Cognition 150 (2016) 213–231

Contents lists available at ScienceDirect

Cognition

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

Original Articles

# The effects of bilingualism on conflict monitoring, cognitive control, and garden-path recovery

Susan E. Teubner-Rhodes [a,b,*], Alan Mishler [c], Ryan Corbett [c], Llorenç Andreu [d], Monica Sanz-Torrent [e], John C. Trueswell [f], Jared M. Novick [b,c,g]

[a] Department of Psychology, University of Maryland, Biology-Psychology Bldg., College Park, MD 20742, USA
[b] Program in Neuroscience and Cognitive Science, University of Maryland, Biology-Psychology Bldg., College Park, MD 20742, USA
[c] Center for Advanced Study of Language, University of Maryland, 7005 52nd Ave, College Park, MD 20742, USA
[d] Grup de Recerca en Cognició i Llenguatge (GRECIL), Universitat Oberta de Catalunya, Rambla de Poble Nou, 156, 08018 Barcelona, Spain
[e] Faculty of Psychology, University of Barcelona, Passeig de la Vall d'Hebron, 171, 08035 Barcelona, Spain
[f] Department of Psychology, University of Pennsylvania, 3401 Walnut Street, Philadelphia, PA 19104, USA
[g] Department of Hearing and Speech Sciences, University of Maryland, Lefrak Hall, College Park, MD 20742, USA

A R T I C L E  I N F O

*Article history:*
Received 18 September 2014
Revised 6 January 2016
Accepted 12 February 2016

*Keywords:*
Bilingualism
Cognitive control
Sentence comprehension
Stimulus ambiguity
Memory

A B S T R A C T

Bilinguals demonstrate benefits on non-linguistic tasks requiring cognitive control—the regulation of mental activity to resolve information-conflict during processing. This "bilingual advantage" has been attributed to the consistent management of two languages, yet it remains unknown if these benefits extend to sentence processing. In monolinguals, cognitive control helps detect and revise misinterpretations of sentence meaning. Here, we test if the bilingual advantage extends to parsing and interpretation by comparing bilinguals' and monolinguals' syntactic ambiguity resolution before and after practicing N-back, a non-syntactic cognitive-control task. Bilinguals outperformed monolinguals on a high-conflict but not a no-conflict version of N-back and on sentence comprehension, indicating that the advantage extends to language interpretation. Gains on N-back conflict trials also predicted comprehension improvements for ambiguous sentences, suggesting that the bilingual advantage emerges across tasks tapping shared cognitive-control procedures. Because the overall task benefits were observed for conflict *and* non-conflict trials, bilinguals' advantage may reflect increased cognitive flexibility.



## 1. Introduction

Balanced bilinguals—people who are equally proficient in two languages—seem to experience a host of cognitive advantages over monolinguals. This so-called "bilingual advantage" is evident across the lifespan: young bilingual children outperform monolinguals on executive function tasks requiring inhibition and focused attention (Bialystok, 1999; Bialystok & Martin, 2004; Kovács & Mehler, 2009; Martin-Rhee & Bialystok, 2008); healthy adult bilinguals are faster than monolinguals on cognitive control tasks (Bialystok, 2006; Costa, Hernández, Costa-Faidella, & Sebastián-Gallés, 2009); and older adult bilinguals exhibit less cognitive decline due to aging than monolinguals (Bialystok, Craik, Klein, & Viswanathan, 2004; Schweizer, Ware, Fischer, Craik, & Bialystok, 2012). In the current research, we are interested in whether there exist broad, domain-general effects of bilingualism on different tasks involving cognitive control—the ability to regulate mental activity to resolve information-conflict during processing. Here, we use the term cognitive control instead of inhibitory control (or inhibition) to describe this process, because conflict could be successfully resolved by inhibiting irrelevant information, by promoting relevant information, or both (Botvinick, Braver, Barch, Carter, & Cohen, 2001). Despite some evidence supporting a bilingual advantage in cognitive control (Bialystok, 2010; Bialystok, Craik, Green, & Gollan, 2009; Bialystok et al., 2004; Costa et al., 2009; Martin-Rhee & Bialystok, 2008; but see also Hilchey & Klein, 2011; Paap & Greenberg, 2013), there are still several unanswered questions regarding its nature, specificity, and extent.

In particular, few studies have tested whether the bilingual advantage cascades into language processing. Provided that the source of bilinguals' cognitive advantage is the systematic control of two languages, these benefits should be observed in the linguistic domain—however, much of the work in this area focuses on the effects of bilingualism in *non-linguistic* contexts. It is also unclear

* Corresponding author at: Department of Otolaryngology, Medical University of South Carolina, 135 Rutledge Ave MSC 550, Charleston, SC 29425, USA.
*E-mail address:* teubnera@musc.edu (S.E. Teubner-Rhodes).

http://dx.doi.org/10.1016/j.cognition.2016.02.011
0010-0277/

how robust the bilingual advantage is to changing task demands, especially given reports of a lack of uniformity in cross-task bilingual performance (Paap & Greenberg, 2013; Paap, Johnson, & Sawi, 2014): Does the advantage emerge consistently across tasks tapping shared cognitive control functions? Do monolinguals 'catch up' to bilinguals during cognitive control practice? The present study aims to address these issues by testing whether healthy, young adult bilinguals outperform monolinguals on a reading task involving syntactic ambiguity resolution—a cognitive control task in the linguistic domain—both before and after brief practice with a recognition-memory task that theoretically taps shared conflict-resolution functions.

We begin by reviewing what is known about the effects of bilingualism on cognitive control and the theoretical accounts of these observed effects. We then discuss how such effects might cascade into on-line sentence processing by providing an account of the role of cognitive control within sentence interpretation. Finally, we present our study, which addresses the open questions raised above.

## 2. What is the effect of bilingualism on cognitive control?

It is striking that the bilingual advantage is observed on *non-linguistic* cognitive control tasks: bilinguals exhibit faster response times (RTs) on (1) the Simon task (Bialystok et al., 2004), in which participants identify a non-spatial attribute of a visual stimulus presented on the same (congruent) or opposite (incongruent) side as the correct response; (2) the Flanker task (Costa et al., 2009), in which participants indicate the direction of an arrow that is flanked by task-irrelevant arrows pointing in the same (congruent) or opposite (incongruent) direction; and (3) the spatial Stroop task (Bialystok, 2006), in which participants indicate the direction of a single arrow that appears on the same (congruent) or opposite (incongruent) side as the correct response. Despite overt dissimilarities, these tasks all involve occasional "conflict trials," where task-irrelevant stimulus features provide misleading information; thus, they all require cognitive control to resolve competition between different sources of information.

In his seminal work, Green (1998) proposed the inhibitory control (IC) model of bilingual language processing, which theorized that a central inhibitory-control mechanism played an important role in bilingual language use by suppressing items from the lexicon not currently in use. For instance, bilinguals might inhibit words from their native language (L1) when speaking their second language (L2). Under this model, bilingualism could strengthen domain-general inhibitory control via extensive practice (Abutalebi & Green, 2008; Bialystok et al., 2009), and bilinguals could then apply their improved control to non-verbal tasks.

However, the IC model does not fully account for the diverse empirical evidence supporting an effect of bilingualism on cognitive control. If bilinguals are better specifically at inhibiting irrelevant information, then they should outperform monolinguals selectively on conflict trials where such inhibition is required. Yet in many studies, bilinguals outperform monolinguals on both congruent *and* incongruent trials (for review, see Hilchey & Klein, 2011). Indeed, in their meta-analysis of bilingual cognitive control studies, Hilchey and Klein (2011) found limited evidence that bilinguals had smaller interference effects than monolinguals, but showed that across studies, bilinguals appeared to enjoy a general advantage as long as the *task* involved conflict processing. On the basis of such evidence, Costa et al. (2009) proposed that bilinguals have superior "conflict monitoring": the ability to detect information conflict and reactively increase cognitive control recruitment (Botvinick et al., 2001). During conflict monitoring, people continuously evaluate input to determine if it contains conflicting sources of information. If so, then cognitive control is recruited to help resolve the competing evidence by inhibiting routine responses or irrelevant information, and/or by promoting correct responses or goal-relevant information; otherwise, cognitive control need not deploy (Botvinick et al., 2001). Cognitive control is thus a sub-component of conflict monitoring that is downstream from monitoring for and detecting conflict (Kerns et al., 2004). Conflict-monitoring demands are high when the input frequently switches between stimuli with and without conflict; people must therefore flexibly recruit cognitive control on a moment-by-moment basis. In such contexts, monitoring facilitates the detection of conflict and subsequent engagement of cognitive control, but it also helps to detect the *absence* of conflict—monitoring occurs continuously because the individual cannot know a priori if a given stimulus will contain conflict. By contrast, in environments where conflict is always or nearly always present, monitoring demands are low because cognitive control can be applied uniformly (Botvinick et al., 2001; Costa et al., 2009).

Consistent with this account, Costa et al. (2009) found that the magnitude of the bilingual advantage was larger on Flanker task versions with approximately equal proportions of conflict and non-conflict trials than versions with relatively unequal proportions. When congruent and incongruent trials occurred equally often, imposing heavy monitoring demands, bilinguals were significantly faster at both trial types. Yet bilinguals performed no differently from monolinguals when the vast majority of trials (92%) were incongruent (Costa et al., 2009); their advantage disappeared when conflict-monitoring demands were low, despite high cognitive control demands. Moreover, brain-imaging research finds that language switching trials and incongruent Flanker trials co-activate overlapping voxels in the anterior cingulate cortex (ACC), a medial-frontal region thought to be involved in monitoring for conflict and signaling adjustments in control (Abutalebi et al., 2012; see Botvinick et al., 2001). This idea is supported by evidence that conflict-related activity in the ACC is reduced when conflict is expected (Carter et al., 2000) and that the ACC responds to cues indicating the conflict-status of an upcoming trial, regardless of whether that status is congruent or incongruent (Aarts, Roelofs, & van Turennout, 2008). Thus, bilinguals' experience of language switching may engage and strengthen the domain-general conflict monitoring system.

Despite this evidence, inconsistencies across the bilingualism literature question the robustness of an effect of bilingualism on conflict monitoring and cognitive control. One problem is that monolinguals often 'catch up' to bilinguals with a small amount of practice (see e.g., Bialystok et al., 2004; Costa et al., 2009). If one session of practice on the Simon or Flanker task is equivalent to a lifetime of bilingual language experience, then the effect of bilingualism on conflict monitoring and cognitive control seems rather weak—perhaps bilinguals reach a functional limit and are unable to improve further. Yet characteristic cognitive control tasks (e.g., Simon, Flanker) typically yield high performance and bilinguals may already be performing at a task ceiling (e.g., accuracy greater than 97% with reaction times faster than 400 ms across several task blocks; Bialystok et al., 2004); thus it may be impossible to observe continued improvements. The current study examines whether monolinguals and bilinguals benefit differentially from cognitive control practice by administering tasks with initially low performance, allowing for greater practice-related changes, potentially even in bilinguals.

A final issue is that a bilingual advantage is observed in some experiments but not in others, with no apparent pattern to its (non-)occurrence (Hilchey & Klein, 2011; Paap & Greenberg, 2013). Indeed, Paap and Greenberg (2013) assessed the stability of bilingual benefits by administering within-subjects a variety of executive function tasks (Simon, Flanker, Antisaccade, Ravens

Progressive Matrices, and Color-Shape Switching) to healthy young monolinguals and bilinguals (see also, Paap et al., 2014). As often as not, bilinguals exhibited a nominal *disadvantage* relative to monolinguals. The authors acknowledged, however, that correlations among these different tasks are rather weak; thus, the inconsistency in bilingual performance may have been because the tasks largely assessed different components of executive function. A current challenge for bilingual research therefore is to demonstrate that a bilingual advantage occurs consistently across tasks that tap a common cognitive control mechanism. To this end, we test whether bilingual benefits manifest in sentence processing when conflict-monitoring demands are high, and if this performance can be tied to conflict-monitoring abilities in a non-syntactic domain.

## 3. Do the effects of bilingualism cascade into on-line sentence processing?

Most investigations of bilingualism's effects on conflict monitoring and cognitive control have been limited to non-linguistic tasks. Yet, if controlled use of two languages enhances conflict monitoring and cognitive control, then bilingualism must impact performance on linguistic tasks involving cognitive control as well. One difficulty with testing this is that bilinguals exhibit *slower* lexical access in each of their languages (Gollan, Montoya, Cera, & Sandoval, 2008; Ivanova & Costa, 2008; Sandoval, Gollan, Ferreira, & Salmon, 2010), perhaps reflecting increased competition across two constituent lexicons (Spivey & Marian, 1999). Yet little is known about the effects of bilingualism on sentence processing after lexical access has occurred. If bilingualism improves conflict monitoring and cognitive control, then we believe that—despite their apparent disadvantages in lexical access—bilinguals should enjoy a sentence processing advantage when behavior must be adjusted—namely, when the environment necessitates monitoring for syntactic conflict and potentially frequent misinterpretation.

This prediction stems directly from evidence that general-purpose cognitive control functions deploy under language processing conditions involving ambiguity (Hsu & Novick, in press; January, Trueswell, & Thompson-Schill, 2009; Novick, Kan, Trueswell, & Thompson-Schill, 2009; Novick, Trueswell, & Thompson-Schill, 2005; Ye & Zhou, 2009). Ambiguity arises frequently in natural language processing: the combinatorial nature of language means that typical utterances can have tens to hundreds of different possible parses (Wasow, Perfors, & Beaver, 2005). In particular, during sentence processing, readers and listeners may recruit cognitive control to revise misinterpretations that arise when multiple, conflicting evidential sources lead them to an incorrect syntactic analysis (Novick et al., 2005). According to constraint-based models of parsing, as readers and listeners perceive input, they rapidly consult multiple, probabilistic sources of information (e.g., lexico-syntactic cues and visual context) to make real-time predictions about sentence meaning (MacDonald, Pearlmutter, & Seidenberg, 1994; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995; Trueswell, Tanenhaus, & Garnsey, 1994). For most sentences, evidential sources converge, and initially favored parses ultimately turn out to be correct. Such sentences should not require conflict resolution even if other parses were initially available, but disfavored. Sometimes, however, an early interpretation of a sentence conflicts with evidence that arrives later in the sentence, which can result in processing difficulty (known as the "garden-path effect"). This forces readers and listeners to revise their incorrect analysis. For example, consider the following sentence from the Washington Post's *Afternoon Buzz* e-mail newsletter, "The new hunting season opens today, with more hunters and more bears allowed to be killed". The reader may initially believe that both hunters and bears can be killed in greater numbers this season. However, as this is implausible, the reader must override this misinterpretation and realize that there are more hunters this season, and these hunters are allowed to kill more bears. Under such conditions, cognitive control may serve to rein-in initial misinterpretations and recover the intended meaning (Novick et al., 2005; Ye & Zhou, 2009). Accordingly, if bilingualism enhances conflict monitoring and cognitive control processes, then it should also improve performance on sentence processing tasks involving syntactic ambiguity.

But how exactly should the effects of bilingualism manifest in syntactic ambiguity resolution? We consider this question in view of the hypothesis that bilinguals have conflict monitoring advantages on non-linguistic tasks (Costa et al., 2009; Hilchey & Klein, 2011). Typical language contexts often contain ambiguous *and* unambiguous sentences, so readers and listeners must monitor for contradictions between their initial interpretation and subsequent input as they cannot know in advance when their initial parse will turn out to be wrong. If bilinguals are better at conflict monitoring, then they should be better at detecting ambiguities and recruiting cognitive control to revise misinterpretations, but also at using converging information sources to efficiently arrive at the correct interpretation even in unambiguous sentences. In other words, good conflict monitors are prepared either to resolve conflict or easily recognize the absence of conflict. Thus, bilinguals should outperform monolinguals on ambiguous *and* unambiguous sentences in linguistic environments that contain both—that is, under conditions when they have to monitor for potential misinterpretations.

