malt, Sloman, Gennari, Shi, & Wang, 1999; Papafragou et al., 2002). For example, language effects on nonverbal behavior vary depending on testing condition, such that effects might disappear under linguistic interference (e.g., Trueswell & Papafragou, 2010), and depending on task demands, appearing only when language provides a suitable strategy to solve the task (e.g., Gennari et al., 2002).

These mixed results suggest that linguistically based categories provide suitable tools to carry out certain nonverbal tasks, but that reliance on these categories varies depending on contextual factors, such as task demands and recent linguistic experience. While the exact nature of the mechanisms underlying transfer from verbal to nonverbal behavior are still under debate (e.g., linguistic intrusion: Papafragou & Selimis, 2010; analogical reasoning: Gentner, 2010; selectivity and enrichment: Landau, Dessalegn, & Goldberg, 2010; label-feedback: Lupyan, 2012, described in more detail below), there is now increasing agreement that such transfer can take place. We follow Casasanto and Lupyan (2015) in referring to this idea as ad hoc category formation. Under this view, conceptual representations are dynamic and context dependent; linguistic experience is one of the factors that can influence which conceptual categories we draw upon to carry out a particular task.

One account that fleshes out the idea of ad hoc formation of conceptual representations is Lupyan's (2012) label-feedback hypothesis. This account seeks to explain why effects of language can be deep, in the sense of affecting even low-level processing, yet task dependent and vulnerable to experimental interventions. According to the label-feedback framework, visual perception is a hybrid visuo-linguistic experience: Verbal labels and conceptual categories are coactivated in a feedback loop whereby visual stimuli (e.g., the picture of a dog) activate linguistic labels (the word "dog") in a bottom-up process, while at the same time, verbal labels in a top-down fashion activate perceptual features that are diagnostic of the category referred to by the label. The label-feedback hypothesis predicts that this feedback loop can be flexibly modulated by relevant manipulations. For instance, verbal interference should suppress or even disrupt the mutual feedback between labels and categories, which is indeed observed (Trueswell & Papafragou, 2010). Conversely, priming participants with verbal labels before exposing them to visual stimuli should facilitate the mutual feedback between label and categories, biasing participants to group the stimuli into categories that are defined along the same dimensions as the linguistic categories with which they have been primed. For example, Lupyan and colleagues showed that labelling novel categories (such as two species of aliens) accelerated category learning and improved accurate classification (Lupyan, Rakison, & McClelland, 2007).

Given this intricate relationship between language and thought, research on linguistic relativity is moving away from the dichotomous question of whether or not language influences nonverbal cognition (cf. Gleitman & Papafragou, 2012), either because a strict distinction between verbal and nonverbal cognition is rejected in the first place (Lupyan, 2012), because linguistically mediated cognition is argued to be sufficiently pervasive in everyday thought to warrant its study (Kersten et al., 2010; Slobin, 1996) or because it is claimed that the real power of language effects might lie precisely in the flexible ways it can affect thought (Landau et al., 2010). Hence, the aim increasingly becomes to understand under what circumstances cognitive processes are affected by language and the extent to which this influence is dependent on task-specific demands. Bilingualism offers an interesting test case in this respect.

How Learning a New Language (Re)shapes Thinking

If speakers of different languages think differently to a certain extent, what happens when you learn a new language? The available literature has documented two major patterns of cognitive restructuring in bilinguals: conceptual convergence and conceptual switching (see Pavlenko, 2005, for theoretically possible patterns). First, research suggests that bilinguals' first language (L1) and their L2 conceptual domains may converge. That is, when compared to monolingual speakers of each of their languages, bilinguals will often show evidence of intermediate conceptual categories. For example, using a forced choice triad paradigm, Cook and colleagues showed that Japanese-English bilinguals who had to categorize complex objects (e.g., a cork pyramid) according to either shape (pyramid) or material (cork), performed in between monolingual patterns: Bilinguals were more likely than native speakers of English to make material-based choices, but were less so than Japanese monolinguals. Moreover, those bilinguals that had lived for a longer period of time in the American context more closely resembled English monolinguals than those bilinguals who had stayed in America for a shorter time, suggesting that bilinguals shift toward L2 categorization patterns as a function of L2 exposure (Cook, Bassetti, Kasai, Sasaki, & Takahashi, 2006).

Similarly, Greek-English and Japanese-English bilinguals' conceptual categories of colors have been found to shift from L1 to L2 patterns as a function of frequency of L2 use and cultural immersion in the L2 country (Athanasopoulos, Damjanovic, Krajciova, & Sasaki, 2011). Related demonstrations of this pattern have also been documented for different language pairs and conceptual domains (Bylund & Athanasopoulos, 2014a; Park & Ziegler, 2014). Findings like these show that the L2 can affect nonverbal behavior and that it tends to do so more strongly with increasing L2 exposure.

Another line of studies suggests that the effect of L2 exposure can also be modulated at shorter time scales, resulting in conceptual switching. For example, bilinguals seem to switch between conceptual representations as a function of the language they are currently using. Athanasopoulos and colleagues found that German-English bilinguals switch their preference regarding goal orientation of motion events as a function of testing language. In a German experimental context, they were more likely to categorize events as tending toward a goal, resembling German monolinguals; while in an English context, they perceived the same events to be less goal oriented, resembling English monolinguals (Athanasopoulos et al., 2015). A related pattern has been documented in studies on temporal cognition (e.g., Miles, Tan, Noble, Lumsden, & Macrae, 2011).

The overall picture that emerges is the following: Bilinguals switch between distinct conceptual representations as a function of language context, performing more like L1 monolinguals in a L1 context and more like L2 monolinguals in a L2 context; yet their patterns tend to fall in between monolingual baselines (but see Filipović, 2011). Before outlining how the present study contributes to the fields of linguistic relativity and bilingual cognition, we briefly review the domain under study—motion events.

Motion Event Cognition

Motion events offer a suitable test case for the linguistic relativity hypothesis, because motion constitutes a universal cognitive domain grounded in visual perception, yet we find variation in how motion events are described across languages (Talmy, 2000). The key crosslinguistic contrast in motion event descriptions concerns the information expressed in the main verb root (Talmy, 2000). In satellite-framed languages, like English or Swedish, the main verb root conveys information about the manner of motion (e.g., The boy *walks*/*jumps*/*runs* up the stairs), while the path or trajectory of motion is conveyed in a verb satellite (e.g., The boy walks *up/down* the stairs). In verb-framed languages, like Spanish, the main verb root instead typically conveys information about the path of motion (El chico *subió/bajó* las escaleras, "The boy *ascended/descended* the stairs"), and manner information is often omitted, although it can potentially be encoded in a gerund (e.g., El chico subió las escaleras *caminando/corriendo*, "The boy ascended the stairs *walking/running*") or in the main verb (El chico *corrió* hacia la casa, "The boy ran towards the house"). In this study, we took advantage of this flexibility of the Spanish system (see Aske, 1989). As a consequence of these crosslinguistic differences in describing motion events, it has been hypothesized that manner of motion will be cognitively more salient for speakers of satellite-framed than for speakers of verb-framed languages (Slobin, 1996).

Much of the work on motion events, both with monolinguals and bilinguals, has been carried out through the theoretical lens of *thinking for speaking* (Slobin, 1996), which claims that language affects the kind of online thinking we recruit while engaged in speech production or comprehension. This line of evidence has shown reliable language effects. For example, when English and Greek speakers are about to describe an unfolding motion event, they visually attend to different aspects of the event as a function of the information each language encodes in the main verb root (Papafragou, Hulbert, & Trueswell, 2008). It has also been shown that speakers of different languages vary in the way they gesture about path and manner when describing motion events, suggesting that they also differ in how they mentally represent these events (Brown & Chen, 2013; McNeill & Duncan, 2000; for motion gesturing in bilinguals, see Stam, 2015).