Relatively few studies have examined the effects of bilingualism on sentence processing. An important exception, however, is an investigation of auditory sentence comprehension, which found that bilinguals had higher comprehension accuracy than monolinguals on "target" sentences with atypical word orders, but only when they had to ignore simultaneously-presented "distracter" sentences (Filippi, Leech, Thomas, Green, & Dick, 2012). This result suggests that bilinguals are better at suppressing interfering linguistic information than monolinguals. However, the bilinguals in that study had primarily acquired their second language after age 10—it is plausible then that they became fluent in a second language *because* they possessed superior linguistic (or cognitive control) abilities. It remains uncertain then whether bilingualism actually improves parsing abilities. In the present study, parsing abilities are tested in early bilinguals who acquired both languages prior to age 10. It is unlikely that such people become bilingual as a result of superior cognitive control, because, by and large, they learn two languages not by choice, but because their particular environment involves simultaneous (or nearly simultaneous) input of two language systems.

## 4. The current study

The present study addressed three open questions in the bilingualism literature. First, do the effects of bilingualism on conflict monitoring and cognitive control emerge consistently across different tasks with shared conflict-resolution demands? Second, does practice on a cognitive control task benefit bilinguals and monolinguals differentially? Finally, does bilingualism affect sentence processing when ambiguity/conflict is present?

We tested Spanish–Catalan bilinguals and Spanish monolinguals on a reading task involving temporary syntactic ambiguity both before and after practice on either a high- or no-conflict version of an N-back recognition-memory task (where N is 3; see Fig. 1 for a study-design schematic). For consistency, the entire experiment was conducted in Spanish for both language groups.

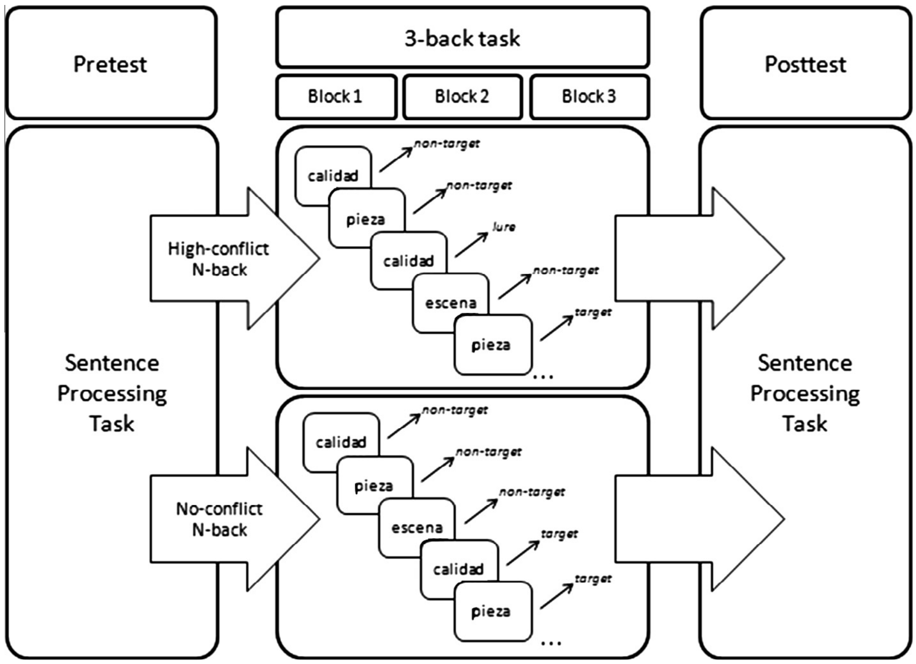


**Fig. 1.** Schematic of the study design. Participants completed a sentence-processing task before they were randomly assigned to *either* a high- *or* no-conflict version of the N-back task. They completed a complementary form of the sentence-processing task after the N-back task. Both N-back versions are depicted: while the no-conflict task (bottom panel) contained only target trials that were 3-back matches and non-target trials that had not appeared before, the high-conflict task (top panel) also included lure trials, items that had appeared before but not in the target 3-back position, thus tapping conflict detection of highly familiar but non-target stimuli. For instance, in the high-conflict task, the second "calidad" is a lure, because it matches the item that had occurred 2 (rather than 3) items previously. In contrast, the same item appears as a target, or 3-back match, in the no-conflict task, which did not include any lures.

The pretest/posttest design allowed us to compare baseline sentence processing abilities and the effects of conflict monitoring and cognitive control practice in bilinguals and monolinguals. It also allowed us to test whether the effect of bilingualism emerges consistently across ostensibly distinct tasks that nevertheless share the need to detect information-conflict and recruit cognitive control.

We specifically chose a recognition-memory task and a sentence-parsing task involving temporary ambiguity because they both appear to recruit a common cognitive control mechanism (Novick et al., 2005). This shared requirement to employ cognitive control when conflict arises is critical, because it drives the need to monitor for conflict and adjust cognitive control to match task demands on a trial-by-trial basis. In our version of the N-back task, participants viewed single words presented in sequence and identified whether the current word matched the one shown three trials back. Importantly, we manipulated conflict monitoring demands through different forms of the task: The high-conflict N-back included *lures*, stimuli that induce familiarity because they were recently presented but not in the critical N-back (i.e., 3-back) position; so, to arrive at the correct position-based response, participants must employ cognitive control to override a familiarity bias (Burgess, Gray, Conway, & Braver, 2011). In contrast, the no-conflict version omitted lure trials, so successful performance only requires recognition memory.

The high-conflict N-back task is demanding and captures individual differences in performance on other cognitive control tasks. For instance, behavioral improvements during long-term training (i.e., 10 sessions over three to six weeks) on this task predict gains in garden-path recovery (Novick, Hussey, Teubner-Rhodes, Harbison, & Bunting, 2014). Moreover, conflict trials on N-back and other, similar recognition-memory tasks (Gray, Chabris, & Braver, 2003; Jonides & Nee, 2006) consistently activate the same neural regions as syntactic ambiguity resolution and prototypical conflict-control tasks like Stroop and Flanker (January et al., 2009; Ye & Zhou, 2009). Thus, the high-conflict N-back task engages cognitive control resources that are also recruited when processing garden-path sentences. By contrast, the no-conflict N-back task omits the conflict trials that drive this cognitive control recruitment.

Our predictions were as follows. First, given the shared cognitive control demands of the sentence processing and the high-conflict N-back tasks, we hypothesized that bilinguals would outperform monolinguals on both tasks. We predicted that bilinguals should be faster and more accurate than monolinguals for both conflict (ambiguous sentences and lures on N-back) and non-conflict (unambiguous and filler sentences and non-lures on N-back) trial types due to an advantage in conflict monitoring that enables more flexible implementation of cognitive control. However, on the no-conflict N-back task, where moment-to-moment adjustments in control are not necessary, we predicted that bilinguals and monolinguals would perform equivalently.

Finally, we hypothesized that *both* bilinguals and monolinguals should benefit from brief practice on the high-conflict N-back task. This hypothesis may appear at odds with previous findings that monolinguals demonstrate greater improvements over the course of cognitive control tasks, ultimately achieving levels of performance comparable to that of bilinguals. However, bilinguals have often started at ceiling performance in prior studies, leaving little room for improvement. We predict that a lifetime of bilingual experience should outweigh the effects of brief cognitive control practice: bilinguals should outperform monolinguals over the course of an entire task when given equal opportunity to improve. Because bilinguals should not start at ceiling on the high-conflict N-back task (average accuracy is typically between 60% and 70%; see Kane, Conway, Miura, & Colflesh, 2007), they are expected to improve with practice, preventing monolinguals from 'catching up.' Bilinguals and monolinguals are also expected to improve in syntactic ambiguity resolution from pretest to posttest; however, these improvements may be larger for participants in the high-conflict N-back group, because only they practiced implementing cognitive control between sentence processing assessments.

## 5. Method

### 5.1. Participants

Participants included healthy adult balanced Spanish–Catalan bilinguals ($N = 59$; 7 males; $M_{age} = 20.78$, $SD_{age} = 3.38$) and Spanish-speaking monolinguals ($N = 51$; 12 males; $M_{age} = 26.51$, $SD_{age} = 5.94$) recruited from the University of Barcelona community. All participants gave informed consent in accordance with the University of Maryland and University of Barcelona IRB-approved protocol. Participants in each language group were randomly assigned to either the high- or no-conflict N-back condition. Participant characteristics are provided in Table 1. Additional details regarding the collection and scoring of education and socioeconomic status (SES) data from a subset of participants are given in Appendix A.

Language status was verified using language questionnaires borrowed from Appendix B in Costa et al. (2009). Such self-report measures of both first- and second-language proficiency have been shown to correlate strongly with objective proficiency assessments (Gollan, Weissberger, Runnqvist, Montoya, & Cera, 2012; Marian, Blumenfeld, & Kaushanskaya, 2007) and ratings by impartial, native language observers (Rosenthal, Wang, Schillinger, Pérez Stable, & Fernandez, 2011; Shameem, 1998). Participants all had some exposure to both Spanish and Catalan because they currently attended University of Barcelona. However, whereas bilinguals had either learned both languages at home or were educated in both languages from childhood,[1] monolinguals had no more than minimal exposure to Catalan (ratings of 1 and 2 on the language use measure) until adolescence or adulthood. Most participants had formal instruction in foreign languages but were not functionally fluent in languages other than Spanish or Catalan.[2]

As shown in Table 1, bilinguals in both conflict groups tended towards Spanish dominance, but reported significantly greater Catalan usage than monolinguals across their lifespan. Their L1 and Spanish proficiency were comparable to that of monolinguals but, as expected, their L2 proficiency was significantly higher. Although bilinguals reported significantly higher other-language proficiency than monolinguals, this proficiency was comparable to monolinguals' L2 proficiency, suggesting similar foreign language proficiency. Importantly, bilinguals and monolinguals in the high- and no-conflict groups did not differ from each other on any measure, except that high-conflict bilinguals had significantly more Catalan use in childhood than no-conflict bilinguals. Note, however, that the bilingual groups reported the same Spanish language proficiency and that their language usage was equivalent in adolescence and adulthood.

SES score distributions showed that distributions did not significantly differ across the high-conflict bilingual, high-conflict monolingual, no-conflict bilingual, and no-conflict monolingual groups (Kruskal–Wallis rank sum test: $H(3) = 0.71$, $p = .87$), suggesting that SES was comparable across groups among those who provided data. However, bilinguals had significantly less education than monolinguals in both the high-conflict and no-conflict groups (see Table 1).

All participants were given the option of receiving payment (12 Euros) or psychology course credit for their participation. More bilinguals ($n = 56$) chose course credit than monolinguals ($n = 15$); however, because participants were allowed to choose, it is unlikely that any observed group differences could be ascribed to motivational factors related to compensation. Also, despite the gender imbalance in the experiment, females accounted for the same high distribution of participants across the two language groups and across the two versions of the N-back task.

An additional 25 participants completed the study, but were dropped from analyses because they did not fit into either language group ($n = 19$), because they were less than 75% accurate on filler sentences or non-target N-back trials ($n = 5$; 2 bilinguals), or because of computer error ($n = 1$; monolingual).

### 5.2. Materials and procedure

#### 5.2.1. Sentence processing task

Participants completed a moving window self-paced reading task (Just, Carpenter, & Wooley, 1982) at pretest and posttest (Fig. 1). In this paradigm, participants are presented with a sentence that is initially masked with a sequence of dashes (each dash denotes a character in the text). The participant presses a button to unmask the first word and ensuing button presses re-mask the previous word while revealing the current one, so that participants read one word at a time. Two initial lists of Spanish sentences were created, consisting of 32 critical items and 64 fillers each (see Appendix B for examples).[3] The critical items were eleven words long and were interpretable as either subject-first or object-first cleft sentences until the seventh, disambiguating word (Betancort, Carreiras, & Sturt, 2009; del Río et al., 2011); however, the subject-first interpretation is strongly preferred. For example:

(1) Este es el general que vigilaba al espía desde la ventana. (Subject-first)
(This is the general who watched the spy from the window.)
(2) Este es el general que vigilaba el espía desde la ventana. (Object-first)
(This is the general who the spy watched from the window.)

In Spanish, the subject-first construction is much more frequent, and the al/el manipulation results in large ambiguity effects: relative to subject-first sentences, object-first constructions elicit increased first-pass and total reading times in the disambiguating region (e.g., el espía), indicating processing difficulty (Betancort et al., 2009). Note that "el" is the masculine nominative singular definite article in Catalan as well as Spanish, and Catalan also uses "a(l)" to mark certain kinds of direct objects (as in Spanish "al" in the sentence above). However, this usage is much more restricted in Catalan than in Spanish. For most speakers of Central Catalan, the dialect spoken in Barcelona, the use of "a" in the sentences that comprised our critical items would be ungrammatical (Escandell-Vidal, 2007; Solà, 1994). In Spanish, processing difficulty for object- versus subject-first constructions is associated with increased activation of neural regions implicated in cognitive control; and, on average, participants incorrectly interpret more than 20% of object-first sentences, compared with only 5–10% misinterpretation of subject-first sentences (del Río et al., 2011). This sug-

[1] One bilingual in the no-conflict group was a native Romanian, but had begun learning both Spanish and Catalan at age 5, was fluent in both languages, and had lived in the Barcelona area since early adolescence. Excluding this subject from analyses did not change the pattern of results.

[2] Two monolinguals (1 high-conflict, 1 no-conflict) and one bilingual (no-conflict) reported fluent proficiency in other languages learned through formal instruction. The pattern of results did not change when these participants were excluded from analyses.

[3] Because our sample included both Spanish and Latin American participants, a native Argentinian psycholinguist and two native Spanish psycholinguists (authors LA and MS-T) reviewed our entire set of sentence materials to ensure that they contained no dialect-specific features.

**Table 1**
Bilingual and monolingual demographics by conflict group.

| | High-conflict group | | | | No-conflict group | | | | High- vs no-conflict | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Biling ($n$ = 32) | Monoling ($n$ = 26) | Biling vs monoling | | Biling ($n$ = 27) | Monoling ($n$ = 25) | Biling vs monoling | | Biling | | Monoling | |
| | | | $t$ | $p$ | | | $t$ | $p$ | $t$ | $p$ | $t$ | $p$ |
| Age | 20.5 (3.2) | 25.5 (5.4) | −4.4 | *** | 21.1 (3.7) | 27.5 (6.4) | −4.5 | *** | −0.6 | .5 | −1.2 | .2 |
| Sex ($n$ males) | 4 | 6 | | | 3 | 6 | | | | | | |
| Education[a] | 3.8 (1.1) | 5.5 (1.8) | −2.6 | * | 3.8 (1.2) | 6.5 (1.4) | −4.5 | *** | −0.0 | ∼1 | −1.3 | .2 |
| SES[a] | 2.0 (0.8) | 1.7 (0.9) | | | 1.8 (0.7) | 1.8 (0.8) | | | | | | |
| L1–L2 use[b] | | | | | | | | | | | | |
| Childhd | 3.9 (1.4) | 1.1 (0.2) | 10.1 | *** | 3.2 (1.5) | 1.0 (0.2) | 7.3 | *** | 2.0 | * | 0.6 | .6 |
| Adolesc | 3.9 (1.3) | 1.1 (0.3) | 11.0 | *** | 3.4 (1.2) | 1.1 (0.3) | 9.0 | *** | 1.5 | .1 | −1.1 | .3 |
| Adulthd[c] | 3.8 (1.3) | 1.7 (0.8) | 7.2 | *** | 3.4 (1.4) | 1.6 (0.6) | 6.6 | *** | 1.0 | .3 | 1.4 | .2 |
| Proficiency[d] | | | | | | | | | | | | |
| L1[e,f] | 4.0 (0.1) | 4.0 (0.2) | −0.3 | .8 | 4.0 (0.1) | 3.9 (0.3) | 1.1 | .3 | −0.2 | .9 | 1.0 | .3 |
| L2 | 3.9 (0.2) | 2.5 (0.8) | 10.0 | *** | 3.9 (0.3) | 2.5 (0.8) | 8.1 | *** | 0.5 | .6 | 0.1 | ∼1 |
| Other | 2.6 (0.6) | 1.8 (0.5) | 5.6 | *** | 2.5 (0.7) | 1.8 (0.8) | 3.7 | *** | 0.3 | .8 | 0.0 | ∼1 |
| Spanish[f] | 3.9 (0.1) | 4.0 (0.2) | −0.7 | .5 | 3.9 (0.1) | 3.9 (0.3) | 0.9 | .4 | −0.3 | .8 | 1.0 | .3 |
| Sp dom ($n$)[g] | 26 | | | | 25 | | | | | | | |
| Sp born ($n$)[h] | 30 | 15 | | | 24 | 12 | | | | | | |

Biling = Bilinguals, Monoling = Monolinguals. SES = socioeconomic status. Childhd = childhood, Adolesc = adolescence, Adulthd = Adulthood. Sp dom = number of Spanish dominant bilinguals (Spanish proficiency ⩾ Catalan). Sp born = number born in Spain.

[a] Different $n$: 10 high-conflict bilinguals, 10 high-conflict monolinguals, 11 no-conflict bilinguals, 9 no-conflict monolinguals. See Appendix A for details.

[b] Language use was rated from 1: *always Spanish* to 7: *always Catalan*. Ratings <4 indicate greater Spanish usage and >4 greater Catalan usage.

[c] Adult language use was not reported for 1 no-conflict monolingual.

[d] Proficiency was rated from 1: *very little* to 4: *perfectly* for spoken comprehension, reading comprehension, speaking (fluency), speaking (pronunciation) and writing.

[e] Catalan was used as L1 for 6 simultaneous bilinguals (3 high-conflict, 3 no-conflict) who received equivalent early language input and reported identical proficiency in both languages.

[f] L1/Spanish proficiency was not reported for 1 no-conflict monolingual.

[g] Excluding Catalan dominant bilinguals did not change the pattern of results.

[h] The remaining 2 high-conflict bilinguals were born in Latin America (1 Chile, 1 Colombia) and moved to Barcelona in childhood ($M_{age}$ = 4.5 years, $SD$ = 2.1). The remaining 11 high-conflict monolinguals were born and raised in Latin America (1 Argentina, 1 Chile, 1 Ecuador, 1 Peru, 7 Colombia) and moved to Barcelona as young adults ($M_{age}$ = 24.6 years, $SD$ = 8.3). Of the remaining 3 no-conflict bilinguals, 1 did not provide their native country, and 2 were born elsewhere in Europe (1 Andorra, 1 Romania) but learned both Spanish and Catalan at home in early childhood and moved to Barcelona at a mean age of 9 ($SD$ = 5.7). The remaining 13 no-conflict monolinguals were born and raised in Latin America (2 Argentina, 4 Chile, 1 Mexico, 6 Colombia) and moved to Barcelona as young adults ($M_{age}$ = 26.6, $SD$ = 8.1).

* $p < .05$.

*** $p < .001$.

gests that participants use cognitive control to overcome a strong subject-first parsing bias to successfully (re)interpret object-first sentences.

Half the critical items in each list contained "al" (marking subject-first) and half contained "el" (object-first). Additionally, we swapped the "al" and "el" conditions in complementary versions of the two lists: subject-first sentences became object-first sentences, and vice versa. Filler sentences were seven to fourteen words long and varied in terms of syntactic structure and complexity. None of the fillers were garden-paths, but sixteen fillers in each list contained a variety of harder-to-process structures, including multiple embedded prepositional phrases, passive verbal constructions, and fronted direct objects. These items helped disguise the critical manipulation by ensuring that object-first sentences were not the only difficult items. A True–False probe (e.g., El general vigilaba al espía (The general watched the spy)) followed each sentence to assess comprehension. The majority of the critical-item probes (75%) were designed to be false, so that participants would have to successfully reanalyze the object-first sentences to respond correctly. That is, for critical items, correct responses on False probes require rejecting the alternative interpretation, whereas a participant could correctly answer a True probe by holding both the subject-first and object-first interpretations in mind. Filler probes were balanced so that overall, each list contained half True and half False probes. True and False probes occurred with the hard fillers in the same proportions as with the rest of the fillers.

Participants saw one list of sentences before completing the N-back task and then a different list afterward. Sentences were presented in pseudorandom order so critical items were never adjacent. List presentation was counterbalanced across participants. Before starting the sentence task, participants were given instructions that included an example of a filler sentence with its probe and the correct response to ensure that they understood the response mapping.

#### *5.2.2. N-back task*

144 four- to eight-letter Spanish nouns and adjectives were selected from the LEXESP database via the BuscaPalabras software tool. Selection criteria were frequency between 20 and 30, familiarity rating between 5 and 7, concreteness rating between 1 and 3.9, and imageability rating between 3.5 and 7 (Davis & Perea, 2005; Sebastián-Gallés, Martí, Cuetos, & Carreiras, 2000).

The N-back task contained three blocks of 96 trials each (see Appendix C). Each block lasted about 6.5 min, was followed by a 1-min break, and used a different set of Spanish words. During the task, word stimuli appeared one-by-one for 2-s each, with a 2-s inter-stimulus interval. Participants judged whether the current item matched or mismatched the item presented *three* trials previously. They were instructed to respond as quickly and accurately as possible, pressing one button for *targets* (i.e., 3-back matches) and another for non-matches using the index finger of each hand. The response mapping was counterbalanced across participants within each language and conflict group. Before starting the task, participants were given instructions that included an example list of 6 words containing 2 lures and 1 target, with the correct response given for each word.

In each block, 3-back targets comprised 50% of the trials. However, in the no-conflict version, all non-match trials were *non-target* words that had not appeared before, whereas in the high-conflict version, 36 out of 48 non-match trials were *lure* items that had appeared recently, but two, four, or five trials previously. While both versions involved maintaining alertness and updating memory, the high-conflict version additionally required participants to override their familiarity for lure items to correctly reject them as non-matches.

### 5.2.3. Task analyses

We conducted multilevel mixed-effects models with participants and items as crossed random effects using R's glmer function (lme4 library, Bates & Sarkar, 2007). Mixed-effects models are preferable to ANOVA because they can be more reliable (Barr, Levy, Scheepers, & Tily, 2013) and because they allow random effects of participants and items to be considered simultaneously (Baayen, 2008). Moreover, mixed-effects models are highly robust to missing and unbalanced data (Baayen, Davidson, & Bates, 2008; Wu, 2010), so they are ideal for the current data where sample size varies across groups. We employed linear models for RT data, but used logistic models for accuracy data because of their binomial distribution. For each analysis, we started with the full structure justified by the design; then, we conducted step-wise comparisons with simpler fixed effects by first removing non-significant interaction terms and then removing variables without significant main effects or interactions. The model with the lowest Akaike Information Criteria (AIC) was considered the best-fitting model and was used to calculate parameter estimates. This model selection procedure minimizes information loss, or the estimated distance between the fitted model and the true structure of the data (Burnham & Anderson, 2002, pp. 60–64). Adding model parameters reduces the AIC (i.e., reduces information loss) until the noise in the parameter estimation exceeds its explanatory power (Burnham & Anderson, 2002, p. 62); thus, our approach favored more complex models unless a parameter added no information. Following the recommendation of Barr et al. (2013), we always used the full random-effects structure justified by the design unless this model (a) failed to converge or (b) contained random slopes that were highly correlated ($r > .9$) with the intercept or with each other. In the former case, interactions between the random slopes terms were removed before fitting the model. In the latter case, the original model's AIC was compared to the AIC when the relevant random slope was removed, and the model with the lower AIC was retained.

We first fitted the mixed-effects models with contrast coded variables, so that the model intercept represented the grand mean of performance across conditions. The parameters for fixed effects in these contrast coded models probe main effects and interactions, much like traditional ANOVA. For example, consider the following model of high-conflict N-back performance: Accuracy $\sim$ (Group + Block $\times$ Type) + (1 + Block + Type | Subject) + (1 + Block | Item). This models the dependent variable, accuracy, as a function of fixed effects of language group, block, trial type, and an interaction between block and trial type (Group + Block $\times$ Type; where + indicates only the main effect is estimated and $\times$ indicates both main effects and the interaction are estimated), by-subject random intercepts and random slopes for block and trial type (1 + Block + Type | Subject), and by-item random intercepts and random slopes for block (1 + Block | Item). In this model, the beta estimate for Group indicates how much the mean accuracy of each language group differs from the grand mean of accuracy; in other words, it evaluates the main effect of language group. The beta estimate for Block 1 $\times$ Target indicates how much the mean accuracy for target trials at block 1 differs from the grand mean of accuracy; examining each of the block and trial type combinations is equivalent to evaluating the interaction of block and trial type.

To follow-up on significant model parameters, we re-coded the best-fitting models using dummy coding. Dummy coding requires that one level of the variable serve as a reference level to which all other levels are compared. This allows for comparisons between individual conditions, much like traditional *t*-tests. For example, the present study had three different blocks in the N-back task. To compare performance between blocks, we could set block 1 as the reference level, and the model parameters would compare block 2 to block 1 and block 3 to block 1. However, to compare performance of blocks 2 and 3 to each other, we would have to re-code the model to have block 2 or block 3 as the reference level. Thus, we applied different coding schemes to the best-fitting models to understand the differences between conditions that gave rise to significant main effects and interactions.

A parameter was considered significant if its $\beta$-estimate was at least twice its standard error, i.e., if the magnitude of its associated $z$- or $t$-statistic (for logistic and linear regression, respectively) was 2 or greater (Gelman & Hill, 2007, p. 40). Because this threshold approximates $\alpha = .05$, we report $p < .05$ for all significant parameters to facilitate comparison with traditional statistical approaches. We report the results only from the best-fitting mixed-effect models.

## 6. Results and discussion

### 6.1. General analyses

Four participants (1 no-conflict monolingual, 1 no-conflict bilingual, and 2 high-conflict bilinguals) initially misunderstood the task instructions for N-back and had abnormally low accuracy on block 1. Consequently, block 1 was removed for these participants, and analyses that computed gains over the course of N-back excluded their data.

Incorrect trials were excluded from response and reading time analyses because they may reflect different underlying cognitive processes than correct trials. This affected 22% of N-back data and 34% of the critical subject- and object-first items for the sentence data. Although these error rates seem high, we anticipated relatively poor accuracy because certain items (i.e., lures on N-back and object-first sentences) were intended to elicit errors. To reduce the effect of outliers, we replaced responses more than 2.5 standard deviations beyond each participant's mean with the 2.5 standard-deviation threshold values.[4] This outlier-resetting procedure affected 2.58% of correct N-back data and 2.76% of correct critical items for the sentence processing data.[5]

We begin by reporting analyses on the N-back data to assess whether we observed a typical bilingual advantage in conflict monitoring, which is a prerequisite to examining whether the advantage extends to a syntactic domain. The high- versus no-conflict N-back manipulation allowed us to test directly whether the effects of bilingualism emerge selectively under high conflict-monitoring demands. Then, we report sentence comprehension accuracy and reading time data to assess whether there is also a bilingual advantage in sentence processing and whether brief cognitive control practice (i.e., the conflict condition of the intervening 3-back task) mediated the relationship between language experience and sentence processing. Notably, if the effects of bilingualism emerge consistently across tasks with shared cognitive control demands, then the sentence processing results should parallel those of the high-conflict N-back task.

### 6.2. N-back performance

We examined accuracy and RT on the N-back task to determine if bilinguals demonstrated better performance on a non-syntactic

[4] The pattern of results was similar when these values were excluded rather than replaced.

[5] Note, however, that for sentence processing data, incorrect trials were excluded after the outlier-resetting procedure so that they were included when computing participants' residualized reading times (see Sentence Processing Results).

cognitive control task than monolinguals and if the groups improve differentially with practice. Because the high- and no-conflict N-back tasks contained different trial types, we conducted mixed-effect models separately for each conflict condition using language group, trial type, block, and their interactions as fixed effects. RT results, which largely corroborated the patterns observed for accuracy, are reported in the Online Supplement.

Average accuracy is reported in Table 2 for both high- and no-conflict conditions.

**Table 2**
Mean (and standard deviation) of proportion of correct responses for the high and no-conflict N-back tasks.

| Trial type | High-conflict | | No-conflict | |
|---|---|---|---|---|
| | Bilingual | Monolingual | Bilingual | Monolingual |
| *Block 1* | | | | |
| Lures | .70 (.12) | .60 (.21) | – | – |
| Non-targets | .96 (.07) | .91 (.12) | .99 (.02) | .97 (.05) |
| Targets | .63 (.16) | .59 (.14) | .68 (.17) | .70 (.17) |
| *Block 2* | | | | |
| Lures | .72 (.18) | .60 (.24) | – | – |
| Non-targets | .94 (.09) | .91 (.13) | .98 (.03) | .97 (.06) |
| Targets | .71 (.16) | .66 (.19) | .76 (.20) | .72 (.22) |
| *Block 3* | | | | |
| Lures | .76 (.17) | .62 (.28) | – | – |
| Non-targets | .96 (.08) | .95 (.09) | .97 (.04) | .98 (.05) |
| Targets | .73 (.21) | .70 (.20) | .78 (.19) | .75 (.20) |

### 6.2.1. High-conflict N-back

Bilinguals were generally more accurate than monolinguals on the high-conflict N-back task (Fig. 2A, left panel). This was confirmed by the model for the high-conflict N-back task (Table 3), which contained a significant fixed effect of language group ($z = 2.43$, $p < .05$). Note that the best-fitting model did <u>not</u> include any interactions with language group, indicating that bilinguals were more accurate than monolinguals regardless of trial type (lures, targets, fillers) or block. Thus, bilinguals exhibited a global accuracy advantage on high-conflict N-back that persisted throughout the duration of the task.

As can be seen in Fig. 2B, all participants were significantly more accurate on non-targets than lures ($z = 11.12$, $p < .05$) or targets ($z = 10.63$, $p < .05$) on the high-conflict task. Lower accuracy on lures than non-targets reflects the expected difficulty associated with the need to resolve interference between the familiarity of the lure and the correct serial-position information.

Additionally, overall accuracy improved over the course of the task for both language groups, which can also be seen in Fig. 2B: bilinguals and monolinguals exhibited significantly higher accuracy at block 3 than at block 1 ($z = 3.90$, $p < .05$), although significant improvement occurred only between the latter blocks (block 1 to block 2: $z = 1.24$, *ns*; block 2 to block 3: $z = 3.55$, $p < .05$). Finally, although participants improved significantly on all three trial types (lures from block 1 to block 3: $z = 2.97$, $p < .05$; targets from block 1 to block 3: $z = 6.10$, $p < .05$; non-targets from block 2 to block 3: $z = 2.38$, $p < .05$), they exhibited significantly greater improvements on targets than lures (from block 1 to block 2: $z = 3.17$, $p < .05$; from block 1 to block 3: $z = 2.97$, $p < .05$) and non-targets (from block 1 to block 2: $z = 2.25$, $p < .05$). Despite this, lure and target accuracy were never significantly different (block 1: $z = 1.19$, *ns*; block 2: $z = -0.60$, *ns*; block 3: $z = -0.61$, *ns*).