The thinking for speaking approach has yielded reliable evidence that speakers of different languages conceptualize motion events differently while speaking and that bilingual speakers' linguistic conceptualizations are affected by all of their languages. However, while thinking for speaking is related to linguistic relativity, one critical difference is that the former targets the effect of language on linguistic conceptualization (as manifested while speaking), whereas the latter places a clear emphasis on language effects on general cognition, as measured in nonverbal tasks (for a comparison of the two approaches, see Athanasopoulos & Bylund, 2013b; Wolff & Holmes, 2011).

For now, the evidence is mixed as to whether motion event cognition beyond speech varies between monolingual speakers of different languages. Some studies have found either no effect of language on how we remember and categorize motion events (Papafragou et al., 2002), or an effect only when participants were linguistically biased (Gennari et al., 2002; Papafragou &

Selimis, 2010), while others have found evidence of such an effect even in the absence of tasks involving overt linguistic behavior (Kersten et al., 2010; Montero-Melis & Bylund, 2016). One explanation of the contradictory results might be that describing events constitutes an inherently more complex form of linguistic labelling than naming objects (Lupyan, 2012) and that crosslinguistic differences in descriptions are probabilistic rather than categorical, showing great within-language variability (Berthele, 2006; Goschler & Stefanowitsch, 2013). Due to this variability, habitual language experience might not bias speakers of different languages to distinct event components in categorical ways, which in turn would limit the crosslinguistic differences one should expect in nonverbal behavior.

Turning to bilingual speakers, those studies that test cognition beyond language use remain few (cf. Athanasopoulos & Bylund, 2013b). In one such study, Kersten et al. (2010) used a supervised learning paradigm, in which participants had to classify alien species either according to path or to manner features (manipulated between subjects). They found that Spanish-English bilinguals performed better in the manner condition when tested in an English-speaking context than when tested in a Spanish-speaking context, suggesting that bilinguals' attention on manner was boosted in English. Filipović (2011) tested whether English-Spanish bilinguals' performance on a recognition task that involved changes in manner varied as a function of language context. Bilinguals' performance resembled that of Spanish monolinguals independently of language context. The lack of effect of language context, however, might be a consequence of how this notion was operationalized in the study: One bilingual group described the target model events in Spanish and carried out the recognition task in English (this time describing the target variant events in English), while the other bilingual group did the opposite. Thus, both bilingual groups used both languages during the task, which might effectively have wiped out any effects induced by language context.

Finally, Lai, Rodriguez, and Narasimhan (2014) tested categorization preferences of motion events by Spanish and English monolinguals and by Spanish-English bilinguals. In each trial, participants first heard a description of the target animation, then watched the animation and had to repeat the description. After that, they had to match a path- or manner-alternate animation with the target clip. Each of the monolingual groups performed the task in their language, while roughly half of the bilinguals carried it out in Spanish and the other half in English. Crucially, Spanish and English descriptions contained the same type of information (path and manner), so that only language, rather than the content of descriptions, was manipulated. Lai and colleagues found that bilinguals tested in Spanish and Spanish monolinguals were more likely than bilinguals tested in English and English monolinguals to judge events as similar on the basis of path information.[2]

In summary, in the domain of motion events, the evidence for language effects on nonverbal behavior remains mixed. The few studies testing bilingual speakers on nonverbal tasks suggest that long-term L2 exposure can affect category learning and event categorization, but not event recognition. What the literature has not explored is the extent to which nonverbal behavior might be affected by L2 linguistic priming.

The Present Study

If the mental representations we recruit to carry out a specific task can be constructed ad hoc, then it should be possible to influence these representations by highlighting different linguistic categories, even while holding language background and currently activated language constant. To test this idea, we adapted a paradigm first used by Billman and colleagues with English monolinguals (Billman & Krych, 1998; Billman, Swilley, & Krych, 2000). During an encoding phase, we primed participants in their L2 by manipulating the semantic content of prime sentences. We then assessed whether semantic priming affected how participants judged event similarity during the test phase.

Some participants were exposed to sentences highlighting path information (path condition) and others to sentences highlighting manner information (manner condition); a third group was not exposed to any L2 priming sentences (control condition). To test perceived event similarity we used an arrangement task, which does not impose categorical choices between manner and path, but instead allows participants to sort events along a number of event dimensions. If the ad hoc proposal is correct in that people form conceptual representations "on the fly" to meet specific task demands, we expect to see that recently primed linguistic categories are more likely to serve as an anchor in similarity arrangements. In this scenario, we would find differences in similarity arrangements between the two primed conditions. The control group should show a nonverbal behavior that falls between the primed conditions, but still somewhat closer to the path-primed group, because participants in the control condition carried out the experiment in Spanish, a verb-framed language in which path is mentioned more often than manner. If, on the other hand, our mental representations of motion events are relatively static and entrenched in memory (possibly through long-term linguistic experience), then they should not be affected by recent L2 exposure. Under this scenario, we would not expect differences in similarity arrangements between any of the three conditions.

One advantage of within-L2 priming—compared to manipulating the language in which the task is carried out—is that it allows us to target path and manner concepts directly. To our knowledge, this is the first attempt to examine whether priming L2 speakers influences their nonverbal behavior. We tested native Swedish speakers of L2 Spanish in the domain of caused motion because of the following crosslinguistic features. First, path and manner information is consistently included in Swedish L1 descriptions of caused motion (Montero-Melis & Bylund, 2016); thus, we may assume a certain homogeneity in participants' L1-based linguistic experience. Second, Spanish allows for more varied structural patterns to describe these events, making it possible to create prime sentences that vary in their semantic content, while holding syntactic structure constant.

Method

Participants

A total of 60 native speakers of Swedish with L2 Spanish took part in the experiment in exchange for payment. All participants were adult, formal learners of Spanish ($M_{age} = 36.1$ years, $SD = 13.6$). Their proficiency, according to the Common European Framework of Reference for languages, was at least that of an "independent user" (B1), as informally assessed by a native speaker of Spanish prior to the experimental session (Council of Europe, 2011). All participants self-rated their Spanish proficiency between 4 and 7 on a 7-point scale where 7 was the maximum rating ($M = 4.9$). Participants were randomly allocated to one of three conditions: path primed, manner primed, or control. An independent measure of participants' Spanish proficiency was obtained by means of a cloze test at the end of each session, after having carried out all the critical experimental tasks. This test was a 280-word text about the Amazonian rainforest in which every seventh word had been removed. The maximum score was 37. The cloze format was chosen because it is generally considered to be a reliable indicator of global language proficiency (McNamara, 2000). Table 1 shows the number of participants per condition and their means and standard deviations on the cloze test. A one-way analysis of variance, with cloze test scores as the dependent measure and prime condition as the independent factor, yielded no significant difference in proficiency between the three conditions, $F(2, 57) = .41, p = .67$.