So far, these results are consistent with a conflict monitoring interpretation of the bilingual advantage. Bilinguals outperformed monolinguals in recognition memory when frequent switching occurred between conflict (lure) and non-conflict (target, non-target) trials; however, this benefit was not selective for lures. This global advantage across all item types would not be predicted if bilinguals merely had better cognitive control, because resolving conflicting information is only necessary on conflict trials.

Additionally, we observed that independent of language group, participants' accuracy (and RTs; see Online Supplement) improved and bilinguals continued to achieve significantly better performance than monolinguals over the three-block course of the experiment, again regardless of trial type. This suggests that the bilingual advantage may not be so transient, as monolinguals' performance did not begin to approximate bilinguals' when both groups had sufficient opportunity to increase performance with practice. In other words, when bilinguals did not start at ceiling, they improved just as much as monolinguals did. The bilingual advantage may therefore be less susceptible to monolingual "catch up" effects than previously understood.

Of course, the conflict monitoring interpretation of this data is only viable if the bilingual advantage disappears when monitoring demands are low. We next report the data from the no-conflict N-back task, which does not require conflict monitoring, to assess the sustainability of this conclusion.

### 6.2.2. No-conflict N-back

As can be seen in Fig. 2A (right panel), bilinguals and monolinguals achieved equivalent accuracy on the no-conflict version of the task (see also Table 2). This was confirmed by the model of accuracy in the no-conflict condition, which included significant model parameters of trial type, a block-by-trial type interaction, and a three-way group, block, and trial type interaction (see Table 3). Note the absence of a significant main effect of group ($z = 0.34$, *ns*), indicating no performance differences between bilinguals and monolinguals. As shown in Fig. 2C, participants were significantly more accurate on non-target than target trials ($z = 14.80$, $p < .05$), but they demonstrated significantly greater improvement on targets than non-targets from block 1 to block 3 ($z = 2.87$, $p < .05$). Indeed, they became significantly more accurate from block 1 to block 3 on target ($z = 4.81$, $p < .05$) but not non-target trials ($z = -0.66$, *ns*); however, this might be attributable to near-ceiling non-target performance at block 1 (see Table 2). Finally, although bilinguals' accuracy was never significantly different from monolinguals' (block 1 targets: $z = -0.37$, *ns*; block 1 non-targets: $z = 1.64$, *ns*; block 2 targets: $z = 0.82$, *ns*; block 2 non-targets: $z = 0.06$, *ns*; block 3 targets: $z = 0.55$, *ns*; block 3 non-targets: $z = -1.24$, *ns*), the three-way interaction indicated that bilinguals improved more on targets and less on non-targets than monolinguals did.

Thus, bilingual performance was comparable to that of monolinguals when lures were removed from the recognition-memory task, reducing conflict monitoring demands: no reliable accuracy (or RT; see Online Supplement) benefit was observed for bilinguals on the no-conflict N-back task. Alongside the data from the high-conflict version of N-back, it appears that bilinguals outperform monolinguals on the <u>*task*</u> that involves conflict, but not selectively on the <u>*trials*</u> that involve conflict. Thus, bilinguals' performance advantage emerges only when the task involves frequent switching between conflict and non-conflict trials, thereby necessitating monitoring and moment-to-moment adjustments in control. The advantage dissolves when this demand is absent from the information-processing context, namely in the no-conflict version of N-back.

### 6.2.3. Summary of N-back performance

On a high-conflict N-back task, we observed a general benefit in accuracy (and RT) for bilinguals across conflict and non-conflict trials. By contrast, bilinguals and monolinguals exhibited equivalent performance on the no-conflict N-back task. This pattern suggests

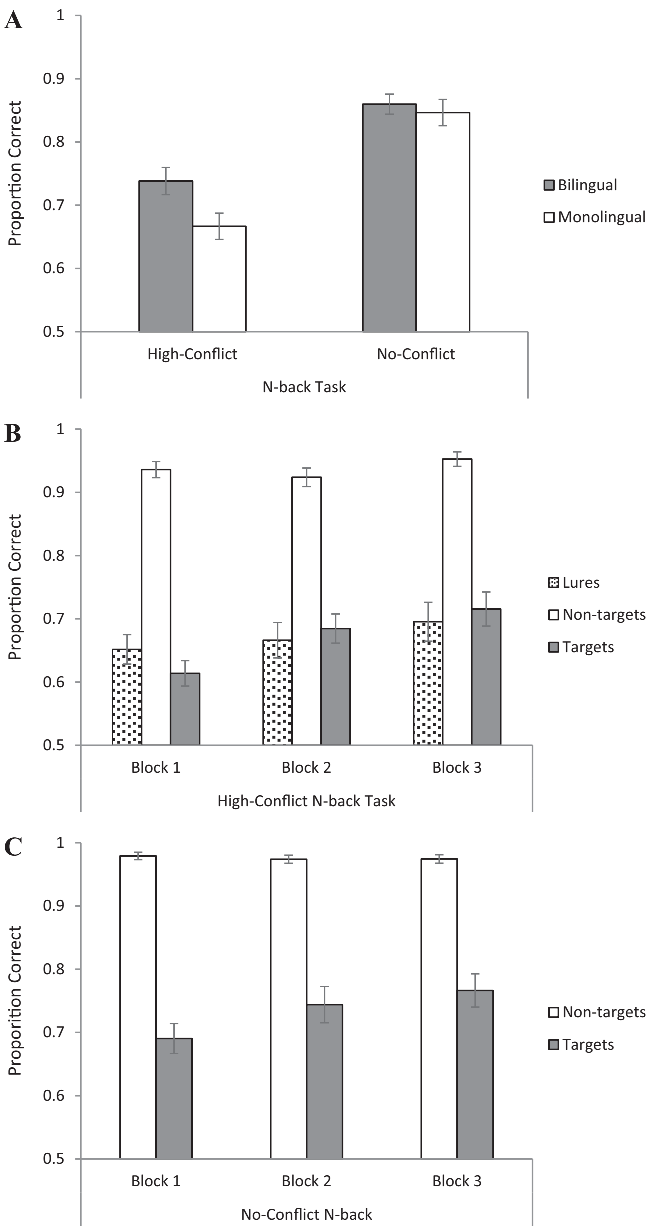


**Fig. 2.** Accuracy on the N-back task. Error bars represent standard error. (A) Accuracy by N-back task version and language group. The high-conflict and no-conflict N-back versions are presented side-by-side to show that the effect of bilingualism emerged only in the high-conflict version. Note, however, that these task versions were not compared statistically because of their differing cognitive demands (see text). (B) Accuracy by block and trial type in the high-conflict N-back task. (C) Accuracy by block and trial type in the no-conflict N-back.

that bilinguals' advantage emerges when conflict monitoring demands are high—namely, when there is frequent switching between conflict and non-conflict trial types, they are better prepared to detect conflict and deploy cognitive control, but also to make use of converging evidence (i.e., when familiarity and position-based information match). However, the effect of bilingualism disappears when monitoring demands are reduced by removing conflict trials. Because the no-conflict N-back task was identical to the high-conflict version except for the inclusion of lures, this indicates that the bilingual advantage is better explained by conflict monitoring than by increased alertness or memory-updating abilities alone. Whereas surface-level task characteristics remained constant across the high- and no-conflict versions (N always equaled 3), the need for moment-to-moment changes in cognitive control were increased in the high-conflict version by interleaving lures with target and non-target items. Finally, bilinguals and monolinguals benefitted equivalently from N-back practice, so that bilinguals maintained their advantage across blocks in the high-conflict condition.

**Table 3**
Model parameters for the best-fitting logistic mixed-effects models of proportion of correct responses for high- and no-conflict N-back.

| Model parameters | Beta estimate (*SE*) | *z*-value |
|---|---|---|
| *High-conflict N-back* | | |
| Intercept | 1.71 (0.11) | 15.61* |
| Language group | 0.16 (0.07) | 2.43* |
| Block: Block 2 | −0.08 (0.05) | −1.57 |
| Block: Block 3 | 0.28 (0.07) | 4.09* |
| Trial type: Lure | −0.83 (0.10) | −7.97* |
| Trial type: Non-target | 1.66 (0.14) | 11.61* |
| Block 2 × Lure | 0.03 (0.05) | 0.57 |
| Block 3 × Lure | −0.09 (0.06) | −1.44 |
| Block 2 × Non-target | −0.16 (0.09) | −1.80 |
| Block 3 × Non-target | 0.07 (0.10) | 0.30 |
| **Block: Block 1** | **−0.20 (0.06)** | **−3.01*** |
| **Trial type: Target** | **−0.83 (0.11)** | **−7.44*** |
| **Block 1 × Target** | **−0.15 (0.06)** | **−2.67*** |
| **Block 2 × Target** | **0.13 (0.05)** | **2.52*** |
| *No-conflict N-back* | | |
| Intercept | 2.93 (0.13) | 22.17* |
| Language group | 0.04 (0.13) | 0.34 |
| Block: Block 2 | 0.12 (0.08) | 1.39 |
| Block: Block 3 | 0.03 (0.08) | 0.43 |
| Trial type | 1.66 (0.11) | 14.80* |
| Group × Block 2 | 0.02 (0.08) | 0.29 |
| Group × Block 3 | −0.15 (0.08) | −1.92 |
| Group × Trial type | −0.01 (0.12) | −0.13 |
| Block 2 × Trial type | −0.02 (0.11) | −0.27 |
| Block 3 × Trial type | −0.16 (0.06) | −2.56* |
| Group × Block 2 × Trial type | −0.06 (0.06) | −0.87 |
| Group × Block 3 × Trial type | −0.18 (0.06) | −2.75* |
| **Block 1 × Trial type** | **0.18 (0.07)** | **2.49*** |
| **Group × Block 1 × Trial type** | **0.23 (0.07)** | **3.24*** |

*Note*. Estimated model parameters for the best-fitting logistic mixed-effects models for N-back accuracy on the high-conflict (Accuracy ∼ (Group + Block × Type) + (1 + Block + Type | Subject) + (1 + Block | Item); AIC = 17,285) and no-conflict (Accuracy ∼ (Group × Block × Type) + (1 + Block + Type | Subject) + (1 + Block | Item); AIC = 9061) tasks using contrast coding. When a variable contains 3 or more levels, not all effects are estimated in a single model. Bold area represents additional significant parameters estimated after refitting the model.
* $p < .05$.

In sum, the N-back results show that, relative to monolinguals, bilinguals enjoy a recognition-memory advantage, but only when the task involves frequent monitoring for information conflict (i.e., familiar but irrelevant lure items). Consistent with previous findings (Costa et al., 2009; Hilchey & Klein, 2011), we show a bilingual advantage across conflict and non-conflict trial types. Finally, we found that the bilingual advantage is robust to practice if the task is sufficiently demanding and bilinguals and monolinguals have equal opportunity to improve.

### 6.3. Sentence processing performance

We analyzed sentence comprehension accuracy and reading times to assess whether bilingualism also improves sentence processing on a task requiring cognitive control. Because ambiguity, and consequently the need for control, occurred unpredictably in the sentence processing task, *all* of the sentences should require conflict monitoring; therefore, we included fillers in addition to subject- and object-first sentences in our analyses of comprehension accuracy. However, fillers were omitted from reading time analyses because they contained a fundamentally different structure than critical sentences, so reading times would not reflect comparable syntactic processing.

#### 6.3.1. Sentence comprehension

Mean sentence comprehension accuracy is reported in Table 4, which shows that bilinguals had better overall sentence comprehension than monolinguals. This was confirmed by the best-fitting model of comprehension accuracy (Table 5), which includes significant parameters for language group (bilingual/monolingual), block (pre/post), sentence type (subject-first/object-first/filler), and a block-by-sentence type interaction (see Table 5). The best-fitting model dropped the effect of N-back conflict condition, indicating that N-back task version did not influence sentence comprehension accuracy.

All participants were less accurate at comprehending object-first than subject-first ($z = -13.90$, $p < .05$) or filler sentences ($z = -14.72$, $p < .05$) and less accurate on subject-first than filler sentences ($z = -3.16$, $p < .05$), replicating the expected ambiguity effect (subject-first > object-first) for these items (Betancort et al., 2009; del Río et al., 2011). Additionally, bilinguals exhibited significantly higher sentence comprehension accuracy than monolinguals ($z = 3.20$, $p < .05$; see Table 4) across all sentence types and assessments (indeed, all interactions with language group dropped out of the best-fitting model), indicating a general bilingual advantage. Finally, overall comprehension accuracy was higher at posttest than pretest ($z = 3.04$, $p < .05$), but participants only made significant gains on object-first sentences ($z = 5.68$, $p < .05$). Still, object-first accuracy remained significantly lower than subject-first ($z = -11.79$, $p < .05$) and filler sentences ($z = -13.24$, $p < .05$) at posttest.

This pattern suggests an effect of bilingualism on sentence processing that is consistent with a conflict monitoring advantage: bilinguals had better reading comprehension than monolinguals irrespective of sentence type or assessment when the task involved switching between garden-path and non-garden-path sentences. Said another way, bilinguals' higher comprehension accuracy was not confined to temporarily ambiguous, object-first sentences, suggesting that the mere presence of syntactic conflict and thus the demand to monitor for it may be driving their comprehension advantage. This interpretation stems from our high-conflict N-back finding, which parallels the current result: both tasks reveal a *global* performance benefit in bilinguals compared to monolinguals.

Furthermore, the sentence comprehension advantage was present at both pre- and posttests, indicating that, bilinguals improve much like monolinguals do (see Table 4) when the opportunity is available (i.e., when they do not start at ceiling). This too is consistent with the N-back observations and therefore implies that the positive effects of lifelong bilingualism are not easily lost once monolinguals are given short-term opportunity to practice and "catch up" to bilinguals.

The type of intervening N-back task (high- or no-conflict) did not impact the magnitude of the ambiguity effect. Instead, all participants improved at processing temporarily ambiguous, object-first sentences. This finding contradicts the prediction that participants who completed the high-conflict N-back task might exhibit greater improvements in object-first sentence comprehension. Instead, this result suggests that repeated contact with similar sentence materials is sufficient to improve syntactic ambiguity resolution, as might be anticipated by exposure-based models of language processing (Long & Prat, 2008; Wells, Christiansen, Race, Acheson, & MacDonald, 2009). It is worth noting, however, that participants only engaged in brief N-back practice (20-min)—longer high-conflict practice may be necessary for benefits to transfer to unpracticed tasks.

**Table 4**
Mean (and standard deviation) of proportion of correct responses to sentence comprehension probes for bilinguals and monolinguals for each sentence type at pretest and posttest.

| Sentence type | Pretest | | Posttest | |
|---|---|---|---|---|
| | Bilingual | Monolingual | Bilingual | Monolingual |
| Subject-first | .90 (.12) | .86 (.13) | .89 (.10) | .87 (.13) |
| Object-first | .42 (.31) | .40 (.30) | .51 (.37) | .47 (.32) |
| Fillers | .92 (.05) | .90 (.06) | .93 (.05) | .89 (.07) |

**Table 5**
Model parameters for the best-fitting logistic mixed-effects model of proportion of correct responses on sentence comprehension probes.

| Model parameters | Beta estimate (*SE*) | *z*-value |
|---|---|---|
| Intercept | 1.64 (0.12) | 13.63* |
| Language group | 0.20 (0.06) | 3.20* |
| Block | 0.10 (0.03) | 3.04* |
| Sentence type: Filler | 1.23 (0.11) | 11.24* |
| Sentence type: Object-first | −1.95 (0.13) | −15.56* |
| Block × Filler | −0.11 (0.03) | −3.13* |
| Block × Object-first | 0.18 (0.04) | 4.94* |
| **Sentence type: Subject-first** | **0.72 (0.09)** | **7.73*** |

*Note.* Model parameters for the best-fitting logistic mixed-effects model for sentence comprehension accuracy using contrast coding (AIC = 13,335; Accuracy ∼ (Group + Block × Type) + (1 + Block + Type | Subject) + (1 + Group + Block | Sentence)). When a variable contains 3 or more levels, not all possible effects are estimated in a single model. Bold area represents additional significant parameters estimated after refitting the model.
* $p < .05$.

#### 6.3.2. Reading times

Only critical items (object- and subject-first sentences) were analyzed, and the final word of each sentence was excluded to prevent wrap-up effects (i.e., longer reading times at sentence-final words that have been attributed to punctuation and clause integration; see Warren, White, & Reichle, 2009) from obscuring the effects of interest or creating spurious effects. As detailed above (see Section 6.1), we first reset each subject's outliers to their 2.5 standard-deviation threshold. We then computed each subject's residual reading times by regressing length and reading times in each region and calculating deviations from the expected reading time. This procedure factors out the effects of word length and individual differences on reading duration (Ferreira & Clifton, 1986; Trueswell et al., 1994). Incorrect trials were excluded prior to statistical analyses.

Residualized reading times were analyzed separately for each word in the sentence using linear mixed-effects models with fixed effects for group (monolingual/bilingual), block (pre/posttest), conflict (high/low), and trial type (subject/object-first), and their interactions. Since the subject- and object-first items were identical up to word 7 (el/al), which was the critical disambiguating region, the primary regions of interest were words 7–10. However, analyses were conducted on all regions to verify that there were no unanticipated effects.

The canonical garden-path effect is evidenced by significant effects of trial type in words 8, 9, and 10 ($|ts| > 4.89$, $p < .05$), reflecting increased reading times for object-first relative to subject-first sentences (see Tables 6 and 7). As expected, there was no effect of trial type prior to word 7, since the sentence types were identical up to this point. Importantly, the absence of group × trial type interactions in the early disambiguating regions (words 7–9) suggests that the garden-path effect was equivalent in bilinguals and monolinguals.

This is somewhat qualified, however, by a significant group × block × conflict × trial type interaction at word 10. Among bilinguals, both the high- and no-conflict groups demonstrated significant cross-assessment reading time improvements on object- and subject-first sentences ($|ts| > 2.15$, $ps < .05$); however, among monolinguals, the high-conflict group improved significantly on object-first ($t = -4.33$, $p < .05$) but not subject-first sentences ($t = -1.87$, *ns*), whereas the no-conflict group improved significantly on subject- ($t = -4.06$, $p < .05$) but not object-first sentences ($t = .02$, *ns*). In other words, bilinguals exhibited cross-assessment decreases in reading times on both sentence types regardless of N-back conflict condition, whereas monolinguals improved selectively on object-first sentences following the high-conflict N-back, but selectively on subject-first sentences following the no-conflict N-back. Thus, no-conflict monolinguals had significantly slower residual reading times at word 10 on object-first sentences at posttest ($M = 30$, $SD = 292$) than high-conflict monolinguals ($M = -25$, $SD = 199$; $t = 2.53$, $p < .05$) or no-conflict bilinguals ($M = -16$, $SD = 222$; $t = 2.08$, $p < .05$).

Participants also exhibited a reliable practice effect: they were faster at posttest than pretest at every word ($|ts| > 5.62$, $p < .05$; see Table 7). There were also significant interactions of group and block at words 1 and 2. At word 1, both bilinguals ($t = -7.01$, $p < .05$) and monolinguals ($t = -3.01$, $p < .05$) demonstrated significant decreases in their reading times from pretest (bilinguals: $M = 34$, $SD = 223$; monolinguals: $M = 21$, $SD = 191$) to posttest (bilinguals: $M = -34$, $SD = 138$; monolinguals: $M = -12$; $SD = 205$), but bilinguals improved significantly more than monolinguals ($t = 2.81$, $p < .05$). Similarly, at word 2, both bilinguals ($t = -6.24$, $p < .05$) and monolinguals ($t = -3.18$, $p < .05$) improved significantly from pretest (bilinguals: $M = 33$, $SD = 159$; monolinguals: $M = 20$, $SD = 183$) to posttest (bilinguals: $M = -37$, $SD = 118$; monolinguals: $M = -21$, $SD = 177$), but bilinguals improved to a greater extent ($t = 2.05$, $p < .05$).

We conducted follow-up analyses on residual reading times for the final word of the sentence to assess group differences in wrap-up effects. Participants were slower at reading object-first sentences ($M = 81$, $SD = 194$) than subject-first sentences ($M = -57$, $SD = 143$), demonstrating an ambiguity effect in the final region. This was confirmed by the best-fitting model of residual reading times (Table 7), which contained a significant effect of trial type ($t = 4.31$, $p < .05$). There was no interaction with language group ($t = 0.41$, *ns*), verifying that bilinguals and monolinguals exhibited comparable online parsing abilities even in the final sentence region.

**Table 6**
Mean (and standard deviation) bilingual and monolingual outlier-reset and residual reading times in milliseconds for the disambiguating regions of correct subject- and object-cleft items, pooled across pretest and posttest.

| Sentence type | Word7 ...*el/al* | Word8 *espía* | Word9 *desde* | Word10 *la*... |
|---|---|---|---|---|
| *Mean (SD) outlier-reset reading times* | | | | |
| Bilingual | | | | |
| Subject | 484 (165) | 686 (304) | 487 (137) | 413 (84) |
| Object | 474 (169) | 766 (423) | 588 (240) | 447 (122) |
| *Difference* | −10 | 80 | 100 | 34 |
| Monolingual | | | | |
| Subject | 471 (149) | 623 (308) | 473 (116) | 413 (114) |
| Object | 467 (148) | 711 (417) | 535 (150) | 455 (138) |
| *Difference* | −5 | 88 | 62 | 42 |
| *Mean (SD) residual reading times* | | | | |
| Bilingual | | | | |
| Subject | 1 (34) | −33 (102) | −29 (38) | −17 (36) |
| Object | −11 (101) | 44 (145) | 66 (163) | 14 (65) |
| *Difference* | −12 | 76 | 95 | 30 |
| Monolingual | | | | |
| Subject | 1 (33) | −30 (71) | −28 (40) | −15 (34) |
| Object | −7 (65) | 43 (112) | 32 (73) | 28 (70) |
| *Difference* | −8 | 73 | 60 | 43 |

*Note*. Raw reading times averaged across all regions of correct sentences did not differ between the language groups for either subject-first (bilinguals, $M = 513$ ms/word; monolinguals, $M = 505$ ms/word; $t(108) = 0.31$, $p = .76$) or object-first sentences (bilinguals, $M = 547$ ms/word; monolinguals, $M = 533$ ms/word; $t(106) = 0.46$, $p = .64$), indicating that bilinguals and monolinguals had comparable reading rates. Negative residual values reflect faster reading times than predicted given word length; positive residuals reflect slower reading times than predicted given word length.

### 6.3.3. Summary of sentence processing performance

We found a small yet reliable effect of bilingualism on overall sentence comprehension accuracy. To our knowledge, this is the first demonstration that the bilingual advantage extends to parsing tasks involving garden-path sentences. Interestingly, this advantage was not specific to temporarily ambiguous, object-first sentences, a finding that parallels the high-conflict N-back results and supports a conflict-monitoring interpretation. Overall, the ambiguity effect was reduced (but not eliminated) for sentence comprehension at posttest, due to selective gains on object-first sentences; nevertheless, the global bilingual advantage persisted across both tests, demonstrating that it is robust to practice effects on sufficiently challenging tasks. However, bilinguals did not differ from monolinguals in their reading times, suggesting that bilinguals' cognitive control advantage may only impact late-stage revision processes (see Section 7.2).

### 6.4. Shared cognitive control demands

As described in Section 4, the high-conflict N-back and sentence processing tasks were selected for their shared cognitive control demands (e.g., Novick et al., 2005, 2014). Conflict monitoring abilities should only impact performance on both tasks consistently if they each require detecting and resolving conflict—a shared demand that is most evident on conflict trials themselves, where cognitive control is needed to override a dominant mental representation. We verified that these tasks shared underlying cognitive control procedures in the current study by testing the effect of gains in lure trial accuracy during N-back (i.e., block 3 minus block 1) on improvement on object-first sentence comprehension across assessments (i.e., posttest minus pretest) using multiple regression. We focused on gain scores because the significant improvement that occurred on both the N-back and sentence processing tasks suggested that the ability to resolve conflict was not constant across blocks. Indeed, lure gains significantly predicted improvements in object-first comprehension accuracy ($t = 2.10$, $p < .05$), accounting for 8% of the variance ($R^2 = .08$; $F(1,54) = 4.43$, $p < .05$). Moreover, after identifying and removing influential observations ($n = 1$; monolingual) with a Cook's distance beyond the conventional threshold (4/N; see UCLA: Statistical Consulting Group, n.d.), the same analysis accounted for 13% of the variance ($R^2 = .13$; $F(1,53) = 8.22$, $p < .01$). The relationship between lure gains and object-first comprehension gains was not significantly different ($z = 0.14$, $p = .89$) for bilinguals ($R^2 = .15$; $t = 2.22$, $p < .05$) versus monolinguals ($R^2 = .12$; $t = 1.80$, $p = .08$). This finding does not reflect general learning ability, as improvements on target trials did not predict gains in object-first comprehension accuracy for either the high-conflict ($R^2 = .03$; $t = 1.35$, $p = .18$) or the no-conflict N-back version ($R^2 = .02$; $t = 1.08$, $p = .28$). Rather, this result provides further support that lure trials and object-first sentences rely on shared cognitive control procedures to resolve conflict.

## 7. General discussion

We observed a bilingual advantage across two tasks sharing a common cognitive control component, namely, a high-conflict N-back task and sentence processing involving syntactic ambiguity resolution. The observation of a bilingual advantage on both tasks

**Table 7**
Linear mixed-effects models of residual sentence reading times in milliseconds by region: Significant model parameters.

| Significant model parameters | Beta estimate (*SE*) | *t*-value |
|---|---|---|
| *Word 1 (Este)* | | |
| Block | −24 (4) | −5.63* |
| Group × Block | −9 (4) | −2.53* |
| *Word 2 (es)* | | |
| Block | −26 (4) | −6.24* |
| Group × Block | −8 (4) | −2.05* |
| *Word 3 (el)* | | |
| Block | −25 (4) | −6.98* |
| *Word 4 (general)* | | |
| Block | −60 (8) | −7.83* |
| *Word 5 (que)* | | |
| Block | −36 (5) | −7.63* |
| *Word 6 (vigilaba)* | | |
| Block | −64 (8) | −7.68* |
| *Word 7 (el/al)* | | |
| Block | −36 (6) | −5.71* |
| *Word 8 (espía)* | | |
| Block | −82 (10) | −8.12* |
| Type | 43 (9) | 4.90* |
| *Word 9 (desde)* | | |
| Block | −35 (5) | −6.82* |
| Type | 36 (5) | 7.58* |
| *Word 10 (la)* | | |
| Block | −27 (4) | −6.39* |
| Type | 20 (4) | 5.09* |
| Group × Block × Interference × Type | 11 (4) | 2.52* |
| Word 11 (*ventata.*) | | |
| Type | −164 (50) | −3.28* |

*Note*. Significant model parameters for the best-fitting linear mixed-effects models for residual sentence reading times for each word in the sentence using contrast coding: *Word 1* (Reading times ∼ (Group × Block × Interference × Type) + (1 + Block + Type | Subject) + (1 + Group + Block + Interference | Item); AIC = 62,130); *Word 2* (Reading times ∼ (Group × Block × Interference × Type) + (1 + Block + Type | Subject) + (1 + Group + Block + Interference | Item); AIC = 60,408); *Word 3* (Reading times ∼ (Group × Block × Interference × Type) + (1 + Block + Type | Subject) + (1 + Block + Interference | Item); AIC = 60,524); *Word 4* (Reading times ∼ (Group × Block × Interference × Type) + (1 + Block | Subject) + (1 + Block + Interference | Item); AIC = 65,516); *Word 5* (Reading times ∼ (Group × Block × Interference × Type) + (1 + Block | Subject) + (1 + Group + Block + Interference | Item); AIC = 62,627); *Word 6* (Reading times ∼ (Group × Block × Interference × Type) + (1 + Block | Subject) + (1 + Block + Interference | Item); AIC = 66,656); *Word 7* (Reading times ∼ (Group × Block × Interference × Type) + (1 + Block | Subject) + (1 + Block + Interference | Item); AIC = 65,008); *Word 8* (Reading times ∼ (Group × Block × Interference × Type) + (1 + Block × Type | Subject) + (1 + Group + Block | Item); AIC = 69,625); *Word 9* (Reading times ∼ (Group × Block × Interference × Type) + (1 | Subject) + (1 + Block | Item); AIC = 66,669); *Word 10* (Reading times ∼ (Group × Block × Interference × Type) + (1 + Block × Type | Subject) + (1 + Group × Block | Item); AIC = 63,644); *Word 11* (Reading times ∼ (Group × Block × Interference × Type) + (1 + Block + Type | Subject) + (1 + Group + Block + Interference | Item); AIC = 71,772). Word 11 was excluded when computing residual reading times for Words 1–10 to reduce the influence of wrap-up effects.
* $p < .05$.

is one of the first demonstrations that bilingualism significantly bolsters performance across tasks relying on common cognitive control processes.

The bilingual advantage manifested in a similar pattern across both tasks, emerging on both conflict trials and non-conflict trials. Because the bilingual advantage consistently extended beyond those trials requiring conflict resolution, our results support the conflict monitoring theory (Costa et al., 2009; Hilchey & Klein, 2011), which characterizes the bilingual advantage as a superior ability to detect conflict and flexibly adjust recruitment of cognitive control. According to this account, the bilingual advantage emerges because the presence of conflict during routine language use heightens monitoring demands, thereby increasing the readiness of cognitive control functions to deploy when necessary. The tasks in the present study may have particularly required good conflict detection, because the external cue for conflict is weak. Specifically, conflict arose in both tasks because the participants' mental representation (familiarity for a word on N-back and bias towards a subject-first interpretation during sentence processing) conflicted with the bottom-up input (actual serial position on N-back and object-first syntactic construction). Contrast this situation with prototypical cognitive control tasks like Stroop and Flanker, where the stimulus itself contains conflicting information. Bilinguals seem to be more adept than monolinguals at detecting that conflict exists and engaging cognitive control appropriately.

We note, however, that the bilingual advantage seems to hinge on the presence of conflict, or irrelevant information that needs to be ignored. When such conflict is removed, as in the no-conflict N-back task, bilinguals and monolinguals perform equivalently. This suggests that information processing conflict is the critical factor underlying the bilingual advantage in conflict monitoring and cognitive control.