Materials

Video Clips

The stimulus set consisted of 32 short animated cartoons, each approximately seven seconds long, depicting caused motion events in which the same

Table 1 Number of participants and descriptive statistics for cloze test scores per prime condition

		Cloze score (0–37)	
Prime condition	*n*	*M*	*SD*
Path-primed	19	22.6	8.8
Manner-primed	21	25.0	9.0
Control (unprimed)	20	24.4	8.2
Total	60	24.1	8.6

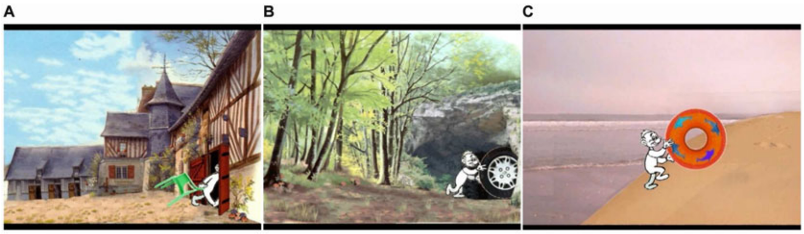

Figure 1 Stills from three target items illustrating the design. The events depicted in (a) and (b) share the same path (both show motion inwards), but they involve different manners (pull + slide vs. push + roll, respectively). Events depicted in (b) and (c) share the same manner (both push + roll), but they involve different paths (inwards vs. upwards, respectively). The events in (a) and (c) do not share any of the critical features. We expected path-primed participants to pay more attention to path-based similarity (a and b) than manner-primed participants, while we expected manner-primed participants to be more sensitive to manner-based similarity (b and c) than path-primed participants.

human-like agent displaced an object along a certain path. The stimuli were originally developed by Hickmann and colleagues (e.g., Hickmann & Hendriks, 2010). Three event components were systematically crossed in the stimuli: the path followed by agent and object (four levels: up, down, across, into); the manner in which the agent manipulated the object (two levels: pushing or pulling); and the manner in which the object moved (two levels: rolling or sliding). The direction of motion (left to right or right to left) was counterbalanced. This design resulted in 16 possible combinations of paths and manners; for each combination, there was one scene progressing from left to right and another from right to left. Figure 1 shows three target items to illustrate the basic design. In addition, stimuli varied with respect to the ground in which the event took place (eight different grounds) and the object that was manipulated (16 different objects), as outlined in Appendix S1 in the Supporting Information online.[3]

Table 2 Primed event components and priming verbs per condition

Path-priming		Manner-priming	
Path concept	Verb	Manner concept	Verb
up	sube "ascends"	push + roll	rueda "rolls"
down	baja "descends"	push + slide	empuja "pushes"
into	entra (en) "enters"	pull + roll	tira (de) "draws"
across	cruza "crosses"	pull + slide	arrastra "drags"

Prime Sentences

For each target item, we created a pair of Spanish priming sentences: One highlighted path information in the main verb and the other highlighted manner information. Prime sentences were constructed by the first author, a native speaker of Spanish, using written corpora of Spanish as a starting point (Davies, 2002; Real Academia Española, n.d.) and double-checking their naturalness informally with other native Spanish speakers. Sentences with manner expressed in the main verb and omitting path are grammatical in Spanish, though not frequent. Research on linguistic priming suggests that less frequent or less expected primes tend to elicit stronger priming (e.g., Bernolet & Hartsuiker, 2010; Ferreira, 2003; Fine & Jaeger, 2013; Jaeger & Snider, 2013). Thus, we expected manner primes to elicit a stronger priming effect than path primes. Table 2 shows the primed path and manner concepts and the verbs used for priming. Other aspects of the event—such as the agent, the object being moved, or the ground where the action took place—were held constant for each pair of prime sentences, as illustrated in Examples 1 and 2.

Example 1. Path priming
El señor <u>*sube*</u> *unos escalones con una televisión*.
The man <u>ascends</u> the stairs with a television.
Example 2. Manner priming
El señor <u>*empuja*</u> *una televisión por unos escalones*.
The man <u>pushes</u> a television along the stairs.

We were interested in how the semantic content of the sentences primed participants in a subsequent nonverbal task. Consequently, the syntactic structure of priming sentences was held constant to rule out the possibility that variation in syntactic structure would account for the priming effect. The only difference in syntactic structure is that some of the Spanish verbs are necessarily followed by a preposition while others are not (this alternation was balanced across priming conditions). All priming sentences are listed in Appendix S1.

Procedure

Participants were tested individually by a native speaker of Spanish. To put participants in a L2 language mode, the experimenter first conversed with them for about 5 minutes, interviewing them about their experience with Spanish and other languages. This was followed by a computerized multiple-choice vocabulary task in which a picture of an object (e.g., a chair) or a ground (e.g., a cave) was shown on the screen together with three numbered Spanish words: the target word and two distractors. The participant had to type the number corresponding to the correct word and received feedback. This task was deliberately made easy in order to familiarize participants with the L2 vocabulary that was needed in the upcoming encoding phase. Accuracy in all groups was near ceiling (path-primed: $M = 98\%$, $SD = 4\%$; manner-primed: $M = 97\%$, $SD = 3\%$; control: $M = 97\%$, $SD = 7\%$).

The main experimental procedure is shown in Figure 2. In the encoding phase, participants had to describe each of the 32 target stimuli in Spanish. Videos were shown in random order. Before each clip, a prompt asked participants to watch the whole clip and then describe it. In the two primed conditions, each clip was preceded by a sentence in Spanish that participants had to read out loud: a path-priming sentence for path-primed participants and a manner-priming sentence for manner-primed participants. Prime sentences were shown as whole sentences in the centre of the screen. The verb in the priming sentences always matched the following scene, but the other elements did not (see Appendix S1). Participants in the control condition described the events without exposure to any prime sentences. Progression throughout the experiment was self-paced.

The test phase followed the encoding phase (after a short break to load the new task) and consisted of a similarity arrangement paradigm (Goldstone, 1994; Kriegeskorte & Mur, 2012). Participants were asked to place the 32 target scenes on the screen according to their similarity. Similar scenes were to be placed near each other and dissimilar scenes far away from each other. The instructions (translated here from Spanish) were: “Your task is to group scenes according to similarity: If the actions are similar, scenes have to be placed near each other. If the actions are different, they should be placed far from each other. The important question is: which actions are more similar and which actions are different?” Participants were informed that they would carry out three arrangement blocks and that they would not be able to move a video clip

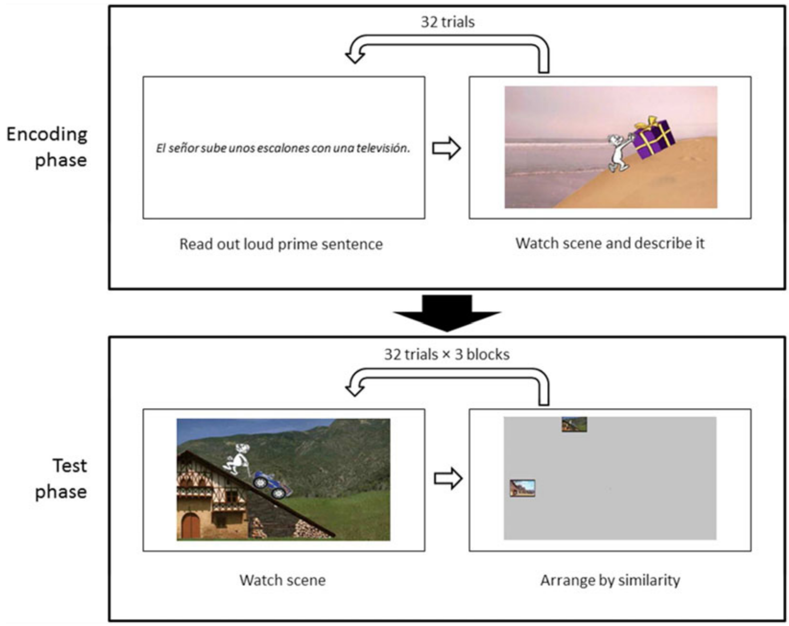

Figure 2 Schematic outline of the experimental procedure. In each of the 32 trials of the encoding phase, primed participants read out loud a prime sentence (path or manner, between subjects) and subsequently described a scene. Participants in the control condition only described the scenes. In the test phase, participants watched a scene and then placed it on the screen by similarity with other scenes. They carried out three consecutive blocks, each time with all 32 items.

once it had been placed. Participants could arrange items freely on the screen (i.e., there was no predefined number of piles or clusters), but were asked not to arrange similar items in rows or columns. After a brief training phase, the actual test phase started. The test phase consisted of three arrangement blocks of 32 video clips each. The progression per block was as follows (see Figure 3). First, a video clip was played on the screen in its entirety (Figure 3a). The participant then moved to a screen where they had to place the scene by clicking with the mouse; upon clicking, a still of the video clip appeared on the screen (Figure 3b). This procedure was repeated until the end of the block (Figure 3c–f). Participants moved forward by clicking on a centered message box to prevent spatial bias in the arrangements.

The order of the items in each block was randomized following the constraint that two consecutive video clips should neither repeat the exact path nor the exact manner, in order to avoid biasing participant attention toward a

Figure 3 Block procedure for similarity arrangement task.

particular contrast. The arrangement task was programmed in E-Basic and run in E-Prime (Schneider, Eschman, & Zuccolotto, 2002). After the test phase, participants were given a cloze test to assess L2 proficiency (Table 1). Finally, participants were asked to explain how they had carried out the nonverbal arrangement task. Participants were generally able to articulate their sorting strategy (e.g., "I looked at whether he was going up or down" or "I used pushing and drawing"). Crucially, however, none of the participants explicitly mentioned the priming sentences during the postexperimental surveys or otherwise linked their strategy to the priming manipulation.

Design

Participants saw and described a set of events in which path and manner varied orthogonally; that is, all levels of path were combined with all levels of manner. So as not to make the events look too homogenous, other aspects of the scenes were varied as well (such as the left-right or right-left direction, the ground where it took place, or the object that was being moved). By arranging the events according to similarity, participants implicitly provided similarity ratings. The dependent measure was the pairwise similarity between the scenes depicted in the video clips and was a continuous measure bounded between 0 (minimal similarity) and 1 (maximal similarity). Similarity was computed in a series of steps. For each final arrangement of clips, we extracted the pairwise distance (in pixels) between the centres of all video clips. This yielded a similarity matrix consisting of 496 entries for each participant and block (a total of 1,488 observations per participant). We then normalized distances in pixels for each participant-block combination, dividing by the maximal pairwise distance for that block, to correct for individual differences in the use of the screen in the arrangement tasks. Finally, we subtracted these normalized distances from 1, so as to obtain a similarity measure between 0 (minimal similarity) and 1 (maximal similarity).

We expected participants in the path-primed condition to be guided by the path dimension to a larger extent than participants in the manner-primed condition; conversely, we expected manner-primed participants to base their arrangements on the manner dimension to a larger extent than path-primed participants. To assess this, we analyzed to what extent a pair of scenes that did or did not share the same value on a specific component (i.e., path, manner) led to an increase in similarity. Consider the three scenes in Figure 1. If path is a relevant dimension for similarity judgements, then two scenes that share the same path (e.g., scenes a and b) should, on average, be more similar than two scenes that do not share path (e.g., scenes a and c, or scenes b and c). We would thus expect an increase in similarity of same path scenes with respect to different path scenes. Exactly the same logic applies to the manner component. Thus, our analyses focused on the increase in similarity of scenes that share a component compared to those that do not. Our critical question was whether this increase differed between priming conditions. All reported analyses were run in *R* (R Development Core Team, 2013). Mixed models were fitted using the *lmer* function from the lme4 library (Bates, Maechler, Bolker, & Walker, 2014). All analyses controlled for the effect of other nontarget variables (e.g., ground, object).

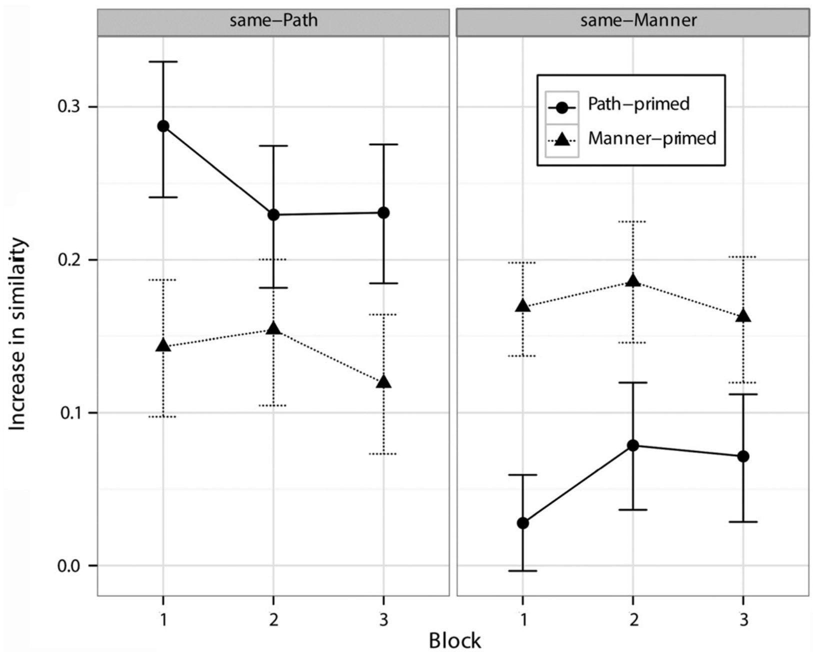

Figure 4 Increase in pairwise event similarity due to shared event components (path in left panel; manner in right panel), expressed as a function of priming condition and block. The *y*-axis indicates the increase in similarity of two events that share the same path (e.g., both *into*) with respect to event pairs that have different paths (e.g., one *into*, the other *up*). The right panel shows corresponding values due to shared manner. Shapes show effect sizes estimated from our analysis. Error bars indicate one standard error (obtained from posterior simulations via *sim* function in R package arm, Gelman and Su, 2014).

Results

Reliance on Path and Manner in the Two Primed Conditions

Reliance on path and manner according to priming condition is plotted in Figure 4, with descriptive statistics shown in Table 3. The left panel in Figure 4 shows that, throughout the three consecutive blocks, two events sharing the same path (e.g., both *into*) as opposed to having different paths (e.g., one *into*, the other *up*) led to a larger increase in similarity ratings for path-primed than for manner-primed participants. In other words, path-primed participants, on average, relied on path to a greater extent than manner-primed participants. Conversely, manner-primed participants relied on manner to a larger extent

Table 3 Mean similarity of event pairs as a function of prime condition

	Path		Manner	
Condition	Different	Same	Different	Same
Path-primed	.46	.71	.51	.54
Manner-primed	.49	.63	.49	.65

Note. Similarity is bounded between 1 (maximal similarity) and 0 (minimal similarity).

than path-primed participants, again throughout the three consecutive blocks (Figure 4 right panel).

To test the statistical significance of these trends, we analyzed the data using mixed-effects modelling (Baayen, Davidson, & Bates, 2008; for introduction, see Jaeger, Graff, Croft, & Pontillo, 2011). The model predicted the continuous dependent variable *similarity between pairs of scenes* as a function of whether or not they shared the same path and the same manner and as a function of the participant's prime condition. Additionally, the model controlled for the effects of arrangement block (1 to 3), ground (e.g., dune), object (e.g., chair), and direction (left-right, right-left). Random effects comprised by-subject intercepts and slopes for all within-subject variables as well as by-item intercepts. We report the maximal model that converged. The model specification in R was: "Sim ~ 1 + (Ground + Object + Direction + Path + Manner) * Block * Condition + (1 + Ground + Object + Direction + (Path + Manner) * Block | Subject) + (1 + Condition | Item)". We chose to include Ground and Object as covariates to control for effects that these features could have on similarity arrangements. (A less conservative model without these covariates yielded qualitatively identical results.) Table 4 describes the final model and coding procedure. A summary of the fixed-effect parameter estimates is given in Appendix S2 in the Supporting Information online. Collinearity among predictors in the model was small, $\kappa < 6$ (Baayen, 2008).