Finally, we found that the bilingual advantage emerged across tasks and was sustained throughout cognitive control practice, suggesting that it is both consistent and robust. It is consistent in that within the same subject groups, bilinguals outperformed monolinguals on two ostensibly different tasks (e.g., recognition memory and sentence comprehension) that nevertheless tap common cognitive control mechanisms, and it is robust because monolinguals did not 'catch up' to bilingual performance over the course of an experiment, when tested on sufficiently challenging tasks. We summarize each of these important findings in greater depth in what follows.

### *7.1. Bilingualism affects conflict monitoring in a recognition memory task*

Bilinguals outperformed monolinguals selectively on the high-conflict version of N-back. This selectivity is critical; if we had also observed an advantage on the no-conflict task, then our results would have suggested that bilinguals had merely been more alert than monolinguals, as conflict monitoring and resolution are not necessary in the total absence of conflict. Instead, we found a bilingual advantage only on an N-back version involving frequent conflict, confirming that the advantage reflects improved conflict monitoring and cognitive control, rather than increased alertness or memory. Said another way, bilinguals do not appear to enjoy an advantage in the recognition or rehearsal aspects of working memory, when information must be maintained for ongoing use in the absence of interfering representations; rather, their advantage emerges only when the demands for conflict monitoring are relatively high, namely when conflict must be detected and resolved throughout a particular task context.

One alternative explanation for the advantage's disappearance on the no-conflict N-back task is that without conflict, the task became too easy, obscuring any group differences in recognition-memory. However, we find this unlikely given the observed pattern of results. Correctly identifying target items evidently taxed memory resources: participants were significantly less accurate and slower on targets than on non-targets, correctly responding on only 73% of targets. Moreover, participants became significantly more accurate and faster on targets with practice, indicating sufficient room for improvement. These results suggest that bilinguals and monolinguals performed equivalently on the no-conflict N-back task not because they were at ceiling, but because they had equivalent recognition-memory abilities.

In contrast to previous studies, we showed that both bilinguals and monolinguals improve markedly during practice on a cognitive control task. This discrepancy with prior work emphasizes the importance of using sufficiently challenging tasks when assessing group differences in cognitive abilities—tasks that are susceptible to ceiling effects will restrict the degree of improvement that is possible over the course of a task, so that lower performing individuals will necessarily "catch up" to higher performing individuals. The finding that both language groups improved was consistent with our prediction that bilinguals and monolinguals should benefit equivalently from practice when initial performance is not at ceiling. Indeed, regardless of language group, participants in the high-conflict condition increased their N-back accuracy by nearly 7%, and monolinguals never achieved bilingual-levels of performance (in either accuracy or RT). This novel finding is important because it suggests that although bilinguals already possess better cognitive control abilities, they are nevertheless able to benefit from further practice. Moreover, it shows that a mere 20 min of cognitive control practice by monolinguals does not produce benefits comparable to those endowed by a lifetime of bilingual experience. Thus, the cognitive benefits of bilingualism may be more robust than previously believed, as monolinguals do not easily 'catch up' to bilinguals with cognitive control practice. However, it remains an open question whether bilinguals' improvements during practice reflect genuine improvements in conflict monitoring and cognitive control or merely task-specific improvements. Future studies could examine whether bilinguals can benefit from long-term cognitive control training and whether such benefits transfer to untrained tasks.

### 7.2. Bilingualism affects conflict monitoring during sentence processing

Bilinguals exhibited a small, non-specific advantage over monolinguals in offline sentence comprehension throughout the study, as evidenced by their higher accuracy on comprehension probes following all sentence types (object-first, subject-first, and filler). However, bilinguals' online sentence processing was not superior to monolinguals'. A bilingual advantage in reading comprehension but not real-time parsing suggests that the observed advantage may impact late-stage semantic-integration processes. However, it is worth noting that prior studies have observed slower lexical access in bilinguals relative to monolinguals (for review, see Bialystok et al., 2009), either because of reduced lexical frequency (Gollan et al., 2008) or because of increased competition for word selection due to interference from the irrelevant language (Sandoval et al., 2010). It is therefore likely that bilinguals suffer a measurable *disadvantage* at the early stages of sentence processing (e.g., lexical retrieval), but their increased cognitive control enables them to compensate in comprehension.

Of relevance to this issue is a recent study that examined bilinguals' and monolinguals' suppression of irrelevant homograph meanings (Paap & Liu, 2014). In that study, participants read sentences containing a within-language homograph or a non-homograph control and decided whether a target word was related to the sentence meaning. Interestingly, bilinguals were slower and less accurate than monolinguals to judge that words related to the contextually-unsupported homograph meaning were unrelated to the sentence (Paap & Liu, 2014). This result suggests that bilinguals had difficulty using sentence context to suppress the irrelevant homograph meaning, arguably contradicting our finding that bilinguals outperformed monolinguals in late-stage semantic integration. However, we believe that the strength of lexical conflict present in the task may determine whether bilingual conflict monitoring and cognitive control advantages can overcome lexical access disadvantages. In Paap and Liu's study, words were presented via rapid serial visual presentation, which may have compounded existing delays in lexical access. Additionally, consider that for bilinguals the homograph is linked not just to two meanings in the test language, but may also activate cross-linguistic translations of those meanings. Thus, bilingualism may not facilitate resolution of lexicosemantic conflict in monolingual contexts because of the additional interference bilinguals face from their other language.

Crucially, in the current research, bilinguals' sentence comprehension advantage was not selective for sentences requiring ambiguity resolution. These results parallel our findings on the N-back task, further corroborating the idea that bilinguals are better at conflict detection and the flexible recruitment of cognitive control. Again, however, we would not expect a global bilingual advantage in sentence comprehension in the absence of temporarily ambiguous sentences. It is easier to manipulate the degree of conflict present in the highly-controlled N-back task than in linguistic material, which is necessarily more naturalistic and hence influenced by frequency effects, degree of ambiguity of each word, and other temporary ambiguities that naturally vary in sentences. It may in fact be impossible to remove conflict from sentence processing altogether, because its incremental nature means that temporary ambiguities routinely arise at multiple points during word and sentence comprehension. Nevertheless, if the bilingual advantage is due to conflict monitoring then it should be modulated by the frequency of conflict. The conflict monitoring theory predicts that the bilingual advantage should be largest when the need to monitor for conflict is high, and prior studies (Costa et al., 2009) have shown that the bilingual advantage disappears on the Flanker task when a high-proportion (92%) of trials are the same type (either conflict or non-conflict). Thus, the relatively low proportion of garden-path sentences in our task may account for the small magnitude of the bilingual advantage in sentence comprehension (and lack thereof in real-time processing). Specifically, the asymmetrical distribution of conflict (17%) and non-conflict (83%) trials in our sentence processing task may reduce monitoring demands, because switching between conflict and non-conflict trials is relatively infrequent. Thus, bilinguals' sentence comprehension advantage may have been relatively small in our study because conflict monitoring demands were relatively low. Future studies should determine whether this advantage could be increased with a higher degree of switching between garden-path and unambiguous sentences.

Bilinguals' sentence processing advantage was limited to offline comprehension—there were no group differences in reading times. This may indicate that the positive effects of bilingualism on cognitive control impact late-stage integration processes rather than online parsing per se. It is also possible that our moving window self-paced paradigm was not sensitive enough to detect group differences in real-time reading effects. Specifically, this paradigm can only assess first-pass times, as readers are not able to return to previous sentence regions that induced processing difficulty. Prior work using more sensitive eye-tracking measures has found that cognitive control training particularly impacts the amount of time spent re-reading materials after encountering disambiguating evidence (Novick et al., 2014). If online syntactic ambiguity resolution is best captured by second-pass or regression-path time measures, then our study would not be able to observe the effects of bilingualism on real-time revision. Future studies should use eye-tracking to examine online parsing differences in bilinguals and monolinguals. Perhaps also, future work could test the effects of bilingualism on real-time sentence revision using the 'visual-world' paradigm during spoken language comprehension (e.g., Chambers & Cooke, 2009; Spivey & Marian, 1999), to evaluate group differences in how bilingual and monolingual listeners coordinate multiple sources of information to inform moment-to-moment processing decisions.

### 7.3. Caveats and limitations

The present study examined bilinguals who speak Spanish and Catalan, two similar languages with substantial lexical overlap. It is possible that the similarity between these two languages influenced the sentence comprehension results. First, written "el" is the masculine nominative definite article in both languages, and both languages utilize "a(l)" to mark certain kinds of direct objects. Thus, there are constructions in both languages in which the presence of "el" or "al" would serve to disambiguate the structure of the sentence. This is potentially important because, in English, experience with relative clause structures increases comprehension of sentences containing relative clauses (Wells et al., 2009). Bilinguals who only encounter a particular syntactic ambiguity in one of their languages have less structure-specific experience than their monolingual counterparts and may therefore have reduced comprehension of such sentences. It is unclear whether the effects of bilingualism on conflict monitoring and cognitive control would override the potential disadvantage conferred by reduced exposure to the language-specific structure. However, although both Spanish and Catalan contain the "el/al" distinction, the use of "a(l)" in Catalan is much more restricted than in Spanish. We expect therefore that the Spanish–Catalan bilinguals in our study had less exposure to the object-first/subject-first disambiguation than monolingual Spanish speakers. In the particular critical items used in our study, the use of "a(l)" in the Catalan equivalents would have been ungrammatical for most speakers of Central Catalan, the variety spoken in Barcelona (Escandell-Vidal, 2007; Solà, 1994). As such, this particular ambiguity is likely infrequent in Catalan, so we do not believe that Catalan speakers had any more experience monitoring for this type of syntactic conflict. Nevertheless, these bilinguals still exhibited better sentence comprehension than the monolinguals.

The similarity between Spanish and Catalan also results in a relatively high incidence of cognates between the two languages. Indeed, in the present study, approximately 25% of the content words in the critical sentences were perfect orthographic cognates. On the one hand, lexical access should be easier for cross-language cognates—bilinguals are faster to name cognates than noncognates in both mixed and single language contexts (Gullifer, Kroll, & Dussias, 2013). Lexical access in Spanish–Catalan bilinguals may therefore be more comparable to that of monolinguals than in bilinguals with more dissimilar languages. However, cognate status does not seem to affect late-stage comprehension processes (Libben & Titone, 2009). Thus, the bilingual advantage observed in offline comprehension is expected to extend to bilinguals whose languages have fewer cognates.

The extent to which the differences we observed between bilinguals' and monolinguals' conflict monitoring and cognitive control abilities can be attributed to bilingual language experience is limited by how comparable the two language groups are in all factors other than language experience. All our participants were healthy, young adults recruited from the same institution, and for the subset of people who provided SES data, there were no significant differences across the language groups (Section 5.1). Because we were not able to collect SES data from all of our participants, we cannot entirely rule out the possibility that, overall, bilinguals and monolinguals came from different socioeconomic backgrounds. However, this seems unlikely, since we have no reason to believe that the participants who provided SES data were not representative of the groups as a whole.

Another relevant demographic factor is age: our monolinguals were, on average, slightly older than our bilinguals (see Table 1). However, this age discrepancy should make it harder to find our effects, because it would predict increased executive function in monolinguals—the opposite pattern of what we observed. The human prefrontal cortex does not reach maturity until the mid- to late-twenties (Miller et al., 2012; Ordaz, Foran, Velanova, & Luna, 2013), and performance on executive function tasks continues to improve until this time (Ordaz et al., 2013). Thus, because of their increased age, monolinguals in our study should have been closer to their developmental peak for conflict monitoring and cognitive control than bilinguals.

Monolingual participants may have had more heterogeneity in dialects than bilinguals, as many were Latin American rather than Spanish natives. This raises concerns about how dialectical variation in frequency and use of an optional cleft sentence construction impacted sentence comprehension. For this reason, all of our sentence materials were reviewed by a native South American psycholinguist and by two native Spanish–Catalan psycholinguists (authors LA and MS-T) during stimulus development. All sentences were judged to be felicitous across dialects. Furthermore, although the use of "a(l)" before direct objects in Spanish varies by dialect, it is generally considered obligatory in all dialects before definite human nouns, which all the nouns in our critical items were. In other linguistic environments, "a(l)" occurs more frequently in dialects of Latin American Spanish than in European Spanish. We would therefore expect our monolingual group to have at least as much exposure to "a(l)" as our bilingual group and to be correspondingly as sensitive or more sensitive to its presence in a sentence. Despite this, the bilingual group outperformed the monolingual group in sentence comprehension.

Another possible difference between the language groups is immigrant status, as a greater number of the bilingual participants than monolingual participants were originally from Spain (see Table 1). Thus, more monolinguals than bilinguals were immigrants (since in Barcelona, the local population is largely bilingual). This would principally be a concern if the two groups differed in terms of education level—when immigrant status has been suggested as an alternative explanation for the bilingual advantage, the bilingual group in question contained more Canadian immigrants, who tend to have more education than native Canadians (Morton & Harper, 2007, 2009). This artifact of immigrant status seems unlikely in our study, given that all participants were students at the University of Barcelona, primarily at the undergraduate level. In fact, our bilinguals had less, not more, education than our monolinguals (see Table 1), as monolinguals were more likely to be graduate students. Additionally, research on the demographic characteristics of study abroad participants indicate that international students tend to come from higher SES backgrounds (Simon & Ainsworth, 2012), which also suggests that, if anything, our bilinguals had a slightly lower SES. Thus, the most parsimonious account of the evidence for a bilingual advantage is that bilingualism, rather than differences in immigrant status, is responsible for the increase in conflict monitoring and cognitive control abilities.

### 7.4. Reconciliation of current findings with mixed evidence for the bilingual advantage

The findings of the present study directly contrast with recent studies that have failed to find a bilingual advantage across a variety of different executive function tasks (Hilchey & Klein, 2011; Paap & Greenberg, 2013; Paap et al., 2014). An explanation of such discrepancies is warranted: why did the advantage emerge consistently across executive function tasks in our study, but not in Paap and Greenberg's (2013), which was explicitly designed to examine the cross-task consistency of the bilingual advantage? We believe that although the tasks in Paap and Greenberg's study (Simon, Flanker, Antisaccade, Ravens Progressive Matrices, and Color-Shape Switching) can all be broadly classified as executive function tasks, they rely on different aspects of executive function and are not actually assessing the same abilities. Miyake and Friedman (2012; see also Friedman et al., 2008; Miyake, Friedman, Emerson, Witzki, & Howerter, 2000) have identified three distinct components of executive function: updating, shifting, and inhibition (i.e., cognitive control), which may be tapped to varying degrees on different executive function tasks. For instance, the Flanker task requires cognitive control to ignore irrelevant-information whereas color-shape switching requires shifting to flexibly change between rules. Additionally, many of these tasks are susceptible to ceiling effects, making it difficult to observe individual differences on these tasks in young adults, who are at their executive function peak. Indeed, previous studies have observed a reduction in color-shape switching costs (Gold, Kim, Johnson, Kryscio, & Smith, 2013) and in the Simon effect (Bialystok et al., 2004) for bilinguals relative to monolinguals in older but not younger adult populations, suggesting that although bilingualism may improve performance on these tasks, it is difficult to detect this advantage in young adults.