There was a main effect of path ($\text{path}_{\text{same-vs-different}}$: $\hat{\beta} = .19$, $SE = .03$, $t = 6.33$, $p < .001$) and manner ($\text{manner}_{\text{same-vs-different}}$: $\hat{\beta} = .12$, $SE = .02$, $t = 4.97$, $p < .001$), indicating that both of these components were used to judge event similarity. Critically, the analysis found that manner-primed participants relied more on manner to judge event similarity than path primed participants, manifested in a significant interaction between the effect of sharing or not manner and prime condition ($\text{manner}_{\text{same-vs-different}} \times \text{prime condition}_{\text{MannerPr-vs-PathPr}}$: $\hat{\beta} = .11$, $SE = .05$, $t = 2.42$, $p < .05$). There was also a trend towards path-primed participants relying more on path than manner-primed participants,

Table 4 Model description

Predictor	Variable type	Levels	Coding
Path	Within-subjects	Same (e.g., both scenes *into*), different (e.g., one scene *into*, other *up*)	Centered*
Manner	Within-subjects	Same (e.g., both scenes *push* + *roll*), different (e.g., one scene *push* + roll, other *pull* + *slide*)	Centered*
Block	Within-subjects	1, 2, 3	Forward coding
Prime condition	Between-subjects	Manner-primed, path-primed	Centered*
Direction	Within-subjects	Same (e.g., both scenes *left-right*), different (e.g., one scene *left-right*, other *right-left*)	Centered*
Ground	Within-subjects	Same (e.g., both scenes *cave*), different (e.g., one scene *cave*, other *barn*)	Centered*
Object	Within-subjects	Same (e.g., both scenes *tyre*), different (e.g., one scene *tyre*, other *table*)	Centered*

*The first level was coded as the positive value, the second as the negative value.

which did not, however, reach significance at the .05 level ($\text{path}_{\text{same-vs-different}} \times \text{prime condition}_{\text{MannerPr-vs-PathPr}}$: $\hat{\beta} = -.11$, $SE = .06$, $t = -1.81$, $p < .10$). No other effects differed significantly between conditions

Comparison With Control Group

A second question was how the two target groups performed in relation to a control group that also had carried out the task in their L2 Spanish, but had not been exposed to L2 priming sentences. Results of this comparison are plotted in Figure 5. Visual inspection suggests that participants in the control condition showed a tendency to perform in between the two manipulated groups, but that they patterned more like path-primed participants than manner-primed participants.

We tested the reliability of these observations by fitting a separate mixed-model, this time including the data from the control group. The model specification is shown in Table 4; the sole difference compared to the model described above is that the factor *prime condition* now consisted of three levels (Path-primed, Control, Manner-primed). We used forward coding for the prime condition factor, such that the first coefficient compared path-primed versus

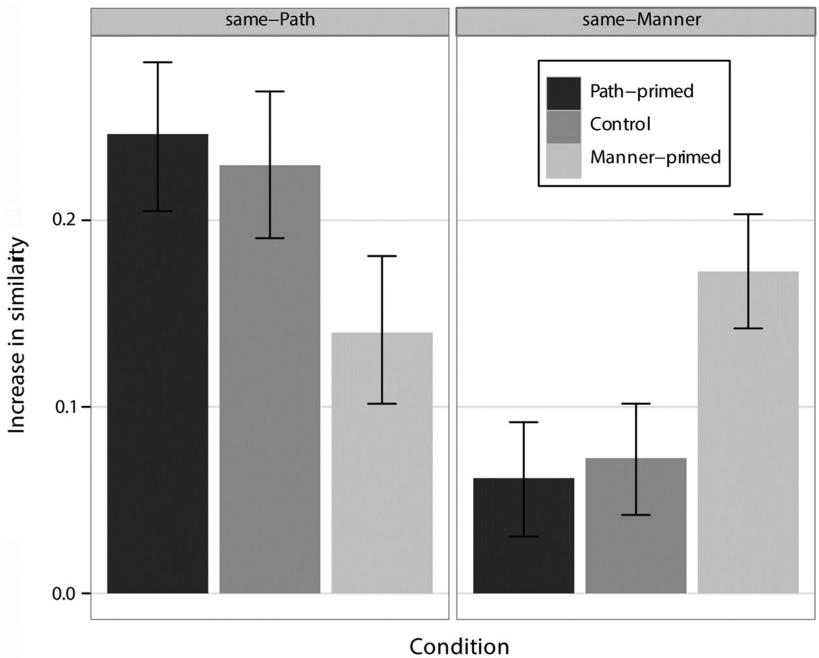

Figure 5 Comparison of primed and control conditions. As in Figure 4, the y-axis shows the increase in pairwise event similarity due to shared event components (path in left panel; manner in right panel), as a function of prime condition (shaded bars). Bars show effect sizes estimated from our analysis. Error bars indicate one standard error (obtained from posterior simulations via *sim* function in R package arm, Gelman and Su, 2014).

control, and the second one control versus manner primed. This coding choice was motivated by our expectation that the control group would show patterns somewhere in between the two manipulated groups. The output for the estimates of fixed-effect parameters is presented in Appendix S3 in the Supporting Information online. Collinearity in the model was moderate, $\kappa = 10.23$.

Again, there were significant main effects of path ($\text{path}_{\text{same-vs-different}}$: $\hat{\beta} = .20$, $SE = .02$, $t = 8.61$, $p < .001$) and manner ($\text{manner}_{\text{same-vs-different}}$: $\hat{\beta} = .10$, $SE = .02$, $t = 5.66$, $p < .001$) on similarity judgments. Turning to the critical interactions, the model revealed no differences between path-primed and control participants in how much they relied on either of the two semantic components ($\text{path}_{\text{same-vs-different}} \times \text{prime condition}_{\text{path-primed-vs-control}}$: $\hat{\beta} = .02$, $SE = .06$, $t = .29$, $p > .10$; $\text{manner}_{\text{same-vs-different}} \times \text{prime condition}_{\text{path-primed-vs-control}}$: $\hat{\beta} = -.01$, $SE = .04$, $t = -.31$, $p > .10$). However, the control group did differ significantly from the manner-primed group in that it relied less on manner (manner$_{\text{same-vs-different}}$ × prime condition$_{\text{control-vs-manner-primed}}$: $\hat{\beta} = -.10$, $SE = .04$, $t = -2.46$, $p < .05$). Control participants also showed a trend toward relying more on path than manner-primed participants, which did not reach statistical significance at the .05 level (path$_{\text{same-vs-different}}$ × prime condition$_{\text{control-vs-manner-primed}}$: $\hat{\beta} = .09$, $SE = .05$, $t = 1.65$, $p < .10$). Thus, overall, control participants showed the same patterns as path-primed participants, and they differed from manner-primed participants in the same way that path-primed participants did, mainly by relying less on manner.

Discussion

Priming Effects on Similarity Assessment

The central finding in the present study is that L2 semantic priming affected event similarity assessment. We found that having to read out loud L2 sentences that highlighted either path or manner during an encoding phase modulated L2 users' reliance on these components in a subsequent nonverbal similarity arrangement task. This argues against a view in which conceptual representations of events are static, thus supporting earlier work on conceptual restructuring in bilinguals. However, it goes further in suggesting that the mental representations we recruit to assess event similarity are at least partly determined in an ad hoc fashion (Casasanto & Lupyan, 2015). When assessing event similarity, we are sensitive to recent short-term linguistic experience highlighting specific event features, even when that exposure likely deviates from previous linguistic experience. While consistent with different accounts that predict task-specific effects of language—such as label feedback (Lupyan, 2012) and linguistic intrusions (Papafragou & Selimis, 2010)—the present study extends previous work, which has mostly focused on the effects of how we habitually encode a semantic domain in language. By choosing a design that explicitly manipulated participants' recent linguistic experience, the present study provides a more direct demonstration of the idea that language effects are dynamic and context dependent. The current results also connect to classic findings in the field of reasoning showing that when people make judgements under uncertainty, they anchor their decisions in recent experience (Tversky & Kahneman, 1974, and follow-up work).

The extent to which path and manner played a role in perceived similarity exhibits several interesting asymmetries. First, participants' similarity ratings were more sensitive to path than to manner (see coefficients for main effects for path and manner in Appendix S2). An overall greater effect of path than manner on similarity ratings replicates previous results with Spanish and

Swedish monolinguals using the same stimuli and a similar arrangement task (Montero-Melis & Bylund, 2016) and might be indicative of a higher salience of path compared to manner (Talmy, 2000), at least in the present stimulus set (see Papafragou & Selimis, 2010, for a discussion of stimuli artifacts). Second, priming condition mainly modulated participants' reliance on the manner component and only marginally affected reliance on path. This could also be indicative of a higher salience of path compared to manner, such that the former is more resistant to priming effects.

Finally, manner priming was stronger than path priming when either was compared to the nonprimed control condition: The control group differed only from the manner-primed, but not from the path-primed group. Two mutually related factors mentioned earlier might contribute to this pattern. First, the effect of path priming might simply have been to exaggerate the habitual Spanish lexicalization pattern of motion events, expressing path in the main verb and only optionally expressing manner. If so, participants in the control condition would have been biased toward path (and away from manner) merely by carrying out the task in Spanish, which was indeed what Lai et al. (2014) found. This would explain why the control group resembled the path-primed group and would also be in line with the results of a study comparing Spanish and Swedish monolinguals on a similar arrangement task, in which Spanish speakers relied less on manner, compared to Swedish speakers (Montero-Melis & Bylund, 2016).

Another related reason why only manner-primed participants differed from the control group is that the Spanish manner-priming sentences might have been less expected by participants because they clashed with the typical Spanish lexicalization pattern. As mentioned in the method section, the manner-priming sentences are grammatical but less frequent than their path-priming counterparts. There is some evidence that the effect of a structural prime is positively correlated with the magnitude of a listener's prediction error or surprisal (e.g., Bernolet & Hartsuiker, 2010; Ferreira, 2003; Fine & Jaeger, 2013; Jaeger & Snider, 2013). As a consequence, we expected stronger effects in the manner-primed than in the path-primed condition because manner-priming sentences were less likely to be predicted by L2 speakers based on their previous linguistic experience in Spanish. Interestingly, Billman and colleagues found precisely the opposite pattern: When priming English monolinguals with bare path or manner verbs, only path-primed participants differed from the control group, whereas manner-primed participants did not (Billman et al., 2000). Because the typical English lexicalization pattern is the opposite of the Spanish pattern (English verbs typically express manner, not path), this strengthens the conclusion that the different effect of path and manner priming, with respect to the control condition, has to do with language-specific expectations.[4]

Implications for Bilingual Cognition and Linguistic Relativity

Put into the broader perspective of bilingual cognition, the effects we found may offer a controlled and miniature real-time demonstration of conceptual restructuring. We isolated a particular semantic contrast that has received much attention in the literature (path vs. manner) and, holding language background and testing language constant, we primed participants with one or the other component while they were describing motion events. This priming then influenced how they assessed similarity between the events. This is indeed a very condensed version of what might be going on in the long process of learning a new language that carves up reality in a different way than our L1. The L2 will redirect our attention toward certain features, and this might play a role in tasks that are not explicitly verbal, at least when the corresponding linguistic categories have been activated through recent linguistic exposure.

More generally, the ad hoc account supported by our data underscores the flexibility of the mental processes underlying nonverbal behavior. This account claims that nonverbal behavior can be flexibly modulated and that language is one factor of influence, be it our first, second, or nth language.[5] There is also growing evidence that language might occupy a privileged position among all factors influencing cognitive processes (Edmiston & Lupyan, 2015; Lupyan & Thompson-Schill, 2012). In particular, an ad hoc account captures the fact that bilinguals seem to switch between different conceptual representations depending on the language they are using at the time. Indeed, priming a language will activate all of our linguistic experience in that language. To the extent that the languages spoken by a bilingual highlight different aspects of reality, we expect a subtle priming effect just by the mere use of one language or the other.

Under an ad hoc account, however, the effect of language on thought would, by hypothesis, not be pervasive. A critical question, then, is whether the effects we found here have any bearing on real-life situations. To answer this question, we may ask: Is language, and the categories it provides, usually as active as in our experiment? We believe that in many situations it might be. Many of our most important social interactions involve language use, be it at work or in our personal lives. Moreover, the particular cognitive process affected here, similarity assessment, is generally held to be central to human cognition and to underlie many of our daily mental operations, from categorization to reasoning and problem solving (Hahn, 2014).

Future Directions

An open question is to what extent ad hoc category formation requires conscious reasoning and, in particular, whether the priming effects observed in the current study are based on implicit or explicit processes. Preliminary evidence that speaks to this question comes from our post-experimental surveys. None of the participants mentioned the priming sentences when explicitly asked to describe their sorting criteria during the postexperimental surveys; that is, they seemed to remain unaware of the manipulation. This is despite the fact that participants in each of the primed conditions showed a greater tendency to use the verbs they were primed with in their retrospective protocols. The present study does not offer conclusive data on this issue, and future research should explore how implicit this type of semantic priming is. Do language learners generally notice contrasts in how their different languages conceptualize a given semantic domain, and does this matter for nonverbal behavior?

A future challenge is to assess whether the language effects we have observed have any long-lasting consequences for our mental representations of caused motion events. For example, will priming effects transfer across a bilingual's languages? Studies on conceptual switching in bilinguals provide indirect evidence that effects do not fully transfer across languages (Athanasopoulos et al., 2015; Miles et al., 2011; see also studies on language-dependent memory, such as Marian & Kaushanskaya, 2007). An extension of the current paradigm to cross-language priming could provide a more direct test of this hypothesis. Additionally, assessing the effects of priming after substantial temporal delay (i.e., in a session on a separate day), possibly with linguistic interference, would assess whether priming only leads to short-term effects within an experiment or whether it induces novel categories that persist over days. This would reveal how deeply (if at all) short-term linguistic experience can affect longer-term mental representations.

We believe that the methodological approach we used here—this miniature experiment of cognitive restructuring—has the potential of offering important insights into bilingual cognition and the relationship between language and thought more generally. The usual correlational approach, in which bilinguals (possibly grouped according to L2 proficiency) are compared to monolingual baselines, by its very nature invites confounds such as the degree of acculturation. Additionally, it is difficult, if not impossible, to know what an individual's linguistic exposure has been to the semantic domain of interest. Randomly assigning participants to conditions in which the relevant linguistic experience is manipulated, as we did here through linguistic priming, offers a protection against confounding variables. Crucially, it also allows researchers to understand causal relations between linguistic experience and nonverbal behavior.

Conclusion

Previous work on linguistic relativity has established that linguistic experience is a factor that influences a wide variety of cognitive processes (see Boroditsky, 2012). Work on bilingual cognition has extended this idea to show that additional language learning continues to shape cognition (see Bylund & Athanasopoulos, 2014b). By showing that L2 priming affects event similarity assessment, the present study supports this previous work and further clarifies that language might exert an influence on the ad hoc cognitive processes underlying nonverbal tasks. The present work thus contributes to our understanding of how learning any number of languages might continue to influence our ways of thinking, while simultaneously highlighting that this influence is dynamic and most likely task specific.