In contrast, the two executive function tasks in our study, N-back with lures and syntactic ambiguity resolution, are hypothesized to recruit shared cognitive control resources (Novick et al., 2005), a hypothesis that is well-supported by their similar neural and behavioral profiles (Gray et al., 2003; January et al., 2009; Jonides & Nee, 2006; Novick et al., 2009, 2014). Indeed, prior research has shown that individuals improving the most during a high-conflict N-back 'intervention' also exhibit the largest gains from pre/post-assessment in resolving syntactic conflict (Novick et al., 2014). The present work further supported the notion that cognitive control processes are required for N-back with lures and syntactic ambiguity resolution by demonstrating that improvements in conflict resolution on N-back predicted gains in object-first sentence comprehension. It is worth noting that some of the tasks that have yielded mixed evidence for a bilingual advantage also tap cognitive control (e.g., Simon, Flanker). However, these tasks may place less of a demand on conflict detection than N-back and sentence comprehension, because the conflict is more apparent: on Simon and Flanker, the stimulus itself contains conflicting information (e.g., surrounding arrows pointing in opposite direction from center arrow on Flanker), whereas on N-back and sentence comprehension, the conflict stems from mental representations that are incompatible with the bottom-up input.

Another advantage to the tasks used in the present study is that they are less susceptible to ceiling effects. Both N-back with lures and sentences with syntactic ambiguity are difficult even for healthy young adults, making it easier to observe group differences in conflict monitoring and cognitive control. Indeed, in our study the bilingual advantage was primarily reflected in accuracy: bilinguals were more accurate than monolinguals on the N-back task and on sentence comprehension probes. Such a result may not be possible on tasks like Simon and Flanker, where accuracy is close to ceiling (Paap & Greenberg, 2013).

### 7.5. Concluding remarks

The existence of a bilingual advantage in conflict monitoring and cognitive control is evident from our study, as well as from some findings in previous research. However, we are only beginning to understand the exact nature and extent of this advantage. If the bilingual advantage is best characterized as superior conflict monitoring, then the mechanisms that would strengthen conflict monitoring in bilinguals need to be delineated. Recent neuroimaging evidence suggests that the processes underlying language switching may be instrumental to the bilingual advantage. Indeed, language switching during a picture-naming task and conflict trials on a Flanker task activate overlapping areas of the ACC (Abutalebi et al., 2012), a region that has been linked to conflict monitoring processes, specifically, detecting conflict and subsequently adjusting control (Botvinick, Nystrom, Fissell, Carter, & Cohen, 1999; Kerns et al., 2004). Because language switching engages the same resources as conflict monitoring, it is plausible that the processing demands associated with switching languages confer a conflict monitoring advantage selectively to bilinguals who must frequently shift between their two languages, as occurs in code-switching communities (Green, 2011; Prior & Gollan, 2011; Valdés Kroff, Dussias, Gerfen, Perrotti, & Bajo, 2016; Verreyt, Woumans, Vandelanotte, Szmalec, & Duyck, 2015). In particular, bilinguals may have to monitor for conflict (1) during listening, to detect when a word conflicts with the expected language (following a language switch) and (2) during production, to detect when the most easily activated lexical item conflicts with the intended language for output.

In conclusion, bilingualism apparently acts as a form of cognitive control training, bestowing measurable advantages in conflict monitoring—the ability to detect unpredictable conflict and flexibly adjust recruitment of cognitive control resources. We demonstrated that this advantage applies not only to recognition memory under high-monitoring demands, but also to sentence processing involving recurrent syntactic ambiguity resolution, suggesting that conflict monitoring operates across syntactic and non-syntactic domains. Moreover, this system continues to be amenable to improvement, as both bilinguals and monolinguals made substantial gains with practice. Taken together, our results support the theory that bilinguals possess a more developed flexible cognitive control system. This increased flexibility is domain-general, underlying bilinguals' heightened detection and resolution of information-conflict during parsing and interpretation (i.e., when syntactic ambiguity is present) and within recognition memory.

## Acknowledgments

The authors would like to thank Cristina Martin and Jennifer Johnson for their assistance with data collection, Sol Lago for reviewing stimulus materials, and Jared Linck for advice on data analysis. This material is based upon work supported, in whole or in part, with funding from the United States Government. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the University of Maryland, College Park and/or any agency or entity of the United States Government. This work was supported in part by NSF-IGERT training grant DGE-0801465.

## Appendix A

### A.1. Socioeconomic status

We did not initially collect information about participants' socioeconomic status (SES); however, recent studies debate whether (see e.g., Morton & Harper, 2009) or not (see e.g., Bialystok, 2009; Calvo & Bialystok, 2014; Engel de Abreu, Cruz-Santos, Tourinho, Martin, & Bialystok, 2012) this factor influences the bilingual advantage. Thus, one-and-a-half years after the study, we invited participants to complete an online survey about their parents' income, occupations, and education levels. Crucially, the subset of participants who responded ($n = 40$) was evenly distributed across the two language and two conflict groups (high-conflict bilinguals: $n = 10$; no-conflict bilinguals: $n = 11$; high-conflict monolinguals: $n = 10$; no-conflict monolinguals: $n = 9$). We scored parental occupations from 1 to 9 on the 9-point Hollingshead Occupational Status Scale (Hollingshead, 1975). Then, we generated a composite score for each subject to determine their

**Table A1**
Parental education and occupation criteria for SES composite scores.

| SES score | Parental education criteria | Parental occupation criteria |
|---|---|---|
| 1: Low SES | Highest parental education level is no more than high school diploma or vocational equivalent | Highest parental occupation is 4 or less on Hollingshead scale |
| 2: Middle SES | At least one parent has an education between an advanced vocational and a college degree | Highest parental occupation is 4–6 on Hollingshead scale |
| 3: High SES | At least one parent has a college degree or better | Highest parental occupation is 7 or greater on Hollingshead scale |

*Note.* Occupational scores of 7 or greater include professions requiring high-level skills (e.g., teachers, engineers, lawyers, or doctors), scores of 4–6 include professions requiring mid-level skills (e.g., clerical workers, dental hygienists, carpenters), and scores less than 4 include low-skill jobs (e.g., bus drivers, bar tenders, janitors).

overall SES; composite measures of parental occupation, education, and income are more stable than income alone (McLoyd, 1998) and have previously been used to examine SES-related differences in cognitive functioning (Noble, Norman, & Farah, 2005). Because several participants ($n = 9$) chose not to report their parents' average annual income, our composite measure was based on parental education and parental occupations. SES composite scores from 1 to 3 were assigned based on the criteria in Table A1, where 1 represents the lowest SES and 3 represents the highest. For the majority of participants ($n = 31$), the scores derived from the education and occupation criteria were in agreement. If, however, these criteria indicated different scores for a particular subject, then the two scores were averaged—for example, if a subject scored a 1 for parental education and a 2 for parental occupation, then his composite SES score would be a 1.5.

## Appendix B

Example sentence items and probes. Critical items are labeled with sentence type for one list version, but type was reversed on the counterbalanced version.

| Type | Item | Probe |
|---|---|---|
| Subject-first | Este es el cardinal que presentó al/el obispo a los creyentes. | El cardinal presentó al obispo./El obispo presentó al cardenal. |
| Subject-first | Este es el general que vigilaba al/el espía desde la ventana. | El espía vigilaba al general./El general vigilaba al espía. |
| Subject-first | Este es el biólogo que visitaba al/el químico cada dos años. | El químico visitaba al biólogo./El biólogo visitaba al químico. |
| Subject-first | Este es el decano que mencionó al/el profesor en su discurso. | El decano mencionó al profesor./El profesor mencionó al decano. |
| Subject-first | Este es el cantante que admira al/el escritor por su elocuencia. | El escritor admira al cantante./El cantante admira al escritor. |
| Subject-first | Esta es la mujer que besaba al/el piloto en el aeropuerto. | El piloto besaba a la mujer./La mujer besaba al piloto. |
| Subject-first | Este es el senador que consultó al/el alcalde sobre la elección. | El alcalde consultó al senador./El senador consultó al alcalde. |
| Subject-first | Este es el político que defendió al/el redactor en el periódico. | El político defendió al redactor./El redactor defendió al político. |
| Object-first | Este es el gerente que fastidiaba el/al constructor con sus preguntas. | El constructor fastidiaba al gerente./El gerente fastidiaba al constructor. |
| Object-first | Este es el cajero que cuestionaba el/al gerente sobre el inventario. | El cajero cuestionaba al gerente./El gerente cuestionaba al cajero. |
| Object-first | Esta es la enfermera que apoyó el/al celador en su trabajo. | El celador apoyó a la enfermera./La enfermera apoyó al celador. |
| Object-first | Este es el motorista que seguía el/al camionero a la distancia. | El motorista seguía al camionero./El caminero seguía al motorista. |
| Object-first | Este es el músico que despertó el/al cantante con la melodía. | El cantante despertó al músico./El músico despertó al cantante. |
| Object-first | Este es el guionista que mencionó el/al productor hace unas semanas. | El guionista mencionó al productor./El productor mencionó al guionista. |
| Object-first | Este es el ladrón que retuvo el/al joyero durante tres horas. | El ladrón retuvo al joyero./El joyero retuvo al ladrón. |
| Object-first | Esta es la niñera que abraza el/al pequeño antes de despedirse. | La niñera abraza al pequeño./El pequeño abraza a la niñera. |
| Filler | El nuevo actor admiraba las películas del famoso director. | El director era poco conocido. |
| Filler | Los árboles del parque al lado de la escuela ocultaban al merodeador. | El merodeador se ocultaba dentro de la escuela. |
| Filler | El zumo empapó el mantel y se filtró por la alfombra. | El mantel se quedó empapado. |
| Filler | La reina quería ser o piloto de avión o médico. | La reina quería ser dentista. |
| Filler | El ministro tomó el avión del empresario durante la | El empresario tomó el avión. |

(*continued*)

| Type | Item | Probe |
| --- | --- | --- |
| | emergencia. | |
| Filler | La familia con perro cuidaba a las mascotas de sus vecinos. | La familia tenía una mascota. |
| Filler | El cachorro jugó con los niños del entrenador toda la tarde. | El entrenador jugó con el cachorro. |
| Filler | El comerciante no confiaba en la justicia después del juicio. | El comerciante confiaba en la justicia. |
| Filler | El avión y el barco impresionaron a los ingenieros. | El barco impresionó a los ingenieros. |
| Filler | Aquel granjero experimentado conduce el tractor nuevo. | El tractor nuevo es conducido por el granjero experimentado. |
| Filler | El coche del médico está mal aparcado frente a la casa. | El coche está aparcado en el hospital. |
| Filler | Luis cortejaba a la nieta de la pescadora con flores y canciones. | Luis cortejaba a la nieta. |
| Filler | Las clientas exigieron una rebaja en el precio después de saber más del producto. | Las clientas estaban satisfechas con el precio. |
| Filler | El nuevo avión fue diseñado por el exitoso ingeniero. | El ingeniero diseñó el avión. |
| Filler | El profesor y el estudiante leyeron el texto juntos. | El profesor leyó el texto solo. |
| Filler | Los prisioneros fueron liberados por los guerrilleros después de un mes en cautiverio. | Los policías liberaron a los prisioneros. |

## Appendix C

Example stimuli lists for high- and no-conflict N-back tasks.

| Item order | N-back version | | | |
| --- | --- | --- | --- | --- |
| | High-conflict | | No-conflict | |
| | Trial type | Stimulus | Trial type | Stimulus |
| 1 | non-target | calidad | non-target | lástima |
| 2 | non-target | pieza | non-target | bloque |
| 3 | lure | calidad | non-target | prenda |
| 4 | non-target | prodigio | non-target | volumen |
| 5 | target | pieza | target | bloque |
| 6 | target | calidad | target | prenda |
| 7 | target | prodigio | target | volumen |
| 8 | target | pieza | non-target | pobreza |
| 9 | non-target | suceso | non-target | canal |
| 10 | lure | calidad | target | volumen |
| 11 | lure | suceso | target | pobreza |
| 12 | lure | prodigio | non-target | salud |
| 13 | lure | pieza | non-target | manía |
| 14 | lure | calidad | non-target | episodio |
| 15 | target | prodigio | non-target | creador |
| 16 | target | pieza | target | manía |
| 17 | target | calidad | target | episodio |
| 18 | lure | pieza | target | creador |
| 19 | lure | calidad | non-target | calidad |
| 20 | non-target | escena | non-target | ritmo |
| 21 | target | pieza | non-target | máquina |
| 22 | target | calidad | non-target | masa |
| 23 | target | escena | non-target | tarea |
| 24 | target | pieza | non-target | claridad |
| 25 | target | calidad | target | masa |
| 26 | target | escena | target | tarea |
| 27 | non-target | cola | target | claridad |
| 28 | target | calidad | target | masa |
| 29 | target | escena | target | tarea |
| 30 | target | cola | non-target | dato |
| 31 | lure | escena | non-target | figura |
| 32 | lure | calidad | non-target | lentitud |
| 33 | target | cola | non-target | animal |
| 34 | lure | calidad | non-target | agente |
| 35 | lure | escena | non-target | medida |

**Appendix C** (*continued*)

| Item order | N-back version | | | |
| --- | --- | --- | --- | --- |
| | High-conflict | | No-conflict | |
| | Trial type | Stimulus | Trial type | Stimulus |
| 36 | target | cola | non-target | dureza |
| 37 | target | calidad | target | agente |
| 38 | lure | cola | target | medida |
| 39 | lure | calidad | target | dureza |
| 40 | lure | escena | non-target | placer |
| 41 | lure | calidad | non-target | dulzura |
| 42 | lure | cola | non-target | detalle |
| 43 | target | escena | non-target | período |
| 44 | lure | cola | target | dulzura |
| 45 | non-target | ocio | target | detalle |
| 46 | target | escena | target | período |
| 47 | target | cola | target | dulzura |
| 48 | target | ocio | target | detalle |
| 49 | target | escena | target | período |
| 50 | target | cola | target | dulzura |
| 51 | lure | escena | non-target | reacción |
| 52 | lure | ocio | non-target | tránsito |
| 53 | target | cola | target | dulzura |
| 54 | target | escena | target | reacción |
| 55 | lure | cola | target | tránsito |
| 56 | lure | escena | non-target | símbolo |
| 57 | non-target | quietud | non-target | núcleo |
| 58 | target | cola | non-target | belleza |
| 59 | lure | quietud | non-target | emoción |
| 60 | lure | escena | non-target | sabor |
| 61 | lure | quietud | target | belleza |
| 62 | lure | cola | target | emoción |
| 63 | target | escena | target | sabor |
| 64 | target | quietud | non-target | quietud |
| 65 | lure | escena | target | emoción |
| 66 | non-target | igualdad | target | sabor |
| 67 | target | quietud | target | quietud |
| 68 | target | escena | target | emoción |
| 69 | target | igualdad | non-target | tensión |
| 70 | target | quietud | non-target | trance |
| 71 | target | escena | non-target | compañía |
| 72 | lure | quietud | non-target | cola |
| 73 | non-target | bloque | target | trance |

(*continued on next page*)

**Appendix C** (*continued*)

| Item order | N-back version | | | |
|---|---|---|---|---|
| | High-conflict | | No-conflict | |
| | Trial type | Stimulus | Trial type | Stimulus |
| 74 | lure | igualdad | target | compañía |
| 75 | target | quietud | target | cola |
| 76 | target | bloque | target | trance |
| 77 | target | igualdad | non-target | ruptura |
| 78 | lure | bloque | non-target | religión |
| 79 | lure | quietud | non-target | peligro |
| 80 | target | igualdad | target | ruptura |
| 81 | target | bloque | target | religión |
| 82 | target | quietud | target | peligro |
| 83 | non-target | belleza | target | ruptura |
| 84 | target | bloque | target | religión |
| 85 | lure | igualdad | non-target | rumor |
| 86 | target | belleza | non-target | peste |
| 87 | target | bloque | non-target | servicio |
| 88 | non-target | unión | non-target | suceso |
| 89 | lure | igualdad | target | peste |
| 90 | lure | belleza | target | servicio |
| 91 | lure | igualdad | target | suceso |
| 92 | lure | bloque | non-target | hallazgo |
| 93 | target | belleza | target | servicio |
| 94 | target | igualdad | non-target | vistazo |
| 95 | target | bloque | target | hallazgo |
| 96 | target | belleza | target | servicio |

## Appendix D. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.cognition.2016.02.011.

## References


Aarts, E., Roelofs, A., & van Turennout, M. (2008). Anticipatory activity in anterior cingulate cortex can be independent of conflict and error likelihood. *Journal of Neuroscience, 28*, 4671–4678.