Final revised version accepted 24 December 2015

Notes

1 We will use the terms "behavior on nonverbal tasks" and "nonverbal behavior" interchangeably. While nonverbal tasks do not involve overt verbal behavior nor do prima facie require covert verbal processes, it is not implied that purely nonlinguistic cognition underlies nonverbal behavior.

2 Lai and colleagues operationalized path as either leftward or rightward motion on the screen. However, to our knowledge, it is not known whether satellite- and verb-framed languages differ in the likelihood of encoding this type of information. Hence, it is not clear why speakers of English and Spanish should differ in their reliance on left- versus rightward motion.

3 Note that grounds and objects must vary as well. For example, the same object cannot both roll and slide. Similarly, events cannot show motion upwards, downwards, inwards, etc., holding the ground constant.

4 Another possibility is that manner primes activated L1 (Swedish) patterns, because in Swedish the main verb typically also conveys manner information (we are thankful to an anonymous reviewer for pointing this out). However, because Swedish descriptions also consistently include verb satellites encoding path information, the structural pattern of manner primes was not the same as the Swedish pattern. Clearly, this question remains open for future research.

5 As one of the anonymous reviewers points out, a qualitatively similar effect as the one we found here would be expected for monolingual Spanish speakers. The demonstration of this effect in L2 speakers is, however, more surprising, not least

because it has been hypothesized that our native language trains us to pay attention to different types of events and that this training "is exceptionally resistant to restructuring in adult second-language acquisition" (Slobin, 1996, p. 89).

References

Aske, J. (1989). Path predicates in English and Spanish: A closer look. In *Proceedings of the Berkeley Linguistic Society* (Vol. *15*, pp. 1–14). Berkeley: University of California.

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

Athanasopoulos, P., & Bylund, E. (2013b). The "thinking" in thinking-for-speaking: Where is it? *Language, Interaction & Acquisition*, *4*, 91–100. doi:10.1075/lia.4.1.05ath

Athanasopoulos, P., Bylund, E., Montero-Melis, G., Damjanovic, L., Schartner, A., Kibbe, A., et al. (2015). Two languages, two minds: Flexible cognitive processing driven by language of operation. *Psychological Science*, *26*, 518–526. doi:10.1177/0956797614567509

Athanasopoulos, P., Damjanovic, L., Krajciova, A., & Sasaki, M. (2011). Representation of colour concepts in bilingual cognition: The case of Japanese blues. *Bilingualism: Language and Cognition*, *14*, 9–17. doi:10.1017/S1366728909990046

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge, UK: Cambridge University Press.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390–412. doi:10.1016/j.jml.2007.12.005

Bates, D. M., Maechler, D., Bolker, B. M., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4 (Version R package version 1.1-7). Retrieved from http://CRAN.R-project.org/package=lme4

Bernolet, S., & Hartsuiker, R. J. (2010). Does verb bias modulate syntactic priming? *Cognition*, *114*, 455–461. doi:10.1016/j.cognition.2009.11.005

Berthele, R. (2006). *Ort und Weg: Die sprachliche Raumreferenz in Varietäten des Deutschen, Rätoromanischen und Französischen* (Vol. *16*). Berlin: Walter de Gruyter.

Billman, D., & Krych, M. (1998). Path and manner verbs in action: Effects of "skipping" or "exiting" on event memory. In M. A. Gernsbacher & S. J. Derry (Eds.), *Proceedings of the Twentieth Annual Conference of the Cognitive Science Society* (pp. 156–161). Mahwah, NJ: Erlbaum.

Billman, D., Swilley, A., & Krych, M. (2000). Path and manner priming: Verb production and event recognition. In L. R. Gleitman & A. K. Joshi (Eds.),

Proceedings of the Twenty-Second Annual Conference of the Cognitive Science Society (pp. 615–620). Mahwah, NJ: Erlbaum.
Boroditsky, L. (2012). How the languages we speak shape the ways we think: The FAQs. In M. Spivey, K. McRae, & M. Joanisse (Eds.), *The Cambridge handbook of psycholinguistics* (pp. 615–632). Cambridge, UK: Cambridge University Press.
Boroditsky, L., Fuhrman, O., & McCormick, K. (2011). Do English and Mandarin speakers think about time differently? *Cognition*, *118*, 123–129. doi:10.1016/j.cognition.2010.09.010
Brown, A., & Chen, J. (2013). Construal of manner in speech and gesture in Mandarin, English, and Japanese. *Cognitive Linguistics*, *24*, 605–631. doi:10.1515/cog-2013-0021
Bylund, E., & Athanasopoulos, P. (2014a). Language and thought in a multilingual context: The case of isiXhosa. *Bilingualism: Language and Cognition*, *17*, 431–441. doi:10.1017/S1366728913000503
Bylund, E., & Athanasopoulos, P. (2014b). Linguistic relativity in SLA: Toward a new research program. *Language Learning*, *64*, 952–985. doi:10.1111/lang.12080
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, MA: MIT Press.
Chen, J.-Y. (2007). Do Chinese and English speakers think about time differently? Failure of replicating Boroditsky (2001). *Cognition*, *104*, 427–436. doi:10.1016/j.cognition.2006.09.012
Cook, V., Bassetti, B., Kasai, C., Sasaki, M., & Takahashi, J. A. (2006). Do bilinguals have different concepts? The case of shape and material in Japanese L2 users of English. *International Journal of Bilingualism*, *10*, 137–152. doi:10.1177/13670069060100020201
Council of Europe. (2011). Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR). Council of Europe. Retrieved from http://www.coe.int/t/dg4/linguistic/Cadre1_en.asp
Davies, M. (2002). Corpus del español. Retrieved June 1, 2014, from http://www.corpusdelespanol.org
Edmiston, P., & Lupyan, G. (2015). What makes words special? Words as unmotivated cues. *Cognition*, *143*, 93–100. doi:10.1016/j.cognition.2015.06.008
Ferreira, V. S. (2003). The persistence of optional complementizer production: Why saying “that” is not saying “that” at all. *Journal of Memory and Language*, *48*, 379–398. doi:10.1016/S0749-596X(02)00523-5
Filipović, L. (2011). Speaking and remembering in one or two languages: Bilingual vs. monolingual lexicalization and memory for motion events. *International Journal of Bilingualism*, *15*, 466–485. doi:10.1177/1367006911403062
Fine, A. B., & Jaeger, T. F. (2013). Evidence for implicit learning in syntactic comprehension. *Cognitive Science*, *37*, 578–591. doi:10.1111/cogs.12022

Gelman, A., & Su, Y.-S. (2014). arm: Data Analysis Using Regression and Multilevel/Hierarchical Models (Version R package version 1.7-07). Retrieved from http://CRAN.R-project.org/package=arm

Gennari, S. P., Sloman, S. A., Malt, B. C., & Fitch, W. T. (2002). Motion events in language and cognition. *Cognition*, *83*, 49–79. doi:16/S0010-0277(01)00166-4

Gentner, D. (2010). Bootstrapping the mind: Analogical processes and symbol systems. *Cognitive Science*, *34*, 752–775. doi:10.1111/j.1551-6709.2010.01114.x

Gleitman, L., & Papafragou, A. (2012). New perspectives on language and thought. In K. J. Holyoak & R. G. Morrison (Eds.), *The Oxford handbook of thinking and reasoning* (pp. 543–568). Oxford, UK: Oxford University Press.

Goldstone, R. L. (1994). An efficient method for obtaining similarity data. *Behavior Research Methods*, *26*, 381–386. doi:10.3758/BF03204653

Goschler, J., & Stefanowitsch, A. (Eds.). (2013). *Variation and change in the encoding of motion events* (Vol. *41*). Amsterdam: John Benjamins.

Hahn, U. (2014). Similarity. *Wiley Interdisciplinary Reviews: Cognitive Science*, *5*, 271–280. doi:10.1002/wcs.1282

Haun, D. B. M., Rapold, C. J., Janzen, G., & Levinson, S. C. (2011). Plasticity of human spatial cognition: Spatial language and cognition covary across cultures. *Cognition*, *119*, 70–80. doi:10.1016/j.cognition.2010.12.009

Hickmann, M., & Hendriks, H. (2010). Typological constraints on the acquisition of spatial language in French and English. *Cognitive Linguistics*, *21*, 189–215. doi:10.1515/COGL.2010.007

Imai, M., Saalbach, H., & Stern, E. (2010). Are Chinese and German children taxonomic, thematic, or shape biased? Influence of classifiers and cultural contexts. *Frontiers in Psychology*, *1*, 194. doi:10.3389/fpsyg.2010.00194

Jaeger, T. F., Graff, P., Croft, W., & Pontillo, D. (2011). Mixed effect models for genetic and areal dependencies in linguistic typology. *Linguistic Typology*, *15*, 281–319. doi:10.1515/LITY.2011.021

Jaeger, T. F., & Snider, N. E. (2013). Alignment as a consequence of expectation adaptation: Syntactic priming is affected by the prime's prediction error given both prior and recent experience. *Cognition*, *127*, 57–83. doi:10.1016/j.cognition.2012.10.013

January, D., & Kako, E. (2007). Re-evaluating evidence for linguistic relativity: Reply to Boroditsky (2001). *Cognition*, *104*, 417–426. doi:10.1016/j.cognition.2006.07.008

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*, 638–653. doi:10.1037/a0020507

Kriegeskorte, N., & Mur, M. (2012). Inverse MDS: Inferring dissimilarity structure from multiple item arrangements. *Frontiers in Psychology*, *3*, 245. doi:10.3389/fpsyg.2012.00245

Lai, V. T., Rodriguez, G. G., & Narasimhan, B. (2014). Thinking-for-speaking in early and late bilinguals. *Bilingualism: Language and Cognition*, *17*, 139–152. doi:10.1017/S1366728913000151

Landau, B., Dessalegn, B., & Goldberg, A. M. (2010). Language and space: Momentary interactions. In V. Evans & P. A. Chilton (Eds.), *Language, cognition and space: The state of the art and new directions* (pp. 51–77). London: Equinox.

Levinson, S. C. (2003). Language and mind: Let's get the issues straight! In D. Gentner & S. Goldin-Meadow (Eds.), *Language in mind: Advances in the study of language and thought* (pp. 25–46). Cambridge, MA: MIT Press.

Li, P., & Gleitman, L. (2002). Turning the tables: Language and spatial reasoning. *Cognition*, *83*, 265–294. doi:10.1016/S0010-0277(02)00009-4

Lupyan, G. (2012). Linguistically modulated perception and cognition: The label-feedback hypothesis. *Frontiers in Psychology*, *3*, 54. doi:10.3389/fpsyg.2012.00054

Lupyan, G., Rakison, D. H., & McClelland, J. L. (2007). Language is not just for talking: Redundant labels facilitate learning of novel categories. *Psychological Science*, *18*, 1077–1083. doi:10.1111/j.1467-9280.2007.02028.x

Lupyan, G., & Thompson-Schill, S. L. (2012). The evocative power of words: Activation of concepts by verbal and nonverbal means. *Journal of Experimental Psychology: General*, *141*, 170–186. doi:10.1037/a0024904

Malt, B. C., Sloman, S. A., Gennari, S., Shi, M., & Wang, Y. (1999). Knowing versus naming: Similarity and the linguistic categorization of artifacts. *Journal of Memory and Language*, *40*, 230–262. doi:10.1006/jmla.1998.2593

Marian, V., & Kaushanskaya, M. (2007). Language context guides memory content. *Psychonomic Bulletin & Review*, *14*, 925–933. doi:10.3758/BF03194123

McNamara, T. (2000). *Language testing*. Oxford, UK: Oxford University Press.

McNeill, D., & Duncan, S. (2000). Growth points in thinking-for-speaking. In D. McNeill (Ed.), *Language and gesture* (pp. 141–161). Cambridge, UK: Cambridge University Press.

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*, 598–604. doi:10.3758/s13423-011-0068-y

Montero-Melis, G., & Bylund, E. (2016). *Getting the ball rolling: The representation of caused motion events in language and cognition*. Manuscript submitted for publication.

Ortega, L. (2013). SLA for the 21st century: Disciplinary progress, transdisciplinary relevance, and the bi/multilingual turn. *Language Learning*, *63*, 1–24. doi:10.1111/j.1467-9922.2012.00735.x

Papafragou, A., Hulbert, J., & Trueswell, J. (2008). Does language guide event perception? Evidence from eye movements. *Cognition*, *108*, 155–184. doi:10.1016/j.cognition.2008.02.007

Papafragou, A., Massey, C., & Gleitman, L. (2002). Shake, rattle, ‘n’ roll: The representation of motion in language and cognition. *Cognition*, *84*, 189–219. doi:16/S0010-0277(02)00046-X

Papafragou, A., & Selimis, S. (2010). Event categorization and language: A cross-linguistic study of motion. *Language and Cognitive Processes*, *25*, 224–260. doi:10.1080/01690960903017000

Park, H. I., & Ziegler, N. (2014). Cognitive shift in the bilingual mind: Spatial concepts in Korean-English bilinguals. *Bilingualism: Language and Cognition*, *17*, 410–430. doi:10.1017/S1366728913000400

Pavlenko, A. (2005). Bilingualism and thought. In J. F. Kroll & A. M. B. De Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 433–453). Oxford, UK: Oxford University Press.

Pinker, S. (1994). *The language instinct: The new science of language and mind*. London: Penguin.

R Development Core Team. (2013). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org

Real Academia Española. (n.d.). Banco de datos (CREA) [en línea]. Corpus de referencia del español actual. Retrieved June 1, 2014, from http://www.rae.es

Roberson, D., Davidoff, J., Davies, I. R. L., & Shapiro, L. R. (2005). Color categories: Evidence for the cultural relativity hypothesis. *Cognitive Psychology*, *50*, 378–411. doi:10.1016/j.cogpsych.2004.10.001

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime reference guide*. Pittsburgh, PA: Psychology Software Tools Inc.

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., UK: Cambridge University Press.

Stam, G. (2015). Changes in thinking for speaking: A longitudinal case study. *The Modern Language Journal*, *99*, 83–99. doi:10.1111/j.1540-4781.2015.12180.x

Talmy, L. (2000). *Toward a cognitive semantics: Typology and process in concept structuring*. Cambridge, MA: MIT Press.

Trueswell, J. C., & Papafragou, A. (2010). Perceiving and remembering events cross-linguistically: Evidence from dual-task paradigms. *Journal of Memory and Language*, *63*, 64–82. doi:16/j.jml.2010.02.006

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*, 1124–1131. doi:10.1126/science.185.4157.1124

Whorf, B. L. (1956). *Language, thought, and reality: Selected writings of Benjamin Lee Whorf*. Cambridge, MA: MIT Press.

Wolff, P., & Holmes, K. J. (2011). Linguistic relativity. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*, 253–265. doi:10.1002/wcs.104

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Appendix S1. Stimulus Description and Priming Sentences.
Appendix S2. Output of Mixed Effects Model (Primed Conditions).
Appendix S3. Output of Mixed Effects Model (All Conditions).