Abutalebi, J., Della Rosa, P. A., Green, D. W., Hernandez, M., Scifo, P., Keim, R., ... Costa, A. (2012). Bilingualism tunes the anterior cingulate cortex for conflict monitoring. *Cerebral Cortex, 22*, 2076–2086.

Abutalebi, J., & Green, D. W. (2008). Control mechanisms in bilingual language production: Neural evidence from language switching studies. *Language and Cognitive Processes, 23*, 557–582.

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. New York, NY: 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.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*, 255–278.

Bates, D. M., & Sarkar, D. (2007). *lme4: Linear mixed-effects models using S4 classes*, R package version 0.999375-42.

Betancort, M., Carreiras, M., & Sturt, P. (2009). The processing of subject and object relative clauses in Spanish: An eye-tracking study. *The Quarterly Journal of Experimental Psychology, 62*, 1915–1929.

Bialystok, E. (1999). Cognitive complexity and attentional control in the bilingual mind. *Child Development, 70*, 636–644.

Bialystok, E. (2006). Effect of bilingualism and computer video game experience on the Simon task. *Canadian Journal of Experimental Psychology, 60*, 68–79.

Bialystok, E. (2009). Claiming evidence from non-evidence: A reply to Morton and Harper. *Developmental Science, 12*, 499–501.

Bialystok, E. (2010). Global-local and trail-making tasks by monolingual and bilingual children: Beyond inhibition. *Developmental Psychology, 46*, 93–105.

Bialystok, E., Craik, F. I. M., Green, D. W., & Gollan, T. H. (2009). Bilingual minds. *Psychological Science in the Public Interest, 10*, 89–219.

Bialystok, E., Craik, F. I. M., Klein, R., & Viswanathan, M. (2004). Bilingualism, aging, and cognitive control: Evidence from the Simon task. *Psychology and Aging, 19*, 290–303.

Bialystok, E., & Martin, M. M. (2004). Attention and inhibition in bilingual children: Evidence from the dimensional change card sort task. *Developmental Science, 7*, 325–339.

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review, 108*, 624–652.

Botvinick, M., Nystrom, L. E., Fissell, K., Carter, C. S., & Cohen, J. D. (1999). Conflict monitoring versus selection-for-action in anterior cingulate cortex. *Nature, 402*, 179–181.

Burgess, G. C., Gray, J. R., Conway, A. R. A., & Braver, T. S. (2011). Neural mechanisms of interference control underlie the relationship between fluid intelligence and working memory span. *Journal of Experimental Psychology: General, 140*, 674–692.

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach* (2nd ed.). New York, NY: Springer.

Calvo, A., & Bialystok, E. (2014). Independent effects of bilingualism and socioeconomic status on language ability and executive functioning. *Cognition, 130*, 278–288.

Carter, C. S., MacDonald, A. M., Botvinick, M., Ross, L. L., Stenger, V. A., Noll, D., & Cohen, J. D. (2000). Parsing executive processes: Strategic vs. evaluative functions of the anterior cingulate cortex. *Proceedings of the National Academy of Sciences of the United States of America, 97*, 1944–1948.

Chambers, C. G., & Cooke, H. (2009). Lexical competition during second-language listening: Sentence context, but not proficiency, constrains interference from the native lexicon. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*, 1029–1040.

Costa, A., Hernández, M., Costa-Faidella, J., & Sebastián-Gallés, N. (2009). On the bilingual advantage in conflict processing: Now you see it, now you don't. *Cognition, 113*, 135–149.

Davis, C. J., & Perea, M. (2005). BuscaPalabras: A program for deriving orthographic and phonological neighborhood statistics and other psycholinguistic indices in Spanish. *Behavior Research Methods, 37*, 665–671.

del Río, D., Maestú, F., López-Higes, R., Moratti, S., Gutiérrez, R., Maestú, C., & del-Pozo, F. (2011). Conflict and cognitive control during sentence comprehension: Recruitment of a frontal network during the processing of Spanish object-first sentences. *Neuropsychologia, 49*, 382–391.

Engel de Abreu, J., Cruz-Santos, A., Tourinho, C. J., Martin, R., & Bialystok, E. (2012). Bilingualism enriches the poor: Enhanced cognitive control in low-income minority children. *Psychological Science, 23*, 1364–1371.

Escandell-Vidal, V. (2007). Acusatiu preposicional i dislocació amb clitic. *Caplletra, 42*, 185–218.

Ferreira, F., & Clifton, C. (1986). The independence of syntactic processing. *Journal of Memory and Language, 25*, 348–368.

Filippi, R., Leech, R., Thomas, M. S. C., Green, D. W., & Dick, F. (2012). A bilingual advantage in controlling language interference during sentence comprehension. *Bilingualism: Language and Cognition, 15*, 858–872.

Friedman, N. P., Miyake, A., Young, S. E., DeFries, J. C., Corley, R. P., & Hewitt, J. K. (2008). Individual differences in executive functions are almost entirely genetic in origin. *JEP: General, 137*, 201–225.

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. New York, NY: Cambridge University Press.

Gold, B. T., Kim, C., Johnson, N. F., Kryscio, R. J., & Smith, C. D. (2013). Lifelong bilingualism maintains neural efficiency for cognitive control in aging. *The Journal of Neuroscience, 33*, 387–396.

Gollan, T. H., Montoya, R. I., Cera, C., & Sandoval, T. C. (2008). More use almost always means a smaller frequency effect: Aging, bilingualism, and the weaker links hypothesis. *Journal of Memory and Language, 58*, 787–814.

Gollan, T. H., Weissberger, G. H., Runnqvist, E., Montoya, R. I., & Cera, C. M. (2012). Self-ratings of spoken language dominance: A Multilingual Naming Test (MINT) and preliminary norms for young and aging Spanish–English bilinguals. *Bilingualism: Language and Cognition, 15*(3), 594–615.

Gray, J. R., Chabris, C. F., & Braver, T. S. (2003). Neural mechanisms of general fluid intelligence. *Nature Neuroscience, 6*, 316–322.

Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition, 1*, 67–81.

Green, D. W. (2011). Language control in different contexts: The behavioral ecology of bilingual speakers. *Frontiers in Psychology, 2*, 1–4.

Gullifer, J. W., Kroll, J. F., & Dussias, P. E. (2013). When language switching has no apparent cost: Lexical access in sentence context. *Frontiers in Psychology, 4*, 278.

Hilchey, M. D., & Klein, R. M. (2011). Are there bilingual advantages on nonlinguistic interference tasks? Implications for the plasticity of executive control processes. *Psychonomic Bulletin & Review, 18*, 625–658.

Hollingshead, A. B. (1975). *Four factor index of social status*. New Haven: Yale University, Department of Sociology.

Hsu, N., & Novick, J. M. (in press). Dynamic engagement of cognitive control modulates recovery from misinterpretation during real-time language processing. *Psychological Science* (in press).

Ivanova, I., & Costa, A. (2008). Does bilingualism hamper lexical access in speech production? *Acta Psychologica, 127*(2), 277–288.

January, D., Trueswell, J. C., & Thompson-Schill, S. L. (2009). Co-localization of Stroop and syntactic ambiguity resolution in Broca's area: Implications for the neural basis of sentence processing. *Journal of Cognitive Neuroscience, 21*, 2434–2444.

Jonides, J., & Nee, D. E. (2006). Brain mechanisms of proactive interference in working memory. *Neuroscience, 139*, 181–193.

Just, M. A., Carpenter, P. A., & Wooley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General, 111*, 228–238.

Kane, M. J., Conway, A. R. A., Miura, T. K., & Colflesh, G. J. H. (2007). Working memory, attention control, and the n-back task: A question of construct validity. *JEP: Learning, Memory, and Cognition, 33*, 615–622.
Kerns, J. G., Cohen, J. D., MacDonald, A. W., Cho, R. Y., Stenger, V. A., & Carter, C. S. (2004). Anterior cingulate conflict monitoring and adjustments in control. *Science, 303*, 1023–1026.
Kovács, A. M., & Mehler, J. (2009). Cognitive gains in 7-month-old bilingual infants. *Proceedings of the National Academy of Sciences of the United States of America, 106*, 6556–6560.
Libben, M. R., & Titone, D. A. (2009). Bilingual lexical access in context: Evidence from eye movements during reading. *JEP: Learning, Memory, and Cognition, 35*, 381–390.
Long, D. L., & Prat, C. S. (2008). Individual differences in syntactic ambiguity resolution: Readers vary in their use of plausibility information. *Memory & Cognition, 36*, 375–391.
MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). Lexical nature of syntactic ambiguity resolution. *Psychological Review, 101*, 676–703.
Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research, 50*, 940–967.
Martin-Rhee, M. M., & Bialystok, E. (2008). The development of two types of inhibitory control in monolingual and bilingual children. *Bilingualism: Language and Cognition, 11*, 81–93.
McLoyd, V. C. (1998). Socioeconomic disadvantage and child development. *American Psychologist, 53*, 185–204.
Miller, D. J., Duka, T., Stimpson, C. D., Schapiro, S. J., Baze, W. B., McArthur, M. J., ... Sherwood, C. C. (2012). Prolonged myelination in human neocortical evolution. *PNAS, 109*, 16480–16485.
Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: Four general conclusions. *Current Directions in Psychological Science, 21*(1), 8–14.
Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., & Howerter, A. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology, 41*, 49–100.
Morton, J. B., & Harper, S. N. (2007). What did Simon say? Revisiting the bilingual advantage. *Developmental Science, 10*, 719–726.
Morton, J. B., & Harper, S. N. (2009). Bilinguals show an advantage in cognitive control – The question is why. *Developmental Science, 12*, 502–503.
Noble, K. G., Norman, M. F., & Farah, M. J. (2005). Neurocognitive correlates of socioeconomic status in kindergarten children. *Developmental Science, 8*, 74–87.
Novick, J. M., Hussey, E., Teubner-Rhodes, S., Harbison, J. I., & Bunting, M. F. (2014). Clearing the garden-path: Improving sentence processing through cognitive control training. *Language, Cognition, and Neuroscience, 29*(2), 186–217. http://dx.doi.org/10.1080/01690965.2012.758297.
Novick, J. M., Kan, I. P., Trueswell, J. C., & Thompson-Schill, S. L. (2009). A case for conflict across multiple domains: Memory and language impairments following damage to ventrolateral prefrontal cortex. *Cognitive Neuropsychology, 26*, 527–567.
Novick, J. M., Trueswell, J. C., & Thompson-Schill, S. L. (2005). Cognitive control and parsing: Reexamining the role of Broca's area in sentence comprehension. *Cognitive, Affective, & Behavioral Neuroscience, 5*, 263–281.
Ordaz, S. J., Foran, W., Velanova, K., & Luna, B. (2013). Longitudinal growth curves of brain function underlying inhibitory control through adolescence. *Journal of Neuroscience, 33*, 18109–18124.
Paap, K. R., & Greenberg, Z. I. (2013). There is no coherent evidence for a bilingual advantage in executive processing. *Cognitive Psychology, 66*, 232–258.
Paap, K. R., Johnson, H. A., & Sawi, O. (2014). Are bilingual advantages dependent upon specific tasks or specific bilingual experiences? *Journal of Cognitive Psychology*. http://dx.doi.org/10.1080/20445911.2014.944914 (Advance online publication).
Paap, K. R., & Liu, Y. (2014). Conflict resolution in sentence processing is the same for bilinguals and monolinguals: The role of confirmation bias in testing for bilingual advantages. *Journal of Neurolinguistics, 27*, 50–74.
Prior, A., & Gollan, T. H. (2011). Good language-switchers are good task-switchers: Evidence from Spanish–English and Mandarin–English bilinguals. *Journal of the International Neuropsychological Society, 17*, 682–691.
Rosenthal, A., Wang, F., Schillinger, D., Pérez Stable, E. J., & Fernandez, A. (2011). Accuracy of physician self-report of Spanish language proficiency. *Journal of Immigrant and Minority Health, 13*, 239–243.
Sandoval, T. C., Gollan, T. H., Ferreira, V. S., & Salmon, D. P. (2010). What causes the bilingual disadvantage in verbal fluency? The dual-task analogy. *Bilingualism: Language and Cognition, 13*, 231–252.
Schweizer, T. A., Ware, J., Fischer, C. E., Craik, F. I. M., & Bialystok, E. (2012). Bilingualism as a contributor to cognitive reserve: Evidence from brain atrophy in Alzheimer's disease. *Cortex, 48*, 991–996.
Sebastián-Gallés, N., Martí, M. A., Cuetos, F., & Carreiras, M. (2000). *LEXESP: Léxico informatizado del español*. Barcelona: Edicions de la Universitat de Barcelona.
Shameem, N. (1998). Validating self-reported language proficiency by testing performance in an immigrant community: The Wellington Indo-Fijians. *Language Testing, 15*, 86–108.
Simon, J., & Ainsworth, J. W. (2012). Race and socioeconomic status differences in study abroad participation: The role of habitus, social networks, and cultural capital. *ISRN Education, 2012*(413896), 1–21.
Solà, J. (1994). *Sintaxi normativa: Estat de la qüestió*. Barcelona: Empúries.
Spivey, M. J., & Marian, V. (1999). Cross-talk between native and second languages: Partial activation of an irrelevant lexicon. *Psychological Science, 10*, 281–284.
Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268*, 1632–1634.
Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influences in parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language, 33*, 285–318.
UCLA: Statistical Consulting Group (n.d.). *R data analysis examples: Robust regression*. <http://www.ats.ucla.edu/stat/r/dae/rreg.htm> Accessed 21.08.13.
Valdés Kroff, J. R., Dussias, P. E., Gerfen, C., Perrotti, L., & Bajo, M. T. (2016). Experience with code-switching modulates the use of grammatical gender during sentence processing. *Linguistic Approaches to Bilingualism*. http://dx.doi.org/10.1075/lab.15010.val.
Verreyt, N., Woumans, E., Vandelanotte, D., Szmalec, A., & Duyck, W. (2015). The influence of language-switching experience on the bilingual executive control advantage. *Bilingualism: Language and Cognition*. http://dx.doi.org/10.1017/S1366728914000352.
Warren, T., White, S. J., & Reichle, E. D. (2009). Investigating the causes of wrap-up effects: Evidence from eye movements and E-Z Reader. *Cognition, 111*, 132–137.
Wasow, T., Perfors, A., & Beaver, D. (2005). The puzzle of ambiguity. In O. Ogun & P. Sells (Eds.), *Morphology and the web of grammar: Essays in memory of Steven G. Lapointe* (pp. 265–282). Stanford, CA: CSLI Publications.
Wells, J. B., Christiansen, M. H., Race, D. S., Acheson, D. J., & MacDonald, M. C. (2009). Experience and sentence processing: Statistical learning and relative clause comprehension. *Cognitive Psychology, 58*, 250–271.
Wu, L. (2010). *Mixed effects models for complex data*. Boca Raton, FL: CRC Press.
Ye, Z., & Zhou, X. (2009). Executive control in language processing. *Neuroscience and Biobehavioral Reviews, 33*, 1168–1177.